arXiv:2603.05480v1 [stat.ML] 5 Mar 2026

奇异贝叶斯模型中的热力学响应函数

Sean Plummer
snplmmr@gmail.com

2026年3月6日

摘要

奇异统计模型——包括混合模型、矩阵分解和神经网络——由于参数不可辨识性和退化Fisher几何而违背正则渐近。尽管奇异学习理论通过实对数典范阈值和奇异涨落等不变量刻画边际似然行为，这些量在操作上仍难以解释。与此同时，广泛使用的准则如WAIC和WBIC似乎与底层奇异几何脱节。我们表明，后验温度化诱导后验分布的单参数形变，其相关可观测量生成热力学响应函数的层级。通用协方差恒等式将回火期望的导数与后验涨落联系起来，从而将WAIC、WBIC和奇异涨落置于统一的响应框架之中。在该框架内，奇异学习理论中的经典量获得自然的热力学解释：实对数典范阈值支配主导自由能斜率，奇异涨落对应于温度化自由能的曲率，而WAIC度量预测波动。我们形式化一个可观测量代数，通过对不可识别方向取商，从而允许在奇异模型中构造结构上有意义的序参量。在典型奇异例子中——包括对称高斯混合模型、降秩回归和过参数化神经网络——我们实证展示了回火下的类相变行为。序参量坍缩，磁化率出现峰值，复杂度度量与后验几何中的结构重组相一致。我们的结果表明，热力学响应理论为解释奇异贝叶斯学习中的复杂度、预测变异性和结构重组提供了一个自然的组织框架。

1 引言

1.1 动机

奇异统计模型出现于多个参数值诱导同一预测分布的情形，如混合模型、低秩分解以及具有对称性或过参数化的神经网络 [7, 10]。在此类设定中，费 Fisher 信息可能退化，经典正则渐近失效：后验质量集中于具有非平凡几何的集合上，标准的“有效维数”启发式变得不可靠。

奇异学习理论通过实对数典范阈值（RLCT）和奇异涨落等不变量，为这些情形下的贝叶斯边际似然与预测性能提供了有原则的渐近描述。这些量例如解释了，为何边际似然可以按照非整数“学习系数”随样本量缩放，以及为何预测复杂度可以不同于正则模型的期望。然而，尽管它们在理论上十分重要，RLCT及相关常数在有限样本中往往难以从操作层面加以解释，也很少被用作实用的诊断工具。

与此同时，广义适用信息准则（WAIC）和广义适用贝叶斯信息准则（WBIC）等广泛使用的准则常被应用于奇异情形，但它们与底层奇异几何的关系可能显得并不透明 [8, 9]。这促使人们寻求一种结构化且可解释的框架，该框架（i）适用于奇异模型，而无需针对特定模型进行渐近推导，并且（ii）以揭示实用复杂度度量所响应的后验几何哪些方面的方式，来组织这些复杂度度量。

本文的核心论点是，贝叶斯模型评估中若干广泛使用的量——包括 WAIC复杂度和奇异涨落——可以理解为由后验温度化产生的热力学响应函数。本研究是更广泛研究的一部分，旨在通过热力学结构理解奇异贝叶斯学习。

1.2 核心思想：将回火视为一种形变

我们的出发点是后验温度化，它定义了后验分布的一参数形变：

$$\pi_\beta(\theta \mid D) \;\propto\; \pi(\theta)\, p(D \mid \theta)^\beta, \qquad \beta > 0.$$

改变 β 可在先验（$\beta \to 0$）与普通后验分布（$\beta = 1$）之间进行插值，从而生成一个受控的分布族，对同一似然景观 [4] 重新加权。回火是一种尤为便利的变形方式，因为它在保持似然景观不变的同时，连续调整似然相对于先验的影响。这使我们能够在不改变底层统计模型的情况下探究后验结构。

我们将可在 π_β 下积的可测函数 $f\colon \Theta \to \mathbb{R}$ 视为**可观测量**。这仅仅意味着，这些量在回火下具有定义良好的后验期望：

$$\mathbb{E}_\beta[f] = \int f(\theta)\, \pi_\beta(\theta \mid D)\, d\theta.$$

在物理学术语中，可观测量是系统的可测量；此处它们是模型参数的后验可测函数。它们的期望概括了后验几何的若干方面，而它们的涨落则编码了结构变异性。一个基本恒等式表明，温度化期望的导数由与对数似然的协方差给出，

$$\frac{d}{d\beta}\,\mathbb{E}_\beta[f] \;=\; \mathrm{Cov}_\beta(f, \ell)\,, \qquad \ell(\theta) := \log p(D \mid \theta),$$

因此，可观测量相对于 β 的灵敏度由后验涨落控制。这提示了一种热力学观点：β 扮演逆温度的角色，ℓ扮演（负）能量的角色，由此产生的协方差和方差定义了响应函数。该恒等式意味着对数似然的涨落控制着后验期望如何响应温度变化，从而将熟悉的贝叶斯量置于热力学响应层级之中。

1.3 贡献与论文组织

本文为奇异贝叶斯学习建立了一个响应函数框架，包含三个组成部分。

可观测量代数。 我们通过将沿不可识别方向取常值的函数商掉，来形式化“可观测量代数”（即那些不改变诱导预测分布的方向）。这将参数化伪影与统计上有意义的量区分开来，并指导奇异模型中序参量的构造。

普适响应恒等式。我们利用协方差恒等式，在回火条件下定义序参量、磁化率和复杂度响应。在这一框架下，WAIC 及相关量表现为 $\beta = 1$ 处基于涨落的响应，而类自由能对象的曲率则为奇异涨落提供了自然的解释。

典型奇异模型中的经验验证。 我们通过实证展示，响应函数能够在三种典型情形下定位后验几何中的结构重组：对称高斯混合模型（对称性破缺）、降秩回归（代数奇异性）以及过参数化神经网络（现代不可辨识性）。在这些例子中，序参量趋于消失，磁化率呈现尖锐峰值，复杂度响应则与相同的类相变行为相吻合。

迈向奇异学习的热力学框架。 本文提出的响应视角，是更广泛探索的一部分，旨在通过热力学结构理解奇异统计模型。后验温度化引入了一个自然的控制参数，其相关涨落揭示了模型的几何性质与预测性质。本文聚焦于这种形变所产生的响应函数，以及它们与 WAIC、WBIC 等实用复杂度度量之间的关系。更广泛而言，这些思想表明，统计物理学中的工具——包括响应理论和重整化概念——或可为奇异贝叶斯学习提供一种有益的组织框架。

