arXiv:2603.04686v1 [math.ST] 5 Mar 2026

增广 van Trees 不等式

Elliot H. Young *[1]

[1]剑桥大学统计实验室，英国

2026年3月

摘要

我们提出了 van Trees 不等式的一种增广形式，与经典 van Trees 不等式相比，它能为估计量的极小极大平方贝叶斯风险提供一致更紧的下界。我们的增广不等式还适用于先验分布的密度在其支撑集边界处不必为零的情形。我们展示了如何利用这一改进，以初等方法证明若干非参数估计目标的极小极大下界；这些下界通常还能达到比通过另一种方法——Le Cam 实验收敛理论和经典 van Trees 不等式——所得结果更尖锐的常数，在某些情况下甚至能得到精确常数。举例而言，我们的增广 van Trees 不等式可用于求得正态误差模型中估计回归函数时的渐近极小极大逐点均方误差：当回归函数为单变量且可微、其导数满足 Lipschitz 条件时，我们得到该量，精确到 1.37 倍的常数因子；而在高维情形下，对于光滑度为 $\beta \in (0,2]$ 的 H¨older 光滑回归函数，我们得到精确常数。这两项结果都无法通过应用经典 van Trees 不等式得到。我们的增广 van Trees 不等式具有灵活性，可为高斯性之外的模型以及平方误差损失之外的损失函数给出下界；我们还能够将这一增广方法纳入适用于非正则模型的 van Trees 不等式推广形式中。

1 引言

van Trees 不等式（van Trees，1968年）是一个广受称道的结果，它给出了给定先验下估计量贝叶斯风险的下界，通常被视为贝叶斯 Cram´er–Rao 界。其在统计理论中的重要性由Gill and Levit （1995年）所推广，他们展示了如何利用该不等式得到极小极大下界。特别地，对于$\sqrt{n}$相合的参数估计问题，van Trees 不等式给出渐近锐利风险下界，从而恢复经典结果，例如参数模型中的 Cram´er–Rao 下界，以及 $\sqrt{n}$相合半参数情形中的局部渐近极小极大性结果（vander Vaart，1998年；GassiatandStoltz，2024年）。

在非参数估计中，van Trees 不等式可以提供一种极为简单的策略来推导极小极大下界，从而绕开 Hajek （1970年）；Le Cam （1972年）；Assouad （1983年）所提出的精巧实验收敛理论。然而，在许多非参数问题中，这些更为精巧的构造可以用来给出在常数方面比 van Trees 不等式所得结果更紧的极小极大下界。

在本文中，我们建立一种增广 *van Trees* 不等式，它既扩大了可容许先验分布的类别，从而严格改进了最坏情形风险的下界。将这一推广应用于极小极大下界的构造时，试图兼取两者之长：既提供与 van Trees 不等式同样极为简单、可直接使用的方法，又能得到更紧的常数，这些常数往往比上文所述更为复杂的实验收敛理论所得的常数更为尖锐。作为例子，我们研究逐点 H¨older 函数估计（见下文第 3 节），其中我们尤其关注在极小极大下界中获得（近乎）尖锐的常数。

*ey244@cam.ac.uk

1.1 van Trees 不等式及我们的贡献概述

为固定记号，考虑参数模型 $(P_t)_{t\in T}$，其由紧区间 $T=[t_1,t_2]$ 标号，其中 $-\infty<t_1<t_2<\infty$。假设 $P_t\ll\nu$ 对所有 $t\in T$ 成立，且存在某个 σ-有限测度 ν，其定义在 $(\mathcal{X},\mathcal{A})$ 上，并令 $p(\cdot,t)$ 表示 P_t 关于 ν 的密度，我们假设该密度是绝对连续的。我们将费舍尔信息定义为

$$\mathcal{I}(t):=\int_{\mathcal{X}}\left(\frac{\partial_t p(x,t)}{p(x,t)}\right)^2 p(x,t)\,d\nu(x),$$

我们假设其对所有 $t\in T$ 都是有限的。给定一个依赖于数据 $\boldsymbol{X}\sim P_t$ 的任意估计量 $\hat{t}(\boldsymbol{X})$，我们主要关注如何获得最坏情况风险的下界

$$\sup_{t\in T}\mathbb{E}_{P_t}\left[(\hat{t}(\boldsymbol{X})-t)^2\right]=\sup_{t\in T}\int_{\mathcal{X}}(\hat{t}(x)-t)^2p(x,t)\,d\nu(x). \tag{1}$$

用于极小极大下界的经典 van Trees 方法（例如 Tsybakov，2008）通过贝叶斯风险给出该量的下界

$$\sup_{t\in T}\mathbb{E}_{P_t}\left[(\hat{t}(\boldsymbol{X})-t)^2\right]\geq\int_T\mathbb{E}_{P_t}\left[(\hat{t}(\boldsymbol{X})-t)^2\right]\mu(t)\,dt=\int_T\int_{\mathcal{X}}(\hat{t}(x)-t)^2p(x,t)\,d\nu(x)\,\mu(t)\,dt, \tag{2}$$

对于 T 上某个合适的先验分布，其具有相对于勒贝格测度绝对连续的密度 μ，且先验信息 $\mathcal{J}(\mu)$ 有限：$=\int_T\frac{(\mu')^2}{\mu}$。请注意，在此处以及全文中，默认取 $1/0=0$，因此例如可以写作 $\int_T\frac{(\mu')^2}{\mu}\mathbb{1}_{\{\mu(\cdot)>0\}}=\int_T\frac{(\mu')^2}{\mu}$。

经典 van Trees 不等式（van Trees，1968）在先验密度绝对连续且在 T 的支撑集边界处为零时，为（2）提供下界。对于任意统计量 $\hat{t}(\boldsymbol{X})$ 以及满足 μ 且 $\mu(t_1)=\mu(t_2)=0$，

$$\int_T\mathbb{E}_{P_t}\left[(\hat{t}(\boldsymbol{X})-t)^2\right]\mu(t)\,dt\geq\frac{1}{\int_T\mathcal{I}(t)\mu(t)\,dt+\mathcal{J}(\mu)}. \tag{3}$$

