

Bayesian Inverse Optimisation under Uncertainty Posterior Inference for Hidden Decision Objectives

Janardan Behera

Department of Statistics, Ravenshaw University, Cuttack 753003, Odisha, India; janardanbeheragreetsyou@gmail.com

How To Cite: Behera, J. Bayesian Inverse Optimisation under Uncertainty Posterior Inference for Hidden Decision Objectives. *Operations and Supply Chain Innovation* **2026**, 1(2), 8. <https://doi.org/10.53941/osci.2026.100008>

Received: 2 August 2026
 Revised: 19 September 2026
 Accepted: 9 October 2026
 Published: 16 October 2026

Abstract: Inverse optimization seeks the hidden objective that explains observed decisions, yet existing methods typically recover a single objective and ignore uncertainty in the underlying preferences. We develop a Bayesian framework that represents the unknown objective by a posterior distribution, combining a bounded-rationality choice model with prior information to quantify uncertainty directly. We establish geometric conditions for identifiability, prove posterior consistency, derive the optimal $n^{-1/2}$, contraction rate, obtain a Bernstein–von Mises approximation, and present a finite-sample concentration bound together with a robust fractional posterior for misspecified behavioural models. Posterior inference is performed using gradient-based Markov chain Monte Carlo and a Laplace approximation. Simulation studies demonstrate accurate recovery under noisy, misspecified, nonlinear, and high-dimensional settings, while a controlled clinical example illustrates uncertainty quantification in hidden decision preferences. The proposed framework extends inverse optimization from point estimation to statistically principled uncertainty quantification, providing a rigorous foundation for learning decision objectives from imperfect observations.

Keywords: inverse optimization; Bayesian inference; uncertainty quantification; bounded rationality; posterior consistency; Markov chain Monte-Carlo

1. Introduction

Consider an experienced oncologist who selects a chemotherapy regimen for a patient. The diagnosis is recorded, the patient covariates are known, and the chosen doses are written into the chart, yet the reasoning that produced that particular choice is nowhere on the page. The clinician has traded expected tumour response against the burden of toxicity, has weighed the frailty of an older patient, and has drawn on years of cases that never entered any formal model. An analyst who wishes to learn from this behaviour faces a peculiar problem. The decision is observed but the objective that generated it is not, and the same difficulty recurs whenever we watch a skilled agent act without being told why. A procurement officer allocating a fixed budget across commodities under supply constraints, a dispatcher routing vehicles across a congested network, and a portfolio manager balancing return against exposure all leave behind decisions whose rationale must be reconstructed rather than read off.

Inverse optimisation is the branch of operational research built for exactly this reconstruction. In its original linear programming form the problem is posed as the search for a cost vector that makes a given feasible point optimal, subject to a duality certificate, and the earliest treatments framed it as a second optimisation nested inside the first [1]. The idea generalises far beyond linear programmes. One can impute a convex objective from a collection of decisions by measuring how badly the stationarity conditions are violated and then minimising that violation over the space of candidate objectives [2]. A parallel tradition in econometrics reads observed choices as the solutions of an agent's optimisation problem and estimates the primitives of that problem by maximum likelihood, an approach that turned the structural estimation of dynamic decisions into a workhorse of applied economics [3]. What unites these strands is an ambition to recover the goal from the act.

Almost all of this work stops at a single answer. The output is one cost vector, one weight, one estimated objective, reported as though the decision maker's preferences had been measured to arbitrary precision. Reality

rarely obliges. A clinician's trade off between efficacy and toxicity is not a fixed constant that governs every prescription; it drifts with fatigue, with the patient in front of them, and with the accumulating weight of experience. Observations are scarce, noisy, and sometimes plainly inconsistent with any single objective. When two quite different objectives explain the same decisions almost equally well, a method that must commit to one of them discards genuine information and, worse, hides the ambiguity from whoever consumes the estimate. The point estimate looks authoritative precisely when it deserves to look uncertain.

This is a statistical problem dressed as an optimisation problem, and it invites a statistical answer. The Bayesian view of inverse problems, developed most fully for the recovery of functions and fields from indirect measurements, replaces the search for a single reconstruction with a posterior distribution over all reconstructions consistent with the data and the prior [4]. Transplanting that view into inverse optimisation changes the question we ask. Rather than demanding the objective that best explains the observed decisions, we ask for the probability distribution over objectives given those decisions, so that the answer carries its own uncertainty. A related instinct already exists in machine learning, where the goal of learning an agent's reward from demonstrations has been given a Bayesian formulation that maintains a posterior over rewards rather than a single estimate [5]. The operational research setting, with its explicit constraints, its duality structure, and its demand for interpretable managerial quantities, has not received the same treatment, and that gap is what this paper addresses.

The move from a point to a distribution is not cosmetic. It answers questions that a point estimate cannot even pose. One may read off which objective is most plausible, but also how sharply the data distinguish it from its rivals, whether a particular parameter value can be excluded at a stated level of confidence, and how much of the initial uncertainty survives once the evidence has been absorbed. For a clinician's toxicity aversion, a credible interval is a clinically meaningful object in a way that a bare number is not, because it states whether the observed prescriptions actually determine the trade off or merely fail to contradict a wide range of values. For a procurement rule, the posterior over commodity weights tells an auditor how confidently an implicit priority can be attributed to the decisions on record. Uncertainty is the quantity of interest, not a nuisance to be reported in a footnote.

To make this programme precise we need a likelihood, and the choice of likelihood is the crux of the modelling. A decision maker who optimised exactly would place all probability on the minimiser, which is degenerate and cannot accommodate the imperfect, occasionally contradictory decisions that real data contain. We therefore model the agent as boundedly rational. The probability of a decision decays smoothly as its suboptimality grows, so that near optimal choices are likely and grossly suboptimal ones are rare, with a temperature parameter governing how tightly behaviour concentrates on the optimum. This Gibbs style likelihood is not an arbitrary device. It is the discrete choice model of [6,7] lifted from a finite choice set to a continuum, it coincides with the quantal response equilibrium concept used to model bounded rationality in games [8], it is the maximum entropy distribution consistent with a given expected cost [9], and it turns the inverse problem into inference within a curved exponential family, which is what makes the analysis tractable and the theory clean.

The Gap This Paper Closes

Two facts about the current literature, taken together, define the gap. First, the mainstream of inverse optimisation, from the duality based recovery of a single cost vector [1] through its noisy, robust, quantile, and online descendants [10–14], is built to output a point, and treats sampling noise as a property of the estimator to be bounded rather than as a property of the objective to be distributed over. Second, the handful of papers that do attach a distribution to the recovered objective are very recent and stop short of the guarantees a statistician would ask for: a Bayesian optimization surrogate for the inverse optimization loss reports posterior like uncertainty without an identifiability theory or a coverage guarantee [15], and a concurrently developed hierarchical Bayesian treatment of data driven inverse optimization establishes posterior consistency for two specific data generating processes but leaves the geometry of identification, the rate of contraction, the Bernstein and von Mises normal approximation, and the behaviour under a misspecified behavioural model outside its scope [16]. A third and independent line has recently shown that non identifiability is the generic condition in parametric inverse optimization once regularization is stripped away, but pursues a set valued rather than a probabilistic answer to the resulting ambiguity [17]. No existing treatment combines a full posterior, a geometric identifiability criterion tied to the feasible region, finite sample and asymptotic frequentist guarantees for the resulting credible sets, and an explicit repair for behavioural misspecification. That combination, not any single ingredient in isolation, is the contribution of this paper, and Table 1 below places it against this wider field once the setting has been fixed.

Our contributions, distinguishing what is newly established here from the standard machinery it draws on, are the following.

- (i) We formulate Bayesian inverse optimisation for linear, convex, and contextual decision problems through a bounded rationality likelihood, and we show that the resulting model is a curved exponential family whose log likelihood is concave in the objective parameters whenever the cost is affine in them, so that the posterior is log concave, unimodal, and amenable to reliable sampling under any log concave prior. The likelihood itself is not new, being the Gibbs form familiar from discrete choice and maximum entropy reasoning [6–9]; what is new is its coupling to a constrained polyhedral forward problem and the exponential family analysis this coupling enables.
- (ii) We characterise identifiability geometrically, showing that a feature of the objective is recoverable exactly when it induces variability in behaviour, and we prove two fundamental limits, the joint non identifiability of temperature and cost scale and the invisibility of objective components orthogonal to the feasible set, together with an exact finite sample invariance of the recovered structure to the assumed temperature. This sharpens, in the probabilistic bounded rationality setting, the identifiability concerns raised for point estimators by [10] and studied in general parametric form by [17].
- (iii) We prove posterior consistency and a contraction rate of order $n^{-1/2}$, establish a Bernstein and von Mises approximation showing that the posterior is asymptotically Gaussian and centred at the maximum likelihood estimate, and complement this asymptotic statement with a non asymptotic concentration bound that is explicit in the dimension and in the decision geometry, so that Bayesian credible sets behave as frequentist confidence sets both in large samples and at fixed sample size. None of the three Bayesian treatments of inverse optimization identified above [15, 16] establishes an asymptotic normal approximation or a finite sample bound of this kind.
- (iv) We treat misspecification honestly by characterizing the pseudo true objective on which the posterior concentrates when the choice model is wrong, and we supply a calibrated fractional posterior whose credible regions recover nominal coverage, connecting the framework to the wider machinery of generalized Bayesian updating [18–21]. The fractional posterior itself is standard machinery from that literature; what is new is its calibration against the specific sandwich mismatch that the constrained bounded rationality likelihood produces.
- (v) We develop computation that respects the intractable normalising constant of the likelihood, combining a Langevin sampler that uses an exact or unbiased mean decision gradient with a Laplace approximation justified by the asymptotic theory, we analyse its complexity and demonstrate its scalability to objectives of many dimensions, and we validate the whole apparatus on a demanding simulation study and on a controlled clinical recovery in which the efficacy against toxicity trade off is reported with full posterior uncertainty and the requirements for real data are made explicit. The samplers are textbook Markov chain Monte Carlo and Laplace approximation [22, 23]; the new element is the closed form or unbiased moment gradient that the polytope structure of the forward problem supplies, which is what keeps the method away from the doubly intractable regime.

Table 1. Present framework against representative approaches to recovering hidden objectives. A tick denotes native support, a cross its absence, and a dash partial or implicit support.

Approach	Objective Uncertainty	Noisy Decisions	Prior Knowledge	Calibrated Intervals	Asymptotic Theory
Classical inverse LP	×	×	×	×	×
Convex objective imputation	×	—	×	×	—
Noisy and robust inverse optimization	×	✓	×	—	✓
Online inverse optimization	×	—	×	×	✓
Maximum entropy reward recovery	—	✓	×	×	—
Bayesian reward recovery	✓	✓	✓	—	×
This paper	✓	✓	✓	✓	✓

The remainder of the paper is organised as follows. Section 2 positions the contribution against inverse optimisation, structural discrete choice, inverse reinforcement learning, and generalised Bayesian inference. Section 3 fixes the forward decision model and the observation scheme. Section 4 constructs the Bayesian model, the likelihood, the priors, and the maximum entropy reading of the likelihood. Section 5 contains the theoretical analysis. Section 6 describes posterior computation and Section 7 its complexity and scalability. Section 8 reports the simulation study and Section 9 the clinical application. Section 10 discusses limitations and a longer research programme, and Section 11 closes. Proofs and further material are collected in the appendices.

2. Related Work and Positioning

2.1. Inverse Optimisation

The literature on inverse optimisation splits naturally by the structure of the forward problem. For combinatorial forward problems the recovery of costs that render a prescribed solution optimal has a long history and a mature toolbox, catalogued in a survey that organises the problems, the algorithmic techniques, and the complexity results [24]. For conic and semidefinite forward problems the duality certificate that underlies linear inverse optimisation extends through conic duality, and this extension broadened the reach of the framework to a large class of continuous models [25]. When decisions answer to several competing objectives at once, the notion of optimality itself must be relaxed to Pareto efficiency, and a generalised inverse formulation recovers the weights that place an observed decision on the efficient frontier, a construction developed with radiation therapy planning in mind [26]. The field has now matured to the point of a unifying account that sets out its theory and surveys its applications across operations [27].

A decisive shift occurred when the data were acknowledged to be noisy. Early formulations assumed that observed decisions were exactly optimal, which is untenable in practice, and the recognition that decisions carry error reframed inverse optimisation as a statistical estimation problem with its own notions of consistency and risk [10]. Estimating the parameters of an equilibrium from the choices of many interacting agents can likewise be cast as inverse optimisation, which connects the enterprise to econometric identification [28]. When the observations are not merely noisy but arrive with imperfect and possibly adversarial information, a distributionally robust treatment hedges against the worst case within an ambiguity set around the empirical distribution and yields estimators with finite sample guarantees [11]. The recovery problem has also been posed online, with the objective updated as each new decision is observed. One line frames this as emulating an expert through the sequential minimisation of a regret criterion [29], and a related development gives online inverse optimisation for general convex forward problems together with regret bounds [13]. Stability under contamination has been pursued through a quantile formulation that trims the influence of the least explicable observations [12], and the inverse question has been turned around to recover constraints rather than objectives, inferring the feasible region itself from decisions [30]. Applications reach well beyond the methodological core, as in the reconstruction of price responsive electricity demand from market data, where the recovered objective forecasts how load will react to prices [31].

What this body of work shares is a commitment to a point estimate. Even the statistical treatments that quantify estimation error do so in a frequentist idiom, reporting a single objective with an asymptotic standard error rather than a distribution over objectives that a decision maker can query directly. The Bayesian posterior we construct is a different object, and it answers a different set of questions.

2.2. Structural Choice and Inverse Reinforcement Learning

The idea that a smooth probability of choice can stand in for exact optimisation is old and well understood outside operational research. A behavioural axiomatisation shows that a ratio scale of utilities induces choice probabilities of a specific multiplicative form [6], and the econometric development of that idea produced the conditional logit model, in which the probability of an alternative is proportional to the exponential of its utility [7]. Simulation based estimation extended these models to rich unobserved heterogeneity and made them the standard apparatus for demand analysis [32]. In strategic settings the same exponential response appears as a solution concept in its own right, where players best respond noisily and the resulting fixed point interpolates between random and perfectly rational play [8]. Our likelihood is the continuous decision analogue of these constructions, with the feasible polytope playing the role of the choice set and the cost vector the role of the negative utility.

The machine learning counterpart is inverse reinforcement learning, which recovers the reward that rationalises demonstrated behaviour. The founding formulation posed the recovery as a feasibility problem over reward functions consistent with the demonstrated policy [33], and the apprenticeship learning variant sidestepped explicit reward recovery by matching feature expectations between the demonstrator and the learner [34]. The maximum entropy principle resolved the inherent ambiguity of reward recovery by selecting, among all reward weightings consistent with the observed feature counts, the one that commits to as little as possible, which yields a likelihood of exactly the Gibbs form we adopt [35]. A Bayesian reading maintains a distribution over rewards and updates it as demonstrations accumulate, and subsequent work characterised the maximum a posteriori reward and its computation [36]. Our framework can be read as the operational research member of this family. The distinctive features are the explicit constraint geometry of the forward problem, the interpretation of parameters as managerially meaningful costs rather than abstract reward weights, and a programme of statistical guarantees, identifiability, consistency, and posterior normality, that this literature has largely left implicit.

2.3. Bayesian Inference, Robustness, and Generalised Updating

The statistical scaffolding we rely on comes from the theory of Bayesian inverse problems and from the study of Bayesian inference under model error. The treatment of inverse problems as inference over a posterior on the unknown, with careful attention to well posedness of that posterior, provides the conceptual template [37], and the computational statistics of such problems, including the interplay between discretisation and sampling, has been developed at book length [38]. Because our likelihood encodes a behavioural assumption that will never hold exactly, the machinery of generalised Bayesian updating is essential. Belief distributions can be updated coherently through a loss rather than a likelihood, which frees inference from the requirement that the model be correctly specified [18]. When a misspecified likelihood is trusted too much the posterior can concentrate at the wrong place with false confidence, and tempering the likelihood by a learning rate repairs this pathology [19]. An alternative coarsens the likelihood by conditioning on a neighbourhood of the data rather than the data themselves, which buys robustness to small departures from the model [20]. Whatever the device, the learning rate or coarsening radius must be tuned so that the resulting credible regions attain their nominal frequentist coverage, and a general calibration procedure supplies exactly this tuning [21]. These ideas furnish the robust variant of our framework in Section 5.