结构安排。第 2 节回顾回火及奇异学习理论的基本背景。第 3 节介绍可观测量代数及示例。第 4 节阐述响应函数，并将 WAIC/WBIC 纳入这一体系。第 5 节将奇异涨落解释为曲率响应。第 6 节介绍实验，第 7 节讨论其影响和局限性。

2 背景

2.1 奇异模型与后验温度化

令 $\Theta \subset \mathbb{R}^d$ 表示一个参数空间，并令

$$p(x \mid \theta), \qquad \theta \in \Theta,$$

构成一个参数统计模型。如果映射

$$\theta \mapsto p(\cdot \mid \theta)$$

不是局部单射，并且在实现真实分布的参数值处，费 Fisher 信息矩阵可能退化，则称该模型为 <style id='1'>奇异的</style>。等价地，不同的参数值可能诱导出相同的预测分布，而似然曲面可能呈现平坦方向或自交。混合模型、低秩矩阵分解，以及具有置换对称性的神经网络都是典型示例。

给定数据 $D = \{x_i\}_{i=1}^n$ 和先验 $\pi(\theta)$，温度化后验分布定义为（ $\beta > 0$ ）：

$$\pi_\beta(\theta \mid D) \;\propto\; \pi(\theta)\, p(D \mid \theta)^\beta, \qquad p(D \mid \theta) = \prod_{i=1}^n p(x_i \mid \theta).$$

普通后验分布对应于 $\beta = 1$，而 $\beta \to 0$ 则恢复先验。改变 β 会产生一个单参数变形，连续地调整似然贡献相对于先验的权重。

定义配分函数

$$Z(\beta) = \int \pi(\theta)\, p(D \mid \theta)^{\beta}\, d\theta,$$

以及相应的自由能

$$F(\beta) = -\frac{1}{\beta} \log Z(\beta).$$

在正则模型中，log $Z(\beta)$的导数具有经典的渐近展开。然而，在奇异模型中，参数到分布的映射的几何性质会从根本上改变 $Z(\beta)$相对于样本量的缩放行为。

2.2 奇异学习理论的基本结果

设 $q(x)$表示真实数据生成分布。定义期望 Kullback–Leibler散度

$$K(\theta) = \int q(x)\, \log \frac{q(x)}{p(x \mid \theta)}\, dx.$$

Let

$$\theta_\star \in \arg\min_{\theta \in \Theta} K(\theta)$$

表示 KL散度的一个极小化点。在可实现情形下， $K(\theta_\star) = 0$ ，且极小化点集可能构成奇异簇，而非孤立点。

令经验负对数似然为

$$L_n(\theta) = -\frac{1}{n} \sum_{i=1}^{n} \log p(x_i \mid \theta).$$

在适当的正则性条件下，奇异学习理论表明，边际似然具有如下渐近展开

$$\log Z(1) = -nL_n(\theta_\star) + \lambda \log n - (m-1) \log \log n + O_p(1),$$

其中 λ 是实对数典范阈值 （RLCT）， m 是其重数 [7]。在正则模型中， $\lambda = d/2$；而在奇异模型中， λ 可以是分数，并反映了 K（θ）在其最小值集合附近的局部代数结构。

另一个关键量是奇异涨落 ν，它决定了后验分布 [8]下对数似然的主导阶涨落。它出现在预测风险展开中，并决定模型的有效复杂度。在正则模型中，$\nu = d/2$；但在奇异模型中，它通常不同于 λ。

广义适用贝叶斯信息准则（WBIC）在逆温度 $\beta_n = 1/\log\ n$下计算对数似然的温度化期望，从而为正则和奇异情形下的边际似然提供渐近近似。

在下文中，我们不依赖详细的渐近展开。相反，这些量充当结构性锚点：实对数典范阈值描述自由能的主导阶标度，奇异涨落控制温度化自由能的类曲率行为，而 WBIC 则选取特定的温度区间。

3 可观测量代数

3.1 可观测量与后验期望

一个可观测量 是一个可测函数

$$f : \Theta \to \mathbb{R}$$

即在温度化后验 $\pi_\beta(\theta \mid D)$ [5] 下可积。其期望

$$\mathbb{E}_\beta[f] = \int f(\theta)\, \pi_\beta(\theta \mid D)\, d\theta$$

因此是良定义的。重要的是，期望仅依赖于 π_β 经映射 $\theta \mapsto p(\cdot \mid \theta)$ 的前推。若两个参数值诱导出相同的预测分布，则仅凭数据无法区分它们对后验平均的贡献。这促使我们识别仅依赖于诱导预测分布的可观测量。

3.2 等价关系与诱导分布空间

定义一个等价关系

$$\theta \sim \theta' \quad \Longleftrightarrow \quad p(\cdot \mid \theta) = p(\cdot \mid \theta').$$

~ ~令

$\Theta = \Theta/\sim$ 表示等价类的集合。 Θ中的每个元素都唯一对应于模型族中的一个预测分布。~温度化后验通过推前在 Θ 上诱导出一个概率测度：

$$\widetilde{\pi}_\beta(A) = \pi_\beta(\{\theta \in \Theta : [\theta] \in A\}).$$

因此，分布不变可观测量的后验期望也可以等价地写成

$$\mathbb{E}_\beta[f] = \int_{\widetilde{\Theta}} \tilde{f}([\theta])\, d\widetilde{\pi}_\beta([\theta]).$$

~无需在 Θ 上添加额外的几何结构；我们仅将其用作预测等价的记账工具。

3.3 分布不变可观测量

可观测量 f 称为分布不变，如果

$$\theta \sim \theta' \;\Rightarrow\; f(\theta) = f(\theta').$$

此类函数下降为良定义的可测函数

$$\tilde{f} : \widetilde{\Theta} \to \mathbb{R}.$$

我们将分布不变可观测量的集合称为可观测量代数。从概念上讲，它们是预测分布空间上的可测函数。这一构造消除了纯粹由不可识别参数方向引起的变化，同时保留了全部后验可测预测结构；这与统计物理学和量子场论中规范不变可观测量的用法相平行，在那里物理量是在模去冗余参数化的意义下定义的。

从等价类到模型像。设

$$\Phi : \Theta \to \mathcal{M}, \qquad \Phi(\theta) := p(\cdot \mid \theta),$$

并定义模型像

$$\mathcal{M} := \Phi(\Theta) = \{\, p(\cdot \mid \theta) : \theta \in \Theta \,\}.$$

根据构造， $\theta \sim \theta'$ 当且仅当 $\Phi(\theta) = \Phi(\theta')$。因此存在一个典范双射

$$\iota : \widetilde{\Theta} \to \mathcal{M}, \qquad \iota([\theta]) := \Phi(\theta),$$