对于 $T=[-1,1]$ 这类先验下 $\mathcal{J}(\mu)$的下确界为 π^2，由 $\mu(t)=\cos^2(\pi t/2)$取得。此外，如果$\mathcal{I}(\cdot)=\mathcal{I}$ 为常数（例如第3节的设定），那么 van Trees 不等式可简化为

$$\sup_{\mu}\int_{-1}^{1}\mathbb{E}_{P_t}\left[(\hat{t}(\boldsymbol{X})-t)^2\right]\mu(t)\,dt\geq\frac{1}{\mathcal{I}+\pi^2}, \tag{4}$$

上确界取遍上述所有有效先验。

我们的增广 *van Trees*不等式在两个关键方面扩展了 van Trees 不等式：

(i) 我们允许先验在 T 的边界处不必为零；并且

(ii) 我们的下界涉及一个增广函数，经过仔细选择后，可保证对最坏情况风险给出一致更紧的下界（1）。

关于(ii)，我们引入两个有用的代表性界，它们可由我们的增广 van Trees 不等式推导出，用于刻画最坏情况风险：

1. 增广 van Trees 1（AVT1）：一种简单易用但并非最优的增广形式为

$$\sup_{t\in[-1,1]}\mathbb{E}_{P_t}\left[(\hat{t}(\boldsymbol{X})-t)^2\right]\geq\max\left(\frac{1}{(\sqrt{\mathcal{I}}+1)^2},\frac{1}{\mathcal{I}+\pi^2}\right). \tag{5}$$

2. 增广 van Trees 2（AVT2）：一种更精确但更复杂的下界为

$$\sup_{t\in[-1,1]}\mathbb{E}_{P_t}\left[(\hat{t}(\boldsymbol{X})-t)^2\right]\geq\frac{1}{\inf_{m>0}(m+1)^2\left\{{}_2F_1\left(-\frac{1}{2},\frac{m}{2},\frac{m}{2}+1;-\frac{\mathcal{I}}{m^2}\right)\right\}^2}, \tag{6}$$

其中 ${}_2F_1$ 是超几何函数。

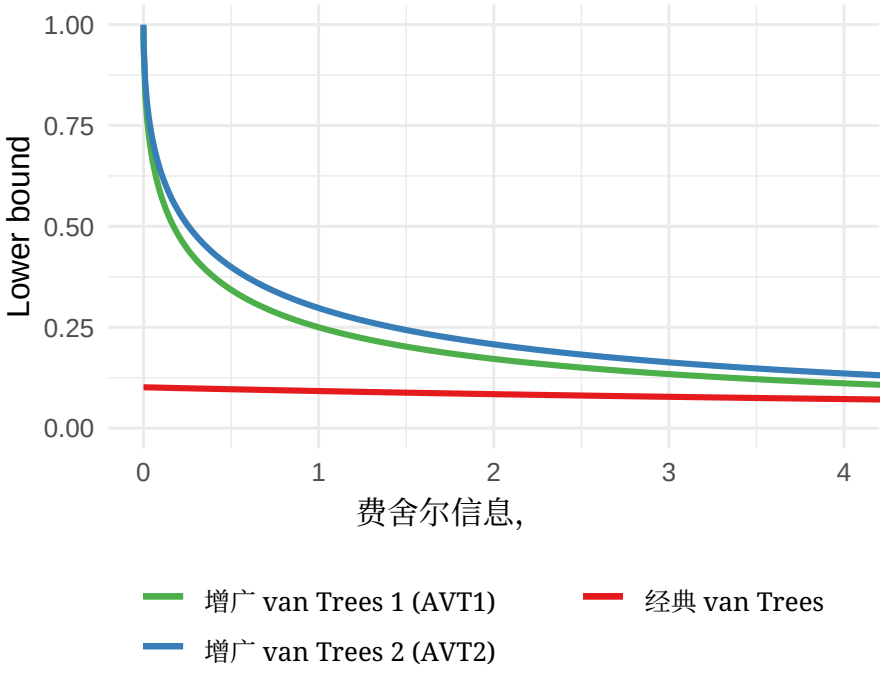


图1：van Trees 不等式（采用最优先验）给出的 $\sup_{t\in T}\mathbb{E}_{P_t}\left[(\hat{}t(\boldsymbol{X})-t)^2\right]$ 下界，以及增广 van Trees 不等式（采用近似最优先验）给出的下界（4）和（6）。

例如，参见Andrews 等人（1999）了解超几何函数的更多详情。这两个下界以及通过经典 van Trees 不等式得到的下界，均见图1。

本文余下部分安排如下。第2 节陈述并证明增广 van Trees 不等式（定理1）。第3 节将这一结果应用于 H¨older 函数估计，得到更精确的极小极大下界，并且在某些情形下给出精确常数。第4 节展示如何将我们的增广机制应用于近期提出的广义 van Trees 不等式（Takatsuand Kuchibhotla，2024年），并概述增广方案的灵活性与适用性。

2 增广 van Trees 不等式

考虑参数模型 $(P_t)_{t\in T}$ 在第1节中。下面的增广 van Trees 不等式对任意（勒贝格）先验给出贝叶斯风险平方的界，并涉及一个额外的辅助增广函数 $\alpha:\ T\to\mathbb{R}$。

定理1（增广 van Trees 不等式）。假设：

(i) $p:\ \mathcal{X}\times T\to[0,\infty)$ 是可测函数，使得 $p(\cdot,t)$ 对每个 $t\in T$ 均为勒贝格密度，函数 $t\mapsto p(x,t)$ 对 ν-几乎所有的 x 绝对连续，并且费舍尔信息

$$\mathcal{I}(t):=\int_{\mathcal{X}}\left(\frac{\partial_t p(x,t)}{p(x,t)}\right)^2 p(x,t)\,d\nu(x),$$