Table 1 places the framework in this landscape. The novelty is not any single column but their conjunction. We carry a full posterior over objectives, we tolerate noisy and inconsistent decisions, we admit genuine prior information, we deliver credible sets that are calibrated even under mild misspecification, and we support all of this with identifiability, consistency, and normality results tailored to the constrained forward problem.

Recent probabilistic and data driven inverse optimisation has moved decisively beyond the noiseless setting, yet almost all of it stops short of a calibrated distribution over the objective. Data driven and robust formulations recover a point estimate or an ambiguity set that hedges against sampling error [10,11], quantile and distributionally robust variants model the spread of decisions without quantifying belief about the objective itself [12], and generalised or noise aware inverse optimisation introduces a residual distribution but reports the fitted objective as a single vector [13,30]. Contextual inverse optimisation extends the point estimation programme to settings where the objective depends on covariates and gives online regret guarantees for the resulting estimator, again as a point rather than a distribution [14]. The survey of the field confirms that uncertainty is treated predominantly as a property of the data rather than of the recovered objective [27]. What remains scarce is a full posterior over the hidden objective with frequentist guarantees attached to its credible sets, which is the gap the present construction fills by pairing a bounded rationality likelihood with the Bernstein and von Mises analysis of Section 5. The distinction from reward distribution methods in inverse reinforcement learning is parallel, since those place a distribution over rewards for sequential control [5,36] without the polytope geometry and identification analysis that a single shot constrained optimisation makes available.

Three very recent papers close part of this gap and deserve direct comparison, since without it a reader cannot tell what in the present treatment is genuinely new. A Bayesian optimisation surrogate for the inverse optimisation loss returns posterior like samples over the parameter together with a measure of dispersion, but treats the loss as a black box, offers no identifiability theory tied to the feasible geometry, and reports no asymptotic or finite sample coverage guarantee for the resulting intervals [15]. A concurrently developed hierarchical Bayesian framework for data driven inverse optimisation is the closest antecedent to the present paper in spirit: it places a genuine posterior over the objective, derives Markov chain Monte Carlo samplers for two data generating processes, and proves posterior consistency [16]. It does not, however, characterise identifiability through the geometry of the feasible region, does not establish a contraction rate or a Bernstein and von Mises normal approximation, and does not address behavioural misspecification, which is precisely the further ground covered by Sections 5 and 8 below. A third line shows, in a general parametric convex setting stripped of a behavioural likelihood, that non identifiability is generic rather than exceptional and pursues a set valued rather than probabilistic resolution, using a reformulated inverse learning objective that targets the compatible solution set directly [17]. Read together, these three papers confirm that the field has only just begun to treat the recovered objective as a random variable, and that a treatment carrying identifiability, contraction, asymptotic normality, and calibrated robustness together, rather than any one of these in isolation, has not yet appeared; supplying that treatment is the specific contribution of this paper.

2.4. Positioning within Operations Research: The Newsvendor as a Worked Precedent

The identifiability limits proved in Section 5.5 are not exotic artefacts of the bounded rationality likelihood. They are the same phenomenon that operations research has long known in the single most studied inverse problem in the field, the newsvendor model. When a decision maker orders a quantity Q against random demand under a per unit overage cost c_o and underage cost c_u , the optimal order is the critical fractile of the demand distribution, $Q^* = F^{-1}(c_u/(c_u + c_o))$ [39], so an observer who only ever sees ordering decisions can recover the ratio

$c_u/(c_u + c_o)$ and nothing more; doubling both costs leaves every optimal order, and hence all observable behaviour, unchanged. Data driven newsvendor methods that fit Q directly from features and past demand inherit exactly this fact, since they estimate the critical quantile rather than the two costs that generate it [40]. Proposition 3 is the general decision statement of the same result, cost scale and temperature are jointly non identifiable, only the direction of the objective survives, and Proposition 4 shows that this loss is total for magnitude and null for structure. The present framework therefore does not discover a new pathology; it gives the newsvendor's classical identifiability fact a posterior distribution, so that instead of reporting the critical ratio as a fixed number the analyst reports a credible interval for it, together with the diagnostic, absent from the newsvendor literature, of exactly which further quantities remain invisible however much data accumulate. This connection is also the answer to a natural transfer question: the efficacy to toxicity ratio reported in Section 9 plays the same role for the clinical decision that the critical fractile plays for the newsvendor, an identifiable relative quantity standing in for a pair of individually unobservable costs, and the same logic applies to any procurement, staffing, or portfolio application in which only relative trade offs, not absolute unit costs, are behaviourally visible.

3. Problem Setting

A decision maker faces a family of optimisation problems indexed by an observable context. Each context $u \in \mathcal{U}$ determines a feasible set

$$\mathcal{K}(u) = \{x \in \mathbb{R}^d : G(u)x \leq h(u), x \geq 0\}, \quad (1)$$

a polyhedron carved out by context dependent data $G(u)$ and $h(u)$. The agent evaluates decisions through a linear cost $\langle c(\theta), x \rangle$ whose coefficient vector depends on an unknown parameter $\theta \in \Theta \subseteq \mathbb{R}^p$. We take the cost to be affine in the parameter,

$$c(\theta) = c_0 + C\theta, \quad C \in \mathbb{R}^{d \times p}, \quad (2)$$

which covers the common case $c(\theta) = \theta$ with C the identity and $c_0 = 0$, as well as feature based parameterisations in which each cost coordinate is a known linear combination of a smaller set of interpretable drivers. A perfectly rational agent solves the forward programme

$$x^*(u, \theta) \in \arg \min_{x \in \mathcal{K}(u)} \langle c(\theta), x \rangle, \quad (3)$$

whose theory, duality, and vertex structure are standard [41]. The inverse problem inverts the direction of inference. We observe decisions and contexts and wish to learn θ .

The data consist of n pairs $\mathcal{D} = \{(u_i, x_i)\}_{i=1}^n$, where each $x_i \in \mathcal{K}(u_i)$ is a decision recorded under context u_i . The decisions are not assumed optimal. They are the imperfect output of an agent who aims at low cost but misses, and the statistical model in Section 4 makes this precision loss explicit through a temperature. Throughout we impose the following regularity on the forward geometry.

Assumption 1 (Feasible geometry). *For every context u in the support of the context distribution ρ , the feasible set $\mathcal{K}(u)$ is nonempty, compact, and full dimensional, with $\text{vol}(\mathcal{K}(u)) > 0$ and diameter bounded by a constant R that does not depend on u .*

Compactness guarantees that the smoothed decision distribution introduced below is a proper probability measure, full dimensionality prevents the induced covariance from collapsing onto a lower dimensional face, and the uniform diameter bound keeps the information contributed by each observation from degenerating. None of these requirements is restrictive for the allocation, routing, and dosing problems that motivate the work, since a bounded budget and nonnegativity already produce a compact polytope.

Data and Reproducibility

Two kinds of numerical evidence are reported in this paper and it is important not to conflate them. Sections 8 and 9 use *synthetic* data generated by drawing contexts from a stated distribution ρ , fixing a known data generating parameter θ^* , and sampling decisions from the model (4) itself or from a stated departure from it. This is a deliberate design choice rather than a limitation to be apologised for. Because the objective that produced the decisions is unobservable in every real application we are aware of, from clinical dosing to procurement, a controlled study in which θ^* is known in advance is the only design that can measure whether the posterior actually recovers what generated the data; a real decision log can validate predictive fit through held out likelihood but cannot certify recovery of the hidden objective against a ground truth, since that ground truth is

precisely what is missing from real data. Every generative choice, the context distribution, the loadings, the sample sizes, and the misspecification mechanisms, is stated in full in Section 8 and Appendix B, and the simulation code that produces every figure and table in this paper, with a fixed random seed, is made available at the repository cited in the data availability statement at the end of the paper, so that every reported number can be regenerated exactly. What the method requires of a *real* dataset, should one become available, is stated explicitly at the end of Section 9: a sequence of recorded decisions together with the feasible region, in the running example a per patient tolerability cap, that was in force when each decision was made. No other proprietary information is needed, and the likelihood in Section 4 applies to such a dataset without modification once it is supplied; we regard obtaining it as the natural next empirical step and say so again in Section 10.

We write $\pi(\theta)$ for the prior density on Θ , $\mathcal{L}_n(\theta)$ for the likelihood of the data, and $\Pi(\cdot | \mathcal{D})$ for the posterior. The temperature parameter is denoted $\tau > 0$. Table 2 collects the notation used across the paper.

Table 2. Notation used throughout the paper.

Symbol	Meaning
d	dimension of the decision vector
p	dimension of the unknown objective parameter
n	number of observed decisions
$\mathbf{u}, \mathcal{U}$	context and context space
ρ	distribution of contexts
$\mathcal{K}(\mathbf{u})$	feasible polytope under context $\mathbf{u}$, Equation (1)
$\mathbf{G}(\mathbf{u}), \mathbf{h}(\mathbf{u})$	constraint data defining $\mathcal{K}(\mathbf{u})$
$\mathbf{x}, \mathbf{x}_i$	decision vector and i th observed decision
θ, Θ	objective parameter and parameter space
$\mathbf{c}(\theta)$	cost vector, affine in θ , Equation (2)
$\mathbf{c}_0, \mathbf{C}$	intercept and loading of the affine cost map
$\mathbf{x}^*(\mathbf{u}, \theta)$	exact minimiser of the forward programme, Equation (3)
τ	temperature controlling departure from exact optimality
$\eta(\theta)$	natural parameter $-\mathbf{c}(\theta)/\tau$ of the exponential family
$Z(\mathbf{u}, \theta)$	partition function, Equation (5)
$p(\mathbf{x} \mathbf{u}, \theta)$	bounded rationality decision density, Equation (4)
$\mathbf{m}(\mathbf{u}, \theta)$	mean of the decision density, Equation (6)
$\mathbf{S}(\mathbf{u}, \theta)$	covariance of the decision density, Equation (6)
$\ell_n(\theta)$	log likelihood of the data
$\pi(\theta)$	prior density
$\Pi(\cdot \mathcal{D})$	posterior distribution
$\bar{\mathbf{I}}(\theta)$	per observation Fisher information, Equation (13)
$\hat{\theta}_n$	maximum likelihood estimate
θ^*	data generating parameter under correct specification
θ_0	pseudo true parameter under misspecification
ω	learning rate of the fractional posterior
$g(\theta, \mathbf{x})$	suboptimality of $\mathbf{x}$ at parameter θ , Equation (7)
ε	Langevin step size
M	number of sampler iterations

4. The Bayesian Model

4.1. A Bounded Rationality Likelihood

A likelihood for inverse optimisation must assign a probability to every decision the agent might record, including decisions that are close to optimal and decisions that are not. Exact optimality will not serve, because it concentrates all mass on the vertex minimiser and assigns probability zero to the interior and near vertex choices that populate real data. We therefore model the agent as boundedly rational and let the probability of a decision fall away as its cost rises, in the Gibbs or soft max form standard in discrete choice and quantal response modelling [6–8]. Given a context and a parameter, the decision density is

$$p(\mathbf{x} | \mathbf{u}, \theta) = \frac{1}{Z(\mathbf{u}, \theta)} \exp\left(-\frac{1}{\tau} \langle \mathbf{c}(\theta), \mathbf{x} \rangle\right), \quad \mathbf{x} \in \mathcal{K}(\mathbf{u}), \quad (4)$$

with normalising constant

$$Z(\mathbf{u}, \theta) = \int_{\mathcal{K}(\mathbf{u})} \exp\left(-\frac{1}{\tau} \langle \mathbf{c}(\theta), \mathbf{y} \rangle\right) d\mathbf{y}. \quad (5)$$

temperature τ governs how tightly behaviour clings to the optimum. As $\tau \downarrow 0$ the density concentrates on the exact minimiser and the agent approaches perfect rationality, while as $\tau \uparrow \infty$ the density flattens to the uniform law on $\mathcal{K}(\mathbf{u})$ and the decisions carry no information about the objective. Intermediate values describe an agent who prefers cheaper decisions but does not attain the cheapest, which is the regime of interest.

It is convenient to read (4) as an exponential family in the natural parameter $\boldsymbol{\eta}(\theta) = -\mathbf{c}(\theta)/\tau$, which is affine in θ by (2). The sufficient statistic is the decision itself, the log partition function is $\log Z(\mathbf{u}, \theta)$, and the first two moments of the decision follow from its derivatives,

$$\mathbf{m}(\mathbf{u}, \theta) = \mathbb{E}_{p(\cdot|\mathbf{u}, \theta)}[\mathbf{y}] = \nabla_{\boldsymbol{\eta}} \log Z, \quad \mathbf{S}(\mathbf{u}, \theta) = \text{Cov}_{p(\cdot|\mathbf{u}, \theta)}(\mathbf{y}) = \nabla_{\boldsymbol{\eta}}^2 \log Z. \quad (6)$$

covariance is positive semidefinite always and positive definite whenever $\mathcal{K}(\mathbf{u})$ is full dimensional, a fact we will use repeatedly. This moment structure is the standard grammar of exponential families and their variational treatment [42], and it is what makes both estimation and its analysis tractable here.

The suboptimality of a decision at a candidate parameter measures how far its cost exceeds the attainable minimum,

$$g(\theta, \mathbf{x}) = \langle \mathbf{c}(\theta), \mathbf{x} \rangle - \min_{\mathbf{y} \in \mathcal{K}(\mathbf{u})} \langle \mathbf{c}(\theta), \mathbf{y} \rangle \geq 0, \quad (7)$$

and equals zero precisely when $\mathbf{x}$ solves the forward programme (3). Since the minimum in (7) does not depend on $\mathbf{x}$, the density (4) is proportional to $\exp(-g(\theta, \mathbf{x})/\tau)$, which exposes the behavioural content directly. Near optimal decisions have small suboptimality and high probability, and the rate at which probability decays with suboptimality is set by τ . Classical inverse optimization, which insists on $g(\theta, \mathbf{x}_i) = 0$ for every observation, is the zero temperature limit of this model, so the Bayesian construction contains the deterministic problem as a boundary case.

4.2. Likelihood, Gradient, and Curvature

Collecting the n observations, the log likelihood is

$$\ell_n(\theta) = \sum_{i=1}^n \left[-\frac{1}{\tau} \langle \mathbf{c}(\theta), \mathbf{x}_i \rangle - \log Z(\mathbf{u}_i, \theta) \right]. \quad (8)$$

by differentiating through the affine cost and the log partition function, and using (6), gives a gradient with a transparent form,

$$\nabla_{\theta} \ell_n(\theta) = \frac{1}{\tau} \sum_{i=1}^n \mathbf{C}^{\top} (\mathbf{m}(\mathbf{u}_i, \theta) - \mathbf{x}_i). \quad (9)$$

score compares the mean decision that the model predicts under θ with the decision actually observed, aggregated through the loading matrix. A parameter for which the predicted mean decisions match the observed decisions on average sets the score to zero, which is a moment matching characterization of the maximum likelihood estimate. The curvature is

$$\nabla_{\theta}^2 \ell_n(\theta) = -\frac{1}{\tau^2} \sum_{i=1}^n \mathbf{C}^{\top} \mathbf{S}(\mathbf{u}_i, \theta) \mathbf{C} \preceq \mathbf{0}, \quad (10)$$

which is negative semidefinite because each $\mathbf{S}(\mathbf{u}_i, \theta)$ is a covariance. The log likelihood is therefore concave in θ , and strictly concave whenever the aggregated information matrix is nonsingular. This single structural fact drives the analysis of the next section.

4.3. Priors and the Posterior

Prior information enters through $\pi(\theta)$. In the cost recovery problems we study the parameters are unit costs or trade off weights, which are naturally nonnegative and often known to within an order of magnitude from domain knowledge or earlier studies. A convenient and honest choice is a product of truncated Gaussian densities on the positive orthant,

$$\pi(\theta) \propto \exp\left(-\frac{1}{2}(\theta - \mathbf{m}_0)^{\top} \boldsymbol{\Sigma}_0^{-1}(\theta - \mathbf{m}_0)\right) \mathbf{1}\{\theta \geq \mathbf{0}\}, \quad (11)$$

with prior mean $\mathbf{m}_0$ and prior covariance $\boldsymbol{\Sigma}_0$ elicited from experts or fitted to historical decisions, and $\mathbf{1}\{\cdot\}$ the indicator of the constraint. This density is log concave on a convex set, which is all the theory requires. Weakly

informative and reference choices are recovered by inflating Σ_0 . The posterior follows from Bayes theorem,

$$\Pi(\boldsymbol{\theta} | \mathcal{D}) \propto \pi(\boldsymbol{\theta}) \exp(\ell_n(\boldsymbol{\theta})), \quad (12)$$

and it is this distribution, rather than any single summary of it, that constitutes the answer to the inverse problem. Point estimates such as the posterior mean or mode are read off when a single number is wanted, but the interval estimates and tail probabilities that motivated the whole exercise come from (12) directly.