这是定义良好的，因为 Φ在等价类上为常数。

命题 1（ $\mathcal{M}$上的可观测表示）。 设 $f : \Theta \to \mathbb{R}$ 是一个可测函数。

1. 若 f 是分布不变的（即 $\theta \sim \theta' \Rightarrow f(\theta) = f(\theta')$），则存在一个可测泛函 $\psi : \mathcal{M} \to \mathbb{R}$，使得

$$f(\theta) = \psi(\Phi(\theta)) \quad \textit{for all } \theta \in \Theta.$$

   等价地， $f = \psi \circ \Phi$。

2. 反过来，任意可测的 $\psi : \mathcal{M} \to \mathbb{R}$ 都定义了一个关于 Θ 的分布不变可观测量 $f = \psi \circ \Phi$ 。

因此，分布不变可观测量恰好是诱导预测分布的可测泛函。

~ ~证明概要。若 f 是分布不变的，则在 Θ 上定义 $f([\theta]) := f(\theta)$，根据不变性，该定义是良定的。令 $\psi := \tilde{f} \circ \iota^{-1}$，即可得到 $f = \psi \circ \Phi$。反向结论显然成立，因为 $\Phi(\theta) = \Phi(\theta')$ 意味着 $\psi(\Phi(\theta)) = \psi(\Phi(\theta'))$. □

该商构造确保响应函数仅依赖于预测分布，因而对不可识别参数方向保持不变。

3.4 可识别与具有统计意义

可辨识性是一种局部性质：某个参数在邻域内可能局部可辨识，但由于对称性而在全局上不可辨识。反之，某些不可识别方向可能不带来任何预测后果。

可观测量代数提取出在预测分布空间上定义良好的函数。在正则模型中，等价类通常是离散的，且 $\mathcal{O}$ 与整个函数空间重合。在奇异模型中，商空间并非平凡，并为构造有意义的序参量提供指引。

3.5 示例

对称高斯混合模型。考虑一个具有对称均值 $\pm\mu$ 且权重相等的双成分高斯混合模型。参数化 (μ) 和 $(-\mu)$ 诱导出相同的预测分布。可观测量 μ 不是分布不变的，而 $|\mu|$ 是。因此，$|\mu|$ 表示 $\mathcal{O}$中的一个定义良好的元素。

降秩回归。设 $B = UV^\top$，其中 $U \in \mathbb{R}^{p\times r}$ 且 $V \in \mathbb{R}^{q\times r}$。对于任意可逆的 $R \in \mathrm{GL}(r)$,

$$(U, V) \mapsto (UR, VR^{-T})$$

使 B 保持不变。U 和 V的各个元素都不是分布不变的，而 B 的函数（例如其奇异值）则是。

具有置换对称性的神经网络。考虑一个具有 H 个隐藏单元的单隐藏层神经网络，

$$f_\theta(x) = \sum_{h=1}^{H} a_h\, \sigma(w_h^\top x),$$

其参数为 $\theta = \{(a_h,\ w_h)\}_{h=1}^{H}$。对于 $\{1, \dots, H\}$的任意置换 π，

$$(a_h, w_h) \mapsto (a_{\pi(h)}, w_{\pi(h)})$$

都会使诱导出的预测函数保持不变。因此，由隐藏单元置换关联的参数配置属于同一个等价类。单个权重向量 w_h 并非分布不变，而诱导出的预测器 f_θ的泛函（例如其在数据分布上的 L^2 范数或其雅可比矩阵的谱性质）则是不变量。

纵观这些例子，可观测量代数识别出具有规范不变性的量，它们在后验温度化下会产生有意义的响应。

4 热力学响应函数

后验温度化在参数空间上引入了一族由单个参数索引的概率测度$\{\pi_\beta(\theta \mid D)\}_{\beta>0}$。因此，可积可观测量$f$通过其温度化期望$\mathbb{E}_\beta[f]$诱导出一个关于逆温度的标量函数。本节旨在形式化描述这些期望如何随 β 变化，并将由此得到的导数和方差整理成一个响应函数层级。下文讨论的所有响应函数均针对分布不变可观测量定义，以确保所得量仅取决于预测分布，而非参数化方式。

我们首先给出一个通用协方差恒等式。该恒等式纯粹基于测度论，只需满足可积性，不需要其他正则性假设。随后我们证明，当 f 是分布不变的（其含义见第 3 节）时，同样的恒等式可以在模型像 $\mathcal{M} = \Phi(\Theta)$ 上以内蕴的方式表示，从而阐明响应函数取决于预测结构，而非参数化。

4.1 通用协方差恒等式

Let

$$\ell(\theta) := \log p(D \mid \theta)$$

表示对数似然。

命题 2 (协方差恒等式**)**。对于在 π_β 下可积的任意可观测量 f，

$$\frac{d}{d\beta}\mathbb{E}_\beta[f] = \mathrm{Cov}_\beta(f, \ell).$$

证明。记作

$$\mathbb{E}_\beta[f] = \frac{\int f(\theta)\pi(\theta)p(D \mid \theta)^\beta d\theta}{\int \pi(\theta)p(D \mid \theta)^\beta d\theta}.$$

对分子和分母分别关于 β 求导，并应用商法则，得到

$$\frac{d}{d\beta}\mathbb{E}_\beta[f] = \mathbb{E}_\beta[f\ell] - \mathbb{E}_\beta[f]\mathbb{E}_\beta[\ell],$$

这恰好就是 $\mathrm{Cov}_\beta(f,\ \ell)$。 □

这个恒等式表明，任何可观测量对温度参数的敏感性都由其与对数似然的协方差决定。因此，后验几何中的涨落决定了可观测量如何响应有效温度的变化。

从参数空间到分布空间。 命题 2 是在参数空间 Θ上陈述的，但在奇异模型中，许多不同的 θ 都对应于同一个预测分布。第 3 节通过模型映射$\Phi(\theta) = p(\cdot \mid \theta)$和模型像 $\mathcal{M} = \Phi(\Theta)$对此进行了形式化。接下来的命题记录了一个有用的推论：当 f 具有分布不变性时，期望$\mathbb{E}_\beta[f]$ 及其 β-导数完全由 $\mathcal{M}$上的推前后验决定。从这个意义上说，响应恒等式是统计模型本身所固有的，而非取决于所选的参数化。

命题 3（响应恒等式可下降至模型像）。设 $\Phi(\theta) = p(\cdot \mid \theta)$，并设 $\mathcal{M} = \Phi(\Theta)$ 为模型像。若 f 是分布不变的，即对于某个可测的 $\psi \colon \mathcal{M} \to \mathbb{R}$，有 $f = \psi \circ \Phi$，则对任意 $\beta > 0$，

$$\mathbb{E}_\beta[f] = \int_{\mathcal{M}} \psi(p)\, d\widetilde{\pi}_\beta(p), \qquad \widetilde{\pi}_\beta := \Phi_{\#}\pi_\beta,$$

