
Beyond Global Divergences: A Local-Mass Perspective on Bayesian Inference

Hanli Xu
UNSW Sydney
z5664653@zmail.unsw.edu.au

Fengxiang He
University of Edinburgh
fhe@ed.ac.uk

Sarat Moka
UNSW Sydney
s.moka@unsw.edu.au

Abstract

Global objectives, such as KL divergence and ELBO, are widely used in Bayesian inference for measuring distributional discrepancy. This paper studies their “local-mass behaviour” that is not directly captured by such objectives. We introduce and use two mathematical tools: (1) Mass Index for recording the polynomial and logarithmic decay scales of local mass, and (2) regularised extended KL (RE-KL), a set-localised divergence that can be formulated in the presence of singular components. Mass Indices help characterise how Bayesian updating changes local mass: (1) power-log likelihood factors shift it explicitly, and (2) parameter-dependent supports, or their smooth softening, may change the local scale through the amount of mass that remains near the parameter value. Using local RE-KL, we prove absolute, relative, and directional inequalities for comparing local small-ball masses under the two KL directions. Together, these results provide a local theoretical account of local mass behaviour. Experiments provide controlled illustrations of the local behaviour. Code is available at <https://github.com/Forsythia0604/Local-Diagnostic-Tools>.

1 Introduction

Global objectives, such as Kullback-Leibler (KL) and the evidence lower bound (ELBO), are standard in Bayesian inference [16, 18, 23, 5]. However, global closeness does not necessarily determine how probability mass behaves near a specified parameter value. Bayesian asymptotic theory has long made this local issue explicit through prior mass and small-ball conditions for posterior contraction, while sparse, singular, and constrained Bayesian models provide common

settings in which the behaviour of mass near special parameter values is substantively important [10, 6, 27]. Such questions are not fully captured by a single global divergence value. A sequence of distributions can converge in KL divergence while still exhibiting different small-ball scaling around a fixed point.

This paper develops a local-mass framework for Bayesian inference. The basic object of study is the small-ball probability $p(B_r(\boldsymbol{\theta}))$, $r \downarrow 0$, where $B_r(\boldsymbol{\theta})$ denotes the ball of radius r centred at a parameter value $\boldsymbol{\theta}$. The asymptotic behaviour of this quantity records how probability mass accumulates around $\boldsymbol{\theta}$.

We use the term *Mass Index* for a concentration-oriented parametrisation of local small-ball behaviour. The object of interest is a probability measure near a fixed parameter value, and the notion is closely related to the local dimension theory of a measure [7, 12, 21]. Specifically, suppose that, as $r \downarrow 0$, $\mu(B_r(\boldsymbol{\theta})) \asymp r^a$. Here d is the ambient dimension, and a is the usual local scaling exponent. The power component of the Mass Index records the normalised reciprocal scale d/a . Thus it reverses the usual local-dimension scale and expresses it as a concentration scale. The logarithmic component records the first log-order correction, when the power scale is well defined. This plays a role analogous to slowly varying corrections in regular variation [14, 15, 4]. The notation therefore distinguishes measures with the same leading polynomial order but different logarithmic local mass behaviour, such as smooth priors and horseshoe-type priors near sparse parameter values.

To compare local masses, we use *regularised extended KL* (RE-KL), a set-localised regularised f -divergence [28, 1, 29]. Its finite recession slope permits an extension to possibly singular measures by separating absolutely continuous and singular contributions. We therefore use RE-KL as a local f -divergence-based tool for comparing small-ball masses, not as a new global divergence. In the absolutely continuous case, it relates to Tsallis-type divergences, includes squared Hellinger

distance at $\alpha = 1/2$, and converges to KL as $\alpha \uparrow 1$.

We prove that Bayesian updating preserves both power and logarithmic local mass scales, if the likelihood is locally regular around the parameter value. Thus Bayesian updating can be decomposed into a prior local scale and a likelihood-induced local correction. When the support is parameter-dependent, the likelihood may contain a hard support constraint that removes part of the prior mass near the parameter value. We prove that now local mass is governed by the fraction of prior mass that survives the constraint in shrinking neighbourhoods. If this surviving fraction decays at a polynomial or logarithmic rate, the posterior Mass Index shifts by the corresponding amount. We also show that smoothing hard constraints may not preserve this local structure: the zero-temperature limit of a softened likelihood and the small-neighbourhood limit need not commute.

We then study local discrepancy between two probability measures. RE-KL gives both absolute and relative bounds on local probability masses. These bounds imply directional sufficient conditions for comparing Mass Indices. The approximation-to-target direction is rigid: sufficiently small normalised local RE-KL forces equality of the local power scale, and also of the logarithmic scale when defined. The target-to-approximation direction is weaker: bounded local divergence prevents the approximation from being locally thinner than the target, but still allows extra local concentration. This expresses the usual asymmetry between KL directions directly in terms of local small-ball mass.

Numerical experiments illustrate the theory. Synthetic examples display regular mass, depletion, cusps, logarithmic corrections, and atoms. A UCI Bayesian logistic-regression experiment shows that regular Bayesian updating can greatly increase posterior mass near the Laplace mean while preserving the leading local power order. A final toy example demonstrates the directionality of local RE-KL, with one direction remaining bounded and the other diverging on shrinking neighbourhoods.

2 Related work

Variational inference and mode coverage. Divergence choice is known to affect the behaviour of variational approximations [18, 23]. Inclusive-KL and related f -divergence methods often encourage broader mass coverage than reverse-KL objectives [23, 24, 22, 28, 29]. These approaches are mainly global. Our use of local RE-KL is closer in spirit to the local analysis of Csiszár ϕ -divergences by Avlogiaris et al. [1]. We study a narrower question: how Bayesian updating and variational approximation affect small-

ball probabilities $p(B_r(\boldsymbol{\theta}))$ as $r \downarrow 0$. This is also related to mode-coverage questions, where global criteria may fail to capture set-level or pointwise coverage [19, 30].

Regular variation and local asymptotic scales.

Our notation draws on two classical sources. The first is regular variation, which provides the language for separating a leading power law from slowly varying, including logarithmic, corrections [14, 15, 4]. The second is the local, or pointwise, dimension theory of measures, where the small-ball behaviour of $p(B_r(\boldsymbol{\theta}))$ is summarised through limits of $\log p(B_r(\boldsymbol{\theta}))/\log r$ [7, 21, 12]. We use these ideas only as a local small-ball bookkeeping device: for a fixed parameter value $\boldsymbol{\theta}$, $p(B_r(\boldsymbol{\theta}))$ records the leading power scale and, when present, the first logarithmic correction of the probability mass.

Bayesian pruning. Bayesian sparsification methods use priors to control how mass is allocated across parameter regions while avoiding excessive shrinkage of large signals [20, 11, 9]. These works motivate our local perspective, but they typically evaluate architecture size, contraction, objective value, or prediction [27, 26, 6]. We instead isolate a measure-level object: the asymptotic scale of local small-ball mass, independent of a particular architecture or optimisation procedure.

3 Preliminaries

Basics. For non-negative functions f and g , with respect to the limit under consideration, $f = O(g)$, $f = \Omega(g)$, and $f \asymp g$ denote upper, lower, and two-sided comparison up to positive constants, respectively. For probability measures p and q , $p \ll q$ means that $q(E) = 0 \Rightarrow p(E) = 0$ for every measurable set E . We write $p \perp q$ if there exists measurable E such that $p(\mathbb{R}^d \setminus E) = 0$ and $q(E) = 0$. These notions are referred to as absolute continuity and mutual singularity, respectively.

Lemma 3.1 (Lebesgue decomposition; Theorem 3.8 [8]). *For probability measures p and q , there are unique measures $q_{\ll p} \ll p$ and $q_{\perp p} \perp p$ such that $q = q_{\ll p} + q_{\perp p}$.*

Definition 3.2 (Total variation [3]). Let p and q be probability measures on a measurable space (Θ, \mathcal{F}) . The total variation distance between p and q is

$$\|p - q\|_{\text{TV}} := \sup_{E \in \mathcal{F}} |p(E) - q(E)|.$$

Definition 3.3 (KL divergence [17]). Let p and q be probability measures on a measurable space (Θ, \mathcal{F}) . If $q \ll p$, the KL divergence from q to p is

$$D_{\text{KL}}(q||p) := \int_{\Theta} \log \left(\frac{dq}{dp} \right) dq.$$

Otherwise, we set $D_{\text{KL}}(q||p) = \infty$.

Problem formulation. Let $\Theta = \mathbb{R}^d$ be the parameter space, and let $\mathcal{D} = (z_1, \dots, z_n) \in \mathcal{Z}^n$ denote the observed data, where $z_i \sim P_{\boldsymbol{\theta}}$, $i = 1, \dots, n$. The likelihood of the observed data under parameter $\boldsymbol{\theta}$ is denoted by $L(\boldsymbol{\theta} \mid \mathcal{D})$. Let π_0 be a prior probability measure on Θ . Whenever the marginal $Z_{\mathcal{D}} := \int_{\Theta} L(\boldsymbol{\theta} \mid \mathcal{D}) \pi_0(d\boldsymbol{\theta}) \in (0, \infty)$, the posterior is defined by $\pi_{\mathcal{D}}(E) = Z_{\mathcal{D}}^{-1} \int_E L(\boldsymbol{\theta} \mid \mathcal{D}) \pi_0(d\boldsymbol{\theta})$. This map from the prior π_0 to the posterior $\pi_{\mathcal{D}}$ is called the Bayesian update.

In this paper, Bayesian inference refers to the construction and representation of the posterior distribution induced by the prior and the observed data. The Bayesian update gives the target posterior $\pi_{\mathcal{D}}$. Since this posterior is often analytically or computationally intractable, one may further replace it by an approximating measure q chosen from a tractable variational family $\mathcal{Q} \subseteq \mathcal{P}(\Theta)$.

The variational approximation problem is to find q^* such that $q^* \in \operatorname{argmin}_{q \in \mathcal{Q}} D_{\text{KL}}(q \parallel \pi_{\mathcal{D}})$. Or equivalently, maximising the evidence lower bound $\mathcal{L}_{\mathcal{D}}(q) := \int_{\Theta} \log L(\boldsymbol{\theta} \mid \mathcal{D}) q(d\boldsymbol{\theta}) - D_{\text{KL}}(q \parallel \pi_0)$.

4 The local mass framework

We now introduce the local mass framework. Throughout this paper, p and q denote two arbitrary probability measures on Θ .

4.1 Mass Index: On local small-ball scale

We now introduce the Mass Index as an order-scale summary of the local small-ball mass $p(B_r(\boldsymbol{\theta}))$, $r \downarrow 0$. The power component records the reciprocal of the sharp local power scale, normalised by the ambient dimension d . The logarithmic component is defined only after the power scale is fixed, and records the first log-power correction. This construction is deliberately weaker than regular variation: it uses upper and lower power gauges, and does not require an exact asymptotic representation or a slowly varying factor.

Definition 4.1 (Power and Logarithmic Mass Indices). Let $B_r(\boldsymbol{\theta})$ be the standard open ball of radius r centred at $\boldsymbol{\theta}$. We adopt the convention that $\sup \emptyset = 0$ and $\inf \emptyset = \infty$.

Power Mass Index. As $r \downarrow 0$, define

$$\begin{aligned} \mathcal{U}_{\text{pow}} &:= \left\{ \eta > 0 : p(B_r(\boldsymbol{\theta})) = O(r^{1/\eta}) \right\}, \\ \mathcal{L}_{\text{pow}} &:= \left\{ \eta > 0 : p(B_r(\boldsymbol{\theta})) = \Omega(r^{1/\eta}) \right\}. \end{aligned}$$

The upper and lower Power Mass Indices are

$$\overline{\text{MI}}_{\text{pow}}(p, \boldsymbol{\theta}) := d \inf \mathcal{U}_{\text{pow}}, \quad \underline{\text{MI}}_{\text{pow}}(p, \boldsymbol{\theta}) := d \sup \mathcal{L}_{\text{pow}}.$$

If $\overline{\text{MI}}_{\text{pow}}(p, \boldsymbol{\theta}) = \underline{\text{MI}}_{\text{pow}}(p, \boldsymbol{\theta})$, then the common value is denoted by $\text{MI}_{\text{pow}}(p, \boldsymbol{\theta})$.

Logarithmic Mass Index. Assume that $\text{MI}_{\text{pow}}(p, \boldsymbol{\theta})$ exists and lies in $(0, \infty)$. Let $a_{p, \boldsymbol{\theta}} := \frac{d}{\text{MI}_{\text{pow}}(p, \boldsymbol{\theta})}$. As $r \downarrow 0$, define

$$\begin{aligned} \mathcal{U}_{\log} &:= \left\{ \eta \in \mathbb{R} : p(B_r(\boldsymbol{\theta})) = O(r^{a_{p, \boldsymbol{\theta}}} (-\log r)^{\eta}) \right\}, \\ \mathcal{L}_{\log} &:= \left\{ \eta \in \mathbb{R} : p(B_r(\boldsymbol{\theta})) = \Omega(r^{a_{p, \boldsymbol{\theta}}} (-\log r)^{\eta}) \right\}. \end{aligned}$$

The upper and lower Logarithmic Mass Indices are

$$\overline{\text{MI}}_{\log}(p, \boldsymbol{\theta}) := \inf \mathcal{U}_{\log}, \quad \underline{\text{MI}}_{\log}(p, \boldsymbol{\theta}) := \sup \mathcal{L}_{\log}.$$

If $\overline{\text{MI}}_{\log}(p, \boldsymbol{\theta}) = \underline{\text{MI}}_{\log}(p, \boldsymbol{\theta})$, then the common value is denoted by $\text{MI}_{\log}(p, \boldsymbol{\theta})$.