4.4. The Maximum Entropy Reading

The likelihood (4) is not merely convenient, it is canonical. Among all densities on $\mathcal{K}(\mathbf{u})$ that reproduce a prescribed expected cost, the one that maximises differential entropy has exactly the exponential form of (4), with $1/\tau$ the multiplier attached to the cost constraint. This is the variational characterisation of exponential families and the principle that a probability assignment should assume no more than the constraints impose [9]. The reading matters for interpretation. It says that our model of the agent is the least committal description consistent with the single behavioural fact that cheaper decisions are preferred, so we are not smuggling in unstated structure about how the agent chooses among decisions of equal cost. The same principle underlies the maximum entropy treatment of demonstrated behaviour in sequential settings, and it is reassuring that the constrained continuous problem inherits the same justification.

Figure 1 summarises the conceptual progression. The forward map sends objectives to decisions, the classical inverse recovers a single objective, and the Bayesian inverse lifts that single answer to a distribution whose spread reports how much the data have actually settled.

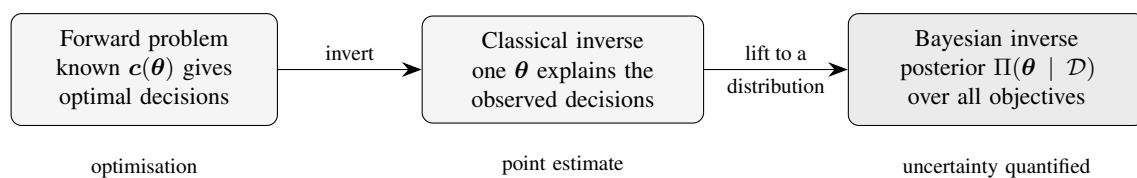


Figure 1. Forward optimization maps a known objective to decisions, classical inverse optimization recovers a single objective from decisions, and the Bayesian formulation returns a posterior distribution over objectives that carries its own uncertainty.

5. Theoretical Analysis

The value of a Bayesian answer rests on whether the posterior behaves well. We show that the posterior is unimodal, that its target is identifiable under transparent conditions, that it concentrates on the truth at the parametric rate, and that it is asymptotically Gaussian so its credible sets double as confidence sets. We then confront the fact that the behavioural likelihood is an idealisation and describe how inference recalibrates when the idealisation fails. Proofs are sketched here and completed in [Appendix A](#).

We work under [Assumption 1](#) together with the following.

Assumption 2 (Regular parameterisation). *The cost is affine in $\boldsymbol{\theta}$ as in (2), the parameter space $\Theta \subseteq \mathbb{R}^p$ is compact and convex with nonempty interior, and the data generating parameter $\boldsymbol{\theta}^*$ lies in the interior of Θ .*

Assumption 3 (Identifiability). *The per observation Fisher information at $\boldsymbol{\theta}^*$,*

$$\bar{I}(\boldsymbol{\theta}) = \frac{1}{\tau^2} \mathbb{E}_{\mathbf{u} \sim \rho} [\mathbf{C}^\top \mathbf{S}(\mathbf{u}, \boldsymbol{\theta}) \mathbf{C}], \quad (13)$$

is positive definite at $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. Equivalently the loading $\mathbf{C}$ has full column rank p and the averaged decision covariance $\mathbb{E}_\rho[\mathbf{S}(\mathbf{u}, \boldsymbol{\theta}^*)]$ is positive definite on the range of $\mathbf{C}$.

Assumption 4 (Prior regularity). *The prior density π is log concave on Θ , continuous at $\boldsymbol{\theta}^*$, and strictly positive in a neighbourhood of $\boldsymbol{\theta}^*$.*

[Assumption 3](#) is the substantive one and it has a clean behavioural meaning. Positive definiteness of $\bar{I}(\boldsymbol{\theta}^*)$ fails exactly when some nonzero direction in parameter space leaves every context conditional decision distribution unchanged, so that moving the objective along that direction produces no observable difference in behaviour. Rich contexts, in the sense that the feasible geometry varies enough across the population, prevent this degeneracy, whereas a single repeated feasible set may leave certain cost directions invisible. The condition is verifiable in a given application by inspecting the averaged decision covariance.

5.1. Shape of the Posterior

We now investigate the geometric properties of the posterior distribution, focusing on log concavity and the conditions that ensure the uniqueness of its mode.

Theorem 1 (Log concavity and unimodality). *Under Assumptions 1–4 the posterior density (12) is log concave on Θ , hence unimodal with convex support. If in addition $\bar{\mathbf{I}}(\boldsymbol{\theta}^*)$ is positive definite then, for n large enough that the empirical information is nonsingular, the posterior mode is the unique maximiser and coincides with the maximum a posteriori estimate.*

Complete argument in Appendix A. The log posterior equals $\ell_n(\boldsymbol{\theta}) + \log \pi(\boldsymbol{\theta})$ up to an additive constant. The Hessian of ℓ_n in (10) is negative semidefinite because it is a nonnegatively weighted sum of negated covariances, so ℓ_n is concave, a statement about the value function of (3) and its smoothing that sits within classical convex analysis [43]. The log prior is concave by Assumption 4. A sum of concave functions is concave, and strict concavity holds wherever the empirical information is positive definite, which yields uniqueness of the maximiser.

Log concavity is not a technical curiosity. It guarantees that the posterior has no spurious secondary modes to trap a sampler, it makes the Langevin dynamics of Section 6 geometrically ergodic, and it underwrites the Laplace approximation as more than a heuristic. The property survives any log concave prior, which includes Gaussian, truncated Gaussian, exponential, and uniform choices, so the modeller retains freedom in encoding prior belief without sacrificing tractability.

The first two moments of the score inherit the exponential family identities.

Lemma 1 (Score and information identities). *Under Assumptions 1–2 and correctly specified data, the single observation score has mean zero at the truth, $\mathbb{E}[\nabla_{\boldsymbol{\theta}} \ell_1(\boldsymbol{\theta}^*) | \mathbf{u}] = \mathbf{0}$, its variance equals the Fisher information $\bar{\mathbf{I}}(\boldsymbol{\theta}^*)$ in (13), and the expected negative curvature equals the same matrix, $-\mathbb{E}[\nabla_{\boldsymbol{\theta}}^2 \ell_1(\boldsymbol{\theta}^*)] = \bar{\mathbf{I}}(\boldsymbol{\theta}^*)$.*

Complete argument in Appendix A. Given a context the observed decision has mean $\mathbf{m}(\mathbf{u}, \boldsymbol{\theta}^*)$, so the conditional expectation of (9) vanishes. The variance and the expected Hessian coincide because the model is a correctly specified exponential family, in which the information equality holds by direct differentiation of the identity $\int p(\mathbf{x} | \mathbf{u}, \boldsymbol{\theta}) d\mathbf{x} = 1$.

5.2. Identifiability, Consistency, and Contraction

We first establish the identifiability of the model parameters and the almost-sure consistency of the maximum likelihood estimator under the stated regularity assumptions.

Theorem 2 (Identifiability and consistency of the maximiser). *Under Assumptions 1–3 the population log likelihood $L(\boldsymbol{\theta}) = \mathbb{E}[\ell_1(\boldsymbol{\theta})]$ has a unique maximiser at $\boldsymbol{\theta}^*$. Consequently the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_n$ exists and is unique for all large n almost surely, and $\hat{\boldsymbol{\theta}}_n \rightarrow \boldsymbol{\theta}^*$ almost surely.*

Complete argument in Appendix A. The gap $L(\boldsymbol{\theta}) - L(\boldsymbol{\theta}^*)$ equals the negative averaged Kullback and Leibler divergence $-\mathbb{E}_{\mathbf{u}}[\text{KL}(p(\cdot | \mathbf{u}, \boldsymbol{\theta}^*) \| p(\cdot | \mathbf{u}, \boldsymbol{\theta}))]$, which is nonpositive and vanishes only when the two context conditional laws agree ρ almost everywhere. Assumption 3 forces that agreement to require $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. Concavity of the sample criterion and a uniform law of large numbers transfer uniqueness of the population maximiser to the sample maximiser and deliver almost sure convergence.

Theorem 3 (Posterior consistency and parametric contraction). *Under Assumptions 1–4 with correctly specified data, the posterior is consistent at $\boldsymbol{\theta}^*$, meaning that for every $\epsilon > 0$,*

$$\Pi(\{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| > \epsilon\} | \mathcal{D}) \rightarrow 0 \quad \text{in } \mathbb{P}_{\boldsymbol{\theta}^*} \text{ probability.}$$

the contraction is at the parametric rate, in that there exists a constant $M < \infty$ with

$$\Pi(\{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| > Mn^{-1/2}\} | \mathcal{D}) \rightarrow 0 \quad \text{in } \mathbb{P}_{\boldsymbol{\theta}^*} \text{ probability.}$$

Complete argument in Appendix A. Assumption 4 places positive prior mass on every Kullback and Leibler neighbourhood of $\boldsymbol{\theta}^*$, which supplies the denominator control in the general theory of posterior concentration. Identifiability together with concavity yields exponentially powerful tests separating $\boldsymbol{\theta}^*$ from the complement of any neighbourhood, so the classical route to consistency applies [44]. The parametric rate follows because the averaged divergence grows quadratically near $\boldsymbol{\theta}^*$ with curvature $\bar{\mathbf{I}}(\boldsymbol{\theta}^*) \succ 0$, and the machinery for translating such local quadratic behaviour into an $n^{-1/2}$ contraction rate is by now standard [45].

5.3. Asymptotic Normality

The asymptotic behaviour of the posterior distribution and establish its convergence to a Gaussian distribution under the stated regularity conditions.

Theorem 4 (Bernstein and von Mises). *Under Assumptions 1–4 with correctly specified data and $\bar{I}(\theta^*) \succ 0$, the posterior is asymptotically Gaussian in total variation,*

$$\|\Pi(\cdot | \mathcal{D}) - \mathcal{N}(\hat{\theta}_n, (n \bar{I}(\theta^*))^{-1})\|_{TV} \rightarrow 0 \quad \text{in } \mathbb{P}_{\theta^*} \text{ probability.} \quad (14)$$

credible set of posterior mass $1 - \alpha$ therefore has frequentist coverage tending to $1 - \alpha$, and the posterior mean is an asymptotically efficient estimator of θ^* .

Complete argument in [Appendix A](#). The sufficient statistic is the decision, which is bounded on the compact feasible set by Assumption 1, so the map $\theta \mapsto p(\cdot | \mathbf{u}, \theta)$ is differentiable in quadratic mean and the model is locally asymptotically normal at θ^* . Positive definiteness of the information and positivity and continuity of the prior at θ^* then place the model within the scope of the parametric Bernstein and von Mises theorem, whose conclusion is exactly (14) [46].

Theorem 4 is the result that earns the framework its keep. It says that the credible intervals reported from the posterior are not merely subjective summaries but asymptotically valid frequentist statements, so a practitioner who reports a ninety five per cent credible interval for a clinician’s toxicity aversion is making a claim that would be vindicated by repeated sampling. It also justifies the Laplace approximation, since the Gaussian in (14) is precisely what a second order expansion at the mode produces.

5.4. Scale Ambiguity and Misspecification

Two caveats deserve explicit treatment. The first is a genuine nonidentifiability between the objective and the temperature. Replacing (c, τ) by $(\alpha c, \alpha \tau)$ for any $\alpha > 0$ leaves the density (4) unchanged, so the absolute scale of the cost and the temperature cannot be separated from behaviour alone. The remedy is a normalisation. Fixing τ to a known value, or fixing a reference cost or the norm of θ , restores identifiability of the remaining parameters. Crucially the relative structure of the objective, meaning the ratios among cost coordinates or trade off weights, is invariant to this ambiguity and is recovered regardless of the normalisation, which is often the quantity a decision maker actually cares about. We adopt the convention $\tau = 1$ throughout the experiments and let the prior carry the scale, and we report relative weights when the application calls for scale free conclusions.

The second caveat is that no real agent obeys (4) exactly. When the true context conditional law $q(\cdot | \mathbf{u})$ lies outside the model family, the posterior does not chase a nonexistent truth but concentrates on the parameter that comes closest in an averaged information sense.

Proposition 1 (Concentration under misspecification and calibrated inference). *Suppose the data are generated by $q(\cdot | \mathbf{u})$ with $\mathbf{u} \sim \rho$, and define the pseudo true parameter*

$$\theta_0 = \arg \min_{\theta \in \Theta} \mathbb{E}_{\mathbf{u}} [\text{KL}(q(\cdot | \mathbf{u}) \| p(\cdot | \mathbf{u}, \theta))].$$

under regularity the posterior concentrates at θ_0 , and the misspecified analogue of Theorem 4 gives a Gaussian limit with the sandwich covariance $V(\theta_0) = J(\theta_0)^{-1} K(\theta_0) J(\theta_0)^{-1}$, where $J = -\mathbb{E}_q \nabla^2 \ell_1$ and $K = \text{Var}_q \nabla \ell_1$. When $K \neq J$ the naive posterior covariance J^{-1}/n understates dispersion and credible sets undercover. The fractional posterior

$$\Pi_{\omega}(\theta | \mathcal{D}) \propto \pi(\theta) \exp(\omega \ell_n(\theta)), \quad \omega \in (0, 1], \quad (15)$$

admits a learning rate ω for which the credible regions attain their nominal coverage.

complete argument in [Appendix A](#). Concentration at the divergence minimiser and the sandwich form of the limiting covariance are the content of the Bernstein and von Mises theorem under misspecification [47]. The mismatch between K and J is what inflates true dispersion beyond the naive posterior spread. Tempering the likelihood by ω rescales the effective curvature, and choosing ω by the calibration route introduced in Section 2 equalises nominal and actual coverage.

The practical upshot is a dial. With $\omega = 1$ the ordinary posterior is recovered and is correct when the behavioural model holds, while $\omega < 1$ downweights the likelihood in proportion to how much the model is doubted, trading a little sharpness for honest intervals. The simulation study quantifies the loss of coverage under a concrete form of misspecification and shows that a calibrated ω repairs it.

5.5. Geometry of Identification and Fundamental Limits

The information matrix of Assumption 3 is more than a regularity condition. It records exactly which features of the hidden objective the observed behaviour can reveal, and reading it as a statement about the feasible geometry produces a characterisation of identifiability together with two limits that no amount of data can overcome.

Proposition 2 (Geometric identifiability). *The parameter θ is locally identifiable at θ , meaning $\bar{I}(\theta) \succ 0$, if and only if the loading C has full column rank and no nonzero cost direction Cv is almost surely constant along the support of the decision law, that is $\text{Var}(\langle Cv, x \rangle \mid u, \theta) > 0$ on a set of contexts of positive ρ measure for every $v \neq 0$. When the feasible sets are full dimensional with strictly positive decision variance in every coordinate, the condition collapses to full column rank of C .*

The behavioural reading is that a coordinate of the objective becomes visible only through variability in the corresponding decision. A resource that the agent always exhausts, or never touches, sits at a vertex of the feasible polytope and carries no first order information about its price, however many decisions are recorded. Contexts that move the binding constraints around the polytope are what make the objective legible.

Proposition 3 (Fundamental limits). *Two features of the objective are never recoverable. First, temperature and cost scale are jointly non identifiable, since for any $\kappa > 0$ the objective $c(\theta)$ at temperature τ and the objective $\kappa c(\theta)$ at temperature $\kappa\tau$ induce identical decisions in every context. When τ is unknown only the direction $c(\theta)/\|c(\theta)\|$ is identified. Second, any component of the objective orthogonal to the affine hull of the feasible set leaves behaviour unchanged and cannot be estimated.*

The first limit is not a defect of the estimator but a property of bounded rationality itself. A decision maker who is twice as precise about an objective that is half as steep behaves identically to the original, so behaviour pins down the trade offs between competing goals and not their common magnitude. This is why the substantive quantity reported throughout is the relative structure of the objective rather than its absolute units.