且热力学响应恒等式可完全在 $\mathcal{M}$上表述：

$$\frac{d}{d\beta}\mathbb{E}_\beta[f] = \mathrm{Cov}_\beta(f, \ell), \qquad \ell(\theta) = \log p(D \mid \theta),$$

其中，f 和 ℓ在等价类上均为常数，因此定义了 $\mathcal{M}$上的可测泛函。

证明概要。第一个恒等式是推前测度的定义。由于$\ell(\theta) = \log p(D \mid \theta)$仅取决于诱导预测分布，因此它是分布不变的。由此，协方差以及关于 β 的微分在 $\mathcal{M}$ 层面上都有良好定义。 □

推论综述。 命题 2–3共同表明，温度调节赋予后验分布一种线性响应结构：期望的导数是协方差，曲率是方差。因此，自然可以将 $(\beta,\ f) \mapsto \mathbb{E}_\beta[f]$视为基本热力学对象，并将序参量、磁化率和复杂度度量定义为由这一协方差结构导出的响应。这些恒等式适用于任意可积可观测量，因此无论统计模型是正则还是奇异的，都适用。下一个定理将概述由此得到的层级关系。图 1示意展示了这些关系。

定理 1（热力学响应层级）。设 $\pi_\beta(\theta|D)$ 为温度化后验分布，设 f 为在 π_β 下可积的任意可观测量。则温度化期望的导数由后验涨落结构支配：

$$\frac{d}{d\beta}\mathbb{E}_\beta[f] = \mathrm{Cov}_\beta(f, \ell).$$

因此：

- *Order parameters correspond to expectations* $m(\beta) = \mathbb{E}_\beta[f]$.

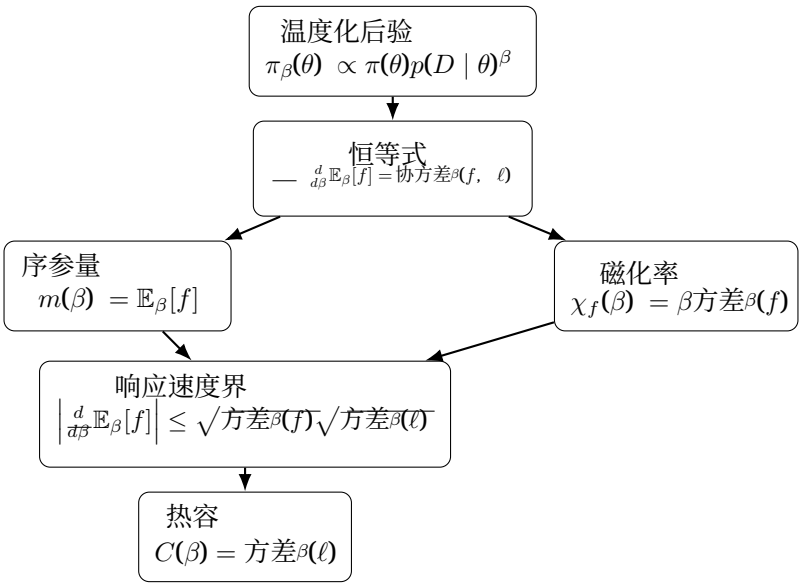

图 1：序参量和磁化率源自协方差恒等式。响应速度界将变化率与涨落幅度联系起来，并与热容自然关联。

- 磁化率对应于涨落幅度 $\chi_f(\beta) = \beta\text{方差}_\beta(f)$。
- 对数配分函数的曲率满足

$$\frac{d^2}{d\beta^2}\log Z(\beta) = \mathrm{Var}_\beta(\ell).$$

因此，温度化后验的所有响应函数都由后验协方差结构生成。此外，高阶导数会生成温度化后验下对数似然的高阶累积量。

序参量。我们从一阶响应开始：分布不变可观测量的期望。给定一个分布不变可观测量 f，定义序参量

$$m(\beta) = \mathbb{E}_\beta[f].$$

序参量追踪后验分布的结构属性如何随 β变化。在奇异模型中，这些量通常会发生剧烈变化，与后验质量的结构重组相关。

磁化率。 可观测量在回火过程中的变异性定义了磁化率

$$\chi_f(\beta) = \beta \operatorname{Var}_\beta(f).$$

磁化率较大，表明可观测量在后验分布下的涨落很强，意味着它对有效温度的变化十分敏感。类比统计物理， $\chi_f(\beta)$中的峰值表示后验分布相对于 β快速重组的区域。

响应速度与涨落约束。 序参量描述后验分布的结构性质，而磁化率则量化其在回火过程中的涨落幅度。接下来一个自然的问题是，这些结构量随逆温度 β变化的速度有多快。协方差恒等式表明，这类变化由对数似然的后验涨落驱动。因此，可观测量对 β 变化的响应速率受到该可观测量与对数似然联合变异性的限制。下面的命题给出了这一响应速度的定量界。

命题 4（响应速度界）。 设 f 是关于 π_β可积的可观测量。那么

$$\left| \frac{d}{d\beta} \mathbb{E}_\beta[f] \right| = |\operatorname{Cov}_\beta(f, \ell)| \le \sqrt{\operatorname{Var}_\beta(f)} \, \sqrt{\operatorname{Var}_\beta(\ell)}.$$

等价地，用磁化率 $\chi_f(\beta) = \beta \operatorname{Var}_\beta(f)$和热容$C(\beta) = \operatorname{Var}_\beta(\ell)$表示，

$$\left| \frac{d}{d\beta} \mathbb{E}_\beta[f] \right| \le \sqrt{\frac{\chi_f(\beta)}{\beta}} \, \sqrt{C(\beta)}.$$

证明。 根据命题 2， $\frac{d}{d\beta}\mathbb{E}_\beta[f] = \operatorname{Cov}_\beta(f, \ell)$。柯西–施瓦茨不等式给出$|\operatorname{Cov}_\beta(f, \ell)| \le \sqrt{\operatorname{Var}_\beta(f)}\sqrt{\operatorname{Var}_\beta(\ell)}$。 □

命题 4 表明，回火过程中序参量的快速变化必然伴随着序参量本身的较大涨落（较大的磁化率）或对数似然的较大涨落（较大的热容），从而为使用这些响应函数的峰值作为结构重组指标提供了定量依据。

注（涨落–不确定性关系）。 命题 4 类似于傅里叶分析和统计力学中的经典不确定性关系：可观测量对控制参数变化的敏感性 β 受到该可观测量与对数似然的联合涨落的限制。