We compare $(\text{MI}_{\text{pow}}(p, \boldsymbol{\theta}), \text{MI}_{\log}(p, \boldsymbol{\theta}))$ using the lexicographic order. Throughout the paper, we denote $\text{MI}_{\text{pow}}^{-1}(p, \boldsymbol{\theta}) := \frac{1}{\text{MI}_{\text{pow}}(p, \boldsymbol{\theta})}$ whenever $\text{MI}_{\text{pow}}(p, \boldsymbol{\theta}) \in (0, \infty)$.

4.1.1 Basic calibrations

The following proposition records the basic local calibrations of the Mass Index.

Proposition 4.2 (Basic local small-ball calibrations). *Let p be a probability measure on \mathbb{R}^d . For some $r_0 > 0$:*

- (1) Local hole. *If $p(B_{r_0}(\boldsymbol{\theta})) = 0$, then $\text{MI}_{\text{pow}}(p, \boldsymbol{\theta}) = 0$.*
- (2) Regular continuous mass. *If p admits a continuous density ρ_p on $B_{r_0}(\boldsymbol{\theta})$, and $\rho_p(\boldsymbol{\theta}) > 0$, then $\text{MI}_{\text{pow}}(p, \boldsymbol{\theta}) = 1$.*
- (3) Power-log asymptotics. *If $p(B_r(\boldsymbol{\theta})) \asymp r^a (-\log r)^b$, $r \downarrow 0$ with $a > 0$, then $\text{MI}_{\text{pow}}(p, \boldsymbol{\theta}) = da^{-1}$, $\text{MI}_{\log}(p, \boldsymbol{\theta}) = b$.*
- (4) Atomic mass. *If $p(\{\boldsymbol{\theta}\}) > 0$, then $\text{MI}_{\text{pow}}(p, \boldsymbol{\theta}) = \infty$.*

Interpretation. Proposition 4.2 gives basic local small-ball calibrations of the Mass Index. A local hole, regular continuous mass, and an atom correspond to $\text{MI}_{\text{pow}} = 0, 1, \infty$, respectively. The factor d normalises the reciprocal exponent against the ordinary full-dimensional scaling r^d .

Table 1 gives representative one-dimensional examples at $\boldsymbol{\theta} = 0$.

4.1.2 Scope of the definition

Although the Horseshoe prior and the Gaussian prior both have $\text{MI}_{\text{pow}} = 1$, the Horseshoe prior is widely used in Bayesian compression [11], and its behaviour near zero is closely connected with an additional logarithmic-order factor. Thus, a logarithmic refinement is needed if one wants to distinguish these two priors beyond their leading power order.

Table 1: Power and Logarithmic Mass Indices at 0.

Distribution	MI_{pow}	MI_{log}
Gaussian	1	0
Student- t_ν	1	0
Laplace	1	0
Horseshoe	1	1
Spike-and-slab	∞	-

The cases $MI_{\text{pow}} = 0$ and $MI_{\text{pow}} = \infty$ correspond to behaviours outside the polynomial scale, such as super-polynomial or logarithmic-type decay. Thus, the Power Mass Index should be viewed as a modest power-scale calibration. It captures the dominant polynomial order when meaningful, but it does not distinguish all finer sub-polynomial or super-polynomial behaviours.

4.2 RE-KL: A set-localised divergence

Avlogiaris et al. [1] introduced local divergences based on Csiszár ϕ -divergences, while Póczos et al. [25] studied global Tsallis- α divergences and their nonparametric estimation. Our construction is related to both: we use a Tsallis-type generator within the extended f -divergence framework [28, 29] and localise it to a measurable set E . The resulting quantity separates absolutely continuous and singular contributions and relates them to small-ball mass behaviour.

Definition 4.3 (Regularised extended Kullback-Leibler divergence). For $\alpha \in (0, 1)$, define

$$f_\alpha(x) := \frac{x^\alpha - \alpha x + \alpha - 1}{\alpha - 1}, \quad x \in [0, \infty).$$

The Lebesgue decomposition (Lemma 3.1) of q with respect to p is $q = q_{\ll p} + q_{\perp p}$. Then the RE-KL divergence on any measurable set E is defined as

$$D_\alpha^E(q||p) := \underbrace{\int_E f_\alpha \left(\frac{dq_{\ll p}}{dp} \right) dp}_{\text{absolutely continuous part}} + \underbrace{\frac{\alpha}{1-\alpha} q_{\perp p}(E)}_{\text{singular penalty}}.$$

4.2.1 Relation to standard divergences

On the whole space Θ with $q \ll p$:

- (1) KL divergence. When $\alpha \uparrow 1$, $D_\alpha^\Theta(q||p) \uparrow D_{\text{KL}}(q||p)$.
- (2) Hellinger distance $\mathcal{H}(q, p)$ [2]. When $\alpha = 1/2$, $D_{1/2}^\Theta(q||p) = 2\mathcal{H}^2(q, p) = 2(1 - \int_\Theta \sqrt{dq/dp} dp)$.
- (3) Tsallis- α divergence T_α [25]. When $\alpha \in (0, 1)$, $D_\alpha^\Theta(q||p) = T_\alpha(q||p) = (\alpha - 1)^{-1} (\int_\Theta q^\alpha(x) p^{1-\alpha}(x) dx - 1)$.

Thus, the present definition can be viewed as a mild localisation of the standard extended f -divergence, iso-

lating the contribution of a measurable set E while keeping track of the corresponding singular part.

4.3 Why global KL is not enough

Global KL divergence need not determine local probability mass. Related but different concerns appear in the GAN literature under mode collapse and incomplete mode coverage [19, 30]. We use this connection only as motivation for separating global closeness from local small-ball behaviour. The result below formulates this distinction through the Power Mass Index and shows that global KL convergence alone need not determine the local power order at θ .

Theorem 4.4 (Sequential discontinuity of the Power Mass Index under KL convergence). *Fix $\theta \in \Theta$. Define $\mathcal{D}_\theta := \{q \in \mathcal{P}(\Theta) : MI_{\text{pow}}(q, \theta) \in (0, \infty)\}$.*

Let $p \in \mathcal{D}_\theta$. Then, for every $k > 0$ with $k \neq MI_{\text{pow}}(p, \theta)$, there exists a sequence $\{q_n\}_{n \geq 1} \subseteq \mathcal{D}_\theta$ such that $D_{\text{KL}}(q_n||p) \rightarrow 0$, but $MI_{\text{pow}}(q_n, \theta) = k$ for every $n \geq 1$. In particular, the functional $\Phi_\theta : \mathcal{D}_\theta \rightarrow (0, \infty)$, $\Phi_\theta(q) := MI_{\text{pow}}(q, \theta)$ is not sequentially continuous at p under global KL convergence.

Interpretation. Theorem 4.4 gives a cautionary example for variational approximation. Even when $D_{\text{KL}}(q_n||p) \rightarrow 0$, global KL convergence alone need not preserve the local small-ball power order near θ . KL-small perturbations may still change the Power Mass Index at θ . This does not mean that larger variational families necessarily cause local distortion. It shows only that, without additional structural restrictions, such stability cannot be derived from the KL objective alone.

Proof sketch. Modify p only on $B_{\rho_n}(\theta)$, where $\rho_n \downarrow 0$, and keep $q_n = p$ outside. Since p has no atom at θ , the modified mass tends to zero. Redistribute this mass so that the local power index at θ equals k . Then $MI_{\text{pow}}(q_n, \theta) = k$ for all n , while the KL cost is confined to a set of vanishing p -mass, so $D_{\text{KL}}(q_n||p) \rightarrow 0$.

5 Local mass preservation theory

Using the two local tools developed in Section 4, we analyse the mathematical structure of local probability mass and local discrepancy under Bayesian update and variational approximation.

5.1 Bayesian updating and local mass scaling

We first study how Bayesian updating changes local mass.

5.1.1 Stability under Bayesian updating

We prove a result that characterises how the likelihood affects the preservation of local mass under Bayesian update. The analysis is based on the following two assumptions.

Assumption 1 (Finite marginal). $0 < Z_{\mathcal{D}} < \infty$.

Assumption 1 ensures that the posterior exists and is well-defined as a probability measure.

Assumption 2 (Well-defined Power Mass Indices). At θ , $\text{MI}_{\text{pow}}(p, \theta)$ exists.

Assumption 2 rules out scale-wise oscillations of small-ball probabilities by requiring the upper and lower Power Mass Indices to coincide.

Example A.1 shows that the MI may fail to exist. It is deliberately artificial: it builds infinite oscillations between two local power laws. Such behaviour is absent from standard priors with atoms, regular densities, finite mixtures, or regularly varying small-ball masses. Hence Assumption 2 rules out pathological oscillations, not ordinary Bayesian models.

Theorem 5.1 (Local likelihood scaling). *Under Assumption 1 and 2 with $p = \pi_0$, suppose that for some $\gamma, \beta \in \mathbb{R}$, the likelihood satisfies*

$$L(\mathbf{x} \mid \mathcal{D}) \asymp \|\mathbf{x} - \theta\|^\gamma (-\log \|\mathbf{x} - \theta\|)^\beta, \quad \mathbf{x} \rightarrow \theta.$$

When $\text{MI}_{\text{pow}}(\pi_0, \theta) \in (0, \infty)$, if $\gamma > -d \text{MI}_{\text{pow}}^{-1}(\pi_0, \theta)$, then $\text{MI}_{\text{pow}}(\pi_{\mathcal{D}}, \theta)$ exists and

$$\text{MI}_{\text{pow}}^{-1}(\pi_{\mathcal{D}}, \theta) = \text{MI}_{\text{pow}}^{-1}(\pi_0, \theta) + d^{-1}\gamma.$$

Furthermore, if $\text{MI}_{\log}(\pi_0, \theta)$ exists, then $\text{MI}_{\log}(\pi_{\mathcal{D}}, \theta)$ exists and

$$\text{MI}_{\log}(\pi_{\mathcal{D}}, \theta) = \text{MI}_{\log}(\pi_0, \theta) + \beta.$$

When $\text{MI}_{\text{pow}}(\pi_0, \theta) = 0$, then $\text{MI}_{\text{pow}}(\pi_{\mathcal{D}}, \theta)$ exists and

$$\text{MI}_{\text{pow}}(\pi_{\mathcal{D}}, \theta) = 0.$$

The condition $\gamma > -d \text{MI}_{\text{pow}}^{-1}(\pi_0, \theta)$ is a natural sufficient condition required by Assumption 1. We therefore restrict the theorem to this non-degenerate case and do not discuss the remaining degenerate cases in this paper.

Interpretation. Theorem 5.1 describes Bayesian updating as a local reweighting of the prior by the likelihood. If the likelihood is locally of power-log order, then the polynomial factor shifts the posterior Power Mass Index, while the logarithmic factor shifts only the posterior Logarithmic Mass Index.

In the regular case, where the likelihood is bounded above and below by positive constants near θ , we have

$\gamma = \beta = 0$. So Bayesian updating preserves both local mass scales. Thus local mass distortion under exact Bayesian updating arises only from non-regular local likelihood behaviour, such as zeros, singularities, logarithmic corrections, or support constraints.

Proof sketch. By Bayes' formula, the problem reduces to estimating the likelihood-weighted prior mass of $B_r(\theta)$. The upper estimates are obtained by decomposing $B_r(\theta)$ into a sequence of shrinking annuli: on each annulus, the likelihood is controlled according to its distance scale, and the upper Mass Index bounds are applied to the prior mass. For the lower estimates, we restrict the integral to an outer annulus. On these annuli, the likelihood has a uniform lower bound, while the discarded inner part is negligible, yielding the matching lower order.

5.1.2 Parameter-dependent support and surviving local mass

In economics, parameter-dependent support arises in certain parametric auction models, search models, among others [13]. In these models, the boundary of the support of the observed data depends on some parameters of interest and on regressor variables. The likelihood can then be written as

$$L(\mathbf{x} \mid \mathcal{D}) = g(\mathbf{x}) \mathbf{1}_E(\mathbf{x}), \quad \text{with } g(\mathbf{x}) \asymp 1, \quad \mathbf{x} \rightarrow \theta. \quad (5.1)$$

The indicator $\mathbf{1}_E$ may remove part, or all, of the local prior mass near θ . The following theorem characterises the effect of such removal on local-mass preservation.

Theorem 5.2 (Acceptance-ratio scaling of local Mass Indices). *Under Assumption 1, assume also that*

$$0 < \text{MI}_{\text{pow}}(\pi_0, \theta) < \infty.$$

The likelihood satisfies (5.1). Define the local acceptance ratio

$$A_{E, \theta}(r) := \frac{\pi_0(B_r(\theta) \cap E)}{\pi_0(B_r(\theta))} \in [0, 1],$$

whenever $\pi_0(B_r(\theta)) > 0$. If

$$A_{E, \theta}(r) \asymp r^\delta (-\log r)^\lambda, \quad r \rightarrow 0,$$

where $\delta > 0$ or $\delta = 0, \lambda \leq 0$, then $\text{MI}_{\text{pow}}(\pi_{\mathcal{D}}, \theta)$ exists and

$$\text{MI}_{\text{pow}}^{-1}(\pi_{\mathcal{D}}, \theta) = \text{MI}_{\text{pow}}^{-1}(\pi_0, \theta) + d^{-1}\delta.$$

Furthermore, if $\text{MI}_{\log}(\pi_0, \theta)$ exists, then $\text{MI}_{\log}(\pi_{\mathcal{D}}, \theta)$ exists and

$$\text{MI}_{\log}(\pi_{\mathcal{D}}, \theta) = \text{MI}_{\log}(\pi_0, \theta) + \lambda.$$

Interpretation. The quantity $A_{E, \theta}(r)$ measures the fraction of the local prior mass around θ that is not

removed by the hard constraint $\mathbf{1}_E$. Since $g(\mathbf{x}) \asymp 1$ near $\boldsymbol{\theta}$, the smooth part of the likelihood does not change the local mass scale. Hence the only local effect of the hard constraint is the multiplicative factor $A_{E,\boldsymbol{\theta}}(r) \asymp r^\delta (-\log r)^\lambda$. Thus the hard constraint adds δ to the local power exponent and λ to the local logarithmic exponent.