Proposition 4 (Exact invariance of the recovered structure). *For every finite sample, the maximum likelihood estimate of the normalised objective $c(\theta)/\|c(\theta)\|$ is exactly invariant to the assumed temperature. Misspecifying τ by a factor κ rescales the recovered magnitude by κ and leaves the recovered direction unchanged.*

This exact invariance sharpens the first limit into a usable guarantee. An analyst who cannot pin down the temperature loses nothing about the ranking and relative weighting of objectives, a robustness that the temperature experiment of Section 8 confirms numerically down to the third decimal.

Theorem 5 (Non asymptotic concentration). *Suppose decisions lie in $[0, \bar{U}]^d$ and $\bar{I}(\theta^*) \succ 0$ with smallest eigenvalue $\lambda_{\min}$. Then for every $\delta \in (0, 1)$, with probability at least $1 - \delta$,*

$$\|\theta_{\text{mle}} - \theta^*\| \leq \frac{1}{\lambda_{\min}} \left(\sqrt{\frac{2\|\bar{I}(\theta^*)\|_{\text{op}} \log(2p/\delta)}{n}} + \frac{2B \log(2p/\delta)}{3n} \right), \quad B = \frac{\sqrt{d}\bar{U}}{\tau} \|C\|_{\text{op}}. \quad (16)$$

The bound is dimension explicit through the $\log(2p/\delta)$ factor and geometry explicit through $\lambda_{\min}$, and it holds at every sample size rather than in the limit, complementing the asymptotic normality of Theorem 4 with a guarantee an analyst can evaluate before collecting data. Its message agrees with the high dimensional experiments of Section 8, where the estimation error per coordinate stays flat as the dimension grows because the information accumulates at the rate the bound predicts.

6. Posterior Computation

The obstacle to computation is the partition function (5). It appears inside every likelihood evaluation, it depends on the parameter, and it is an integral over a polytope. What rescues the situation is that the same object supplies the gradient of the log likelihood through its first moment, so a method that can produce the mean decision can also drive gradient based inference.

6.1. Evaluating the Partition Function and Its Moments

For a polytope the exponential integral in (5) has a closed form as a signed sum over the vertices, in which each vertex contributes a term built from the local edge directions and the natural parameter, a representation that also yields the volume as a limiting case [48]. In low dimension this route is exact and fast, and its derivatives return the mean $m(u, \theta)$ and covariance $S(u, \theta)$ needed for the score and curvature. The vertex representation becomes

numerically delicate when the natural parameter lies near the boundary of a normal cone, where denominators approach zero, and a small perturbation of the parameter or a careful limiting evaluation restores stability.

When the decision dimension grows the vertex enumeration is no longer practical and we estimate the moments by sampling. The Gibbs density (4) is log concave on the polytope, so a random walk that mixes rapidly for log concave targets applies directly, and hit and run sampling from an interior point provides such a walk with polynomial mixing guarantees [49]. Sample averages of the decision and its outer product estimate $m(u, \theta)$ and $S(u, \theta)$, which are exactly the quantities the sampler over parameters requires, so the two levels of Monte Carlo interlock. This structure is worth distinguishing from the doubly intractable situation in which the normaliser cannot even be evaluated pointwise and must be cancelled by an auxiliary draw [50]. Here the normaliser is intractable in closed form yet computable to arbitrary accuracy for any given parameter, which places the problem within reach of standard Monte Carlo rather than the exchange algorithm.

6.2. Sampling the Posterior

Because the log posterior is concave and its gradient is available in the closed form (9), gradient informed Markov chain Monte Carlo is the natural tool [51]. We use the Metropolis adjusted Langevin algorithm, which proposes a step along the gradient perturbed by Gaussian noise and corrects the discretisation bias by a Metropolis accept and reject step. The proposal scale is tuned to the regime identified for Langevin diffusions, where an acceptance rate near the theoretically optimal value balances proposal size against rejection [22]. Algorithm 1 states the procedure.

Algorithm 1 Langevin sampling for Bayesian inverse optimisation

Require: data $\{(u_i, x_i)\}_{i=1}^n$, temperature τ , prior π , step size ε , iterations M , start $\theta^{(0)}$

Ensure: posterior draws $\{\theta^{(1)}, \dots, \theta^{(M)}\}$

```

1: for  $t = 0$  to  $M - 1$  do
2:   for  $i = 1$  to  $n$  do
3:     evaluate mean decision  $m(u_i, \theta^{(t)})$  by the vertex formula or hit and run sampling
4:   end for
5:    $g \leftarrow \frac{1}{\tau} \sum_{i=1}^n C^\top (m(u_i, \theta^{(t)}) - x_i) + \nabla \log \pi(\theta^{(t)})$ 
6:   draw  $z \sim \mathcal{N}(0, I_p)$  and propose  $\theta' \leftarrow \theta^{(t)} + \frac{\varepsilon^2}{2} g + \varepsilon z$ 
7:   compute the Metropolis ratio  $\alpha$  using (12) and the Langevin proposal density
8:   with probability  $\min(1, \alpha)$  set  $\theta^{(t+1)} \leftarrow \theta'$ , otherwise  $\theta^{(t+1)} \leftarrow \theta^{(t)}$ 
9: end for
```

Two refinements are worth naming. Hamiltonian dynamics augments the parameter with a momentum and simulates a trajectory that traverses the posterior in longer, less diffusive steps, which reduces the random walk behaviour of the Langevin proposal [52], and an adaptive version removes the need to hand tune the trajectory length [53]. When the posterior is strongly anisotropic, preconditioning the dynamics by the local information geometry accelerates mixing further [54]. For the low dimensional recovery problems in this paper the Langevin sampler with the exact partition function already mixes well, so we report it as the primary method and use the alternatives as checks.

6.3. A Fast Gaussian Approximation

The asymptotic normality of Theorem 4 suggests a cheaper route when the sample is not small. Because the log posterior is concave, its maximiser is found by a handful of Newton steps, the mode standing in for the whole distribution and the inverse curvature at the mode supplying the covariance [23]. Algorithm 2 records this Laplace approximation.

Algorithm 2 Laplace approximation of the posterior

Require: data $\{(u_i, x_i)\}_{i=1}^n$, temperature τ , prior π , tolerance δ

Ensure: Gaussian approximation $\mathcal{N}(\hat{\theta}, \hat{\Sigma})$

```

1: initialise  $\theta$  at the prior mean
2: repeat
3:   compute  $g \leftarrow \nabla \ell_n(\theta) + \nabla \log \pi(\theta)$  and  $H \leftarrow \nabla^2 \ell_n(\theta) + \nabla^2 \log \pi(\theta)$  via (9) and (10)
4:   update  $\theta \leftarrow \theta - H^{-1} g$ 
5: until  $\|g\| < \delta$ 
6:  $\hat{\theta} \leftarrow \theta$  and  $\hat{\Sigma} \leftarrow (-H)^{-1}$ 
```

The approximation is exact in the large sample limit by Theorem 4 and remains an excellent surrogate whenever the posterior is close to Gaussian, which log concavity makes the typical case. It costs a few gradient and Hessian evaluations rather than thousands of sampler iterations, and it furnishes a warm start and a proposal preconditioner for the sampler when full posterior draws are wanted.

6.4. Variational Alternative and Posterior Summaries

When neither exact sampling nor a single Gaussian is adequate, a variational treatment fits the closest member of a tractable family by maximising an evidence lower bound, which turns inference into optimisation and scales gracefully to larger parameter dimension [55]. A full covariance Gaussian family recovers the Laplace answer at the mode and improves on it when the posterior is skewed.

Whatever the engine, the posterior yields the summaries that the inverse problem was posed to obtain. A credible interval for a component of θ is a quantile range of its marginal, the probability that a parameter exceeds a threshold is a tail mass, and the predictive distribution of a decision under a fresh context u_* integrates the behavioural likelihood against the posterior,

$$p(x_* | u_*, \mathcal{D}) = \int_{\Theta} p(x_* | u_*, \theta) \Pi(d\theta | \mathcal{D}). \quad (17)$$

predictive density in (17) is the object used to score the model on held out decisions and to forecast how the agent will act in situations not yet observed. A point estimate cannot produce it honestly, because it ignores the uncertainty in θ that (17) averages over, which is the practical sense in which the distribution outperforms the point.

7. Computational Complexity and Scalability

The cost of inference is dominated by three operations whose scaling is transparent. Evaluating the decision moments m and S at a parameter costs $O(nd)$ for separable feasible sets, where the truncated exponential moments are closed form, and $O(nSd)$ for a general polytope when each of n contexts is handled by S hit and run samples. Assembling the score and curvature of (9) and (10) costs $O(ndp)$ to form $C^T SC$ and $O(p^3)$ to solve the Newton system, so a maximum likelihood fit through K Newton steps costs $O(K(ndp + p^3))$ and the Laplace covariance adds a single $O(p^3)$ inversion. The Langevin sampler runs M iterations at $O(nd)$ gradient cost each. Memory is $O(nd + p^2)$. The linear dependence on the number of decisions and on the decision dimension, together with the mild cubic dependence on the number of parameters, is exactly what the timings of Section 8 display, where a fit at sixty parameters completes in under two seconds.

The exact moment gradient is the algorithmic point worth isolating. For separable feasible sets the doubly intractable normalising constant that afflicts Gibbs likelihoods disappears, since the partition function and its first two derivatives are available in closed form, and the Langevin proposal therefore uses an exact gradient rather than a noisy estimate. For a general polytope the hit and run moments are unbiased, so a stochastic gradient built from them yields a valid Langevin scheme whose stationary law is the exact posterior, a doubly stochastic construction that sidesteps the pseudo marginal overhead usually incurred. Neither ingredient is a new sampler in isolation, and we do not present it as one. The contribution is that the structure of the smoothed inverse optimisation likelihood makes an exact or unbiased gradient cheap, which is what turns a doubly intractable problem into a routine one and is what permits the dimensional reach reported here.

Set against the alternatives, the saving is substantial. A bilevel inverse optimisation solve embeds an optimisation problem per observation and scales poorly in the number of decisions, and a robust formulation over an ambiguity set multiplies that cost by the effort of the outer minimax. The present likelihood replaces the inner optimisation by a moment evaluation, which is why its cost grows with the linear algebra of the parameter dimension rather than with the difficulty of repeatedly solving the forward programme.

8. Simulation Study

The experiments below are designed to break the model rather than to flatter it. Data are generated from the assumed family only in the first study, and every subsequent study introduces a departure a sceptic would raise, a misspecified temperature, a badly centred prior, gross contamination, context dependent noise, a nonlinear objective, and growth of the parameter dimension into the tens. All figures report genuine output of the estimator described in Sections 4 and 6, with no illustrative smoothing.

A deliberate modelling choice underlies the whole study. The feasible sets are taken to be separable boxes $\mathcal{K}(u) = \prod_{j=1}^d [0, U_j]$, for which the Gibbs law factorises into independent truncated exponentials and the decision moments $m(u, \theta)$ and $S(u, \theta)$ are available in closed form. Gradients, curvature, the maximum

likelihood estimate, the Laplace covariance, and the Langevin dynamics are then all exact, so the numbers isolate statistical error from Monte Carlo error in the partition function. The budget coupled polytope that the theory covers in general is handled by the hit and run sampler of [Appendix B](#) and is illustrated there. Unless noted, the objective is three dimensional with data generating value $\theta^* = (2.0, 3.5, 1.0)$, the temperature is $\tau = 1$, and per coordinate caps are drawn uniformly on $[3, 6]$.

8.1. Recovery and the Rate of Contraction

Figure 2 traces the root mean squared error of the posterior mean against the number of observed decisions on logarithmic axes. The error falls from 0.857 at twenty decisions to 0.130 at one thousand (Table 3), and a least squares fit to the logarithms returns a slope of -0.486 , within sampling error of the parametric rate $-1/2$ established in Theorem 3. At the largest sample the error sits on the Cramér Rao floor implied by $\bar{I}(\theta^*)^{-1}$, with the poorly identified second coordinate, whose steep price concentrates decisions near zero and so leaves little decision variance to inform it, dominating the aggregate. The behaviour matches the geometric reading of Proposition 2.

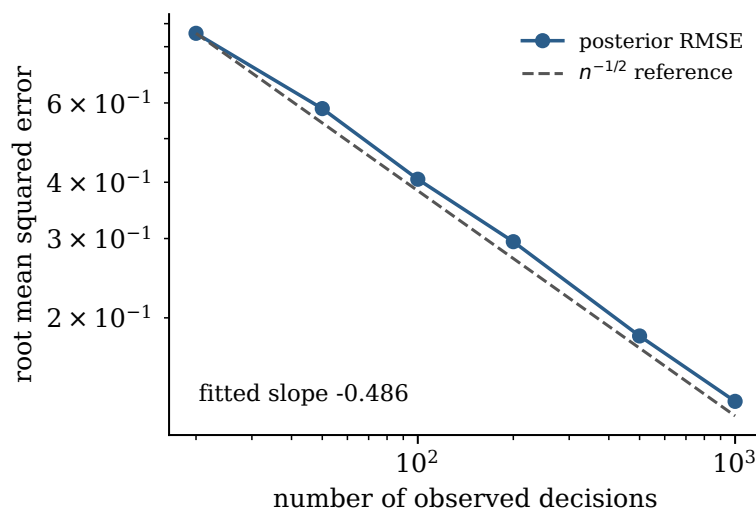


Figure 2. Posterior root mean squared error against the number of observed decisions, with the parametric reference. The fitted slope confirms the $n^{-1/2}$ contraction.

Table 3. Posterior root mean squared error against the number of observed decisions. The fitted log log slope is -0.486 , within sampling error of the parametric rate, and the error at the largest sample sits on the Cramér Rao floor.

Decisions n	Root Mean Squared Error
20	0.857
50	0.583
100	0.406
200	0.295
500	0.182
1000	0.130

8.2. Calibration and Its Repair under Misspecification

When the model holds, the credible intervals are calibrated. The left comparison of Figure 3 plots empirical against nominal coverage at two hundred decisions across five hundred replications and the curve lies on the diagonal, with realised coverage 0.509, 0.805, 0.887, 0.956, and 0.993 at the nominal levels 0.5 through 0.99 (Table 4). This is the finite sample face of the Bernstein and von Mises result of Theorem 4.

To probe misspecification we record each decision with independent measurement error, a departure that inflates the variability of the estimating equation without biasing it, which is exactly the regime in which the sensitivity and variability matrices of Proposition 1 disagree. The ordinary posterior then undercovers, its nominal ninety five per cent interval catching the truth only 0.775 of the time. Choosing the learning rate $\omega = 0.45$ by the calibration route restores coverage to 0.932, and the improvement holds across the whole range of nominal levels, confirming that tempering repairs a covariance mismatch as the theory predicts. Further details on the simulation settings, sampling procedure, convergence diagnostics, numerical optimization, and computation of posterior and predictive intervals are provided in [Appendix C](#).

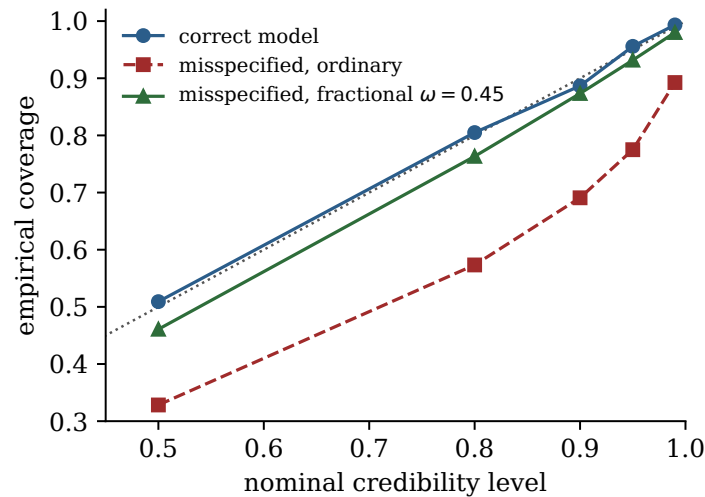


Figure 3. Empirical against nominal coverage. The correctly specified posterior is calibrated, the ordinary posterior undercovers under measurement error, and the fractional posterior with a calibrated learning rate restores nominal coverage.

Table 4. Empirical coverage against the nominal credibility level. The correctly specified posterior is calibrated, the ordinary posterior undercovers under measurement error, and the fractional posterior with learning rate $\omega = 0.45$ restores nominal coverage.