4.2 复杂度响应

某些响应函数与广泛使用的贝叶斯复杂度度量直接对应。

WAIC复杂度。定义

$$p_{\mathrm{WAIC}} = \sum_{i=1}^{n} \mathrm{Var}_{\beta}(\log p(y_i \mid \theta)) \,.$$

该量衡量后验分布中的预测方差，并在 WAIC [8, 3] 中充当有效复杂度校正。由于它是预测对数似然项的方差，WAIC复杂度自然表现为温度化后验中的二阶响应量。

由于 $\log p(y_i \mid \theta)$ 仅取决于诱导预测分布，WAIC复杂度对不可识别参数方向保持不变。如果两个参数值 θ 和 θ' 诱导出相同的预测分布，那么

$$\log p(y_i \mid \theta) = \log p(y_i \mid \theta').$$

因此，纯粹沿不可识别方向发生的后验变异不会对 WAIC 复杂度项产生贡献。从这个意义上说，WAIC 衡量的是预测波动，而非参数波动。这种不变性与第 3 节的可观测量框架自然契合：WAIC 完全由分布不变可观测量决定。

热容。 对数似然的方差

$$C(\beta) = \mathrm{Var}_{\beta}(\ell)$$

在热力学中发挥着类似热容的作用。它衡量温度化后验下对数似然的涨落。在统计物理中，热容衡量系统能量的涨落，并且通常会在相变附近达到峰值，此时系统会在相互竞争的构型之间重新组织 [5, 2]。这里也会出现类似现象。当后验质量在参数空间中那些诱导出定性不同预测行为的区域之间转移时，对数似然在温度化过程中可能会强烈波动。这些波动会使 $C(\beta)$ 出现峰值。在奇异模型中，这种重新组织常发生在退化参数构型附近，此处多个预测结构共存（例如混合分量合并或秩坍缩）。因此，热容峰值可作为后验几何中结构转变的经验指标。

4.3 统一视角

序参量、磁化率与复杂度度量均源于回火下期望的导数。协方差恒等式表明，这些量最终都源自后验分布的同一涨落结构。在此框架下，WAIC和WBIC等常用准则可解释为在特定温度下求值的特定响应函数。综合来看，这些响应函数区分了预测结构变化与参数冗余。热容

检测后验支撑集的重组，而WAIC则度量与不可识别参数方向无关的预测变异性。

5 奇异涨落作为曲率响应

上文建立的响应函数框架，为奇异学习理论中的经典量提供了自然的解释。具体而言，奇异涨落可以视为温度化自由能的曲率响应。

5.1 自由能曲率

回顾温度化配分函数

$$Z(\beta) = \int \pi(\theta)\, p(D \mid \theta)^{\beta} d\theta,$$

以及相应的自由能

$$F(\beta) = -\frac{1}{\beta} \log Z(\beta).$$

对 $\log Z(\beta)$ 求导可得

$$\frac{d}{d\beta} \log Z(\beta) = \mathbb{E}_{\beta}[\ell], \qquad \ell(\theta) = \log p(D \mid \theta).$$

对其求二阶导数可得

$$\frac{d^2}{d\beta^2} \log Z(\beta) = \mathrm{Var}_{\beta}(\ell).$$

因此，对数似然的方差对应于对数配分函数关于温度参数的曲率。用热力学语言来说，这个量起到热容的作用。

从统计学角度看，这种曲率量化了不同参数区域对数据提供相互竞争解释的程度。如果后验分布集中在单一预测构型附近，那么对数似然在 π_β 下几乎保持不变，曲率也较小。相反，当多个结构不同的解释并存时——例如不同的混合分配、秩或网络表示——对数似然在后验样本之间会有显著涨落，从而产生较大的曲率。

5.2 奇异涨落

奇异学习理论引入了量 ν，称为奇异波动，该量控制着对数似然的后验变异性，并出现在

预测风险展开中。在正则模型中，$\nu = d/2$；而在奇异模型中，ν 通常不同于实对数典范阈值 λ。

在响应框架中，ν 可解释为大样本极限下，对数似然波动对温度化自由能曲率的主导阶贡献。由于这些波动对应于对数配分函数的曲率，奇异涨落便发挥着与后验分布几何结构相关的曲率响应作用。这表明，ν 衡量的是多个不同的参数配置对数据给出相互竞争的解释时所产生的预测不稳定性。

5.3 与 WAIC 和 WBIC 的关系

WAIC 通过衡量逐点对数似然项的后验方差 [3] 来估计预测风险。如第 4 节所述，该方差源于决定自由能曲率的同一涨落结构。WBIC 在逆温度为 $\beta_n = 1/\log n$ 的温度化后验分布下计算对数似然的期望。该温度对应于奇异学习理论中的边际似然渐近性质显现的状态。从响应视角来看，WBIC 探测的是自由能景观，其中由 λ 和 ν 决定的曲率效应在该温度下决定模型复杂度。这些联系共同表明，WAIC、WBIC 和奇异涨落属于由后验温度化所产生的统一响应函数层级，概述见表 1。

6 实验

我们通过实证研究，考察响应函数框架在三类典型奇异统计模型中的表现：

1. 混合模型中的对称性破缺 2. 降秩回归中的
秩坍缩 3. 神经网络中的隐藏单元冗余

这三个例子对应于产生奇异统计模型的三种典型机制：对称性、代数约束和参数冗余。在每种情况下，我们都在温度化后验分布下评估可观测量

$$\pi_\beta(\theta) \propto \pi(\theta) p(D|\theta)^\beta$$

在跨越多个数量级的逆温度对数网格 β 上取值，以同时涵盖先验主导区间和似然主导区间。这些实验考察序参量、磁化率和预测复杂度如何随后验集中而变化。每个温度下的后验样本均通过哈密顿蒙特卡洛获取，并采用针对该温度的似然缩放。

在所有实验中，我们报告三个量：

Response level	Thermodynamic quantity	Statistical quantity	Interpretation
Order parameter	$m(\beta) = \mathbb{E}_\beta[f]$	Structural observable	Effective model structure
Susceptibility	$\chi_f(\beta) = \beta \operatorname{Var}_\beta(f)$	Structural fluctuation	Sensitivity of structure to tempering
Heat capacity	$C(\beta) = \operatorname{Var}_\beta(\ell)$	Log-likelihood fluctuation	Posterior competition between explanations
Predictive response	p_{WAIC}	WAIC complexity	Predictive variability across posterior
Free-energy slope	RLCT λ	Marginal likelihood scaling	Leading asymptotic complexity
Free-energy curvature	Singular fluctuation ν	Predictive risk correction	Curvature of tempered free energy
Special temperature probe	WBIC ($\beta_n = 1/\log n$)	WBIC estimator	Marginal likelihood approximation regime

Table 1: Hierarchy of thermodynamic response quantities generated by posterior tempering and their relationship to classical quantities from singular learning theory.