Proof sketch. The proof is a direct analogue of Theorem 5.1. Since $g(\mathbf{x}) \asymp 1$ near $\boldsymbol{\theta}$, Bayes' formula gives

$$\pi_{\mathcal{D}}(B_r(\boldsymbol{\theta})) \asymp \pi_0(B_r(\boldsymbol{\theta}) \cap E) = A_{E,\boldsymbol{\theta}}(r)\pi_0(B_r(\boldsymbol{\theta})).$$

Thus $A_{E,\boldsymbol{\theta}}(r) \asymp r^\delta (-\log r)^\lambda$ shifts the local power and logarithmic exponents by δ and λ , respectively. The condition on δ guarantees that the posterior power exponent remains positive. The identities follow.

In practice, hard constraints, indicator factors, and parameter-dependent supports are often smoothed for optimisation or sampling. Such softening is not locally innocuous: it can change small-ball mass scales and hence the local mass concentration induced by the hard formulation.

Corollary 5.3 (Softening and non-commuting local limits). *Let $\{s_\tau\}_{\tau>0}$ be positive continuous functions. Assume that there exists $C > 0$ such that $0 < s_\tau \leq C$ for every $\tau > 0$, and that*

$$\lim_{\tau \downarrow 0} s_\tau(\mathbf{x}) = \mathbf{1}_E(\mathbf{x}) \quad \text{for } \pi_0\text{-a.e. } \mathbf{x}.$$

Assume that Assumption 1 holds for the hard likelihood $L = \mathbf{1}_E$. Assume also that $0 < \text{MI}_{\text{pow}}(\pi_0, \boldsymbol{\theta}) < \infty$. For each $\tau > 0$, denote by π_τ the posterior obtained from the softened likelihood s_τ , and denote by $\pi_{\mathcal{D}}$ the hard posterior obtained from $\mathbf{1}_E$. Then

$$\|\pi_\tau - \pi_{\mathcal{D}}\|_{\text{TV}} \rightarrow 0, \quad \tau \downarrow 0.$$

If

$$A_{E,\boldsymbol{\theta}}(r) \asymp r^\delta (-\log r)^\lambda, \quad r \downarrow 0,$$

with $\delta > 0$, then

$$\text{MI}_{\text{pow}}(\pi_{\mathcal{D}}, \boldsymbol{\theta}) \neq \lim_{\tau \downarrow 0} \text{MI}_{\text{pow}}(\pi_\tau, \boldsymbol{\theta}).$$

Furthermore, assume that $\text{MI}_{\log}(\pi_0, \boldsymbol{\theta})$ exists. If

$$A_{E,\boldsymbol{\theta}}(r) \asymp (-\log r)^\lambda, \quad r \downarrow 0,$$

with $\lambda < 0$, then

$$\text{MI}_{\log}(\pi_{\mathcal{D}}, \boldsymbol{\theta}) \neq \lim_{\tau \downarrow 0} \text{MI}_{\log}(\pi_\tau, \boldsymbol{\theta}).$$

Interpretation. This corollary shows that the local limit $r \rightarrow 0$ and the zero-temperature limit $\tau \downarrow 0$ need not commute. For fixed $\tau > 0$, the soft factor s_τ is locally positive near $\boldsymbol{\theta}$, so the local MI is preserved. In the

hard limit, the constraint may retain only a fraction $A_{E,\boldsymbol{\theta}}(r) \asymp r^\delta (-\log r)^\lambda$ of the local prior mass, and therefore changes the local mass scale unless $\delta = \lambda = 0$. Thus softening is not locally innocuous: it may fail to reproduce the local mass-concentration structure of the hard-constrained model.

5.2 Local RE-KL analysis for Mass Indices

MI provides a way to characterise the local mass of a probability distribution. The local RE-KL bounds below give local bounds and sufficient conditions for comparing the mass behaviour of two measures.

Local RE-KL bounds. We first show how the absolutely continuous component and the singular component jointly contribute to the absolute error.

Theorem 5.4 (Absolute error bound). *For any measurable set E ,*

$$|q(E) - p(E)| \leq 2\sqrt{C_\alpha D_\alpha^E(q||p)},$$

where $C_\alpha = \max\{1, \alpha^{-1}(1 - \alpha)\}$.

Interpretation. Theorem 5.4 gives a local absolute error control for the mass assigned to a measurable region E . Small local RE-KL discrepancy on E forces $q(E)$ and $p(E)$ to be close in additive mass. Importantly, the bound remains meaningful even when q has a p -singular component: singular mass does not make the right-hand side automatically infinite, but is charged through the finite recession penalty in $D_\alpha^E(q||p)$.

Proof sketch. Decompose q into its p -absolutely continuous and p -singular parts. For the former, rewrite the local mass error in Hellinger form and apply Cauchy-Schwarz. The RE-KL term dominates the resulting quantity up to C_α . The singular part is already charged by the recession term in $D_\alpha^E(q||p)$. Combining these two estimates yields the bound.

Theorem 5.4 leads to Corollary 5.5 directly.

Corollary 5.5. *Assume that for some $\delta > 0$, $D_\alpha^{B_r(\boldsymbol{\theta})}(q||p) = O(r^\delta)$, $r \rightarrow 0$. We have*

$$|q(B_r(\boldsymbol{\theta})) - p(B_r(\boldsymbol{\theta}))| = O(r^{\delta/2}).$$

We next derive a relative bound for the local mass ratio $q(E)/p(E)$.

Theorem 5.6 (Relative error bound). *Denote the normalised RE-KL divergence as $\overline{D}_\alpha^E(q||p) := D_\alpha^E(q||p)/p(E)$ whenever $p(E) > 0$. Then*

$$f_\alpha\left(\frac{q(E)}{p(E)}\right) \leq \overline{D}_\alpha^E(q||p).$$

Furthermore, we can obtain a relative error bound:

$$\left| 1 - \frac{q(E)}{p(E)} \right| \leq \frac{2}{\alpha} \sqrt{\overline{D}_\alpha^E(q||p)} \sqrt{\overline{D}_\alpha^E(q||p) + 2}.$$

Interpretation. Theorem 5.6 upgrades absolute error control to relative error control. This is needed for MI: when $p(E)$ is itself small, absolute error is not enough to prevent $q(E)$ from deviating too much in relative terms, whereas controlling $q(E)/p(E)$ prevents asymptotic local under- or over-coverage.

Proof sketch. Jensen’s inequality controls the absolutely continuous part, while the recession slope of f_α absorbs the singular part. Then a scalar lower bound for f_α near its minimum at 1 converts this control into a quadratic inequality for $|1 - q(E)/p(E)|$, whose solution gives the displayed relative error bound.

Theorem 5.6 leads to Corollary 5.7 directly.

Corollary 5.7. *Suppose $p(B_r(\boldsymbol{\theta})) > 0$ for all sufficiently small r . Assume that for some $\delta > 0$, $\overline{D}_\alpha^{B_r(\boldsymbol{\theta})}(q||p) = O(r^\delta)$, $r \rightarrow 0$. We have*

$$\left| 1 - \frac{q(B_r(\boldsymbol{\theta}))}{p(B_r(\boldsymbol{\theta}))} \right| = O(r^{\delta/2}).$$

Directional local RE–KL implications for Mass Indices.

We next use Theorem 5.6 to compare the local small-ball scales of two measures. In variational approximation, $p = \pi_{\mathcal{D}}$ denotes the posterior and q denotes the approximation. The comparison of interest is the one-sided relation $\text{MI}_{\text{pow}}(q, \boldsymbol{\theta}) \geq \text{MI}_{\text{pow}}(p, \boldsymbol{\theta})$. The following result gives directional sufficient conditions for this relation, and for equality in the stronger case.

Theorem 5.8 (Directional sufficient conditions for Power Mass Index comparison). *Assume $p(B_r(\boldsymbol{\theta})) > 0$ and $q(B_r(\boldsymbol{\theta})) > 0$ for all sufficiently small $r > 0$. Suppose Assumption 2 holds with p and q .*

(i) *If $\limsup_{r \rightarrow 0} \overline{D}_\alpha^{B_r(\boldsymbol{\theta})}(q||p) < 1$,*

then $\text{MI}_{\text{pow}}(q, \boldsymbol{\theta}) = \text{MI}_{\text{pow}}(p, \boldsymbol{\theta})$. If $\text{MI}_{\log}(p, \boldsymbol{\theta})$ exists, then $\text{MI}_{\log}(q, \boldsymbol{\theta}) = \text{MI}_{\log}(p, \boldsymbol{\theta})$.

(ii) *If $\limsup_{r \rightarrow 0} \overline{D}_\alpha^{B_r(\boldsymbol{\theta})}(p||q) < \infty$, then $\text{MI}_{\text{pow}}(q, \boldsymbol{\theta}) \geq \text{MI}_{\text{pow}}(p, \boldsymbol{\theta})$. Furthermore, if $\text{MI}_{\text{pow}}(q, \boldsymbol{\theta}) = \text{MI}_{\text{pow}}(p, \boldsymbol{\theta})$, and both $\text{MI}_{\log}(p, \boldsymbol{\theta})$ and $\text{MI}_{\log}(q, \boldsymbol{\theta})$ exist, then $\text{MI}_{\log}(q, \boldsymbol{\theta}) \geq \text{MI}_{\log}(p, \boldsymbol{\theta})$.*

Interpretation. The two parts describe different strengths of local information. The $q||p$ direction is rigid: a merely finite value of $\overline{D}_\alpha^{B_r(\boldsymbol{\theta})}(q||p)$ is not enough at the power scale, and one needs the subcritical condition $\limsup_{r \rightarrow 0} \overline{D}_\alpha^{B_r(\boldsymbol{\theta})}(q||p) < 1$. Once this condition is imposed, there is no remaining freedom at the leading small-ball scale: q and p have the same power index,

and also the same logarithmic index whenever the latter is defined for p .

The $p||q$ direction is weaker but more permissive. Finiteness of $\limsup_{r \rightarrow 0} \overline{D}_\alpha^{B_r(\boldsymbol{\theta})}(p||q)$ is enough only for a one-sided comparison: it precludes q from being locally thinner than p at the polynomial scale, but it still allows q to assign more mass near $\boldsymbol{\theta}$. Thus this direction leaves room for local over-concentration of the approximation. This asymmetry is consistent with the usual distinction between KL directions in variational approximation [22, 23, 24]. These are conditional local implications, not a characterisation of minimisers of the global variational objective.

Proof sketch. Apply the local RE–KL mass bound to $B_r(\boldsymbol{\theta})$. In the $q||p$ direction, the subcritical bound forces equality of the power scale, and then of the logarithmic scale. In the $p||q$ direction, finite normalised divergence only rules out polynomially smaller $q(B_r(\boldsymbol{\theta}))$, giving the one-sided Power Mass Index bound. When the power scales coincide, the logarithmic comparison gives the final claim.

Example A.2 illustrates that the two directions encode genuinely different local comparisons. The condition in the $p||q$ direction permits the conclusion that q has at least as large a local power mass index as p , while the reverse condition is stronger here and is not satisfied.

6 Experiments

We give three small-scale checks of the proposed local quantities. These experiments are not intended as scalable diagnostics for large neural networks, but as controlled sanity checks. For the real-data experiment, we use three UCI binary tasks: Breast Cancer, Iris 0 vs 1, and Wine 0 vs 1. Each model is Bayesian logistic regression with four PCA covariates and an intercept, so $d = 5$. We use a Gaussian prior with scale 2, approximate the posterior by a Laplace Gaussian, set $\boldsymbol{\theta}_0$ to the Laplace posterior mean, and estimate Euclidean small-ball masses by Sobol quadrature over $r \in [0.03, 0.45]$. UCI results are averaged over five seeds. Full experimental details are given in Appendix C.

Experiment 1: Synthetic calibration. Figure 1 shows the basic regimes used in the theory. A regular Gaussian follows the reference law $p(B_r(\boldsymbol{\theta}_0)) \asymp r^d$. The power-law examples $p(B_r(\boldsymbol{\theta}_0)) \asymp r^{d+\beta}$ show how the parameter β changes the local small-ball mass order. The log-singular and spike-slab examples illustrate logarithmic correction and atomic mass. The exact curves used in this calibration are listed in Appendix C.1.

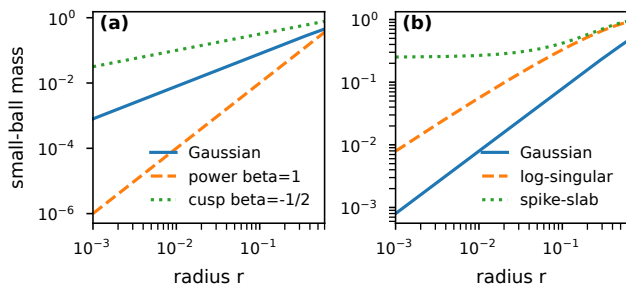


Figure 1: Synthetic small-ball calibration.