是有限的，且满足 $\int_T\mathcal{I}(t)\,dt<\infty$。

(ii) 先验 μ 是 T 上的勒贝格密度，并且是绝对连续函数。

(iii) 增广函数 $\alpha:\ T\to\mathbb{R}$ 是绝对连续的，且满足 $\alpha(t_1)=\alpha(t_2)=0$，并且

$$\max\left\{\int_T\alpha,\ \int_T\frac{\alpha^2}{\mu},\ \int_T\frac{(\alpha')^2}{\mu}\right\}<\infty.$$

那么，对于任意可测函数 $\hat{t}: \mathcal{X} \to \mathbb{R}$，贝叶斯风险满足

$$\int_T \mathbb{E}_{P_t}\left[(\hat{t}(\boldsymbol{X}) - t)^2\right] \mu(t)\, dt \geq \frac{\left(\int_T \alpha\right)^2}{\int_T \frac{\mathcal{I}\alpha^2 + (\alpha')^2}{\mu}}. \tag{7}$$

取 $\alpha = \mu$ 可恢复经典 van Trees 不等式（3）。引入辅助增广函数 α 后，可以得到更紧的下界，例如通过在适当的增广函数类 α上对（7）进行优化即可获得。与经典 van Trees 不等式及其推广（van Trees，1968；Gassiat and Stoltz，2024年；Takatsu and Kuchibhotla，2024年）不同，我们不要求先验密度 μ 在 T的边界处为零，而是由增广函数 α承担这一余量。事实上，这使我们能够为任意绝对连续先验密度获得贝叶斯风险平方的下界。这种额外的灵活性使我们能够将更大比例的先验质量集中在 T的边界附近，而对于给定统计量而言，这通常对应于最难区分的参数值，从而提高下界（2）的紧致性（示例见后文图2 ）。因此，将定理1 的增广 van Trees 不等式与（2）结合，并对有效先验进行优化，可得到最坏情况误差的如下下界。

定理 *2.* 考虑满足定理*1*中假设 *(i)* 的分布类 $(P_t)_{t\in T}$，并假设 $\mathcal{I}$ 在 T 上绝对连续。令 $\mathcal{A}$ 为任意一类增广函数 $\alpha: T \to \mathbb{R}$，满足：*(i)* 在 T 上绝对连续，且其弱导数为 α'，使得 $(\alpha')^2$ 在 T 上绝对连续；*(ii)* 满足 $\alpha(t_1) = \alpha(t_2) = 0$。那么

$$\sup_{t\in T} \mathbb{E}_{P_t}\left[(\hat{t}(\boldsymbol{X}) - t)^2\right] \geq \sup_{\alpha\in\mathcal{A}} \left(\frac{\int_T \alpha}{\int_T \sqrt{\mathcal{I}\alpha^2 + (\alpha')^2}}\right)^2. \tag{8}$$

定理 2 是柯西–施瓦茨不等式的一个简单推论；对于给定的 $\alpha \in \mathcal{A}$，使下界（7）达到最大的先验为

$$\mu(t) \propto \sqrt{\mathcal{I}(t)\alpha^2(t) + \alpha'(t)^2}. \tag{9}$$

因此请注意，取任何一类包含函数 $\mathcal{A}$ 的增广函数 $\alpha(t) = \text{余弦}^2(\pi t/2)$，即可保证相应的下界（8）优于经典 van Trees 不等式。尽管对式（8）中所有合适增广函数取上确界所对应的变分问题会导出相当复杂的欧拉–拉格朗日方程，但我们仍可通过选择合理的增广函数试探函数类（例如以下两种情形）得到较为理想的界。

例 3 (增广 van Trees 1)。 假设 $\mathcal{I}(\cdot) = \mathcal{I}$ 为常数，且 $T = [-1, 1]$。取函数类 $\mathcal{A} = \{\alpha(t) = \mathbb{1}_{\{|t|\leq 1-\delta\}} + \left(\frac{1-t}{\delta}\right)\mathbb{1}_{\{1-\delta<|t|\leq 1\}} : \delta \in (0, 1)\}$。则 $\int_T \alpha \geq (1-\delta)$ 和 $\int_T \sqrt{\mathcal{I}\alpha^2 + (\alpha')^2} \leq (1-\delta)\sqrt{\mathcal{I}} + \sqrt{\delta^2\mathcal{I} + 1} \leq \sqrt{\mathcal{I}} + 1$。取 $\delta \searrow 0$，我们得到

$$\sup_{t\in[-1,1]} \mathbb{E}_{P_t}\left[(\hat{t}(\boldsymbol{X}) - t)^2\right] \geq \frac{1}{(\sqrt{\mathcal{I}} + 1)^2}.$$

例4 (增广 van Trees 2)。 再次假设 $\mathcal{I}(\cdot) = \mathcal{I}$ 为常数，且 $T = [-1, 1]$。考虑增广函数类 $\mathcal{A} = \{\alpha(t) = (1 - |t|)^m : \ m > 0\}$。则（见附录 *A.4*）不等式 (8)化为

$$\sup_{t\in[-1,1]} \mathbb{E}_{P_t}\left[(\hat{t}(\boldsymbol{X}) - t)^2\right] \geq \frac{1}{\inf_{m>0}(m+1)^2\left\{{}_2F_1\left(-\frac{1}{2}, \frac{m}{2}, \frac{m}{2}+1, -\frac{\mathcal{I}}{m^2}\right)\right\}^2}.$$