- 序参量 $m(\beta)$，用于描述模型中活跃组成部分的有效数量。在每项实验中，选择序参量来代表一种可观测量，以反映模型所用的有效结构自由度。

- 磁化率

$$\chi(\beta) = \beta \operatorname{Var}_\beta(m)$$

  用于衡量序参量的涨落

- WAIC复杂度 $p_{\mathrm{WAIC}}(\beta)$

磁化率发挥热力学响应函数的作用，并能检测后验结构中的转变。

在三项实验中，我们观察到一致的响应层级（图2–4）：序参量随温度平滑变化，磁化率在结构转变附近出现峰值，而冗余结构消失时，预测复杂度随之降低。

6.1 混合模型中的对称性破缺

第一个实验研究了一个具有可交换分量的简单高斯混合模型。混合模型具有置换对称性：交换分量标签不会改变似然。当分量发生重叠时，这种对称性会产生奇异后验几何。

随着 β 增加，后验质量集中到只有一部分混合分量实际处于激活状态的构型上。序参量衡量的是实际被占据的分量数量。

结果显示出三个阶段：

1. 低 β：各分量的贡献大致相等
2. 中等 β：对称性开始破缺
3. 高 β：单一构型占主导

该行为如图2所示，图中给出了序参量、磁化率和WAIC复杂度随逆温度的变化。磁化率在对称区制与对称破缺区制之间的转变附近出现峰值，表明分量分配的涨落达到最大。

6.2 降秩回归中的秩坍缩

第二个实验考察降秩回归（RRR），这是一种经典的奇异模型，其参数空间包含有界秩矩阵。设 $Y = XB + \epsilon$，其中 B 被约束为具有秩 r。当 B 的奇异值趋近于零时会出现奇异点，导致参数空间中的某些方向变得不可辨识。

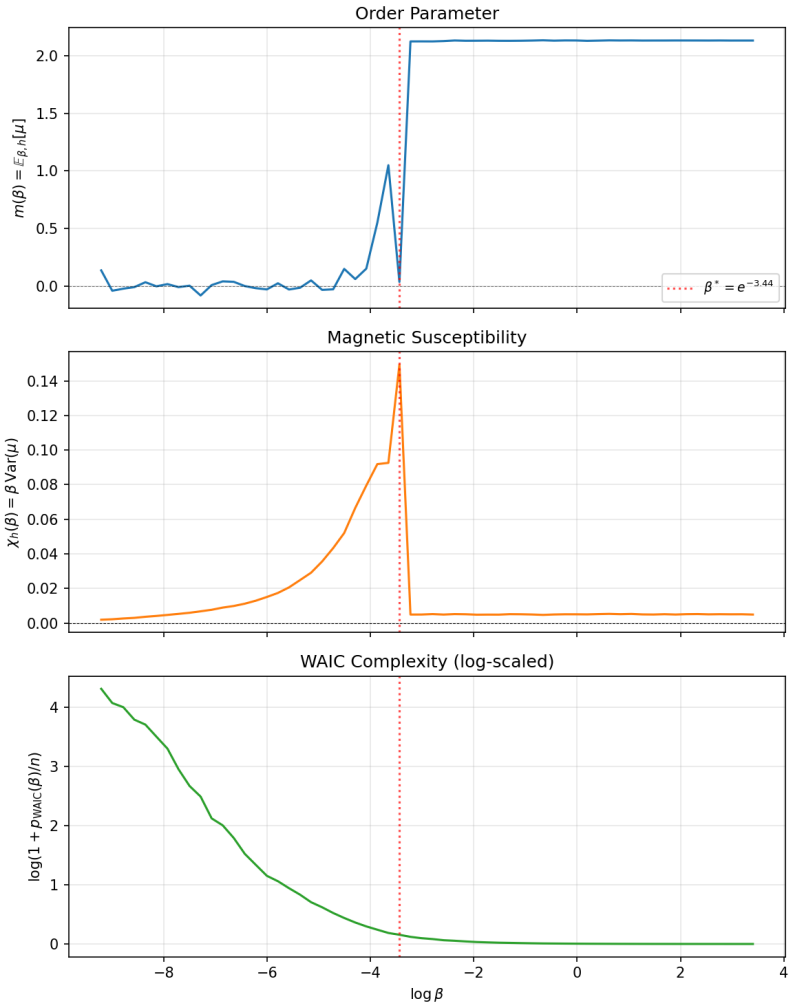

图 2：混合对称性破缺实验的响应层级。上图显示序参量 $m(\beta) = \mathbb{E}_\beta[|\mu|]$，它衡量后验分布对某一分量均值相对于对称构型的偏好。在低逆温度下，后验分布会探索两个对称模态。随着 β 增加，后验分布集中于一个模态，从而产生自发对称性破缺。中图显示磁化率$\chi(\beta) = \beta\mathrm{Var}(|\mu|)$，它在转变附近达到峰值，此时后验分布在对称构型之间波动。下图显示 WAIC复杂度 $\log(1 + p_{\mathrm{WAIC}}(\beta)/n)$。随着后验分布逐渐集中，预测方差降低，表明对称性破缺后预测不确定性减小。竖直虚线标示磁化率达到最大值时的温度。

序参量由估计系数矩阵的有效秩定义。随着 β 增大，后验分布更倾向于与数据一致的较低秩结构。我们再次观察到，在秩降低发生之处出现磁化率峰值，如图3所示。这反映了具有不同有效维数的模型之间存在强烈的后验涨落。

6.3 神经网络中的隐藏单元坍缩

第三个实验研究在教师–学生回归任务上训练的双层神经网络。该模型为

$$f(x) = \sum_{j=1}^{H} a_j \phi(w_j x + b_j) + c.$$

神经网络表现出若干奇异性来源：隐藏单元的置换对称性、 a_j 与 w_j之间的缩放简并，以及隐藏单元 [1, 6]的冗余。为刻画活跃单元，我们定义振幅

$$s_j = |a_j||w_j|$$

并计算活跃单元的有效数目

$$N_{\mathrm{eff}} = \frac{\left(\sum_j s_j\right)^2}{\sum_j s_j^2}.$$

该量近似等于对网络输出有实质贡献的隐藏单元数目。磁化率的计算方式为

$$\chi(\beta) = \beta \, \mathrm{Var}_\beta(N_{\mathrm{eff}}).$$

由此得到的响应层级如图4所示。随着 β 增加，冗余单元逐渐坍缩，有效活跃单元的数量随之减少。磁化率在不同活跃单元数量的多种构型共存时出现显著峰值。WAIC复杂度呈现出类似趋势，证实预测变异性在序参量涨落最大的同一区域达到最高。