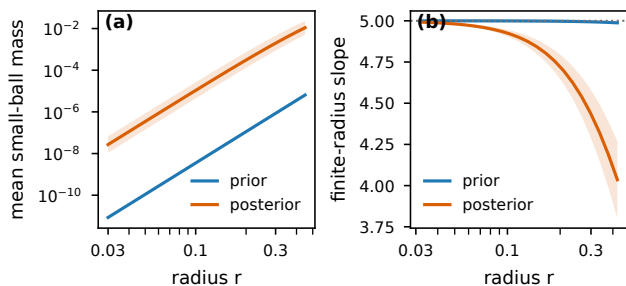


Figure 2: UCI prior/posterior small-ball masses and finite-radius slopes.

 Table 2: Small-radius slopes fitted over $r \leq 0.06$.

Dataset	n	prior	posterior
Breast Cancer	398	5.000 ± 0.000	4.969 ± 0.005
Iris 0 vs 1	70	5.000 ± 0.000	4.999 ± 0.000
Wine 0 vs 1	91	5.000 ± 0.000	4.993 ± 0.001

Experiment 2: UCI Bayesian sanity check. Figure 2 shows that posterior small-ball mass around the Laplace mean is much larger than prior mass. As shown in Table 2, at the smallest radii, both prior and posterior slopes remain close to $d = 5$. This matches the regular case: Bayesian updating changes the local density level, but not the leading local power order. The posterior slope decrease at larger radii is a finite-radius effect. Dataset preprocessing, posterior fitting, quadrature, and slope computation are described in Appendix C.2.

Experiment 3: Directionality of local RE-KL. Figure 3 illustrates that local RE-KL control is directional. On the same shrinking neighbourhoods, the $p \parallel q$ direction stays bounded while the $q \parallel p$ direction diverges. Thus the order of the two arguments determines which local mismatch is detected. The explicit construction is given in Appendix C.3.

7 Conclusion

This paper studied local small-ball mass $p(B_r(\theta))$ as a measure-level quantity in Bayesian inference. The aim was not to propose a new global divergence or a

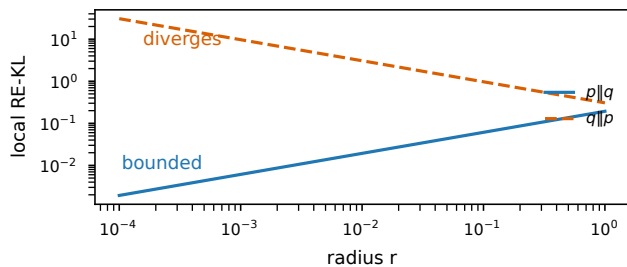


Figure 3: Directionality of local RE-KL.

scalable diagnostic method, but to make precise how local mass scales can change under Bayesian updating and approximation.

We made three concrete contributions. First, we used the Mass Index to record the leading power order and the first logarithmic correction of local small-ball probabilities. Second, we derived local scale identities for Bayesian updating, showing how regular likelihoods preserve the local scale, how power-log likelihood factors shift it, and how parameter-dependent supports act through the surviving fraction of prior mass. Third, we used local RE-KL to prove absolute, relative, and directional bounds for comparing small-ball masses, yielding sufficient conditions for equality or one-sided comparison of Mass Indices under the two KL directions.

These results should be read as local asymptotic scale comparisons rather than a characterisation of global variational minimisers. The experiments provide finite-radius illustrations of the quantities and regimes appearing in the theory.

References

- [1] G Avlogiaris, A Micheas, and K Zografos. On local divergences between two probability measures. *Metrika*, 79(3):303–333, 2016.
- [2] Rudolf Beran. Minimum hellinger distance estimates for parametric models. *The annals of Statistics*, pages 445–463, 1977.
- [3] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.
- [4] Nicholas H Bingham, Charles M Goldie, and Jef L Teugels. *Regular variation*, volume 27. Cambridge university press, 1989.
- [5] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017.
- [6] Ismaël Castillo, Johannes Schmidt-Hieber, and Aad van der Vaart. Bayesian linear regression with

- sparse priors. *The Annals of Statistics*, 43(5):1986–2018, 2015.
- [7] Kenneth J Falconer and KJ Falconer. *Techniques in fractal geometry*, volume 3. Wiley Chichester, 1997.
- [8] Gerald B Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 1999.
- [9] Vincent Fortuin, Adrià Garriga-Alonso, Sebastian W. Ober, Florian Wenzel, Gunnar Ratsch, Richard E Turner, Mark van der Wilk, and Laurence Aitchison. Bayesian neural network priors revisited. In *International Conference on Learning Representations*, 2022.
- [10] Subhashis Ghosal, Jayanta K Ghosh, and Aad W Van Der Vaart. Convergence rates of posterior distributions. *Annals of Statistics*, pages 500–531, 2000.
- [11] Soumya Ghosh, Jiayu Yao, and Finale Doshi-Velez. Model selection in bayesian neural networks via horseshoe priors. *Journal of Machine Learning Research*, 20(182):1–46, 2019.
- [12] Yanick Heurteaux. Dimension of measures: the probabilistic approach. *Publicacions Matemàtiques*, 51(2):243–290, 2007.
- [13] Keisuke Hirano and Jack R Porter. Asymptotic efficiency in parametric structural models with parameter-dependent support. *Econometrica*, 71(5):1307–1338, 2003.
- [14] Jovan Karamata. Sur un mode de croissance régulière des fonctions. *Mathematica (Cluj)*, 4:38–53, 1930.
- [15] Jovan Karamata. Sur un mode de croissance régulière. théorèmes fondamentaux. *Bulletin de la Société Mathématique de France*, 61:55–62, 1933.
- [16] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [17] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.
- [18] Yingzhen Li and Richard Turner. Rényi Divergence Variational Inference. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.
- [19] Zinan Lin, Ashish Khetan, Giulia Fanti, and Sewoong Oh. Pacgan: The power of two samples in generative adversarial networks. *Advances in neural information processing systems*, 31, 2018.
- [20] Christos Louizos, Karen Ullrich, and Max Welling. Bayesian compression for deep learning. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [21] Pertti Mattila. *Geometry of sets and measures in Euclidean spaces: fractals and rectifiability*, volume 44 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999.
- [22] Declan McNamara, Jackson Loper, and Jeffrey Regier. Sequential Monte Carlo for Inclusive KL Minimization in Amortized Variational Inference. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, pages 4312–4320. PMLR, April 2024.
- [23] Thomas P. Minka. Divergence measures and message passing. Technical Report MSR-TR-2005-173, Microsoft Research, 2005.
- [24] Christian Naesseth, Fredrik Lindsten, and David Blei. Markovian score climbing: Variational inference with $\text{kl}(p||q)$. *Advances in Neural Information Processing Systems*, 33:15499–15510, 2020.
- [25] Barnabás Póczos and Jeff Schneider. On the estimation of α -divergences. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 609–617. JMLR Workshop and Conference Proceedings, 2011.
- [26] Nicholas G Polson and Veronika Ročková. Posterior Concentration for Sparse Deep Learning. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.
- [27] Aad W. van der Vaart and J. Harry van Zanten. Rates of contraction of posterior distributions based on gaussian process priors. *The Annals of Statistics*, 36(3):1435–1463, 2008.
- [28] Neng Wan, Dapeng Li, and Naira Hovakimyan. F-divergence variational inference. *Advances in neural information processing systems*, 33:17370–17379, 2020.
- [29] Mingtian Zhang, Thomas Bird, Raza Habib, Tianlin Xu, and David Barber. Variational f-divergence Minimization, December 2024. arXiv:1907.11891 [stat.ML].

- [30] Peilin Zhong, Yuchen Mo, Chang Xiao, Pengyu Chen, and Changxi Zheng. Rethinking generative mode coverage: A pointwise guaranteed approach. *Advances in Neural Information Processing Systems*, 32, 2019.

A Examples and Counterexamples

Example A.1 (Oscillating small-ball order). This example shows why the upper and lower Power Mass Indices need not coincide. We construct a probability measure whose small-ball mass behaves like r^a along one sequence of radii, but like r^b along another sequence of radii, where $0 < a < b$.

For simplicity, take $\Theta = \mathbb{R}^d$. The same construction works whenever Θ contains a sufficiently small line segment starting from θ . Fix $0 < a < b$, and set

$$t_n := \left(\frac{2b}{a}\right)^n, \quad n = 0, 1, 2, \dots$$

We define an increasing continuous function ϕ on large positive values of t by prescribing its values on the sequence (t_n) , and then interpolating linearly between consecutive points:

$$\phi(t_{2m}) = at_{2m}, \quad \phi(t_{2m+1}) = bt_{2m+1}.$$

The prescribed values are strictly increasing. Indeed,

$$bt_{2m+1} > at_{2m}, \quad at_{2m+2} = 2bt_{2m+1}.$$

Hence the piecewise-linear interpolation is increasing.

Now define, for small $r > 0$,

$$m(r) := \exp\{-\phi(\log(1/r))\}.$$

As $r \downarrow 0$, we have $\log(1/r) \rightarrow \infty$. Since ϕ is increasing and tends to infinity, $m(r) \downarrow 0$. Also, as a function of r , $m(r)$ is increasing. Thus m can be used as the distribution function of radial mass near θ .

Choose $\varepsilon > 0$ such that $m(\varepsilon) < 1$. By the Lebesgue–Stieltjes construction, there exists a finite measure ν on $(0, \varepsilon]$ such that

$$\nu((0, r]) = m(r), \quad 0 < r \leq \varepsilon.$$

Push this measure forward by the map

$$s \mapsto \theta + se_1,$$

where e_1 is the first coordinate vector. Finally, place the remaining mass $1 - m(\varepsilon)$ outside $B_\varepsilon(\theta)$. This defines a probability measure p .

Since m is continuous, the measure ν has no atoms. Therefore no mass is placed on spheres centred at θ . Consequently, for every $0 < r \leq \varepsilon$,

$$p(B_r(\theta)) = m(r).$$

Let

$$r_n := e^{-t_n}.$$

Then $\log(1/r_n) = t_n$. Along the even subsequence, we obtain

$$p(B_{r_{2m}}(\theta)) = \exp\{-\phi(t_{2m})\} = \exp\{-at_{2m}\} = r_{2m}^a.$$

Along the odd subsequence, we obtain

$$p(B_{r_{2m+1}}(\theta)) = \exp\{-\phi(t_{2m+1})\} = \exp\{-bt_{2m+1}\} = r_{2m+1}^b.$$

Moreover, by construction, the ratio $\phi(t)/t$ stays between a and b on each interpolation interval. Hence, for all sufficiently small r ,

$$r^b \leq p(B_r(\theta)) \leq r^a.$$

The even subsequence shows that the upper power order cannot be better than a , while the odd subsequence shows that the lower power order cannot be better than b . Therefore

$$\overline{\text{MI}}_{\text{pow}}(p, \theta) = \frac{d}{a}, \quad \underline{\text{MI}}_{\text{pow}}(p, \theta) = \frac{d}{b}.$$

Since $a < b$, these two values are different. Hence the Power Mass Index at θ does not exist.

Example A.2 (Directionality of local RE–KL conditions). This example illustrates that the local RE–KL conditions are directional. We work at the point 0. Let p and q be probability measures on $[-1, 1]$ with densities

$$\frac{dp}{dx}(x) = \frac{1}{2}, \quad \frac{dq}{dx}(x) = \frac{1}{4}|x|^{-1/2}.$$

Both are probability measures, since

$$\int_{-1}^1 \frac{1}{2} dx = 1, \quad \int_{-1}^1 \frac{1}{4}|x|^{-1/2} dx = 1.$$

For $0 < r < 1$, writing $B_r = B_r(0) = (-r, r)$, we have

$$p(B_r) = \int_{-r}^r \frac{1}{2} dx = r,$$

whereas

$$q(B_r) = \int_{-r}^r \frac{1}{4}|x|^{-1/2} dx = r^{1/2}.$$

Therefore

$$\text{MI}_{\text{pow}}(p, 0) = 1, \quad \text{MI}_{\text{pow}}(q, 0) = 2.$$

Thus q has more local mass near 0 than p , in the power-order sense.

We first consider the direction $p||q$. Since

$$\frac{dp}{dq}(x) = 2|x|^{1/2},$$

the likelihood ratio tends to 0 as $x \rightarrow 0$. Moreover, on B_r ,

$$0 \leq \frac{dp}{dq}(x) \leq 2r^{1/2}.$$

Hence the ratio dp/dq converges uniformly to 0 on B_r as $r \downarrow 0$. By continuity of f_α at 0,

$$f_\alpha\left(\frac{dp}{dq}(x)\right) \longrightarrow f_\alpha(0) \quad \text{uniformly on } B_r.$$

Consequently,

$$\overline{D}_\alpha^{B_r}(p||q) := \frac{1}{q(B_r)} \int_{B_r} f_\alpha\left(\frac{dp}{dq}\right) dq \longrightarrow f_\alpha(0) = 1.$$

Thus the local RE–KL condition in the direction $p||q$ is satisfied. The resulting one-sided conclusion is exactly the expected one:

$$\text{MI}_{\text{pow}}(q, 0) \geq \text{MI}_{\text{pow}}(p, 0).$$

The reverse direction behaves differently. In the direction $q||p$, we have

$$\frac{dq}{dp}(x) = \frac{1}{2}|x|^{-1/2},$$

which diverges as $x \rightarrow 0$. For large t ,

$$f_\alpha(t) \sim \frac{\alpha}{1-\alpha} t.$$

Therefore, near 0,

$$f_\alpha\left(\frac{dq}{dp}(x)\right) \asymp \frac{dq}{dp}(x).$$

It follows that

$$\int_{B_r} f_\alpha\left(\frac{dq}{dp}\right) dp \asymp \int_{B_r} \frac{dq}{dp} dp = q(B_r).$$

After normalisation by $p(B_r)$, we obtain

$$\overline{D}_\alpha^{B_r}(q||p) := \frac{1}{p(B_r)} \int_{B_r} f_\alpha\left(\frac{dq}{dp}\right) dp \asymp \frac{q(B_r)}{p(B_r)} = \frac{r^{1/2}}{r} = r^{-1/2} \longrightarrow \infty.$$

Hence the corresponding $q||p$ local RE–KL condition fails. This shows that the two directions are not interchangeable: $p||q$ controls a different local comparison from $q||p$.