上文得到的两个下界绘制在图1中。

2.1 向其他损失函数的推广

van Trees 不等式通常仅适用于平方误差风险。下面我们将（增广）van Trees 不等式推广到一般的 L_p 损失函数。

定理 5. 采用定理 *1* 和*2*中的设定。假设 $p, q > 1$ 满足 $1/p + 1/q = 1$。另假设条件得分函数 $\rho_t(x) := \frac{\partial_t p_{(x,t)}}{p(x,t)}$ 满足 $\mathbb{E}_{P_t}[|\rho_t(\boldsymbol{X})|^q]$ 有限且可积。则

$$\sup_{t\in T} \mathbb{E}_{P_t}\left[|\hat{t}(\boldsymbol{X}) - t|^p\right] \geq \sup_{\alpha\in\mathcal{A}} \left(\frac{\left|\int_T \alpha(t)\, dt\right|}{\int_T \left\{\mathbb{E}_{P_t}\left[|\alpha'(t) + \alpha(t)\rho_t(\boldsymbol{X})|^q\right]\right\}^{1/q} dt} \right)^p,$$

其中，上确界取遍所有增广函数 $\mathcal{A}$，如定理*2*所示。

取2 即可得到定理 $p = q = 2$，其中

$$\mathbb{E}_{P_t}\left[\{\alpha'(t) + \alpha(t)\rho_t(\boldsymbol{X})\}^2\right] = \alpha'(t)^2 + \alpha^2(t)\mathbb{E}_{P_t}[\rho_t^2(\boldsymbol{X})] = \alpha'(t)^2 + \alpha^2(t)\mathcal{I}(t).$$

3 极小极大函数估计的应用

我们将增广 van Trees 不等式应用于点态估计 H¨older 类中函数的问题，该类的指数为 $\beta \in (0, 2]$，常数为 $L > 0$,

$$\mathcal{H}(\beta, L) := \begin{cases} \left\{f : \mathbb{R}^d \to \mathbb{R},\ |f(x) - f(x_0)| \leq L\|x - x_0\|_2^\beta\right\} & \text{if } \beta \in (0, 1], \\ \left\{f : \mathbb{R}^d \to \mathbb{R} \text{ once differentiable},\ \|\nabla f(x) - \nabla f(x_0)\|_2 \leq L\|x - x_0\|_2^{\beta-1}\right\} & \text{if } \beta \in (1, 2], \end{cases}$$

对于一般维数 $d \in \mathbb{N}$。假设我们可以获得 n 个独立同分布对 $(X_1, Y_1) \ldots, (X_n, Y_n) \in \mathbb{R}^d \times \mathbb{R}$，其中

$$Y_i = f(X_i) + \varepsilon_i, \tag{10}$$

其中 $\varepsilon_i \,|\, X_i \overset{\text{独立同分布}}{\sim} N(0, \sigma^2)$，回归函数为 $f \in \mathcal{H}(\beta, L)$，且 $X_i \sim P_X$ 其中 P_X 关于勒贝格测度绝对连续，其 Radon-Nikodym 导数也绝对连续。我们将构造 $(\mathcal{H}(\beta, L), \delta)$ 上极小极大平方风险的下界，其中半范数为$\delta(f_1, f_2) = |f_1(x_0) - f_2(x_0)|$，对某个 $x_0 \in \mathbb{R}^d$，即

$$\inf_{\hat{f}} \sup_{f\in\mathcal{H}(\beta,L)} \mathbb{E}_f\left[\left(\hat{f}(x_0) - f(x_0)\right)^2\right],$$

其中下确界取遍数据的所有 Borel 可测函数。增广 van Trees 不等式（AVT2 形式（6））可用于得到关于极小极大风险的如下近乎尖锐的刻画。

定理 6（逐点 H¨older 函数估计）。考虑模型（10）。对任意 $\beta \in (0, 2], d \in \mathbb{N}, L > 0, \sigma^2 > 0, x_0 \in \mathbb{R}^d$,

$$\frac{1}{1.69} \leq \liminf_{n\to\infty} \frac{\inf_{\hat{f}} \sup_{f\in\mathcal{H}(\beta,L)} \mathbb{E}_f\left[\left(\hat{f}(x_0) - f(x_0)\right)^2\right]}{\left(\frac{d^d(\beta+d)^{2\beta}\Gamma^{2\beta}(1+d/2)}{\pi^{\beta d}\beta^{2\beta}(2\beta+d)^d(1\vee\beta)^{2d}}\right)^{1/(2\beta+d)} \left(\frac{L^{d/\beta}\sigma^2}{p_X(x_0)n}\right)^{2\beta/(2\beta+d)}}$$
$$\leq \limsup_{n\to\infty} \frac{\inf_{\hat{f}} \sup_{f\in\mathcal{H}(\beta,L)} \mathbb{E}_f\left[\left(\hat{f}(x_0) - f(x_0)\right)^2\right]}{\left(\frac{d^d(\beta+d)^{2\beta}\Gamma^{2\beta}(1+d/2)}{\pi^{\beta d}\beta^{2\beta}(2\beta+d)^d(1\vee\beta)^{2d}}\right)^{1/(2\beta+d)} \left(\frac{L^{d/\beta}\sigma^2}{p_X(x_0)n}\right)^{2\beta/(2\beta+d)}} \leq 1.$$

固定设计、有限样本以及局部渐近极小极大界同样可借助这一方法得到。普适常数 1.69（定理 6 ）使上述结果对每一个$\beta \in (0, 2], d \in \mathbb{N}$ 都成立，但在具体情形中可以改进。例如当 $(\beta, d) = (2, 1)$ 时，对应于具有 Lipschitz 导数的一元可微回归函数，我们得到

$$\frac{1}{1.37} \leq \liminf_{n\to\infty} \frac{\inf_{\hat{f}} \sup_{f\in\mathcal{H}(2,L)} \mathbb{E}_f\left[\left(\hat{f}(x_0) - f(x_0)\right)^2\right]}{\frac{3^{4/5}}{2^2\cdot 5^{1/5}} \cdot \left(\frac{L^{1/2}\sigma^2}{p_X(x_0)n}\right)^{4/5}} \leq \limsup_{n\to\infty} \frac{\inf_{\hat{f}} \sup_{f\in\mathcal{H}(2,L)} \mathbb{E}_f\left[\left(\hat{f}(x_0) - f(x_0)\right)^2\right]}{\frac{3^{4/5}}{2^2\cdot 5^{1/5}} \cdot \left(\frac{L^{1/2}\sigma^2}{p_X(x_0)n}\right)^{4/5}} \leq 1.$$