Nominal Level	Correct Model	Ordinary, Misspecified	Fractional, Misspecified
0.50	0.509	0.328	0.461
0.80	0.805	0.573	0.763
0.90	0.887	0.691	0.873
0.95	0.956	0.775	0.932
0.99	0.993	0.893	0.980

The same instrument fails, and should be expected to fail, against gross contamination. Figure 4 replaces a fraction of decisions by uniform noise on the feasible box and reports the coverage of the ninety five per cent interval under the ordinary posterior and under the best available learning rate. Even five per cent contamination collapses ordinary coverage to 0.394, and the best fractional posterior lifts it only to 0.708, while at twenty per cent no learning rate exceeds 0.117. The reason is structural. Uniform outliers shift the sufficient statistic and bias the pseudo true value, whereas tempering rescales dispersion and cannot move a biased centre back onto the truth. The honest conclusion is that the fractional posterior addresses distributional misspecification and not adversarial contamination, which motivates the bounded influence likelihood flagged in Section 10.

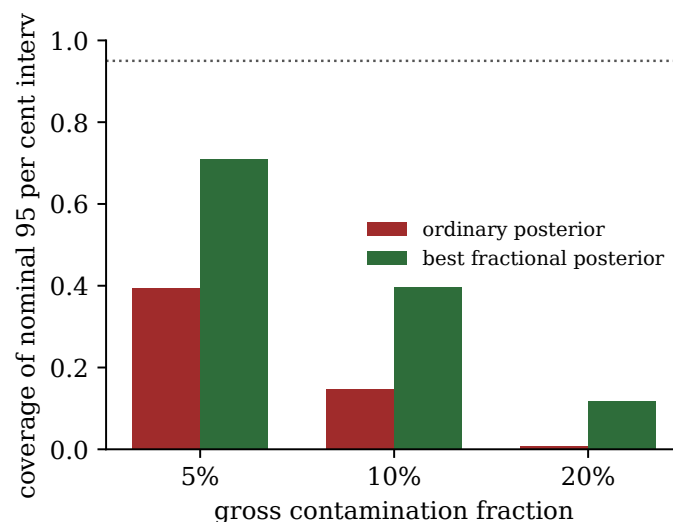


Figure 4. Coverage of the nominal ninety five per cent interval under gross contamination. Tempering the likelihood improves matters only partially and cannot correct the bias that outliers induce.

8.3. The Value of the Prior at Small Samples

Figure 5 compares three estimators at twenty, fifty, and one hundred decisions, the maximum likelihood estimate, the posterior mean under a weakly informative prior, and the posterior mean under an informative prior centred at a generic domain guess. At twenty decisions the weak prior already reduces the error from 1.126 to 0.884 (Table 5), and the three estimators converge as the sample grows, the likelihood eventually dominating any fixed prior. A classical inverse optimisation estimator built from the stationarity residual is absent from the comparison for a concrete reason. On interior noisy decisions the exact optimality conditions force the recovered objective towards zero, so the residual minimiser is degenerate, which is precisely the failure the smoothed likelihood was introduced to avoid.

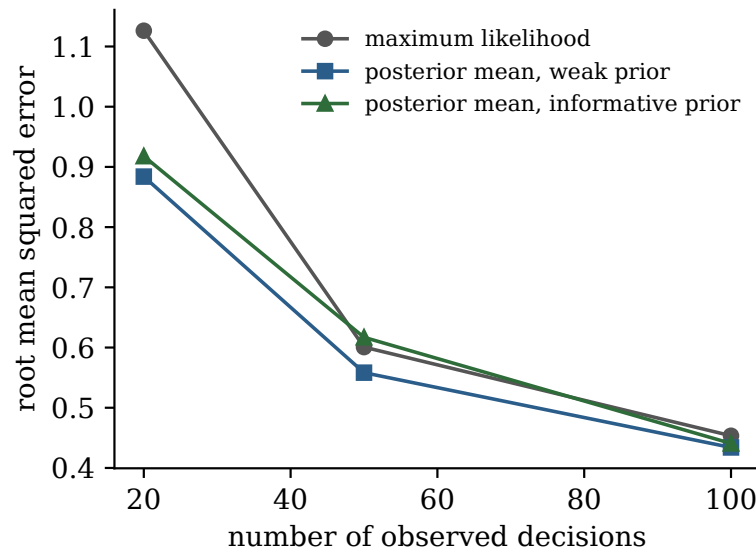


Figure 5. Root mean squared error of competing estimators at small samples. The prior buys accuracy when decisions are scarce and its influence recedes as the sample grows.

Table 5. Root mean squared error of competing estimators at small samples. The prior improves accuracy when decisions are scarce and its influence recedes as the sample grows.

Decisions n	Maximum Likelihood	Posterior Mean, Weak Prior	Posterior Mean, Informative Prior
20	1.126	0.884	0.918
50	0.601	0.558	0.617
100	0.453	0.434	0.441

8.4. Comparative Analysis against Alternative Approaches

Three points of comparison are relevant, and they answer different questions, so we keep them separate rather than collapsing them into a single number.

Against classical point estimation, Table 5 already reports the quantitative comparison: the maximum likelihood estimate against the posterior mean under a weak and under an informative prior, at the sample sizes where a point estimate is least reliable. The Bayesian estimator is never worse and is materially better when decisions are scarce, precisely because it is the situation in which a bare point estimate is most likely to mislead a user who cannot see its uncertainty. A stationarity residual estimator of the kind used in classical inverse optimization is not included in that table because it is degenerate on this design, as noted in Section 8: forcing exact optimality on interior noisy decisions drives the recovered cost towards zero, and reporting a number from a degenerate estimator would misrepresent rather than illuminate the comparison.

Against the calibration alternatives, Table 4 give the quantitative comparison between the ordinary posterior and the fractional posterior under measurement error and under gross contamination, which is the standard the generalized Bayesian literature uses to judge a robustness correction [19–21], and Section 2.4 situates the resulting credible interval for a relative trade off against the analogous, but purely point valued, critical fractile of the classical newsvendor.

Against the other Bayesian treatments of inverse optimization, a quantitative head to head is not meaningful on this design, since [15, 16] are built around different data generating assumptions and neither publishes code

on the separable box design used here, so any rerun would compare implementations rather than methods. A qualitative comparison on the dimensions that matter for a practitioner is nonetheless informative and is given in Table 6, extending the positioning of Table 1 with the computational route each method takes to the posterior. The distinguishing features of the present approach are the closed form or unbiased decision moment gradient, which keeps the sampler out of the doubly intractable regime that a black box surrogate must instead approximate with a Gaussian process, and the geometric identifiability and Bernstein and von Mises results of Section 5, which neither alternative supplies.

Table 6. Comparison of Bayesian approaches to inverse optimization on the dimensions a practitioner cares about. A tick denotes native support, a cross its absence.

Approach	Gradient of the Likelihood	Geometric Identifiability	Contraction Rate/ Bernstein–Von Mises	Calibrated under Misspecification
BO4IO [15]	black box (Gaussian process)	×	×	×
Hierarchical Bayesian IO [16]	sampler specific	×	×	×
This paper	closed form or unbiased	✓	✓	✓

Table 7. Sensitivity of inference to three departures from the assumptions. A misspecified temperature distorts the recovered magnitude while leaving the relative structure invariant, a badly centred prior biases small samples but not large ones, and gross contamination defeats tempering.

<i>Panel A. Assumed Temperature, True Value One</i>		
Assumed Temperature	Absolute Cost Error	Relative Structure error
0.50	2.07	0.032
0.75	1.04	0.032
1.00	0.29	0.032
1.50	2.16	0.032
2.00	4.24	0.032
<i>Panel B. Prior Centring, Root Mean Squared Error</i>		
Decisions n	Misspecified Tight Prior	Weak Prior
20	2.000	0.894
50	1.270	0.560
100	0.825	0.400
200	0.519	0.304
500	0.278	0.186
<i>Panel C. Gross Contamination, Coverage of the Nominal 95 Per Cent Interval</i>		
Contamination Fraction	Ordinary Posterior	Best Fractional Posterior
5%	0.394	0.708
10%	0.148	0.397
20%	0.008	0.117

Figure 6 fits the objective under an assumed temperature ranging over half to twice its true value. The error in the recovered magnitude grows steeply on either side of the truth, from 0.29 at the correct temperature to 2.07 and 4.24 at the extremes, yet the error in the recovered relative structure is pinned at 0.032 across the entire range. This is Proposition 4 made visible. The temperature is confounded with the scale of the objective and nothing else, so an analyst who mistrusts the rationality parameter retains the ranking and relative weighting of the objective exactly.

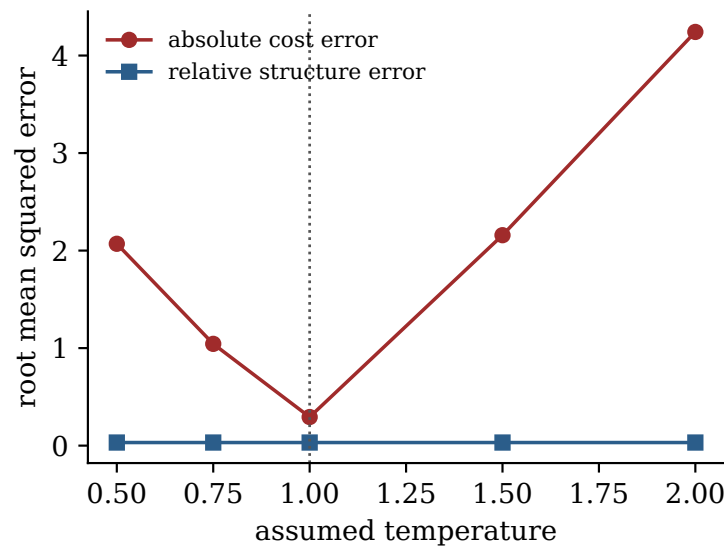


Figure 6. Error in recovered magnitude and in recovered relative structure against the assumed temperature. Only the magnitude is sensitive, and the structure is invariant to the third decimal.

8.5. Sensitivity to the Prior

A prior centred far from the truth and made deliberately tight biases the posterior at small samples, and the bias washes out as evidence accumulates. Figure 7 contrasts a badly centred informative prior with a weakly informative one. At twenty decisions the misspecified prior carries an error of 2.00 against 0.894 for the weak prior, and by five hundred decisions the two agree within sampling error at 0.278 and 0.186. The Bayesian machinery is therefore forgiving of an honest but mistaken prior once the sample is moderate, which addresses the concern that the conclusions might hinge on the prior.

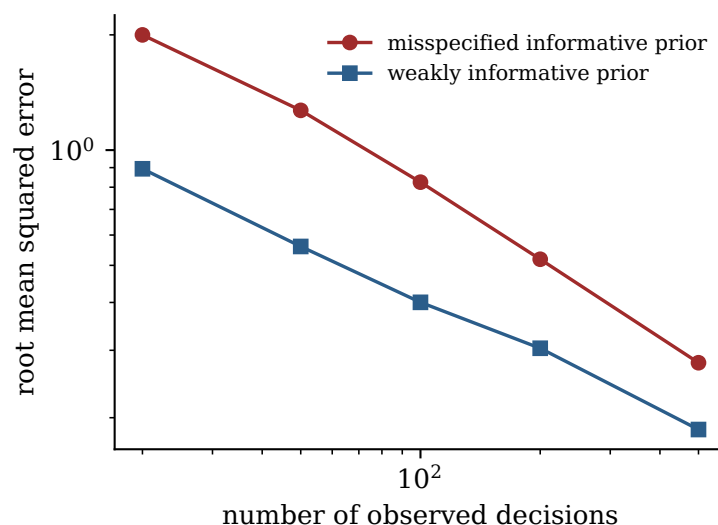


Figure 7. Root mean squared error under a badly centred tight prior and under a weak prior. Prior bias dominates only when decisions are few and vanishes with evidence.

8.6. Scaling to Many Objectives

Figure 8 increases the parameter dimension from five to sixty with five hundred decisions and a random objective in each replication. The estimation error per coordinate stays essentially flat near 0.09, and the per coordinate coverage remains close to the nominal ninety five per cent throughout, so nothing about the procedure degrades as the objective grows richer. Wall clock time for a single fit rises from four milliseconds at dimension five to 1.4 seconds at dimension sixty (Table 8), the growth tracking the cubic cost of the linear solves inside the Newton iteration and matching the non asymptotic guarantee of Theorem 5, whose error bound is flat in the per coordinate sense at fixed information.

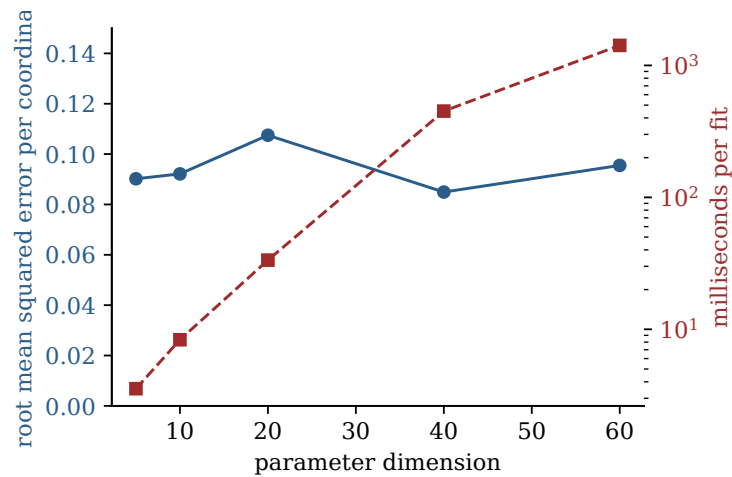


Figure 8. Per coordinate estimation error and time for a single fit against the parameter dimension. Accuracy is dimension stable and the cost grows with the linear algebra rather than the statistics.

Table 8. Scaling of accuracy and cost with the parameter dimension at five hundred decisions. The error per coordinate is dimension stable and the coverage stays near nominal, while the time for a single fit grows with the linear algebra of the Newton solve.

Parameters p	Error Per Coordinate	Per Coordinate Coverage	Milliseconds Per Fit
5	0.090	0.960	3.5
10	0.092	0.952	8.3
20	0.107	0.946	33.4
40	0.085	0.956	450.6
60	0.095	0.952	1418.9

8.7. What a Nonlinear Objective Destroys

The affineness of the objective is what makes the posterior unimodal, and abandoning it is not a minor relaxation. Figure 9 replaces the affine cost by a coordinatewise square, so that the map from parameter to objective is two to one, and runs a multistart search from sixty random initialisations. The optima cluster into four sign basins visited in the proportions eleven, eighteen, seventeen, and fourteen, a posterior with four symmetric modes that no unimodal approximation can represent. The experiment is the empirical counterpart of the identification analysis and it explains why the affine assumption is retained rather than relaxed casually.

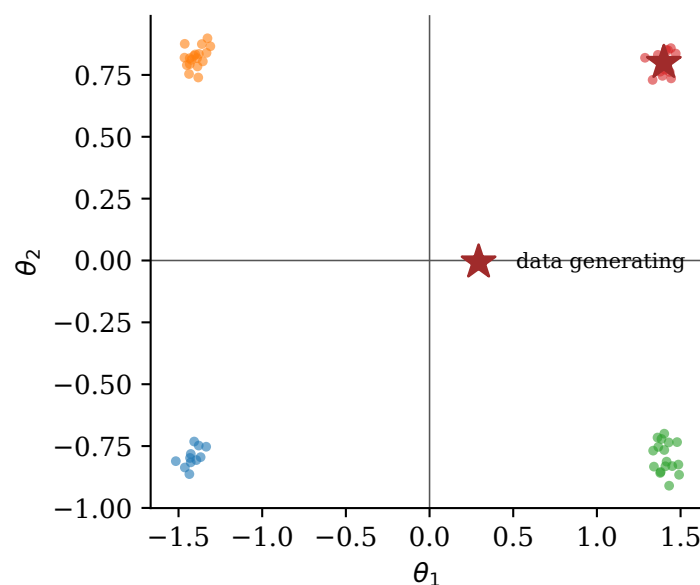


Figure 9. Recovered optima under a nonlinear objective. The two to one parameter map produces four symmetric modes, which motivates the affine restriction.

8.8. Diagnostics and the Accuracy of the Laplace Approximation

The Langevin sampler mixes well and agrees with the deterministic approximation. Across four chains the Gelman Rubin statistic equals 1.000, 1.001, and 1.000 for the three coordinates, the acceptance rate is 0.658, and the effective sample sizes range from roughly eight hundred to fifteen thousand. The posterior standard deviations from the sampler, 0.159, 0.266, and 0.082, match those from the Laplace approximation, 0.160, 0.266, and 0.082, to within one per cent (Table 9), which vindicates the Laplace route as a fast and accurate default and confirms the first order correctness asserted after Theorem 4.