B Proofs

Lemma B.1 (Gauge identification). *Let $m(r) \geq 0$ be a small-ball mass function and let $\ell(r) := -\log r$.*

(i) *Assume $a > 0$. If, for every $\varepsilon > 0$,*

$$m(r) = O(r^{a-\varepsilon}), \quad m(r) = \Omega(r^{a+\varepsilon}), \quad r \downarrow 0,$$

then the Power Mass Index exists and equals d/a .

(ii) *Assume that the power exponent is $a > 0$. If, for some $b \in \mathbb{R}$ and every $\varepsilon > 0$,*

$$m(r) = O(r^a \ell(r)^{b+\varepsilon}), \quad m(r) = \Omega(r^a \ell(r)^{b-\varepsilon}), \quad r \downarrow 0,$$

then the Logarithmic Mass Index exists and equals b .

Proof. For (i), the upper bounds imply that every exponent strictly below a is an upper power gauge, while the lower bounds imply that every exponent strictly above a is a lower power gauge. Therefore the sharp threshold in exponent form is a . Since Definition 4.1 parametrises power bounds as $r^{1/\eta}$, this gives

$$\inf \mathcal{U}_{\text{pow}} = \sup \mathcal{L}_{\text{pow}} = \frac{1}{a},$$

and hence $\text{MI}_{\text{pow}} = d/a$.

For (ii), the definition of the logarithmic gauge is already taken at the identified power scale a . The upper bounds imply that every logarithmic exponent strictly above b is an upper log gauge, and the lower bounds imply that every logarithmic exponent strictly below b is a lower log gauge. Thus

$$\inf \mathcal{U}_{\text{log}} = \sup \mathcal{L}_{\text{log}} = b,$$

so $\text{MI}_{\text{log}} = b$. □

Lemma B.2 (Radial power transform has bounded KL). *Let μ be a probability measure on $[0, \rho)$, and write $F(r) := \mu([0, r])$. Fix $t > 0$, and let ν be the probability measure on $[0, \rho)$ determined by*

$$\nu([0, r]) = F(r)^t, \quad 0 < r < \rho.$$

Then $\nu \ll \mu$, and

$$D_{\text{KL}}(\nu \parallel \mu) \leq \log t + \frac{1}{t} - 1.$$

Proof. Let λ be Lebesgue measure on $[0, 1]$, and let λ_t have density

$$\frac{d\lambda_t}{d\lambda}(u) = tu^{t-1}, \quad 0 < u < 1.$$

Let Q be a quantile map that pushes λ forward to μ . Then the measure defined above is the pushforward $Q_{\#}\lambda_t$. Since $\lambda_t \ll \lambda$, we have $\nu \ll \mu$. By data processing for KL divergence,

$$D_{\text{KL}}(\nu \parallel \mu) \leq D_{\text{KL}}(\lambda_t \parallel \lambda).$$

The right-hand side is explicit:

$$D_{\text{KL}}(\lambda_t \parallel \lambda) = \int_0^1 tu^{t-1} \log(tu^{t-1}) du = \log t + \frac{1}{t} - 1.$$

This proves the claim. □

Proof of Proposition 4.2

Proof. Let

$$m(r) := p(B_r(\boldsymbol{\theta})), \quad \ell(r) := -\log r.$$

(1) Local hole. If $p(B_{r_0}(\boldsymbol{\theta})) = 0$, then $m(r) = 0$ for all $0 < r < r_0$. Hence $m(r) = O(r^{1/\eta})$ for every $\eta > 0$, so $\inf \mathcal{U}_{\text{pow}} = 0$. On the other hand, $m(r) = \Omega(r^{1/\eta})$ holds for no $\eta > 0$. By the convention $\sup \emptyset = 0$, we also have $\sup \mathcal{L}_{\text{pow}} = 0$. Thus

$$\text{MI}_{\text{pow}}(p, \boldsymbol{\theta}) = 0.$$

(2) Regular continuous mass. Suppose that p admits a continuous density ρ_p on $B_{r_0}(\boldsymbol{\theta})$ and $\rho_p(\boldsymbol{\theta}) > 0$. Then, for some $r_1 \in (0, r_0)$ and constants $0 < c < C < \infty$,

$$c \leq \rho_p(x) \leq C, \quad x \in B_{r_1}(\boldsymbol{\theta}).$$

Therefore, for $0 < r < r_1$,

$$c \text{Vol}(B_r(\boldsymbol{\theta})) \leq p(B_r(\boldsymbol{\theta})) \leq C \text{Vol}(B_r(\boldsymbol{\theta})).$$

Since $\text{Vol}(B_r(\boldsymbol{\theta})) \asymp r^d$, we get

$$p(B_r(\boldsymbol{\theta})) \asymp r^d.$$

The definition of the Power Mass Index then gives

$$\text{MI}_{\text{pow}}(p, \boldsymbol{\theta}) = 1.$$

(3) Power-log asymptotics. Assume

$$p(B_r(\boldsymbol{\theta})) \asymp r^a \ell(r)^b, \quad r \downarrow 0.$$

For every $\varepsilon > 0$, logarithmic factors are slower than powers, hence

$$r^a \ell(r)^b = O(r^{a-\varepsilon}), \quad r^a \ell(r)^b = \Omega(r^{a+\varepsilon}).$$

Lemma B.1(i) gives

$$\text{MI}_{\text{pow}}(p, \boldsymbol{\theta}) = \frac{d}{a}.$$

At the corresponding power scale $a_{p, \boldsymbol{\theta}} = a$, the two-sided asymptotic also gives, for every $\varepsilon > 0$,

$$p(B_r(\boldsymbol{\theta})) = O(r^a \ell(r)^{b+\varepsilon}), \quad p(B_r(\boldsymbol{\theta})) = \Omega(r^a \ell(r)^{b-\varepsilon}).$$

Lemma B.1(ii) gives

$$\text{MI}_{\log}(p, \boldsymbol{\theta}) = b.$$

(4) Atomic mass. If $p(\{\boldsymbol{\theta}\}) > 0$, then $m(r) \geq p(\{\boldsymbol{\theta}\}) > 0$ for every $r > 0$. Hence $m(r) = O(r^{1/\eta})$ holds for no $\eta > 0$, so $\inf \mathcal{U}_{\text{pow}} = \infty$. Also $m(r) = \Omega(r^{1/\eta})$ holds for every $\eta > 0$, so $\sup \mathcal{L}_{\text{pow}} = \infty$. Therefore

$$\text{MI}_{\text{pow}}(p, \boldsymbol{\theta}) = \infty.$$

The proof is complete. □

Proof of the RE-KL relations

Proof. Assume first that $q \ll p$, and write

$$R := \frac{dq}{dp}.$$

Then

$$D_\alpha^\ominus(q||p) = \int_\Theta f_\alpha(R) dp = \frac{\int_\Theta R^\alpha dp - 1}{\alpha - 1} = \frac{1 - \int_\Theta R^\alpha dp}{1 - \alpha}.$$

If both measures are represented by densities with respect to a common dominating measure, this is exactly

$$(\alpha - 1)^{-1} \left(\int q(x)^\alpha p(x)^{1-\alpha} dx - 1 \right),$$

which is the Tsallis- α form used in the main text.

For $\alpha = 1/2$,

$$f_{1/2}(x) = (\sqrt{x} - 1)^2.$$

Therefore

$$D_{1/2}^\Theta(q\|p) = \int_\Theta (\sqrt{R} - 1)^2 dp = 2 \left(1 - \int_\Theta \sqrt{R} dp \right) = 2\mathcal{H}^2(q, p),$$

under the Hellinger convention used in the paper.

It remains to justify the monotone KL limit. For $x > 0$, differentiating

$$f_\alpha(x) = \frac{\alpha x - x^\alpha + 1 - \alpha}{1 - \alpha}$$

with respect to α gives

$$\partial_\alpha f_\alpha(x) = \frac{x - x^\alpha \{1 + (1 - \alpha) \log x\}}{(1 - \alpha)^2} = \frac{x^\alpha}{(1 - \alpha)^2} (x^{1-\alpha} - 1 - \log x^{1-\alpha}).$$

The elementary inequality $u - 1 - \log u \geq 0$, $u > 0$, therefore implies $\partial_\alpha f_\alpha(x) \geq 0$. At $x = 0$, $f_\alpha(0) = 1$ for all $\alpha \in (0, 1)$, so monotonicity is trivial. Moreover, for each fixed $x \geq 0$, a Taylor expansion of x^α at $\alpha = 1$, with the endpoint values interpreted by continuity, gives

$$f_\alpha(x) \uparrow x \log x - x + 1, \quad \alpha \uparrow 1.$$

By monotone convergence,

$$\lim_{\alpha \uparrow 1} D_\alpha^\Theta(q\|p) = \int_\Theta (R \log R - R + 1) dp = \int_\Theta R \log R dp = D_{\text{KL}}(q\|p),$$

because $\int R dp = 1$. If $q \not\ll p$, then the singular part is charged by

$$\frac{\alpha}{1 - \alpha} q_{\perp p}(\Theta),$$

which diverges as $\alpha \uparrow 1$, matching $D_{\text{KL}}(q\|p) = \infty$. This proves the stated relations. □

Proof of Theorem 4.4

Proof. Let

$$m := \text{MI}_{\text{pow}}(p, \boldsymbol{\theta}) \in (0, \infty).$$

Since the Power Mass Index of p at $\boldsymbol{\theta}$ exists, Definition 4.1 gives

$$\inf \mathcal{U}_{\text{pow}}^p = \sup \mathcal{L}_{\text{pow}}^p = \frac{m}{d},$$

where the superscript indicates that the corresponding sets are defined with respect to p .

Fix $k > 0$ with $k \neq m$, and set

$$t := \frac{m}{k} > 0.$$

Choose a sequence $\rho_n \downarrow 0$, and define

$$E_n := B_{\rho_n}(\boldsymbol{\theta}), \quad a_n := p(E_n).$$

Because $m < \infty$, the measure p has no atom at $\boldsymbol{\theta}$, and $a_n \rightarrow 0$. Since $m > 0$, there is no local hole around $\boldsymbol{\theta}$, so $a_n > 0$ for all sufficiently large n . Passing to this tail of the sequence, define the conditional probability measure

$$p_n(A) := \frac{p(A \cap E_n)}{a_n}.$$

Let

$$R(x) := \|x - \boldsymbol{\theta}\|, \quad F_n(r) := p_n(B_r(\boldsymbol{\theta})), \quad 0 < r < \rho_n.$$

Let μ_n be the radial distribution of R under p_n . To keep the open-ball convention used in the Mass Index definition, we use half-open radial intervals and write

$$\mu_n([0, r)) = F_n(r) = p_n(B_r(\boldsymbol{\theta})).$$

This convention avoids any ambiguity from possible mass on spheres. Define a new radial probability measure ν_n by

$$\nu_n([0, r)) = F_n(r)^t, \quad 0 < r < \rho_n.$$

Disintegrate p_n with respect to the radial variable:

$$p_n(dx) = \mu_n(ds)K_n(s, d\omega), \quad s = \|x - \boldsymbol{\theta}\|,$$

where $K_n(s, \cdot)$ is a regular conditional angular distribution. Define

$$\eta_n(dx) := \nu_n(ds)K_n(s, d\omega).$$

Then $\eta_n \ll p_n$, and, for $0 < r < \rho_n$,

$$\eta_n(B_r(\boldsymbol{\theta})) = \nu_n([0, r)) = F_n(r)^t = \left(\frac{p(B_r(\boldsymbol{\theta}))}{a_n} \right)^t.$$

By Lemma B.2,

$$D_{\text{KL}}(\nu_n \| \mu_n) \leq C_t := \log t + \frac{1}{t} - 1 < \infty.$$

Since η_n and p_n have the same conditional angular distributions, the Radon–Nikodym derivative is

$$\frac{d\eta_n}{dp_n}(x) = \frac{d\nu_n}{d\mu_n}(\|x - \boldsymbol{\theta}\|) \quad p_n\text{-a.e.}$$

Therefore

$$\begin{aligned} D_{\text{KL}}(\eta_n \| p_n) &= \int \log \left(\frac{d\eta_n}{dp_n} \right) d\eta_n \\ &= \int \log \left(\frac{d\nu_n}{d\mu_n} \right) d\nu_n = D_{\text{KL}}(\nu_n \| \mu_n) \leq C_t. \end{aligned}$$

This bound is uniform in n .