为了大致理解增广 van Trees 不等式为何优于经典 van Trees，不妨注意到，用贝叶斯风险界定最坏情形风险的紧致程度，取决于先验质量有多少

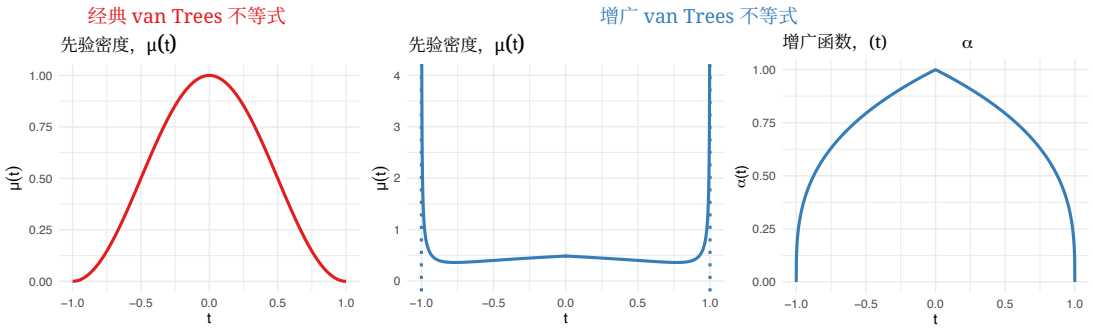


图2：经由经典 van Trees 不等式与增广 van Trees 不等式得到的“最优”先验密度，设定为极小极大 H¨older 函数估计，见第3 节，其中$(\beta, d) = (2,1)$。对于经典 van Trees 不等式，最优先验为 $\mu(t) = \cos^2(\pi t/2)$，且无增广。对于增广 van Trees 不等式（AVT2），我们通过取增广函数得到一个下界：$\alpha(t) = (1 - |t|)^m$，其中 $m > 0$，以及相应的最优先验(9)。

集中在子路径上最难区分的点附近（P_t）$t\in[-1,1]$；通常是两个极端点 ± 1。增广 van Trees 不等式允许 $\mu(\pm 1)$ 任意取值，因此允许更多先验质量集中在 ± 1 的邻域中；相比之下，经典 van Trees 不等式要求 $\mu(\pm 1) = 0$；示例见图 2 。

此外，在高维情形 $d \to \infty$ 下，对于任意 $\beta \in (0, 2]$, $L > 0$, $\sigma > 0$ 以及（光滑的）协变量设计 p_X，我们可以得到逐点 H¨older 函数估计的精确渐近极小极大风险。下文中，我们用 $\mathcal{B}_0(1)$ 表示以 0 为中心、半径为 1 的欧几里得球。

定理 7 （高维情形下的精确风险）。 假设 X 关于勒贝格测度绝对连续，其 *Radon-Nikodym* 导数为 p_X。此外，设 $p_X(\cdot)/p_X(x_0) \in \mathcal{H}(\alpha, L_X)$，其中某个 $\alpha \in (0, 1]$, $L_X > 0$。那么，对于任意满足 $\frac{1}{C} \le \frac{p_X(x_0)1/\mathrm{vol}\mathcal{B}_1(0)}{} \le C$ 的 $x_0 \in \mathbb{R}^d$，存在一个仅依赖于 x_0 的常数 $C \ge 1$；并且对于任意 $\beta \in (0, 2]$, $\sigma^2 > 0$,

$$\lim_{\substack{n,d\to\infty \\ (\log n)/d\to\infty}} \frac{\inf_{\hat{f}} \sup_{f\in\mathcal{H}(\beta,L)} \mathbb{E}_f\left[\left(\hat{f}(x_0) - f(x_0)\right)^2\right]}{\frac{1}{(2\pi e)^\beta (1\vee\beta)^2} d^\beta \left(\frac{\sigma^2 L^{d/\beta}}{p_X(x_0) n}\right)^{2\beta/(2\beta+d)}} = 1 \tag{11}$$

协变量密度的附加条件例如在如下情形中成立： X ∼均匀分布 $\mathcal{B}_0(1)$ 以及 $x_0 \in \mathrm{int}\ \mathcal{B}_0(1)$，常数为 $(C, \alpha, L_X) = (1, 1, 1/(1 - \|x_0\|_2))$。注意，上述结果并不能通过应用 van Trees 不等式得到，该不等式对 (11) 只能给出 π^2 的上界。

增广 van Trees 不等式的另一个优点是其应用简便。类似的、能够给出显式常数的极小极大下界往往需要对特定模型进行相当繁琐的处理；例如参见 Donoho 和 Liu （1991年）；Fan （1993年）中特定模型的显式常数。然而与之相反，增广 van Trees 不等式能够给出同样尖锐的常数，同时实现起来却极为简单；只要能够量化参数子模型的（渐近）费舍尔信息，相应的下界便立即得出。因此，这使我们能够在此考虑如下宽泛设定，即 H¨older 函数估计，且光滑性可为任意 $\beta \in (0, 2]$ 以及任意维数 $d \in \mathbb{N}$。因此，这些结果也可以自然地加以推广，从而为具有有限费舍尔信息且密度足够光滑的任意误差分布提供下界。

4 增强广义 van Trees 不等式

在近期的研究中，Takatsu andKuchibhotla (2024年) 提出了广义 van Trees 不等式。该不等式通过纳入一个额外的绝对连续近似函数 ϕ: $T \to \mathbb{R}$对经典 van Trees 界进行了扩展。这一扩展能够为不可微泛函以及非正则统计模型中的估计问题推导极小极大下界和局部渐近极小极大下界。我们灵活的增广策略同样可以纳入这一广义框架，从而在这些非正则情形下再次得到更尖锐的常数，如下述定理所示。