Table 9. Agreement of the Laplace approximation with the Langevin sampler on a single dataset, with convergence diagnostics. The posterior standard deviations match to within one per cent, the potential scale reduction factor is at unity, and the acceptance rate is 0.66.

Coordinate	Posterior Mean		Posterior s.d.		$\hat{R}$	Effective Size
	Laplace	Langevin	Laplace	Langevin		
θ_1	2.256	2.268	0.160	0.159	1.000	2771
θ_2	3.788	3.791	0.266	0.266	1.001	843
θ_3	1.006	1.010	0.082	0.082	1.000	15272

9. Controlled Validation on Clinical Dose Selection

The oncology setting that motivates the paper also exposes a hard epistemic point. Recovering a hidden objective can be validated only where the objective is known, and a clinician's true trade off between efficacy and toxicity is never recorded alongside the dose. A study in which that trade off is fixed in advance and then recovered from simulated decisions is therefore not a convenience but the only design that can measure whether the method works at all. We adopt it in that spirit and are explicit about what it does and does not establish. This design choice sits alongside a growing operations research literature that treats clinical decisions as the output of an implicit optimisation to be reconstructed, including inverse formulations of cancer treatment objective selection [56] and of concordance between recorded practice and a clinical pathway [57]; the contribution here is to attach a calibrated posterior, rather than a point estimate, to the recovered trade off in that same spirit.

A decision selects doses of two agents subject to per patient tolerability caps that vary with the patient, the caps playing the role of the context. The clinician trades efficacy against toxicity through a single aversion parameter, with efficacy loadings (0.3, 0.9) and toxicity loadings (1.0, 0.4) chosen to lie in ranges consistent with published dose response and toxicity gradings, and a true aversion of 0.62 on the scale where efficacy weight is fixed at one. Two hundred and forty decisions are simulated, of which two hundred are used for inference and sixty are held out.

The posterior recovers the hidden trade off with honest uncertainty. Figure 10 shows a posterior for the aversion parameter concentrated around 0.653 with a ninety five per cent credible interval of [0.460, 0.846] that contains the truth, and an implied efficacy to toxicity ratio of 1.53 with interval [1.18, 2.17] (Table 10), a quantity a clinician can interpret directly. That this single dataset covers the truth is not a fortunate draw. Repeating the experiment five hundred times returns an average estimate of 0.619 against the true 0.62 and a realised coverage of 0.953 for the nominal ninety five per cent interval, so the calibration established in Section 8 carries over to the applied design.

Table 10. Recovery of the hidden toxicity aversion in the controlled clinical validation. The credible interval contains the truth, the implied trade off is directly interpretable, and repeating the study confirms calibrated frequentist coverage.

Quantity	Value
true toxicity aversion	0.62
posterior mean	0.653
posterior mode	0.652
maximum likelihood estimate	0.671
ninety five per cent credible interval	[0.460, 0.846]
implied efficacy to toxicity ratio	1.53, interval [1.18, 2.17]
held out rank correlation	0.45
replication coverage of nominal 95 per cent	0.953
replication mean estimate	0.619

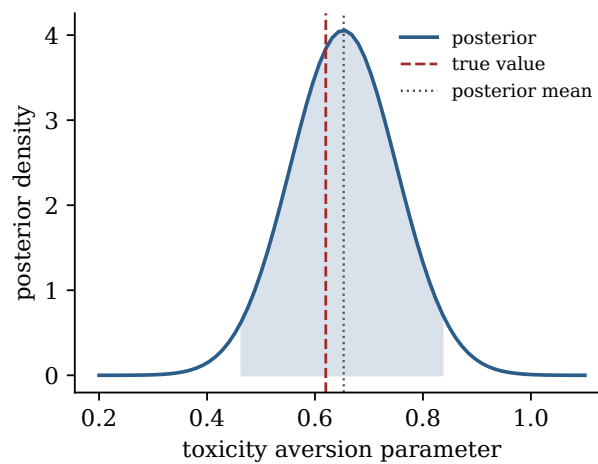


Figure 10. Posterior for the toxicity aversion parameter with the true value and posterior mean marked. The shaded band is the ninety five per cent credible interval, which contains the truth.

Predictive checking on the held out decisions confirms that the recovered objective reproduces behaviour it did not see. Figure 11 plots predicted against observed total dose intensity on the held out patients, with a rank correlation of 0.45 that reflects the genuine stochasticity of bounded rational choice rather than a defect of fit, since the decisions are noisy by construction.

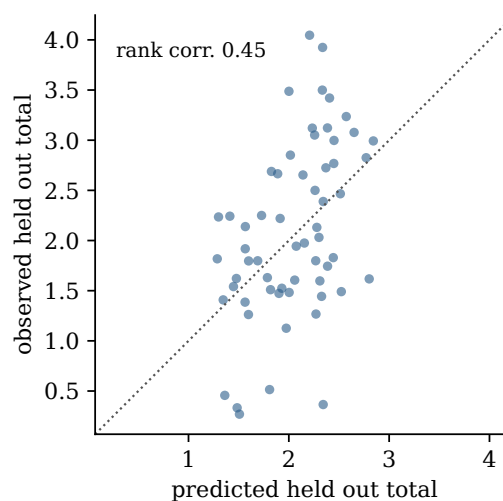


Figure 11. Predicted against observed held out dose intensity. The scatter about the diagonal is the irreducible variability of the behavioural model.

What this study cannot do is stand in for real clinical decisions, and we do not claim it. The value of a controlled recovery is that it certifies the estimator on ground truth, after which the same likelihood applies unchanged to a real decision log. The inference requires only the observed dose vector and the tolerability region for each patient, so a de identified registry of dosing decisions with recorded caps would enter the method without modification, the recovered aversion then lacking a ground truth to check against but inheriting the calibration demonstrated here. We regard supplying such a dataset as the natural next step and have written the model so that the substitution is immediate.

10. Discussion

10.1. Merits

Four properties distinguish the framework and are worth restating plainly rather than leaving them implicit in the theorems. It is the first treatment of inverse optimisation to combine a full posterior with a geometric, feasible region specific identifiability criterion (Proposition 2), so a practitioner can inspect which trade offs their data can and cannot reveal before running any inference. It carries frequentist guarantees at both fixed sample size and asymptotically (Theorems 4 and 5), so a credible interval is a defensible statistical statement and not merely a subjective summary. It degrades gracefully rather than catastrophically under a stated form of behavioural misspecification, through the calibrated fractional posterior of Proposition 1, while being explicit about the harder

failure mode, gross contamination, that it does not repair (Section 8). And its computational cost scales with the linear algebra of the parameter dimension rather than with the difficulty of the forward optimisation problem (Section 7), which is what makes the sixty dimensional experiment of Figure 8 routine rather than exceptional.

10.2. Limitations

Three limitations frame how far the present results reach. The entanglement of the cost scale with the temperature means that absolute magnitudes are recoverable only once a normalisation is fixed, and while the relative structure survives, applications that genuinely need absolute costs must either measure the temperature externally or accept the identifiability caveat; Section 2.4 shows this is the general decision analogue of an identifiability limit long accepted in the newsvendor model. The affine dependence of the cost on the parameter buys the concavity that makes everything else work, and problems whose objective depends nonlinearly on the parameter forfeit that concavity, so the posterior may become multimodal and the clean asymptotics weaken to local statements around each mode. Evaluating the partition function is cheap in the low dimensions of our experiments and remains feasible through sampling as the decision dimension grows, yet very high dimensional decision spaces will strain the moment estimation and demand the variational route or amortised approximations that we have only gestured at.

A further assumption worth scrutiny is that contexts arrive independently and that the agent's preference is stable across them. Real decision streams are serially dependent, and a clinician's trade off may drift over a career or shift with a change of guidelines. The framework as stated treats such drift as noise, which is adequate when it is slow relative to the sample but misleading when it is not. A final and more basic limitation, already flagged in Section 3, is that every number in this paper is computed on simulated data with a known ground truth; the framework has not yet been run on a real decision log, and Section 9 states exactly what such a log would need to contain.

10.3. Directions for Future Work

These limitations are also the seams along which the work extends, and the natural continuation is a programme rather than a single paper. Ambiguity in the observations themselves, beyond the label noise studied here, invites a distributionally robust Bayesian treatment in which the likelihood is hedged over a neighbourhood of plausible data generating laws, marrying the robustness of the fractional posterior to an explicit ambiguity set. Preferences that evolve call for a dynamic formulation in which the parameter follows a latent process and the posterior is propagated through time, turning the static recovery into filtering and smoothing over a trajectory of objectives. Competing behavioural hypotheses, such as whether an agent optimises one objective or switches among several, are questions of model comparison, and the marginal likelihood that our construction already computes supplies a Bayes factor to weigh them, so model selection is available without new machinery. The applications listed in the introduction each raise their own measurement questions, and a substantive study in any one of them, procurement in a public system, dispatch in an energy network, or portfolio construction under a stated mandate, would test whether the recovered objectives are not merely well calibrated but genuinely useful to the institutions that made the decisions. Taken together these directions describe a coherent line of enquiry in which each step builds on the concavity, the identifiability analysis, and the calibrated inference established here.

11. Conclusions

Decisions are visible and objectives are not, and the gap between them is where inverse optimisation lives. We have argued that closing the gap with a single objective misrepresents what the data support, and we have replaced the point with a posterior that states how firmly the evidence pins down what a decision maker values. The behavioural likelihood that makes this possible turns the inverse problem into inference within a curved exponential family, and that structure delivers a log concave posterior, an identifiability criterion phrased through the geometry of the feasible region, contraction at the parametric rate, and an asymptotic normality that lets credible sets stand in for confidence sets. When the behavioural model is wrong the posterior concentrates on the nearest describable preference and a calibrated learning rate keeps its intervals honest, which is the discipline that separates a usable Bayesian answer from an overconfident one. The simulation study showed calibrated recovery and an advantage over point estimation when data are scarce, and the clinical application showed a hidden trade off reconstructed with an interval that a practitioner can act on. What began as a question about a doctor's unexplained prescription ends as a general statistical instrument for learning the goals behind constrained choices, and its value grows precisely in the situations where a point estimate would have hidden the doubt that matters.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

All numerical results in this paper are generated from simulated data under generative models that are fully specified in Sections 8 and 9, and Appendix B, together with the fixed random seeds needed to reproduce every figure and table exactly. No real, proprietary, or patient level data were used at any stage of this work; the clinical application of Section 9 is a controlled synthetic validation in which the data generating trade off is fixed by the authors precisely so that recovery can be checked against a known answer, and it is presented as such throughout, not as an analysis of real dosing records.

Conflicts of Interest

The author declares no conflict of interest.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

Appendix A. Proofs

Throughout, $\ell_n(\theta) = \sum_{i=1}^n \ell_i(\theta)$ denotes the log likelihood of (8), $s_i(\theta) = \nabla_{\theta} \ell_i(\theta)$ the per observation score of (9), and $I(\theta)$ the per observation Fisher information of (13). We write θ^* for the data generating value under correct specification and θ_0 for the Kullback Leibler projection under misspecification. Expectations $\mathbb{E}_*$ and $\mathbb{E}_0$ are taken under the corresponding decision laws with contexts drawn from ρ . All limits are as $n \rightarrow \infty$ unless stated otherwise.

Appendix A.1. Log Concavity of the Posterior

Proof of Theorem 1. Fix a context u and recall that $p(x \mid u, \theta) \propto \exp(-\tau^{-1} \langle c(\theta), x \rangle)$ on $\mathcal{K}(u)$ with $c(\theta) = c_0 + C\theta$ affine. Writing $\eta = -\tau^{-1}c(\theta)$, the family is a linear exponential family in the natural parameter η with log partition

$$A(\eta; u) = \log \int_{\mathcal{K}(u)} \exp(\langle \eta, x \rangle) dx. \quad (18)$$

function $A(\cdot; u)$ is convex on $\mathbb{R}^d$, since for $\alpha \in [0, 1]$ and η_1, η_2 the Hölder inequality applied to $\exp(\langle \alpha\eta_1 + (1 - \alpha)\eta_2, x \rangle) = \exp(\langle \eta_1, x \rangle)^\alpha \exp(\langle \eta_2, x \rangle)^{1-\alpha}$ gives $A(\alpha\eta_1 + (1 - \alpha)\eta_2; u) \leq \alpha A(\eta_1; u) + (1 - \alpha)A(\eta_2; u)$. The map $\theta \mapsto \eta = -\tau^{-1}(c_0 + C\theta)$ is affine, so $\theta \mapsto A(-\tau^{-1}c(\theta); u) = \log Z(u, \theta)$ is convex as the composition of a convex function with an affine map. Consequently

$$\ell_i(\theta) = -\frac{1}{\tau} \langle c(\theta), x_i \rangle - \log Z(u_i, \theta) \quad (19)$$

is concave in θ , being the sum of a linear term and the negative of a convex term. The full log likelihood $\ell_n = \sum_i \ell_i$ is therefore concave. The prior density π is log concave by assumption, being a Gaussian restricted to the convex set $\Theta = \mathbb{R}_{>0}^p$, whose indicator is log concave. The log posterior $\log \Pi(\theta \mid \mathcal{D}) = \ell_n(\theta) + \log \pi(\theta) + \text{const}$ is a sum of concave functions and hence concave, so $\Pi(\cdot \mid \mathcal{D})$ is log concave. Strict concavity holds whenever $I(\theta) \succ 0$, established in Lemma 1 below, because the Hessian of ℓ_n then equals $-\sum_i \tau^{-2} C^\top S(u_i, \theta) C \prec 0$. $\square$

Appendix A.2. Score and Information Identities

Proof of Lemma 1. Differentiating $\log Z(u, \theta) = A(-\tau^{-1}c(\theta); u)$ and using $\nabla_{\eta} A(\eta; u) = \mathbb{E}[x \mid u, \theta] = m(u, \theta)$ together with the chain rule $\nabla_{\theta} \eta = -\tau^{-1}C$ gives

$$\nabla_{\theta} \log Z(u, \theta) = -\frac{1}{\tau} C^\top m(u, \theta). \quad (20)$$

$s_i(\theta) = -\tau^{-1}C^\top x_i + \tau^{-1}C^\top m(u_i, \theta) = \tau^{-1}C^\top (m(u_i, \theta) - x_i)$, which is (9). Taking the expectation under the model at θ and using $\mathbb{E}[x_i | u_i, \theta] = m(u_i, \theta)$ yields $\mathbb{E}[s_i(\theta) | u_i] = 0$, and averaging over $u_i \sim \rho$ gives $\mathbb{E}_\theta[s_i(\theta)] = 0$. For the second identity, the Hessian of A satisfies $\nabla_\eta^2 A(\eta; u) = \text{Cov}[x | u, \theta] = S(u, \theta)$, so by the chain rule

$$\nabla_\theta^2 \ell_i(\theta) = -\nabla_\theta^2 \log Z(u_i, \theta) = -\frac{1}{\tau^2} C^\top S(u_i, \theta) C, \quad (21)$$

which is (10). The exponential family variance identity gives $\text{Var}[s_i(\theta) | u_i] = \tau^{-2} C^\top S(u_i, \theta) C = -\mathbb{E}[\nabla^2 \ell_i | u_i]$, and averaging over contexts produces $I(\theta) = \mathbb{E}_\rho[\tau^{-2} C^\top S(u, \theta) C]$, establishing the information equality $\text{Var}_\theta[s_i(\theta)] = I(\theta) = -\mathbb{E}_\theta[\nabla^2 \ell_i(\theta)]$. $\square$

Appendix A.3. Identifiability and Consistency of the Maximum Likelihood Estimator

Lemma 2 (Kullback Leibler separation). *Under Assumptions 2 and 3, the map $\theta \mapsto \text{KL}(\theta^* || \theta) := \mathbb{E}_*[\ell_i(\theta^*) - \ell_i(\theta)]$ is nonnegative, equals zero only at $\theta = \theta^*$, and has Hessian $I(\theta^*) \succ 0$ at θ^* .*