Now define

$$q_n(A) := p(A \cap E_n^c) + a_n \eta_n(A \cap E_n).$$

Then q_n is a probability measure, $q_n \ll p$, and $q_n = p$ on E_n^c . For $0 < r < \rho_n$,

$$q_n(B_r(\boldsymbol{\theta})) = a_n \eta_n(B_r(\boldsymbol{\theta})) = a_n^{1-t} p(B_r(\boldsymbol{\theta}))^t.$$

The positive constant a_n^{1-t} does not affect the local O - or Ω -classes as $r \downarrow 0$. Therefore, for every $\eta > 0$,

$$q_n(B_r(\boldsymbol{\theta})) = O(r^{1/\eta}) \iff p(B_r(\boldsymbol{\theta})) = O(r^{1/(t\eta)}),$$

and similarly

$$q_n(B_r(\boldsymbol{\theta})) = \Omega(r^{1/\eta}) \iff p(B_r(\boldsymbol{\theta})) = \Omega(r^{1/(t\eta)}).$$

Thus

$$\inf \mathcal{U}_{\text{pow}}^{q_n} = \frac{1}{t} \inf \mathcal{U}_{\text{pow}}^p = \frac{1}{t} \frac{m}{d} = \frac{k}{d},$$

and

$$\sup \mathcal{L}_{\text{pow}}^{q_n} = \frac{1}{t} \sup \mathcal{L}_{\text{pow}}^p = \frac{1}{t} \frac{m}{d} = \frac{k}{d}.$$

Consequently,

$$\text{MI}_{\text{pow}}(q_n, \boldsymbol{\theta}) = k$$

for every n .

Finally, since $q_n = p$ outside E_n and the conditional law inside E_n is changed from p_n to η_n ,

$$D_{\text{KL}}(q_n \| p) = a_n D_{\text{KL}}(\eta_n \| p_n) \leq a_n C t.$$

As $a_n = p(B_{\rho_n}(\boldsymbol{\theta})) \rightarrow 0$, we obtain

$$D_{\text{KL}}(q_n \| p) \rightarrow 0.$$

Thus $q_n \in \mathcal{D}_{\boldsymbol{\theta}}$, $D_{\text{KL}}(q_n \| p) \rightarrow 0$, and $\text{MI}_{\text{pow}}(q_n, \boldsymbol{\theta}) = k$ for every n . This proves the theorem. \square

Proof of Theorem 5.1

Proof. Write

$$m_0(r) := \pi_0(B_r(\boldsymbol{\theta})), \quad \ell(r) := \log \frac{1}{r}.$$

First assume that

$$0 < \text{MI}_{\text{pow}}(\pi_0, \boldsymbol{\theta}) < \infty, \quad a := d \text{MI}_{\text{pow}}^{-1}(\pi_0, \boldsymbol{\theta}).$$

The condition in the theorem is exactly

$$a + \gamma > 0.$$

Let

$$I(r) := \int_{B_r(\boldsymbol{\theta})} \|x - \boldsymbol{\theta}\|^\gamma \left(\log \frac{1}{\|x - \boldsymbol{\theta}\|} \right)^\beta \pi_0(dx).$$

Since $0 < \text{MI}_{\text{pow}}(\pi_0, \boldsymbol{\theta}) < \infty$, the prior has no atom at $\boldsymbol{\theta}$, so the value of the integrand at $\boldsymbol{\theta}$ is irrelevant. The local likelihood assumption gives constants $0 < c < C < \infty$ and $r_0 > 0$ such that, for $x \in B_{r_0}(\boldsymbol{\theta}) \setminus \{\boldsymbol{\theta}\}$,

$$c \|x - \boldsymbol{\theta}\|^\gamma \left(\log \frac{1}{\|x - \boldsymbol{\theta}\|} \right)^\beta \leq L(x | \mathcal{D}) \leq C \|x - \boldsymbol{\theta}\|^\gamma \left(\log \frac{1}{\|x - \boldsymbol{\theta}\|} \right)^\beta.$$

By Bayes' formula,

$$\pi_{\mathcal{D}}(B_r(\boldsymbol{\theta})) = Z_{\mathcal{D}}^{-1} \int_{B_r(\boldsymbol{\theta})} L(x | \mathcal{D}) \pi_0(dx).$$

Since $Z_{\mathcal{D}} \in (0, \infty)$, the normalising constant is fixed, and therefore

$$\pi_{\mathcal{D}}(B_r(\boldsymbol{\theta})) \asymp I(r), \quad r \downarrow 0.$$

It remains to identify the power and logarithmic orders of $I(r)$.

For the power order, the definition of the Power Mass Index gives, for every sufficiently small $u > 0$,

$$m_0(r) = O(r^{a-u}), \quad m_0(r) = \Omega(r^{a+u}).$$

We claim that, for every $\varepsilon > 0$,

$$I(r) = O(r^{a+\gamma-\varepsilon}), \quad I(r) = \Omega(r^{a+\gamma+\varepsilon}).$$

For the upper bound, choose $u, v > 0$ such that

$$u + v < \varepsilon, \quad a + \gamma - u - v > 0.$$

Since logarithmic factors are slower than powers,

$$\left(\log \frac{1}{t}\right)^\beta \leq C_v t^{-v}$$

for all sufficiently small $t > 0$, with C_v independent of the annulus index used below. Let $r_j = e^{-j}r$, and set

$$A_j := B_{r_j}(\boldsymbol{\theta}) \setminus B_{r_{j+1}}(\boldsymbol{\theta}), \quad j \geq 0.$$

Then $B_r(\boldsymbol{\theta}) \setminus \{\boldsymbol{\theta}\} = \bigcup_{j \geq 0} A_j$. On A_j , $r_{j+1} \leq \|x - \boldsymbol{\theta}\| < r_j$. Hence, after absorbing the fixed factor $e^{|\gamma-v|}$ when $\gamma - v < 0$,

$$\|x - \boldsymbol{\theta}\|^\gamma \left(\log \frac{1}{\|x - \boldsymbol{\theta}\|}\right)^\beta \leq C r_j^{\gamma-v},$$

where the constant is uniform in j and in all sufficiently small r . Therefore

$$\begin{aligned} I(r) &\leq C \sum_{j=0}^{\infty} r_j^{\gamma-v} \pi_0(A_j) \\ &\leq C \sum_{j=0}^{\infty} r_j^{\gamma-v} m_0(r_j) \\ &\leq C \sum_{j=0}^{\infty} r_j^{a+\gamma-u-v} \leq C r^{a+\gamma-u-v}. \end{aligned}$$

This proves $I(r) = O(r^{a+\gamma-\varepsilon})$.

For the lower bound, choose $u, v > 0$ and $s > 1$, with s sufficiently close to one, such that

$$s(a-u) > a+u, \quad u + (s-1) \max\{\gamma, 0\} + v < \varepsilon.$$

Let

$$A(r) := B_r(\boldsymbol{\theta}) \setminus B_{r^s}(\boldsymbol{\theta}).$$

The prior power bounds imply

$$\pi_0(A(r)) = m_0(r) - m_0(r^s) \geq cr^{a+u} - Cr^{s(a-u)} \geq cr^{a+u}$$

for all sufficiently small r . On $A(r)$,

$$\|x - \boldsymbol{\theta}\|^\gamma \geq Cr^{\gamma+(s-1)\max\{\gamma, 0\}}, \quad \left(\log \frac{1}{\|x - \boldsymbol{\theta}\|}\right)^\beta \geq Cr^v.$$

Consequently,

$$I(r) \geq Cr^{a+\gamma+u+(s-1)\max\{\gamma, 0\}+v}.$$

Because

$$u + (s-1) \max\{\gamma, 0\} + v < \varepsilon,$$

we have

$$r^{a+\gamma+u+(s-1)\max\{\gamma, 0\}+v} = \Omega(r^{a+\gamma+\varepsilon}),$$

and hence $I(r) = \Omega(r^{a+\gamma+\varepsilon})$. The preceding two estimates satisfy Lemma B.1(i) with exponent $a + \gamma$. Therefore

$$d \text{MI}_{\text{pow}}^{-1}(\pi_{\mathcal{D}}, \boldsymbol{\theta}) = a + \gamma = d \text{MI}_{\text{pow}}^{-1}(\pi_0, \boldsymbol{\theta}) + \gamma.$$

Equivalently,

$$\text{MI}_{\text{pow}}^{-1}(\pi_{\mathcal{D}}, \boldsymbol{\theta}) = \text{MI}_{\text{pow}}^{-1}(\pi_0, \boldsymbol{\theta}) + d^{-1}\gamma.$$

Now assume that $\text{MI}_{\log}(\pi_0, \boldsymbol{\theta})$ exists, and set

$$b := \text{MI}_{\log}(\pi_0, \boldsymbol{\theta}), \quad \lambda := a + \gamma > 0.$$

Then, for every sufficiently small $u > 0$,

$$m_0(r) = O(r^a \ell(r)^{b+u}), \quad m_0(r) = \Omega(r^a \ell(r)^{b-u}).$$

We show that, for every $\varepsilon > 0$,

$$I(r) = O(r^\lambda \ell(r)^{b+\beta+\varepsilon}), \quad I(r) = \Omega(r^\lambda \ell(r)^{b+\beta-\varepsilon}).$$

For the upper bound, use the same annuli A_j , and choose $u \in (0, \varepsilon)$. On A_j , the relations $r_{j+1} \leq \|x - \boldsymbol{\theta}\| < r_j$ and $\log(1/\|x - \boldsymbol{\theta}\|) \asymp \ell(r_j)$ give

$$\|x - \boldsymbol{\theta}\|^\gamma \left(\log \frac{1}{\|x - \boldsymbol{\theta}\|} \right)^\beta \leq C r_j^\gamma \ell(r_j)^\beta,$$

with a constant uniform in j and in all sufficiently small r . Hence

$$\begin{aligned} I(r) &\leq C \sum_{j=0}^{\infty} r_j^\gamma \ell(r_j)^\beta m_0(r_j) \\ &\leq C \sum_{j=0}^{\infty} r_j^{a+\gamma} \ell(r_j)^{b+\beta+u} \\ &= C r^\lambda \sum_{j=0}^{\infty} e^{-j\lambda} (\ell(r) + j)^{b+\beta+u}. \end{aligned}$$

For $c = b + \beta + u$, the elementary geometric estimate

$$\sum_{j=0}^{\infty} e^{-j\lambda} (\ell(r) + j)^c \leq C \ell(r)^c$$

holds uniformly for all sufficiently small r . Indeed, if $c < 0$, then $(\ell(r) + j)^c \leq \ell(r)^c$; if $c \geq 0$, then $(\ell(r) + j)^c \leq \ell(r)^c (1 + j)^c$, and $\sum_{j \geq 0} e^{-j\lambda} (1 + j)^c < \infty$. Therefore

$$I(r) \leq C r^\lambda \ell(r)^{b+\beta+u}.$$

This gives the desired upper logarithmic bound.

For the lower bound, write $\gamma_+ := \max\{\gamma, 0\}$. Choose $u > 0$ and $K > 0$ such that

$$Ka > 2u, \quad u + K\gamma_+ < \varepsilon.$$

Set

$$\rho(r) := r \ell(r)^{-K}, \quad A(r) := B_r(\boldsymbol{\theta}) \setminus B_{\rho(r)}(\boldsymbol{\theta}).$$

Using the lower and upper logarithmic bounds for m_0 ,

$$\begin{aligned} \pi_0(A(r)) &= m_0(r) - m_0(\rho(r)) \\ &\geq cr^a \ell(r)^{b-u} - Cr^a \ell(r)^{b+u-Ka} \geq cr^a \ell(r)^{b-u} \end{aligned}$$

for all sufficiently small r . On $A(r)$,

$$\|x - \boldsymbol{\theta}\|^\gamma \geq Cr^\gamma \ell(r)^{-K\gamma_+}, \quad \log \frac{1}{\|x - \boldsymbol{\theta}\|} \asymp \ell(r).$$

Therefore

$$I(r) \geq Cr^{a+\gamma} \ell(r)^{b+\beta-u-K\gamma_+} = \Omega(r^\lambda \ell(r)^{b+\beta-\varepsilon}).$$

The upper and lower logarithmic bounds satisfy Lemma B.1(ii) at the already identified power scale λ . Hence

$$\text{MI}_{\log}(\pi_{\mathcal{D}}, \boldsymbol{\theta}) = b + \beta = \text{MI}_{\log}(\pi_0, \boldsymbol{\theta}) + \beta.$$

It remains to treat the case $\text{MI}_{\text{pow}}(\pi_0, \boldsymbol{\theta}) = 0$. In this case, for every $M > 0$,

$$m_0(r) = O(r^M), \quad r \downarrow 0.$$

Fix any $N > 0$. Choose $v > 0$, and then choose M so large that

$$M + \gamma - v > N, \quad M + \gamma - v > 0.$$

Using the same annular decomposition as above and the bound $m_0(r_j) = O(r_j^M)$, we obtain

$$\begin{aligned} I(r) &\leq C \sum_{j=0}^{\infty} r_j^{\gamma-v} m_0(r_j) \\ &\leq C \sum_{j=0}^{\infty} r_j^{M+\gamma-v} \leq C r^{M+\gamma-v} = O(r^N). \end{aligned}$$

Thus $\pi_{\mathcal{D}}(B_r(\boldsymbol{\theta})) = O(r^N)$ for every $N > 0$. Consequently the posterior small-ball mass is faster than every polynomial power. Hence every positive power is an upper gauge and no positive power is a lower gauge, which gives

$$\text{MI}_{\text{pow}}(\pi_{\mathcal{D}}, \boldsymbol{\theta}) = 0.$$

The proof is complete. □

Proof of Theorem 5.2

Proof. Set

$$m_0(r) := \pi_0(B_r(\boldsymbol{\theta})), \quad m_{\mathcal{D}}(r) := \pi_{\mathcal{D}}(B_r(\boldsymbol{\theta})), \quad \ell(r) := -\log r.$$

By the theorem hypothesis, $0 < \text{MI}_{\text{pow}}(\pi_0, \boldsymbol{\theta}) < \infty$. Set

$$a := d \text{MI}_{\text{pow}}^{-1}(\pi_0, \boldsymbol{\theta}).$$

Since $g(x) \asymp 1$ as $x \rightarrow \boldsymbol{\theta}$, there exist constants $0 < c_g < C_g < \infty$ and $r_0 > 0$ such that

$$c_g \leq g(x) \leq C_g, \quad x \in B_{r_0}(\boldsymbol{\theta}).$$

By Bayes' formula and the finite marginal assumption, the normalising constant only changes small-ball masses by a fixed multiplicative constant. Hence, for $0 < r < r_0$,

$$m_{\mathcal{D}}(r) \asymp \pi_0(B_r(\boldsymbol{\theta}) \cap E) = A_{E, \boldsymbol{\theta}}(r) m_0(r).$$

Using the assumed acceptance-ratio scaling,

$$m_{\mathcal{D}}(r) \asymp r^{\delta} \ell(r)^{\lambda} m_0(r), \quad r \downarrow 0. \tag{B.1}$$

Because $A_{E, \boldsymbol{\theta}}(r) \in [0, 1]$, the regime stated in the theorem is the non-degenerate regime compatible with this two-sided comparison.