定理8（增广广义 *van Trees* 不等式）。设 $T \subset \mathbb{R}^d$ 是一个闭子集，且 $(P_t)_{t\in T}$是一个参数模型，其中对所有 $t \in T$， $P_t \ll \nu$都成立；这里， ν是 $(\mathcal{X}, \mathcal{A})$ 上的一个 σ-有限测度。再次记 $p(\cdot, t)$ 为 P_t 关于 ν 的密度，它关于 t 绝对连续，且对ν-几乎所有 x，其费舍尔信息 $\mathcal{I}(t) := \int_{\mathcal{X}} \frac{\partial_t p(x,t)}{p(x,t)} \frac{\partial_t p(x,t)}{p(x,t)}^\top d\nu(x)$ 在 T 上可积，且算子范数为 $\|\mathcal{I}(t)\|_{\mathrm{op}} < \infty$。进一步假设：

(i)（先验密度）先验密度 μ 在 T上绝对连续。

(ii)（逼近函数）逼近函数 $\phi: T \to \mathbb{R}^k$ 绝对连续，且其几乎处处导数为 $\nabla\phi: T \to \mathbb{R}^{k\times d}$。

(iii)（增广函数）增广函数 $\alpha: T \to \mathbb{R}$ 绝对连续，并且对于 T边界上的所有 $t \in \partial T$，均有$\alpha(t) = 0$ 。

那么，对于任意绝对连续函数 $\psi: T \to \mathbb{R}^k$，其几乎处处导数为 $\nabla\psi: T \to \mathbb{R}^{k\times d}$；对于任意可测函数 $\hat{\psi}: \mathcal{X} \to \mathbb{R}^k$，以及任意向量范数 $\|\cdot\|: \mathbb{R}^k \to \mathbb{R}_+$及其对偶范数$\|\cdot\|_*$，

$$\sup_{t\in T} \mathbb{E}_{P_t}\left[\left\|\hat{\psi}(\mathbf{X}) - \psi(t)\right\|^2\right] \geq \sup_{\mu,\alpha}\left(\sup_{\|u\|_*\leq 1} \left\|\Gamma^{1/2}_{\mu,\phi,\alpha} u\right\| - \left(\int_T \|\psi(t) - \phi(t)\|^2 \mu(t)\, dt\right)^{1/2}\right)_+^2, \tag{12}$$

其中 $(\cdot)_+ := \max(\cdot, 0)$，并且

$$\Gamma_{\mu,\phi,g} := \left(\int_T \nabla\phi(t)\alpha(t)\, dt\right)^\top \left(\int_T \frac{\mathcal{I}(t)\alpha^2(t) + \nabla\alpha(t)\nabla\alpha(t)^\top}{\mu(t)}\right)^{-1} \left(\int_T \nabla\phi(t)\alpha(t)\, dt\right),$$

并且 上确界是在满足以下条件的所有先验上取的 μ ： *(i)* 以及所有满足条件的增广函数 α 足 *s-*足 *(iii)*。

与定理 1 中一样，我们同样不要求先验密度满足所谓“良好”的性质，即不要求先验密度按照 Gassiat andStoltz (2024年)；Takatsu and Kuchibhotla (2024年) 的做法在 T 的边界处趋于零。在第 3 节等正则估计情形下，最优逼近函数仍为 $\phi = \psi$，从而得到定理 1 中标准的增广 van Trees 不等式（讨论见附录A.9）。

5 讨论

我们提出了 van Trees 不等式的一个实用扩展，它适用于密度在其支撑集边界处不趋于零的先验，并且对于任何固定先验都能给出更紧的下界。因此，这个增广 van Trees 不等式能够为最坏情况风险提供更紧的界。我们将该不等式应用于极小极大 H¨older 函数估计问题，得到该问题中接近尖锐的常数，以及高维情形下的精确常数。我们还将（增广）van Trees 不等式扩展到均方误差以外的损失函数，以及非正则模型中的估计问题。

增广 van Trees 不等式提供了一种极其简单且易于理解的方法，用于推导非参数极小极大界；与更复杂的实验收敛理论相比，它往往能达到具有竞争力（甚至更优）的常数。在构造均方误差的极小极大下界时，问题便简化为量化模型类中一条路径的费舍尔信息。因此，我们认为，增广 van Trees 不等式是统计学家工具箱中一项非常有用的工具，可在多种情形下获得简单而有力的极小极大下界。

Funding。 EHY 由欧洲研究理事会高级资助 101019498 支持 .

参考文献

Andrews, G. E., Askey, R., and Roy, R. (1999). *Special Functions*. Cambridge University Press. Assouad, P. (1983). 关于估计的两点意见。科学院院报´（法兰西科学院）´，科学。´系列 *1*，数学´。, 296(23):1021–1024. Donoho, D. L. and Liu, R. C. (1991). 收敛速率的几何化，III。《统计年鉴》, 19:668–701. Fan, J. (1993). 局部线性回归及其极小极大效率。《统计年鉴》, 21:196–216. Gassiat, E. and Stoltz, G. (2024). Hajek 和 Le Cam 思想下的 van Trees 不等式。《统计科学》, 39(4):644 – 653. Gill, R. D. and Levit, B. Y. (1995). van Trees 不等式的应用：贝叶斯 Cramér-Rao 界。《伯努利》, 1(1-2):59 – 79. Hajek, J. (1970). 正则估计量极限分布的刻画。概率论与相关领域杂志¨, 14(4):323–330. Le Cam, L. (1972). 实验的极限。载于第六届伯克利数学统计与概率研讨会论文集，第 1 卷，第 245–261 页。Takatsu, K. and Kuchibhotla, A. K. (2024). 广义 van Trees 不等式：非光滑泛函和非正则统计模型的局部极小极大界。*arXiv* 预印本 *2405.06437*。