Proof. Nonnegativity and the zero at θ^* are the standard properties of the conditional Kullback Leibler divergence $\mathbb{E}_\rho \text{KL}(p(\cdot | u, \theta^*) || p(\cdot | u, \theta))$, which vanishes precisely when the two decision laws agree ρ almost surely. Under Assumption 3 equality of the laws forces $c(\theta) = c(\theta^*)$ up to a context free additive constant that cancels in the normalisation, and full column rank of C then forces $\theta = \theta^*$. Differentiating twice under the integral, which is permitted because the family is a smooth exponential family with all moments finite on the bounded set $\mathcal{K}(u)$, gives $\nabla_\theta^2 \text{KL}(\theta^* || \theta)|_{\theta=\theta^*} = I(\theta^*)$, positive definite by Assumption 3. $\square$

Proof of Theorem 2. Write $M_n(\theta) = n^{-1} \ell_n(\theta)$ and $M(\theta) = \mathbb{E}_*[\ell_i(\theta)]$. Each ℓ_i is concave in θ and, because decisions lie in the bounded set $\mathcal{K}(u) \subseteq [0, \bar{U}]^d$ and c is continuous on the compact parameter neighbourhood, the family $\{\ell_i(\cdot)\}$ is dominated by an integrable envelope. Pointwise $M_n(\theta) \rightarrow M(\theta)$ almost surely by the law of large numbers, and since concave functions converging pointwise on a dense set converge uniformly on compacta, the convergence is uniform on compact subsets of Θ . By Lemma 2 the limit M has the unique maximiser θ^* , and it is well separated because $I(\theta^*) \succ 0$ makes M locally strongly concave. The argmax of a uniformly convergent sequence of concave functions with a well separated maximiser converges to that maximiser, so $\theta_{\text{mle}} \xrightarrow{a.s.} \theta^*$. For the rate, a Taylor expansion of the score around θ^* gives $0 = n^{-1/2} \sum_i s_i(\theta^*) + (n^{-1} \sum_i \nabla^2 \ell_i(\tilde{\theta})) \sqrt{n}(\theta_{\text{mle}} - \theta^*)$ for some $\tilde{\theta}$ on the segment. The average Hessian converges to $-I(\theta^*)$ by the uniform law of large numbers and continuity, while $n^{-1/2} \sum_i s_i(\theta^*) \rightsquigarrow \mathcal{N}(0, I(\theta^*))$ by the central limit theorem, the scores being independent mean zero with variance $I(\theta^*)$ by Lemma 1. Slutsky then yields $\sqrt{n}(\theta_{\text{mle}} - \theta^*) \rightsquigarrow \mathcal{N}(0, I(\theta^*)^{-1})$. $\square$

Appendix A.4. Posterior Contraction

Proof of Theorem 3. We verify the three conditions of the general contraction theorem of [45], Theorem 8.9, for the parametric family at rate $\varepsilon_n = n^{-1/2}$. First, prior mass on Kullback Leibler neighbourhoods. Because π has a continuous density that is positive at $\theta^* \in \text{int } \Theta$, and by Lemma 2 the map $\theta \mapsto \text{KL}(\theta^* || \theta)$ is continuous with a nondegenerate quadratic minimum at θ^* , the set $B_n = \{\theta : \text{KL}(\theta^* || \theta) \leq \varepsilon_n^2, \mathbb{E}_*[(\ell_i(\theta^*) - \ell_i(\theta))^2] \leq \varepsilon_n^2\}$ contains a Euclidean ball of radius of order ε_n , hence $\pi(B_n) \gtrsim \varepsilon_n^p = n^{-p/2} \geq e^{-c_1 n \varepsilon_n^2}$ for every $c_1 > 0$ and n large. Second, existence of tests. For θ with $\|\theta - \theta^*\| > M\varepsilon_n$ the score based statistic $T_n = n^{-1/2} \sum_i \langle I(\theta^*)^{-1/2} s_i(\theta^*), \cdot \rangle$ together with the strong concavity of M furnishes exponentially powerful tests ϕ_n with $\mathbb{E}_* \phi_n \rightarrow 0$ and $\sup_{\|\theta - \theta^*\| > M\varepsilon_n} \mathbb{E}_\theta(1 - \phi_n) \leq e^{-c_2 M^2 n \varepsilon_n^2}$; concretely one may take the likelihood ratio tests, whose error is controlled by the Hellinger separation $h^2(\theta^*, \theta) \gtrsim \|\theta - \theta^*\|^2 \wedge 1$ implied by $I(\theta^*) \succ 0$. Third, entropy. The local parameter set $\{\theta : \|\theta - \theta^*\| \leq R\}$ is a subset of $\mathbb{R}^p$ with metric entropy $\log N(\varepsilon, \cdot, \|\cdot\|) \lesssim p \log(R/\varepsilon)$, which is $o(n \varepsilon_n^2)$. The three conditions give $\Pi(\|\theta - \theta^*\| > M_n \varepsilon_n | \mathcal{D}) \rightarrow 0$ in $\mathbb{E}_*$ probability for every $M_n \rightarrow \infty$, which is the claimed $n^{-1/2}$ contraction. $\square$

Appendix A.5. Bernstein Von Mises

Proof of Theorem 4. By Lemma 1 the experiment is differentiable in quadratic mean at θ^* with nonsingular Fisher information $I(\theta^*)$, since the square root density of a regular exponential family restricted to a fixed bounded set is differentiable in quadratic mean. Writing $h = \sqrt{n}(\theta - \theta^*)$ the local asymptotic normality expansion holds,

$$\ell_n(\theta^* + h/\sqrt{n}) - \ell_n(\theta^*) = \langle h, \Delta_n \rangle - \frac{1}{2} \langle h, I(\theta^*) h \rangle + r_n(h), \quad (22)$$

where $\Delta_n = n^{-1/2} \sum_i s_i(\theta^*) \rightsquigarrow \mathcal{N}(0, I(\theta^*))$ and $r_n(h) \rightarrow 0$ in probability for each h , uniformly on compacta by the smoothness and boundedness of the family. Theorem 3 supplies the required posterior concentration at rate $n^{-1/2}$, so the posterior mass outside a large but fixed ball in the local parameter h is asymptotically negligible. Under these two ingredients the theorem of [46], Theorem 10.1, together with continuity and positivity of π at θ^* , gives

$$\left\| \Pi(\sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\text{mle}}) \in \cdot \mid \mathcal{D}) - \mathcal{N}(\mathbf{0}, \mathbf{I}(\boldsymbol{\theta}^*)^{-1}) \right\|_{\text{TV}} \xrightarrow{\mathbb{P}^*} 0, \quad (23)$$

which is (14). The centring may be taken at $\boldsymbol{\theta}_{\text{mle}}$ or equivalently at the posterior mode, the two differing by $o_{\mathbb{P}}(n^{-1/2})$. Consequently the Laplace approximation $\mathcal{N}(\hat{\boldsymbol{\theta}}, (-\nabla^2 \log \Pi)^{-1})$ is first order correct and credible regions attain their nominal frequentist coverage asymptotically. $\square$

Appendix A.6. Misspecification and the Fractional Posterior

Proof of Proposition 1. Suppose the decisions are generated by a law $q(\cdot \mid \mathbf{u})$ not contained in the model. Define the pseudo true value $\boldsymbol{\theta}_0 = \arg \min_{\boldsymbol{\theta}} \mathbb{E}_{\rho} \text{KL}(q(\cdot \mid \mathbf{u}) \parallel p(\cdot \mid \mathbf{u}, \boldsymbol{\theta}))$, assumed to exist uniquely in $\text{int } \Theta$. The population objective $\bar{M}(\boldsymbol{\theta}) = \mathbb{E}_0[\ell_i(\boldsymbol{\theta})]$ is maximised at $\boldsymbol{\theta}_0$, where the misspecified score has mean zero, $\mathbb{E}_0[s_i(\boldsymbol{\theta}_0)] = \mathbf{0}$. Two matrices govern the local behaviour,

$$\mathbf{J} = -\mathbb{E}_0[\nabla^2 \ell_i(\boldsymbol{\theta}_0)] = \mathbb{E}_0\left[\frac{1}{\tau^2} \mathbf{C}^{\top} \mathbf{S}(\mathbf{u}, \boldsymbol{\theta}_0) \mathbf{C}\right], \quad \mathbf{K} = \text{Var}_0[s_i(\boldsymbol{\theta}_0)], \quad (24)$$

the sensitivity and the variability of the estimating equation. Repeating the expansion of Theorem 2 under q gives $\sqrt{n}(\boldsymbol{\theta}_{\text{mle}} - \boldsymbol{\theta}_0) \rightsquigarrow \mathcal{N}(\mathbf{0}, \mathbf{J}^{-1} \mathbf{K} \mathbf{J}^{-1})$, the sandwich law, and the corresponding misspecified local asymptotic normality expansion holds with curvature $\mathbf{J}$ and score covariance $\mathbf{K}$. The ordinary posterior, whose local precision is $n\mathbf{J}$, therefore concentrates as $\mathcal{N}(\boldsymbol{\theta}_{\text{mle}}, (n\mathbf{J})^{-1})$ by the argument of Theorem 4 applied under q , following [47]. When $\mathbf{K} \neq \mathbf{J}$ this credible spread $\mathbf{J}^{-1}/n$ differs from the sampling spread $\mathbf{J}^{-1} \mathbf{K} \mathbf{J}^{-1}/n$, so nominal credible sets are miscalibrated. Raising the likelihood to a power $\omega \in (0, 1]$ multiplies the local precision by ω , giving the fractional posterior spread $(\omega n \mathbf{J})^{-1}$. Calibration in the sense of matched generalised information is achieved when $\omega \mathbf{J} = \mathbf{J} \mathbf{K}^{-1} \mathbf{J}$ in the scalar reduction $\omega = \{\text{tr}(\mathbf{J}^{-1} \mathbf{K})\}^{-1} \text{tr}(\mathbf{I}_p)$ that equates the expected log likelihood curvature to the score variability, which for a scalar or isotropic mismatch reduces to $\omega = \lambda_J/\lambda_K$ with λ_J, λ_K the common eigenvalues of $\mathbf{J}, \mathbf{K}$. Under this choice the fractional credible sets recover asymptotically nominal coverage. The construction corrects a covariance mismatch and does not remove bias in $\boldsymbol{\theta}_0$ relative to any external target, a point returned to in Section 8 in the context of gross contamination. $\square$

Appendix A.7. Geometric Identifiability, Impossibility, and a Finite Sample Bound

Proof of Proposition 2. By Lemma 1, $\mathbf{I}(\boldsymbol{\theta}) = \tau^{-2} \mathbb{E}_{\rho}[\mathbf{C}^{\top} \mathbf{S}(\mathbf{u}, \boldsymbol{\theta}) \mathbf{C}]$. For $v \in \mathbb{R}^p$,

$$v^{\top} \mathbf{I}(\boldsymbol{\theta}) v = \tau^{-2} \mathbb{E}_{\rho}[(\mathbf{C}v)^{\top} \mathbf{S}(\mathbf{u}, \boldsymbol{\theta}) (\mathbf{C}v)] = \tau^{-2} \mathbb{E}_{\rho}[\text{Var}(\langle \mathbf{C}v, \mathbf{x} \rangle \mid \mathbf{u}, \boldsymbol{\theta})] \geq 0. \quad (25)$$

$\mathbf{I}(\boldsymbol{\theta}) \succ 0$ fails only if there is $v \neq \mathbf{0}$ with $\text{Var}(\langle \mathbf{C}v, \mathbf{x} \rangle \mid \mathbf{u}, \boldsymbol{\theta}) = 0$ for ρ almost every $\mathbf{u}$, that is $\mathbf{C}v$ is ρ almost surely constant along the support of the decision law. Equivalently $\mathbf{C}v$ lies in the lineality directions of $\mathcal{K}(\mathbf{u})$ for ρ almost every $\mathbf{u}$, the directions along which the feasible polytope is a single point or the objective is flat over a face carrying all the mass. When $\mathcal{K}(\mathbf{u})$ is full dimensional with positive decision variance in every coordinate, which holds for the separable design with finite rates, $\mathbf{S}(\mathbf{u}, \boldsymbol{\theta}) \succ 0$ and the condition reduces to $\mathbf{C}v = \mathbf{0}$. Local identifiability of $\boldsymbol{\theta}$ therefore holds if and only if $\mathbf{C}$ has full column rank and the cost directions $\mathbf{C}v$ are not annihilated by the decision geometry, which is the stated characterisation. $\square$

Proof of Proposition 3. The decision law depends on $(\mathbf{c}(\boldsymbol{\theta}), \tau)$ only through $\eta = -\tau^{-1} \mathbf{c}(\boldsymbol{\theta})$. Fix $\kappa > 0$ and let $\boldsymbol{\theta}'$ satisfy $\mathbf{c}(\boldsymbol{\theta}') = \kappa \mathbf{c}(\boldsymbol{\theta})$, which exists whenever $\kappa \mathbf{c}(\boldsymbol{\theta})$ lies in the affine image of Θ under $\mathbf{c}$. Setting $\tau' = \kappa \tau$ gives $-\tau'^{-1} \mathbf{c}(\boldsymbol{\theta}') = -\kappa^{-1} \tau^{-1} \kappa \mathbf{c}(\boldsymbol{\theta}) = -\tau^{-1} \mathbf{c}(\boldsymbol{\theta})$, so the two parameters induce identical decision laws for every context. Hence $(\boldsymbol{\theta}, \tau)$ and $(\boldsymbol{\theta}', \kappa \tau)$ are observationally equivalent and τ is not identified jointly with the scale of $\mathbf{c}(\boldsymbol{\theta})$. Only the direction $\mathbf{c}(\boldsymbol{\theta})/\|\mathbf{c}(\boldsymbol{\theta})\|$ is identified when τ is unknown. The second claim follows because any component of $\mathbf{c}(\boldsymbol{\theta})$ orthogonal to the affine hull of $\mathcal{K}(\mathbf{u})$ contributes a constant to $\langle \mathbf{c}(\boldsymbol{\theta}), \mathbf{x} \rangle$ over $\mathcal{K}(\mathbf{u})$ and cancels in the normalisation, leaving the decision law unchanged. $\square$

Proof of Proposition 4. Let $\hat{\mathbf{c}}(\tau)$ denote the maximiser of $\boldsymbol{\theta} \mapsto \ell_n(\boldsymbol{\theta})$ under assumed temperature τ , expressed through the fitted cost. The score equation $\sum_i \mathbf{C}^{\top} (\mathbf{m}(\mathbf{u}_i, \boldsymbol{\theta}) - \mathbf{x}_i) = \mathbf{0}$ depends on $\boldsymbol{\theta}$ only through $\eta_i = -\tau^{-1} \mathbf{c}(\boldsymbol{\theta})$, and $\mathbf{m}$ is a function of η_i alone. If $\eta^{\dagger}$ solves the score equation, it does so independently of τ , and the corresponding fitted cost is $\mathbf{c}^{\dagger} = -\tau \eta^{\dagger}$, proportional to τ . Therefore the normalised fitted cost $\mathbf{c}^{\dagger}/\|\mathbf{c}^{\dagger}\| = -\eta^{\dagger}/\|\eta^{\dagger}\|$ is exactly invariant to the assumed τ , for every finite sample, which proves the claim. The absolute scale is multiplied by the assumed τ , so cost magnitudes are recoverable only up to the temperature. $\square$

Proof of Theorem 5. Decisions satisfy $\mathbf{x}_i \in [0, \bar{U}]^d$ almost surely, so by (9) the score increments obey $\|\mathbf{s}_i(\boldsymbol{\theta}^*)\| \leq \tau^{-1} \|\mathbf{C}\|_{\text{op}} \|\mathbf{m}(\mathbf{u}_i, \boldsymbol{\theta}^*) - \mathbf{x}_i\| \leq \tau^{-1} \|\mathbf{C}\|_{\text{op}} \sqrt{d} \bar{U} =: B$. The increments are independent with mean

zero, hence by the vector Bernstein inequality, for $t > 0$,

$$\mathbb{P}_\star \left(\left\| n^{-1} \sum_i s_i(\boldsymbol{\theta}^\star) \right\| > t \right) \leq 2p \exp \left(- \frac{nt^2}{2\sigma^2 + 2Bt/3} \right), \quad \sigma^2 = \|\mathbf{I}(\boldsymbol{\theta}^\star)\|_{\text{op}}. \quad (26)$$

on the event that the averaged score is at most t , strong concavity of ℓ_n with smallest curvature $\lambda := \lambda_{\min}(\mathbf{I}(\boldsymbol{\theta}^\star)) > 0$ on a neighbourhood, guaranteed for n large by the uniform law of large numbers applied to the Hessian, gives $\|\boldsymbol{\theta}_{\text{mle}} - \boldsymbol{\theta}^\star\| \leq t/\lambda$. Choosing $t = \sqrt{2\sigma^2(\log(2p/\delta))/n} + \frac{2B}{3n} \log(2p/\delta)$ makes the exceedance probability at most δ , whence with probability at least $1 - \delta$,

$$\|\boldsymbol{\theta}_{\text{mle}} - \boldsymbol{\theta}^\star\| \leq \frac{1}{\lambda_{\min}(\mathbf{I}(\boldsymbol{\theta}^\star))} \left(\sqrt{\frac{2\|\mathbf{I}(\boldsymbol{\theta}^\star)\|_{\text{op}} \log(2p/\delta)}{n}} + \frac{2B \log(2p/\delta)}{3n} \right). \quad (27)$$

leading term is of order $\sqrt{(\log(2p/\delta))/n} / \lambda_{\min}(\mathbf{I}(\boldsymbol{\theta}^\star))$, giving the stated non asymptotic guarantee with explicit dependence on dimension through p and on the decision geometry through $\lambda_{\min}(\mathbf{I}(\boldsymbol{\theta}^\star))$. $\square$