6.4 实证发现总结

在所有三个模型中，我们都观察到相同的定性行为：

1. 序参量揭示了后验分布中有多少有效自由度处于活跃状态。2. 磁化率在不同后验结构之间发生转变附近达到峰值。

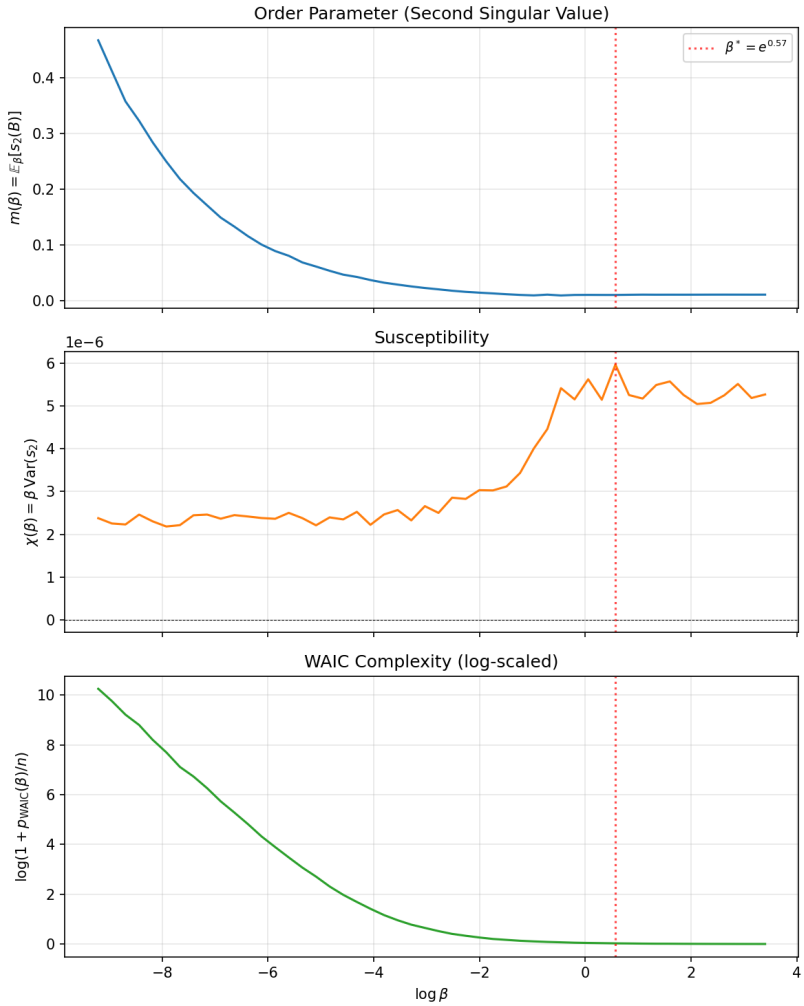

图3：降秩回归的响应层级。序参量$m(\beta) = \mathbb{E}_\beta[s_2(B)]$ 跟踪回归矩阵的第二个奇异值。随着β 增加，后验集中会推动第二个奇异值趋近于零，表明模型坍缩为低秩模型。磁化率 $\chi(\beta) = \beta\mathrm{Var}(s_2)$ 衡量有效秩的涨落，并在发生秩坍缩的温度附近达到峰值。随着后验分布消除参数空间中的冗余方向，WAIC复杂度会降低。磁化率与预测复杂度之间的对应关系说明，奇异结构如何控制预测变异性。

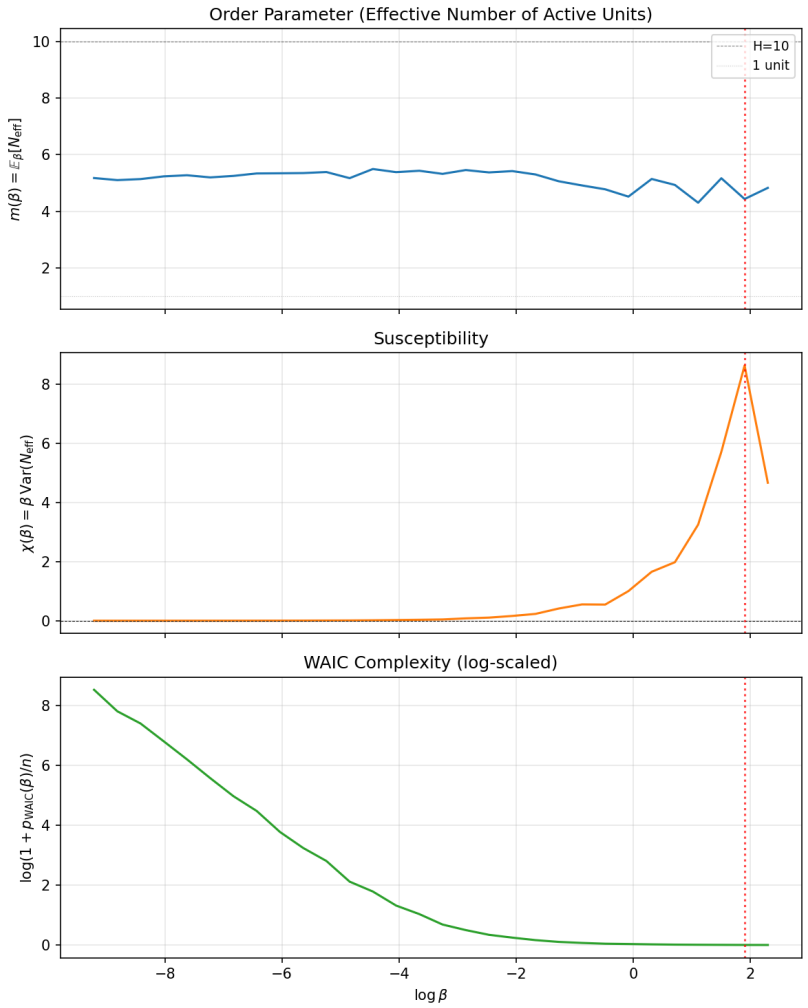

图4：神经网络隐藏单元坍缩实验的响应层级。序参量 $m(\beta) = \mathbb{E}_\beta[N_{\text{eff}}]$ 用于衡量有效活动隐藏单元数。尽管网络包含 $H = 10$ 个单元，随着 β 增大，后验分布倾向于更小的有效单元数。由于对称性和缩放简并，冗余隐藏单元会变为非活跃状态。当具有不同活跃单元数的多种构型共存时，磁化率$\chi(\beta) = \beta\,\text{Var}(N_{\text{eff}})$达到峰值。此区域对应于网络表示后验不确定性最大的情况。随着冗余单元坍缩，WAIC复杂度降低，这表明网络内部表示不稳定时，预测不确定性最高。