For every $\varepsilon > 0$, the prior Power Mass Index gives

$$m_0(r) = O(r^{a-\varepsilon}), \quad m_0(r) = \Omega(r^{a+\varepsilon}).$$

Since logarithmic factors are sub-polynomial, for every $\eta > 0$,

$$\ell(r)^{\lambda} = O(r^{-\eta}), \quad \ell(r)^{\lambda} = \Omega(r^{\eta}).$$

Given $\varepsilon > 0$, apply the prior bounds with $\varepsilon/2$ and the last display with $\eta = \varepsilon/2$. Then (B.1) gives

$$m_{\mathcal{D}}(r) = O(r^{a+\delta-\varepsilon}), \quad m_{\mathcal{D}}(r) = \Omega(r^{a+\delta+\varepsilon}).$$

By Lemma B.1(i), the posterior power exponent is $a + \delta$, and

$$d\text{MI}_{\text{pow}}^{-1}(\pi_{\mathcal{D}}, \boldsymbol{\theta}) = a + \delta = d\text{MI}_{\text{pow}}^{-1}(\pi_0, \boldsymbol{\theta}) + \delta.$$

Equivalently,

$$\text{MI}_{\text{pow}}^{-1}(\pi_{\mathcal{D}}, \boldsymbol{\theta}) = \text{MI}_{\text{pow}}^{-1}(\pi_0, \boldsymbol{\theta}) + d^{-1}\delta.$$

Now assume that $\text{MI}_{\text{log}}(\pi_0, \boldsymbol{\theta})$ exists and set

$$b := \text{MI}_{\text{log}}(\pi_0, \boldsymbol{\theta}).$$

Then, for every $\varepsilon > 0$,

$$m_0(r) = O(r^a \ell(r)^{b+\varepsilon}), \quad m_0(r) = \Omega(r^a \ell(r)^{b-\varepsilon}).$$

Combining these bounds with (B.1) gives

$$m_{\mathcal{D}}(r) = O(r^{a+\delta} \ell(r)^{b+\lambda+\varepsilon}), \quad m_{\mathcal{D}}(r) = \Omega(r^{a+\delta} \ell(r)^{b+\lambda-\varepsilon}).$$

The posterior power exponent has already been identified as $a + \delta$. Hence Lemma B.1(ii) gives

$$\text{MI}_{\text{log}}(\pi_{\mathcal{D}}, \boldsymbol{\theta}) = b + \lambda = \text{MI}_{\text{log}}(\pi_0, \boldsymbol{\theta}) + \lambda.$$

The proof is complete. □

Proof of Corollary 5.3

Proof. For each $\tau > 0$, define

$$Z_{\tau} := \int s_{\tau}(x) \pi_0(dx), \quad Z_E := \int \mathbf{1}_E(x) \pi_0(dx).$$

For each $\tau > 0$, the softened normalising constant satisfies $0 < Z_{\tau} \leq C$. By Assumption 1 applied to the hard likelihood $L = \mathbf{1}_E$, we have $Z_E \in (0, \infty)$. By uniform boundedness, pointwise convergence $s_{\tau} \rightarrow \mathbf{1}_E$ π_0 -a.e., and dominated convergence,

$$Z_{\tau} \rightarrow Z_E.$$

For any measurable set A ,

$$\pi_{\tau}(A) = Z_{\tau}^{-1} \int_A s_{\tau}(x) \pi_0(dx), \quad \pi_{\mathcal{D}}(A) = Z_E^{-1} \int_A \mathbf{1}_E(x) \pi_0(dx).$$

Therefore

$$\|\pi_{\tau} - \pi_{\mathcal{D}}\|_{\text{TV}} \leq |Z_{\tau}^{-1} - Z_E^{-1}| Z_{\tau} + Z_E^{-1} \int |s_{\tau} - \mathbf{1}_E| d\pi_0,$$

and both terms converge to zero. Hence

$$\|\pi_{\tau} - \pi_{\mathcal{D}}\|_{\text{TV}} \rightarrow 0.$$

Now fix $\tau > 0$. Since s_{τ} is positive and continuous, there are constants $0 < c_{\tau} < C_{\tau} < \infty$ and $r_{\tau} > 0$ such that

$$c_{\tau} \leq s_{\tau}(x) \leq C_{\tau}, \quad x \in B_{r_{\tau}}(\boldsymbol{\theta}).$$

Thus

$$\pi_{\tau}(B_r(\boldsymbol{\theta})) \asymp \pi_0(B_r(\boldsymbol{\theta})), \quad r \downarrow 0,$$

for each fixed $\tau > 0$. Consequently

$$\text{MI}_{\text{pow}}(\pi_{\tau}, \boldsymbol{\theta}) = \text{MI}_{\text{pow}}(\pi_0, \boldsymbol{\theta}), \quad \tau > 0.$$

If $A_{E, \boldsymbol{\theta}}(r) \asymp r^{\delta} \ell(r)^{\lambda}$ with $\delta > 0$, then Theorem 5.2 applied to the hard posterior gives

$$d\text{MI}_{\text{pow}}^{-1}(\pi_{\mathcal{D}}, \boldsymbol{\theta}) = d\text{MI}_{\text{pow}}^{-1}(\pi_0, \boldsymbol{\theta}) + \delta.$$

This differs from the softened exponent $d\text{MI}_{\text{pow}}^{-1}(\pi_0, \boldsymbol{\theta})$. Hence

$$\text{MI}_{\text{pow}}(\pi_{\mathcal{D}}, \boldsymbol{\theta}) \neq \lim_{\tau \downarrow 0} \text{MI}_{\text{pow}}(\pi_{\tau}, \boldsymbol{\theta}).$$

For the logarithmic statement, assume $\delta = 0$, $\lambda < 0$, and that $\text{MI}_{\log}(\pi_0, \boldsymbol{\theta})$ exists. For each fixed $\tau > 0$, the local positivity of s_{τ} gives

$$\text{MI}_{\log}(\pi_{\tau}, \boldsymbol{\theta}) = \text{MI}_{\log}(\pi_0, \boldsymbol{\theta}).$$

For the hard posterior, Theorem 5.2 gives

$$\text{MI}_{\log}(\pi_{\mathcal{D}}, \boldsymbol{\theta}) = \text{MI}_{\log}(\pi_0, \boldsymbol{\theta}) + \lambda.$$

Since $\lambda < 0$, the two logarithmic indices differ, and therefore

$$\text{MI}_{\log}(\pi_{\mathcal{D}}, \boldsymbol{\theta}) \neq \lim_{\tau \downarrow 0} \text{MI}_{\log}(\pi_{\tau}, \boldsymbol{\theta}).$$

The proof is complete. □

Proof of Theorem 5.4

Proof. Let $q = q_{\ll p} + q_{\perp p}$, and write

$$R := \frac{dq_{\ll p}}{dp}.$$

For the restrictions of q and p to E , define the local Hellinger-type quantity

$$H_E^2(q, p) := p(E) + q(E) - 2 \int_E \sqrt{R} dp.$$

Using the Lebesgue decomposition, this can be written as

$$H_E^2(q, p) = \int_E (\sqrt{R} - 1)^2 dp + q_{\perp p}(E).$$

We first spell out the finite-measure Cauchy–Schwarz step. Since

$$\int_E \sqrt{R} dp \leq \sqrt{q_{\ll p}(E)} \sqrt{p(E)} \leq \sqrt{q(E)p(E)},$$

we have

$$\begin{aligned} H_E^2(q, p) &= p(E) + q(E) - 2 \int_E \sqrt{R} dp \\ &\geq p(E) + q(E) - 2\sqrt{p(E)q(E)} = \left(\sqrt{q(E)} - \sqrt{p(E)} \right)^2. \end{aligned}$$

Hence

$$\left| \sqrt{q(E)} - \sqrt{p(E)} \right| \leq H_E(q, p),$$

and therefore

$$|q(E) - p(E)| \leq (\sqrt{q(E)} + \sqrt{p(E)}) H_E(q, p) \leq 2H_E(q, p),$$

because $p(E), q(E) \leq 1$.

It remains to compare $H_E^2(q, p)$ with $D_{\alpha}^E(q||p)$. We prove the scalar inequality

$$(\sqrt{x} - 1)^2 \leq C_{\alpha} f_{\alpha}(x), \quad x \geq 0, \quad C_{\alpha} := \max \left\{ 1, \frac{1 - \alpha}{\alpha} \right\}.$$

Set $t = \sqrt{x}$. If $0 < \alpha \leq 1/2$, then $C_{\alpha} = (1 - \alpha)/\alpha$, and the desired inequality is equivalent to

$$2\alpha t - t^{2\alpha} - 2\alpha + 1 \geq 0.$$

The derivative of the left-hand side is $2\alpha(1 - t^{2\alpha-1})$, which is negative on $(0, 1)$ and positive on $(1, \infty)$. Since the value at $t = 1$ is zero, the inequality follows. If $1/2 \leq \alpha < 1$, then $C_\alpha = 1$, and the desired inequality is equivalent to

$$(2\alpha - 1)t^2 + 2(1 - \alpha)t \geq t^{2\alpha}.$$

This follows from weighted AM–GM, because the weights $2\alpha - 1$ and $2(1 - \alpha)$ are non-negative and sum to one. The endpoint $x = 0$ is covered by continuity.

Consequently,

$$\int_E (\sqrt{R} - 1)^2 dp \leq C_\alpha \int_E f_\alpha(R) dp.$$

Moreover, since $C_\alpha \geq (1 - \alpha)/\alpha$,

$$q_{\perp p}(E) \leq C_\alpha \frac{\alpha}{1 - \alpha} q_{\perp p}(E).$$

Therefore

$$H_E^2(q, p) \leq C_\alpha \left\{ \int_E f_\alpha(R) dp + \frac{\alpha}{1 - \alpha} q_{\perp p}(E) \right\} = C_\alpha D_\alpha^E(q \| p).$$

Combining the two displays yields

$$|q(E) - p(E)| \leq 2\sqrt{C_\alpha D_\alpha^E(q \| p)}.$$

This proves the theorem. □

Proof of Corollary 5.5

Proof. Apply Theorem 5.4 with $E = B_r(\theta)$. If

$$D_\alpha^{B_r(\theta)}(q \| p) = O(r^\delta),$$

then

$$|q(B_r(\theta)) - p(B_r(\theta))| \leq 2\sqrt{C_\alpha D_\alpha^{B_r(\theta)}(q \| p)} = O(r^{\delta/2}).$$

□

Proof of Theorem 5.6

Proof. Assume $p(E) > 0$. Let $q = q_{\ll p} + q_{\perp p}$, and write

$$R := \frac{dq_{\ll p}}{dp}, \quad x := \frac{1}{p(E)} \int_E R dp, \quad h := \frac{q_{\perp p}(E)}{p(E)}.$$

Then

$$\frac{q(E)}{p(E)} = x + h.$$

Since

$$f'_\alpha(u) = \frac{\alpha}{1 - \alpha} (1 - u^{\alpha-1}) \leq \frac{\alpha}{1 - \alpha},$$

we have

$$f_\alpha(x + h) \leq f_\alpha(x) + \frac{\alpha}{1 - \alpha} h.$$

By Jensen's inequality,

$$f_\alpha(x) \leq \frac{1}{p(E)} \int_E f_\alpha(R) dp.$$

Combining these bounds gives

$$f_\alpha\left(\frac{q(E)}{p(E)}\right) \leq \frac{D_\alpha^E(q \| p)}{p(E)} = \overline{D}_\alpha^E(q \| p).$$

This proves the first claim.

For the relative error bound, we derive the scalar inequality

$$f_\alpha(y) \geq \frac{\alpha|y-1|^2}{2(1+|y-1|)}, \quad y \geq 0.$$

Since $f_\alpha(1) = f'_\alpha(1) = 0$ and $f''_\alpha(u) = \alpha u^{\alpha-2}$, Taylor's formula with integral remainder gives, for $0 \leq y \leq 1$,

$$f_\alpha(y) = \alpha \int_y^1 (u-y)u^{\alpha-2} du \geq \frac{\alpha}{2}(1-y)^2 \geq \frac{\alpha(1-y)^2}{2(1+1-y)}.$$

For $y = 1 + s \geq 1$, the same formula gives

$$f_\alpha(1+s) = \alpha \int_0^s (s-u)(1+u)^{\alpha-2} du.$$

Because $\alpha > 0$, $(1+u)^{\alpha-2} \geq (1+u)^{-2}$. Hence

$$f_\alpha(1+s) \geq \alpha \int_0^s \frac{s-u}{(1+u)^2} du = \alpha\{s - \log(1+s)\}.$$

The elementary bound

$$s - \log(1+s) \geq \frac{s^2}{2(1+s)}, \quad s \geq 0,$$

follows by differentiating the difference between the two sides. Therefore the same scalar inequality holds for $y \geq 1$.