Tsybakov, A. B.（2008）。《非参数估计导论》。施普林格纽约。

van der Vaart, A. W.（1998年）。《渐近统计学》。剑桥大学出版社。

van Trees, H.L.（1968）。《信号检测》—《信号参数估计》，第4章，第239–422页。约翰·威立父子出版公司有限公司。

A 结果证明

A.1 定理1的证明

请注意，定理1 是定理8的推论。然而，由于定理1 这一情形较为简单，我们仍在此给出这个受限情形的简要证明，这已足以支持第3节中的所有结果。

定理的证明 *1*。 定义

$$\zeta(x,t) := \frac{\partial_t\big(p(x,t)\alpha(t)\big)}{p(x,t)\mu(t)},$$

其中 $\partial_t(g(x,t))$ 表示 g 关于 t 的导数： $\mathcal{X}\times T \to \mathbb{R}$。然后由柯西–施瓦茨不等式，

$$\int_T \mathbb{E}_{P_t}\big[(\hat{t}(\boldsymbol{X})-t)^2\big]\mu(t)\,dt \geq \left(\int_T\int_{\mathcal{X}} p(x,t)\mu(t)\zeta^2(x,t)\,d\nu(x)\,dt\right)^{-1}\left(\int_T\int_{\mathcal{X}}(\hat{t}(x)-t)p(x,t)\mu(t)\zeta(x,t)\,d\nu(x)\,dt\right)^2. \quad (13)$$

我们将依次化简（13）的分子和分母。首先注意，由于 $p(x,t)$, $\mu(t)$ 和 $\alpha(t)$ 关于 $\alpha(t_1) = \alpha(t_2) = 0$ 绝对连续，

$$\int_T p(x,t)\mu(t)\zeta(x,t)\,dt = \int_T \partial_t\big(p(x,t)\alpha(t)\big)\,dt = 0, \quad (14)$$

对 ν-几乎所有 x。因此

$$\int_T \int_{\mathcal{X}} \hat{t}(x) p(x,t) \mu(t) \zeta(x,t)\, d\nu(x)\, dt = 0. \tag{15}$$

通过类似的论证（也利用绝对连续性）以及分部积分，

$$\int_T t\, \partial_t \big(p(x,t)\alpha(t)\big)\, dt = -\int_T p(x,t)\alpha(t)\, dt,$$

thus

$$\int_T \int_{\mathcal{X}} t p(x,t) \mu(t) \zeta(x,t)\, d\nu(x)\, dt = -\int_T \int_{\mathcal{X}} p(x,t)\, d\nu(x) \alpha(t)\, dt = -\int_T \alpha(t)\, dt. \tag{16}$$

(13) 的分子项因此可简化为

$$\left(\int_T \int_{\mathcal{X}} (\hat{t}(x) - t) p(x,t) \mu(t) \zeta(x,t)\, d\nu(x)\, dt\right)^2 = \left(\int_T \alpha(t)\, dt\right)^2 \tag{17}$$

应用 (15) 和 (16) 得到。

(13) 的分母项为，

$$\begin{aligned}
&\int_T \int_{\mathcal{X}} p(x,t) \mu(t) \zeta^2(x,t)\, d\nu(x)\, dt \\
&= \int_T \int_{\mathcal{X}} \frac{\big(\partial_t p(x,t)\alpha(t) + p(x,t)\alpha'(t)\big)^2}{p(x,t)\mu(t)}\, d\nu(x)\, dt \\
&= \int_T \int_{\mathcal{X}} \frac{(\partial_t p(x,t))^2}{p(x,t)} \frac{\alpha^2(t)}{\mu(t)}\, d\nu(x)\, dt + 2\int_T \int_{\mathcal{X}} \partial_t p(x,t)\, d\nu(x) \frac{\alpha(t)\alpha'(t)}{\mu(t)}\, dt + \int_T \int_{\mathcal{X}} p(x,t) \frac{\big(\alpha'(t)\big)^2}{\mu(t)}\, d\nu(x)\, dt \\
&= \int_T \mathcal{I}(t) \frac{\alpha^2(t)}{\mu(t)}\, dt + \int_T \frac{\alpha'(t)^2}{\mu(t)}\, dt,
\end{aligned} \tag{18}$$

因为交叉项为零；首先注意，由詹森不等式可得

$$\int_{\mathcal{X}} \big|\partial_t p(x,t)\big|\, d\nu(x) \le \sqrt{\mathcal{I}(t)} < \infty,$$

对所有 $t \in T$ 均成立，因此我们可以应用控制收敛定理，得到

$$\int_{\mathcal{X}} \partial_t p(x,t)\, d\nu(x) = \frac{\partial}{\partial t} \int_{\mathcal{X}} p(x,t)\, d\nu(x) = 0.$$

结合（13）、（17）和（18），我们得到所需结果；

$$\int_T \mathbb{E}_{P_t}\big[(\hat{t}(\boldsymbol{X}) - t)^2\big] \mu(t)\, dt \ge \frac{\big(\int_T \alpha(t)\, dt\big)^2}{\int_T \frac{\mathcal{I}(t)\alpha^2(t) + \alpha'(t)^2}{\mu(t)}\, dt}.$$

□

A.2 定理2的证明

定理2的证明。 固定某个 $\alpha \in \mathcal{A}$。根据柯西–施瓦茨不等式，

$$\left(\int_T \frac{\mathcal{I}\alpha^2 + (\alpha')^2}{\mu}\right) = \left(\int_T \mu\right)\left(\int_T \frac{\mathcal{I}\alpha^2 + (\alpha')^2}{\mu}\right) \ge \left(\int_T \sqrt{\mathcal{I}\alpha^2 + (\alpha')^2}\right)^2,$$