3. WAIC复杂度反映预测变异性，并与磁化率较大的区域相对应。

这些结果支持这样一种解释：WAIC 等信息准则对应于调温后验分布的响应函数。因此，奇异学习现象自然地呈现于热力学响应层级之中。

7 讨论

7.1 奇异学习的热力学解释

结果表明，奇异统计模型中的后验几何呈现出一种热力学响应结构。在三种性质迥异的奇异机制中——混合模型中的对称性破缺、降秩回归中的秩坍缩，以及神经网络中的隐藏单元冗余——我们观察到了相同的定性行为。随着逆温度 β 升高，后验质量集中到参数空间中维度较低的区域，在这些区域里，模型实际上使用的自由度更少。在这些转变附近，后验分布在相互竞争的构型之间波动，从而使磁化率 $\chi(\beta) = \beta \mathrm{Var}_\beta(m)$ 出现峰值。这种行为与统计物理中的相变十分相似：宏观可观测量会随温度变化而剧烈响应。在统计学情境中，控制参数 β决定后验集中，而非物理能量，但由此产生的响应结构似乎与之类似。

重要的是，响应框架并不依赖渐近展开或特定于模型的代数分析。相反，它识别出一些可观测量，这些量在回火下的行为反映了驱动奇异学习理论的同一几何结构。从这个意义上说，响应函数提供了一种有限样本诊断，可作为实对数典范阈值和奇异涨落等渐近不变量的对应物。

7.2 作为响应函数的信息准则

实验还提示了对预测信息准则的一种热力学解释。WAIC复杂度项衡量逐点对数似然贡献的后验方差，因此量化了预测性能的涨落。经验上，我们观察到，WAIC复杂度在结构序参量磁化率最大的相同温度附近变化最快。这支持了这样一种解释：预测信息准则可视为温度化后验分布的响应函数。从这一视角看，WAIC 和 WBIC 等量描述了预测性能如何响应后验集中程度的变化。

7.3 可观测量与可识别结构

第 3 节介绍可观测量框架，其核心动机之一是：奇异模型的参数空间中存在一些方向，它们并不对应不同的概率分布。置换对称性、缩放简并和冗余参数化都会引入对预测不可见的规范方向。可观测量是定义在概率分布空间而非参数空间上的函数。由于它们仅取决于诱导预测分布，可观测量对这些不可识别方向保持不变。这一视角阐明了为何即使费 Fisher 信息矩阵是奇异的，WAIC 等量仍然有明确定义。

7.4 实际意义

从响应函数的视角审视奇异模型，也可能带来实际洞见。易感性峰值标示后验分布几何结构快速变化的区域，这些区域可能对应于模型选择边界，或有效模型结构之间的转变。例如，隐藏单元坍缩实验表明，在冗余单元变为非活跃状态的转变附近，有效活跃单元数量会出现强烈波动。此类行为可为理解过参数化模型中的容量与冗余提供有用的诊断依据。

7.5 局限性

本文所呈现的实验具有说明性，而非穷尽性。它们表明若干典型奇异模型呈现出一致的定性行为，但尚未对响应函数给出完整的理论刻画。此外，在较宽的温度范围内探索温度化后验，对大型模型而言计算代价可能很高。开发这些响应量的更高效估计量，仍是一个有待解决的问题。

7.6 未来方向

这一视角引出了若干理论方向。首先，响应函数所依据的协方差恒等式表明，可能存在一种作用于可观测量的更一般响应算子。理解该算子的谱性质，或许能够揭示后验几何如何组织奇异模型。其次，与统计物理的类比表明，重整化群思想可能有助于洞察有效模型结构如何随尺度而变化 [11]。最后，将这一框架扩展到奇异模型中的不确定性量化，或许能够为理解过参数化系统中的预测变异性提供新的工具。

8 结论

本文基于定义在预测分布空间上的可观测量，提出了关于奇异统计模型的热力学视角。通过聚焦于对不可识别参数方向保持不变的函数，我们获得了一种后验结构表示；该表示仅取决于所诱导的概率模型，而非特定的参数化方式。

在这一可观测量框架中，使用逆温度参数 β 对后验分布进行回火，会生成一套类似于统计物理学中所研究的响应函数层级结构。序参量描述有效模型结构，磁化率衡量该结构中的涨落，而 WAIC 等预测信息准则则会自然地作为与预测方差相关的响应量出现。

针对混合模型、降秩回归和神经网络的实验表明，不同类别的奇异模型呈现出一致的定性行为。在每种情况下，后验几何中的转变都表现为磁化率峰值以及预测复杂度的变化。

这些观察表明，热力学响应理论或许能够为理解现代统计模型中的奇异学习现象提供一个有用的组织框架。特别是，响应函数为奇异学习理论中的渐近不变量提供了一个有限样本诊断对应物。更广泛地说，这些结果表明，由后验分布回火诱导的热力学响应结构，可能为探究奇异统计模型的几何提供一条超越渐近理论的实用途径。

AI 致谢

作者在准备本文期间使用大型语言模型（ChatGPT）作为写作和编辑助手。该系统用于帮助整理提纲、改进论述并编辑草稿。所有科学思想、数学论证、实验设计、代码和解释均由作者提出并核实。

参考文献

[1] Shun-ichi Amari. 自然梯度在学习中高效运作。*NeuralComputation*, 10:251–276，1998。[2] Herbert Callen.《热力学与热统计学导论》。Wiley，1985。[3] Andrew Gelman、Jessica Hwang 和 Aki Vehtari. 理解贝叶斯模型的预测信息准则。*Statistics and Computing*，24:997–1016，2014。

[4] Radford Neal. 退火重要性采样。*Statistics and Computing*，11:125–139，2001。[5] R. K. Pathria 和 Paul Beale. 统计力学。Elsevier，2011。[6] Sumio Watanabe. 不可辨识学习机的代数分析。*Neural Computation*，13:899–933，2001。[7] Sumio Watanabe. 代数几何与统计学习理论。剑桥大学出版社，2009。[8] Sumio Watanabe. 贝叶斯交叉验证与广义适用信息准则的渐近等价。机器学习研究杂志，11:3571–3594，2010。[9]Sumio Watanabe. 一种广义适用贝叶斯信息准则。机器学习研究杂志，14:867–897，2013。[10] Sumio Watanabe. 贝叶斯统计数学。CRC出版社，2018。[11] Kenneth Wilson. 重整化群：临界现象与近藤问题。《现代物理评论》，47:773–840，1975。