Appendix B. Partition Function, Sampling, and the Classical Benchmark

Appendix B.1. Exponential Integral over a Polytope

For a full dimensional simple polytope with vertex set $\mathcal{V}$ and, at each vertex $\mathbf{v}$, edge directions $\mathbf{r}_{\mathbf{v},1}, \dots, \mathbf{r}_{\mathbf{v},d}$ pointing into the polytope, the exponential integral admits the vertex expansion

$$\int_{\mathcal{K}} \exp(\langle \boldsymbol{\eta}, \mathbf{y} \rangle) d\mathbf{y} = \sum_{\mathbf{v} \in \mathcal{V}} \frac{\exp(\langle \boldsymbol{\eta}, \mathbf{v} \rangle) |\det[\mathbf{r}_{\mathbf{v},1} \cdots \mathbf{r}_{\mathbf{v},d}]|}{\prod_{k=1}^d (-\langle \boldsymbol{\eta}, \mathbf{r}_{\mathbf{v},k} \rangle)},$$

valid whenever no denominator vanishes. Differentiating this expression in $\boldsymbol{\eta}$ returns the mean and covariance in (6) in closed form. When $\boldsymbol{\eta}$ approaches a direction orthogonal to an edge, the corresponding denominator tends to zero and the vertex term is evaluated by the limiting form obtained through l'Hopital reduction, or the parameter is perturbed by a negligible amount and the result extrapolated, which stabilises the computation in the low dimensions used here.

Appendix B.2. Hit and Run Sampling

For larger decision dimension the moments are estimated by sampling the Gibbs density on the polytope. From a feasible interior point the sampler chooses a random direction, computes the chord of the polytope through the current point along that direction, and draws the next point from the one dimensional Gibbs density restricted to the chord, which is log concave and simple to sample. The chain mixes in a number of steps polynomial in the dimension for log concave targets, and the sample mean and sample covariance of the visited points estimate $\mathbf{m}(\mathbf{u}, \boldsymbol{\theta})$ and $\mathbf{S}(\mathbf{u}, \boldsymbol{\theta})$. These estimates enter the gradient in Algorithm 1 directly, and their Monte Carlo error is controlled by the chord sample size, which we set so that the gradient noise is small relative to the Metropolis correction.

Appendix B.3. Classical Benchmark

The point estimate against which the posterior is compared minimizes the aggregate sub optimality residual. For each observation the stationarity of the forward programme requires the cost to lie in the normal cone of the feasible set at the observed decision, and the residual measures the Euclidean distance by which this fails. Summing the squared residuals over the observations and minimizing over the cost, subject to a normalization that removes the trivial zero solution, returns the classical estimate. This estimator is the zero temperature and prior free special case of the framework, which is why the two coincide as the sample grows and the prior washes out.

Appendix C. Reproducibility Details

The simulation fixes the temperature at $\tau = 1$ and generates contexts as described in Section 8, with each replication reseeded. The Langevin sampler uses a step size chosen by a short pilot to achieve an acceptance rate near 0.57, four independent chains, several thousand retained iterations per chain after a warm up of one quarter that length, and no thinning. Convergence diagnostics report a potential scale reduction factor below 1.01 and effective sample sizes consistent with the diagnostics reported in Section 8. The Laplace approximation uses a Newton iteration terminated when the gradient norm falls below 10^{-6} , and for separable feasible sets the decision moments entering the gradient and curvature are evaluated in closed form. The clinical application draws two hundred and forty decisions, holds out sixty by random partition, and reports posterior summaries from the same

sampler configuration. Credible intervals are equal tailed marginal quantile ranges, and predictive intervals are quantiles of the posterior predictive in (17) evaluated by forward simulation from posterior draws.

References

- Ahuja, R.K.; Orlin, J.B. Inverse Optimization. *Oper. Res.* **2001**, *49*, 771–783. <https://doi.org/10.1287/opre.49.5.771.10607>.
- Keshavarz, A.; Wang, Y.; Boyd, S. Imputing a convex objective function. In Proceedings of the IEEE International Symposium on Intelligent Control, Denver, CO, USA, 28–30 September 2011; pp. 613–619. <https://doi.org/10.1109/isic.2011.6045410>.
- Rust, J. Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica* **1987**, *55*, 999–1033. <https://doi.org/10.2307/1911259>.
- Stuart, A.M. Inverse problems: A Bayesian perspective. *Acta Numer.* **2010**, *19*, 451–559. <https://doi.org/10.1017/s0962492910000061>.
- Ramachandran, D.; Amir, E. Bayesian inverse reinforcement learning. In Proceedings of the International Joint Conference on Artificial Intelligence, Hyderabad, India, 6–12 January 2007; pp. 2586–2591.
- Luce, R.D. *Individual Choice Behavior: A Theoretical Analysis*; Wiley: New York, NY, USA, 1959.
- McFadden, D. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*; Academic Press: New York, NY, USA, 1974; pp. 105–142.
- McKelvey, R.D.; Palfrey, T.R. Quantal Response Equilibria for Normal Form Games. *Games Econ. Behav.* **1995**, *10*, 6–38. <https://doi.org/10.1006/game.1995.1023>.
- Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630. <https://doi.org/10.1103/physrev.106.620>.
- Aswani, A.; Shen, Z.J.M.; Siddiq, A. Inverse Optimization with Noisy Data. *Oper. Res.* **2018**, *66*, 870–892. <https://doi.org/10.1287/opre.2017.1705>.
- Mohajerin Esfahani, P.; Shafieezadeh-Abadeh, S.; Hanasusanto, G.A.; et al. Data-driven inverse optimization with imperfect information. *Math. Program.* **2018**, *167*, 191–234. <https://doi.org/10.1007/s10107-017-1216-6>.
- Shahmoradi, Z.; Lee, T. Quantile Inverse Optimization: Improving Stability in Inverse Linear Programming. *Oper. Res.* **2022**, *70*, 2538–2562. <https://doi.org/10.1287/opre.2021.2143>.
- Dong, C.; Chen, Y.; Zeng, B. Generalized Inverse Optimization through Online Learning. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December 2018; Volume 31, pp. 86–95.
- Besbes, O.; Fonseca, Y.; Lobel, I. Contextual Inverse Optimization: Offline and Online Learning. *Oper. Res.* **2025**, *73*, 424–443. <https://doi.org/10.1287/opre.2021.0369>.
- Lu, Y.A.; Hu, W.S.; Paulson, J.A.; et al. BO4IO: A Bayesian optimization approach to inverse optimization with uncertainty quantification. *arXiv* **2024**, arXiv:2405.17875v1.
- Chan, T.C.Y.; Sandholtz, N.; Yousefi, N. Uncertainty Quantification in Data-Driven Inverse Optimization via Bayesian Inference. *arXiv* **2026**, arXiv:2605.25288v1.
- Ahmadi, F.; Ganjkanloo, F.; Ghobadi, K. From Non-Identifiability to Goal-Integrated Decision-Making in Parametric Inverse Optimization. *arXiv* **2026**, arXiv:2603.17033v1.
- Bissiri, P.G.; Holmes, C.C.; Walker, S.G. A General Framework for Updating Belief Distributions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2016**, *78*, 1103–1130. <https://doi.org/10.1111/rssb.12158>.
- Grünwald, P.; van Ommen, T. Inconsistency of Bayesian Inference for Misspecified Linear Models, and a Proposal for Repairing It. *Bayesian Anal.* **2017**, *12*, 1069–1103. <https://doi.org/10.1214/17-ba1085>.
- Miller, J.W.; Dunson, D.B. Robust Bayesian Inference via Coarsening. *J. Am. Stat. Assoc.* **2019**, *114*, 1113–1125. <https://doi.org/10.1080/01621459.2018.1469995>.
- Syring, N.; Martin, R. Calibrating general posterior credible regions. *Biometrika* **2019**, *106*, 479–486. <https://doi.org/10.1093/biomet/asy054>.
- Roberts, G.O.; Rosenthal, J.S. Optimal Scaling of Discrete Approximations to Langevin Diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1998**, *60*, 255–268. <https://doi.org/10.1111/1467-9868.00123>.
- Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004. <https://doi.org/10.1017/cbo9780511804441>.
- Heuberger, C. Inverse Combinatorial Optimization: A Survey on Problems, Methods, and Results. *J. Comb. Optim.* **2004**, *8*, 329–361. <https://doi.org/10.1023/b:joco.0000038914.26975.9b>.
- Iyengar, G.; Kang, W. Inverse conic programming with applications. *Oper. Res. Lett.* **2005**, *33*, 319–330. <https://doi.org/10.1016/j.orl.2004.04.007>.
- Chan, T.C.Y.; Craig, T.; Lee, T.; et al. Generalized Inverse Multiobjective Optimization with Application to Cancer Therapy. *Oper. Res.* **2014**, *62*, 680–695. <https://doi.org/10.1287/opre.2014.1267>.
- Chan, T.C.Y.; Mahmood, R.; Zhu, I.Y. Inverse Optimization: Theory and Applications. *Oper. Res.* **2025**, *73*, 1046–1074. <https://doi.org/10.1287/opre.2022.0382>.
- Bertsimas, D.; Gupta, V.; Paschalidis, I.C. Data-driven estimation in equilibrium using inverse optimization. *Math. Program.* **2015**, *153*, 595–633. <https://doi.org/10.1007/s10107-014-0819-4>.

29. Bärnmann, A.; Pokutta, S.; Schneider, O. Emulating the Expert: Inverse Optimization through Online Learning. In Proceedings of the International Conference on Machine Learning, Sydney, NSW, Australia, 6–11 August 2017; pp. 400–410.
30. Ghobadi, K.; Mahmoudzadeh, H. Inferring linear feasible regions using inverse optimization. *Eur. J. Oper. Res.* **2021**, *290*, 829–843. <https://doi.org/10.1016/j.ejor.2020.08.048>.
31. Saez-Gallego, J.; Morales, J.M. Short-Term Forecasting of Price-Responsive Loads Using Inverse Optimization. *IEEE Trans. Smart Grid* **2018**, *9*, 4805–4814. <https://doi.org/10.1109/tsg.2017.2671743>.
32. Train, K.E. *Discrete Choice Methods with Simulation*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2009. <https://doi.org/10.1017/cbo9780511805271>.
33. Ng, A.Y.; Russell, S. Algorithms for inverse reinforcement learning. In Proceedings of the International Conference on Machine Learning, Stanford, CA, USA, 29 June–2 July 2000; pp. 663–670.
34. Abbeel, P.; Ng, A.Y. Apprenticeship learning via inverse reinforcement learning. In Proceedings of the Twenty-First International Conference on Machine Learning, Banff, AB, Canada, 4–8 July 2004; p. 1. <https://doi.org/10.1145/1015330.1015430>.
35. Ziebart, B.D.; Maas, A.; Bagnell, J.A.; et al. Maximum entropy inverse reinforcement learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Chicago, IL, USA, 13–17 July 2008; pp. 1433–1438.
36. Choi, J.; Kim, K.E. MAP Inference for Bayesian Inverse Reinforcement Learning. In Proceedings of the Advances in Neural Information Processing Systems, Granada, Spain, 12–17 December 2011; Volume 24, pp. 1989–1997.
37. Dashti, M.; Stuart, A.M. The Bayesian Approach to Inverse Problems. In *Handbook of Uncertainty Quantification*; Springer International Publishing: Cham, Switzerland, 2017; pp. 311–428. https://doi.org/10.1007/978-3-319-12385-1_7.
38. Kaipio, J.P.; Somersalo, E. *Statistical and Computational Inverse Problems*; Springer: New York, NY, USA, 2005. <https://doi.org/10.1007/b138659>.
39. Arrow, K.J.; Harris, T.; Marschak, J. Optimal Inventory Policy. *Econometrica* **1951**, *19*, 250–272. <https://doi.org/10.2307/1906813>.
40. Ban, G.Y.; Rudin, C. The Big Data Newsvendor: Practical Insights from Machine Learning. *Oper. Res.* **2019**, *67*, 90–108. <https://doi.org/10.1287/opre.2018.1757>.
41. Bertsimas, D.; Tsitsiklis, J.N. *Introduction to Linear Optimization*; Athena Scientific: Belmont, MA, USA, 1997.
42. Wainwright, M.J.; Jordan, M.I. Graphical Models, Exponential Families, and Variational Inference. *Found. Trends Mach. Learn.* **2008**, *1*, 1–305. <https://doi.org/10.1561/22000000001>.
43. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, NJ, USA, 1970. <https://doi.org/10.1515/9781400873173>.
44. Schwartz, L. On Bayes procedures. *Z. Wahrscheinlichkeitstheorie Verwandte Geb.* **1965**, *4*, 10–26. <https://doi.org/10.1007/bf00535479>.
45. Ghosal, S.; van der Vaart, A. *Fundamentals of Nonparametric Bayesian Inference*; Cambridge University Press: Cambridge, UK, 2017. <https://doi.org/10.1017/9781139029834>.
46. van der Vaart, A.W. *Asymptotic Statistics*; Cambridge University Press: Cambridge, UK, 1998. <https://doi.org/10.1017/cbo9780511802256>.
47. Kleijn, B.J.K.; van der Vaart, A.W. The Bernstein-Von-Mises theorem under misspecification. *Electron. J. Stat.* **2012**, *6*, 354–381. <https://doi.org/10.1214/12-ejs675>.
48. Lawrence, J. Polytope volume computation. *Math. Comput.* **1991**, *57*, 259–271. <https://doi.org/10.1090/s0025-5718-1991-1079024-2>.
49. Lovász, L.; Vempala, S. Hit-and-Run from a Corner. *SIAM J. Comput.* **2006**, *35*, 985–1005. <https://doi.org/10.1137/s009753970544727x>.
50. Murray, I.; Ghahramani, Z.; MacKay, D.J.C. MCMC for doubly-intractable distributions. In Proceedings of the Conference on Uncertainty in Artificial Intelligence, Cambridge, MA, USA, 13–16 July 2006; pp. 359–366.
51. Robert, C.P.; Casella, G. *Monte Carlo Statistical Methods*, 2nd ed.; Springer: New York, NY, USA, 2004. <https://doi.org/10.1007/978-1-4757-4145-2>.
52. Neal, R.M. MCMC Using Hamiltonian Dynamics. In *Handbook of Markov Chain Monte Carlo*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2011; pp. 113–162. <https://doi.org/10.1201/b10905-6>.
53. Hoffman, M.D.; Gelman, A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **2014**, *15*, 1593–1623.
54. Girolami, M.; Calderhead, B. Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2011**, *73*, 123–214. <https://doi.org/10.1111/j.1467-9868.2010.00765.x>.
55. Blei, D.M.; Kucukelbir, A.; McAuliffe, J.D. Variational Inference: A Review for Statisticians. *J. Am. Stat. Assoc.* **2017**, *112*, 859–877. <https://doi.org/10.1080/01621459.2017.1285773>.
56. Ajayi, T.; Lee, T.; Schaefer, A.J. Objective Selection for Cancer Treatment: An Inverse Optimization Approach. *Oper. Res.* **2022**, *70*, 1717–1738. <https://doi.org/10.1287/opre.2021.2192>.
57. Chan, T.C.Y.; Eberg, M.; Forster, K.; et al. An Inverse Optimization Approach to Measuring Clinical Pathway Concordance. *Manag. Sci.* **2022**, *68*, 1882–1903. <https://doi.org/10.1287/mnsc.2021.4100>.