Set

$$y := \frac{q(E)}{p(E)}, \quad s := |1-y|, \quad z := \overline{D}_\alpha^E(q\|p).$$

The first part gives $f_\alpha(y) \leq z$, so

$$\frac{\alpha s^2}{2(1+s)} \leq z.$$

Solving this quadratic inequality for $s \geq 0$ yields

$$s \leq \frac{z}{\alpha} + \sqrt{\left(\frac{z}{\alpha}\right)^2 + \frac{2z}{\alpha}}.$$

The last display is bounded by

$$\frac{2}{\alpha} \sqrt{z} \sqrt{z+2}.$$

Therefore

$$\left|1 - \frac{q(E)}{p(E)}\right| \leq \frac{2}{\alpha} \sqrt{\overline{D}_\alpha^E(q\|p)} \sqrt{\overline{D}_\alpha^E(q\|p) + 2}.$$

This proves the theorem. □

Proof of Corollary 5.7

Proof. Set

$$t_r := \overline{D}_\alpha^{B_r(\boldsymbol{\theta})}(q\|p).$$

By assumption,

$$t_r = O(r^\delta), \quad r \downarrow 0,$$

and hence $t_r \rightarrow 0$. Applying Theorem 5.6 with $E = B_r(\boldsymbol{\theta})$ gives

$$\left|1 - \frac{q(B_r(\boldsymbol{\theta}))}{p(B_r(\boldsymbol{\theta}))}\right| \leq \frac{2}{\alpha} \sqrt{t_r} \sqrt{t_r + 2} = O(t_r^{1/2}) = O(r^{\delta/2}).$$

□

Proof of Theorem 5.8

Proof. Let

$$E_r := B_r(\boldsymbol{\theta}), \quad m_p(r) := p(E_r), \quad m_q(r) := q(E_r).$$

By assumption, $m_p(r) > 0$ and $m_q(r) > 0$ for all sufficiently small r . We use two elementary facts about f_α . First, f_α is continuous on $[0, \infty)$, satisfies $f_\alpha(1) = 0$, $f_\alpha(0) = 1$, and diverges as $x \rightarrow \infty$. Hence, for every $\lambda < 1$, the sublevel set

$$\{x \geq 0 : f_\alpha(x) \leq \lambda\}$$

is contained in an interval $[c, C]$ with $0 < c < C < \infty$. Second, for every finite M , the sublevel set

$$\{x \geq 0 : f_\alpha(x) \leq M\}$$

is bounded above.

We first prove (i). Suppose

$$\limsup_{r \downarrow 0} \overline{D}_\alpha^{E_r}(q||p) < 1.$$

Choose $\lambda < 1$ such that

$$\overline{D}_\alpha^{E_r}(q||p) \leq \lambda$$

for all sufficiently small r . By Theorem 5.6,

$$f_\alpha\left(\frac{m_q(r)}{m_p(r)}\right) \leq \overline{D}_\alpha^{E_r}(q||p) \leq \lambda.$$

The first scalar fact gives constants $0 < c < C < \infty$ such that

$$c \leq \frac{m_q(r)}{m_p(r)} \leq C$$

for all sufficiently small r . Equivalently,

$$m_q(r) \asymp m_p(r), \quad r \downarrow 0.$$

Thus the upper and lower power gauge classes of p and q coincide. For any small-ball function, $\sup \mathcal{L}_{\text{pow}} \leq \inf \mathcal{U}_{\text{pow}}$. Since Assumption 2 gives equality of these two quantities for the relevant measures, the common gauge classes identify

$$\text{MI}_{\text{pow}}(q, \boldsymbol{\theta}) = \text{MI}_{\text{pow}}(p, \boldsymbol{\theta}).$$

If $\text{MI}_{\log}(p, \boldsymbol{\theta})$ exists, then the common power exponent is

$$a := d \text{MI}_{\text{pow}}^{-1}(p, \boldsymbol{\theta}) = d \text{MI}_{\text{pow}}^{-1}(q, \boldsymbol{\theta}).$$

Since $m_q(r) \asymp m_p(r)$, for every $\eta \in \mathbb{R}$,

$$m_p(r) = O(r^a (-\log r)^\eta) \iff m_q(r) = O(r^a (-\log r)^\eta),$$

and similarly

$$m_p(r) = \Omega(r^a (-\log r)^\eta) \iff m_q(r) = \Omega(r^a (-\log r)^\eta).$$

Hence the logarithmic upper and lower gauge classes also coincide. Since the Logarithmic Mass Index of p exists, and the common gauge classes force the same upper and lower logarithmic thresholds for q ,

$$\text{MI}_{\log}(q, \boldsymbol{\theta}) = \text{MI}_{\log}(p, \boldsymbol{\theta}).$$

We now prove (ii). Suppose

$$\limsup_{r \downarrow 0} \overline{D}_\alpha^{E_r}(p||q) < \infty.$$

Then there exists $M < \infty$ such that

$$\overline{D}_\alpha^{E_r}(p\|q) \leq M$$

for all sufficiently small r . Applying Theorem 5.6 with p and q interchanged gives

$$f_\alpha\left(\frac{m_p(r)}{m_q(r)}\right) \leq \overline{D}_\alpha^{E_r}(p\|q) \leq M.$$

By the second scalar fact, $m_p(r)/m_q(r)$ is bounded above. Therefore there exists $C < \infty$ such that

$$m_q(r) \geq C^{-1}m_p(r)$$

for all sufficiently small r .

Consequently, every lower power bound for m_p is also a lower power bound for m_q :

$$\mathcal{L}_{\text{pow}}(p, \boldsymbol{\theta}) \subseteq \mathcal{L}_{\text{pow}}(q, \boldsymbol{\theta}).$$

Taking suprema and using the dimensional normalisation in Definition 4.1 gives

$$\sup \mathcal{L}_{\text{pow}}(q, \boldsymbol{\theta}) \geq \sup \mathcal{L}_{\text{pow}}(p, \boldsymbol{\theta}).$$

Assumption 2 identifies these lower thresholds with the corresponding Power Mass Indices, because $\sup \mathcal{L}_{\text{pow}} = \inf \mathcal{U}_{\text{pow}}$ for both measures. Therefore

$$\text{MI}_{\text{pow}}(q, \boldsymbol{\theta}) \geq \text{MI}_{\text{pow}}(p, \boldsymbol{\theta}).$$

Finally, assume

$$\text{MI}_{\text{pow}}(q, \boldsymbol{\theta}) = \text{MI}_{\text{pow}}(p, \boldsymbol{\theta})$$

and that both Logarithmic Mass Indices exist. Let

$$a := d \text{MI}_{\text{pow}}^{-1}(p, \boldsymbol{\theta}) = d \text{MI}_{\text{pow}}^{-1}(q, \boldsymbol{\theta}).$$

The inequality $m_q(r) \geq C^{-1}m_p(r)$ implies that every lower logarithmic bound for m_p at the common power scale is also a lower logarithmic bound for m_q . Hence

$$\mathcal{L}_{\log}(p, \boldsymbol{\theta}) \subseteq \mathcal{L}_{\log}(q, \boldsymbol{\theta}).$$

Taking suprema and using the existence of both Logarithmic Mass Indices gives

$$\text{MI}_{\log}(q, \boldsymbol{\theta}) \geq \text{MI}_{\log}(p, \boldsymbol{\theta}).$$

The proof is complete. □

C Additional experimental details

This appendix gives the concrete settings used for the experiments in Section 6. All experiments are finite-radius illustrations. We do not estimate asymptotic Mass Indices directly.

C.1 Experiment 1: synthetic calibration

Experiment 1 plots closed-form small-ball curves over

$$r \in [10^{-3}, 0.6].$$

The purpose is to visualize representative local regimes rather than to fit a statistical model.

For the regular reference case, we use the one-dimensional Gaussian small-ball mass

$$p_{\text{Gauss}}(B_r(0)) = \text{erf}\left(\frac{r}{\sqrt{2}}\right).$$

This behaves as r for small r .

For the local-depletion example, we plot

$$p_{\text{dep}}(B_r(0)) = r^2.$$

This has faster small-radius decay than the regular reference.

For the cusp example, we plot

$$p_{\text{cusp}}(B_r(0)) = \sqrt{r}.$$

This has slower small-radius decay than the regular reference.

For the logarithmically corrected example, we plot

$$p_{\text{log}}(B_r(0)) = r(1 - \log r).$$

This has the same leading power order as the regular reference, but differs by a logarithmic factor.

For the spike-slab example, we use spike weight 0.25 and slab scale 0.35, giving

$$p_{\text{spike}}(B_r(0)) = 0.25 + 0.75 \operatorname{erf}\left(\frac{r}{0.35\sqrt{2}}\right).$$

The nonzero spike produces a small-ball mass that does not vanish as $r \rightarrow 0$.

C.2 Experiment 2: UCI Bayesian logistic-regression sanity check

Experiment 2 uses three binary classification datasets: Breast Cancer, Iris 0 vs 1, and Wine 0 vs 1. For Iris and Wine, only classes 0 and 1 are retained. Each dataset is repeated over five random seeds,

$$\{0, 1, 2, 3, 4\}.$$

The seed controls the stratified train-test split. Conditional on the training set, preprocessing, Laplace optimisation, and quadrature settings are fixed. For each dataset and seed, we use a stratified train-test split with test fraction 0.30. Only the training set is used for the Bayesian logistic-regression fit.

The covariates are standardized on the training set and then projected to four principal components using PCA. An intercept is added, so the parameter dimension is

$$d = 5.$$

The model is Bayesian logistic regression. Given design matrix X , labels $y_i \in \{0, 1\}$, and parameter $\theta \in \mathbb{R}^5$, the negative log posterior objective is

$$\sum_i [\log(1 + \exp(x_i^\top \theta)) - y_i x_i^\top \theta] + \frac{1}{2\sigma_0^2} \|\theta\|^2, \quad \sigma_0 = 2.$$

Equivalently, the prior is

$$\theta \sim N(0, 4I_5).$$

The posterior is approximated by a Laplace Gaussian approximation. The posterior mean $\hat{\theta}$ is obtained by minimizing the negative log posterior. We use a trust-region Newton optimizer with analytic gradient and Hessian. If this optimizer fails, we use an L-BFGS-B fallback. The Laplace covariance is the inverse Hessian at $\hat{\theta}$, with eigenvalues clipped below 10^{-10} for numerical stability. We set

$$\theta_0 = \hat{\theta}.$$

For each run, prior and posterior Euclidean small-ball masses are estimated for

$$B_r(\theta_0) = \{\theta : \|\theta - \theta_0\|_2 \leq r\}$$

over a geometric grid of 24 radii in

$$[0.03, 0.45].$$

The prior covariance is $4I_5$, while the posterior covariance is the Laplace covariance. The prior mean is 0, and the posterior mean is θ_0 .

The small-ball masses are estimated by Sobol quadrature over the Euclidean unit ball. The default number of Sobol points is

$$2^{15} = 32768$$

per run. The unit-ball Sobol points are generated in dimension $d + 1$. The first d coordinates are transformed into Gaussian directions and normalized to obtain directions on the sphere. The final coordinate is transformed as $U^{1/d}$ to obtain the radial component. For a radius r , the quadrature points are

$$\theta_0 + rz_j, \quad z_j \in B_1(0).$$

For a Gaussian distribution $N(m, \Sigma)$, the estimated log small-ball mass is

$$\log |B_1(0)| + d \log r + \log \left[\frac{1}{M} \sum_{j=1}^M \varphi_{\Sigma}(\theta_0 + rz_j - m) \right],$$

where $M = 32768$, and φ_{Σ} denotes the Gaussian density with covariance Σ .

Finite-radius slopes are computed by first differences on the log-log scale:

$$s_k = \frac{\log p(B_{r_{k+1}}(\theta_0)) - \log p(B_{r_k}(\theta_0))}{\log r_{k+1} - \log r_k}.$$

The slope location is reported at the geometric midpoint

$$(r_k r_{k+1})^{1/2}.$$

The main-text table reports small-radius slopes fitted over $r \leq 0.06$.

Across datasets and seeds, the plotted aggregate curves use means and standard errors over all runs. In total, the aggregate UCI figure uses

$$3 \times 5 = 15$$

runs.

C.3 Experiment 3: local RE-KL directionality

Experiment 3 is a one-dimensional toy example designed to show that local RE-KL control is directional. On $(-1, 1)$, define

$$p(x) = \frac{1}{2}, \quad q(x) = \frac{1}{4}|x|^{-1/2}.$$

Both are probability densities on $(-1, 1)$. Around $\theta_0 = 0$, their small-ball masses are

$$p(B_r(0)) = r, \quad q(B_r(0)) = \sqrt{r}.$$

The plotted local RE-KL quantities use mass-ratio weighted conditional terms. The conditional constants are

$$C_{p||q} = \log 2 - \frac{1}{2}, \quad C_{q||p} = 1 - \log 2.$$

The plotted finite-radius quantities are

$$D_r(p||q) = \sqrt{r} \left(\log 2 - \frac{1}{2} \right),$$

and

$$D_r(q||p) = \frac{1 - \log 2}{\sqrt{r}}.$$

Therefore,

$$D_r(p||q) \rightarrow 0, \quad D_r(q||p) \rightarrow \infty \quad (r \rightarrow 0).$$

This example shows that the two directions can behave differently on the same shrinking neighbourhoods. Hence the direction in a local RE-KL assumption is substantive.