当且仅当等号成立时

$$\mu \propto \sqrt{\mathcal{I}\alpha^2 + (\alpha')^2}$$

由于 μ 是密度，因此最优先验 μ_α^* 是

$$\mu_\alpha^* := \frac{\sqrt{\mathcal{I}\alpha^2 + (\alpha')^2}}{\int_T \sqrt{\mathcal{I}\alpha^2 + (\alpha')^2}}, \tag{19}$$

根据假设， $\mathcal{I}$、 α和α' 均绝对连续，因此它也绝对连续。由于 α 可微且其导数绝对连续，max $\{\int_T |\alpha|, \int_T |\alpha'|\} < \infty$。应用定理 1 并取$\mu = \mu^*_\alpha$,

$$\sup_{\mu} \frac{\left(\int_T \alpha\right)^2}{\int_T \frac{\mathcal{I}\alpha^2+(\alpha')^2}{\mu}} = \frac{\left(\int_T \alpha\right)^2}{\int_T \frac{\mathcal{I}\alpha^2+(\alpha')^2}{\mu^*_\alpha}} = \left(\frac{\int_T \alpha}{\int_T \sqrt{\mathcal{I}\alpha^2+(\alpha')^2}}\right)^2.$$

对所有 $\alpha \in \mathcal{A}$ 取上确界，即得所需结果。 □

A.3 示例 3结果的证明

示例 3的证明。 取 $\delta \in (0,1)$。考虑增广函数 $\alpha\colon [-1,1] \to \mathbb{R}$ 定义为

$$\alpha(t) := \begin{cases} e^{-k|t|} & \text{if } |t| \le 1-\delta, \\ \frac{1}{\delta} e^{-(1-\delta)k}(1-|t|) & \text{if } 1-\delta < |t| \le 1. \end{cases}$$

根据构 造， α 是绝对连续的，且 $\alpha(1) = \alpha(-1) = 0$。另定义 $f\colon [0,1] \to \mathbb{R}$ ，其定义为

$$f(t) := \frac{\alpha(t)}{-\alpha'(t)} = \begin{cases} \frac{1}{k} & \text{if } t \in [0, 1-\delta], \\ \frac{1-t}{\delta} & \text{if } t \in (1-\delta, 1]. \end{cases}$$

由于 α 限制在定义域 $[0,1]$ 上且 f 单调，我们可以写成

$$f(\alpha(t)) = \begin{cases} \delta e^{(1-\delta)k}\alpha(t) & \text{if } \alpha(t) \in \left[0, e^{-(1-\delta)k}\right], \\ \frac{1}{k} & \text{if } \alpha(t) \in \left(e^{-(1-\delta)k}, 1\right]. \end{cases}$$

Then

$$\begin{aligned} \frac{\int_{-1}^1 \alpha(t)\,dt}{\int_{-1}^1 \sqrt{\mathcal{I}\alpha^2(t)+\alpha'(t)^2}\,dt} = \frac{\int_0^1 \alpha(t)\,dt}{\int_0^1 \sqrt{\mathcal{I}\alpha^2(t)+\alpha'(t)^2}\,dt} &= \frac{\int_0^1 f(\alpha)\,d\alpha}{\int_0^1 \sqrt{\mathcal{I}(f(\alpha))^2+1}\,d\alpha} \\ &= \frac{\frac{\delta}{2}e^{-(1-\delta)k} + \frac{1}{k}\left(1-e^{-(1-\delta)k}\right)}{\sqrt{\frac{\mathcal{I}}{k^2}+1}\left(1-e^{-(1-\delta)k}\right) + e^{-(1-\delta)k}\int_0^1 \sqrt{\delta^2\mathcal{I}u^2+1}\,du}. \end{aligned}$$

取 $\delta \searrow 0$ ，我们得到

$$\begin{aligned} \sup_{\alpha} \frac{\int_{-1}^1 \alpha(t)\,dt}{\int_{-1}^1 \sqrt{\mathcal{I}\alpha^2(t)+\alpha'(t)^2}\,dt} &\ge \sup_{k>0} \lim_{\delta \searrow 0} \frac{\int_{-1}^1 \alpha_\delta(t)\,dt}{\int_{-1}^1 \sqrt{\mathcal{I}\alpha_\delta^2(t)+\alpha_\delta'(t)^2}\,dt} \\ &= \sup_{k>0} \frac{1}{\sqrt{\mathcal{I}+k^2}+\frac{k}{e^k-1}} \ge \lim_{k \searrow 0} \frac{1}{\sqrt{\mathcal{I}+k^2}+\frac{k}{e^k-1}} = \frac{1}{\sqrt{\mathcal{I}}+1}. \end{aligned}$$

□

A.4 示例 4 结论的证明

示例 4 的结论证明。首先注意 $\alpha(1) = \alpha(-1) = 0$并且 α是绝对连续的。此外，$\int_{-1}^1 |\alpha| = 2\int_0^1 (1-t)^m\,dt = \frac{2}{m+1} \le 2$并且 $\int_{-1}^1 |\alpha'| = -2\int_0^1 \alpha' = 2(\alpha(0)-\alpha(1)) = 2$。类 $\mathcal{A}$因此满足定理 2 的假设。然后直接计算可得 $\int_{-1}^1 \alpha = \frac{2}{m+1}$且

$$\int_{-1}^1 \sqrt{\mathcal{I}\alpha^2+(\alpha')^2} = 2\int_0^1 \sqrt{\mathcal{I}t^{2m}+m^2t^{2(m-1)}}\,dt = 2\int_0^1 t^{m-1}\sqrt{\mathcal{I}t^2+m^2}\,dt = 2{}_2F_1\left(\tfrac{-1}{2}, \tfrac{m}{2}, \tfrac{m}{2}+1; \tfrac{-\mathcal{I}}{m^2}\right),$$

结合定理2即得该结果。 □

A.5 定理5的证明

定理5的证明。我们采用与定理1的证明相同的记号。然后由H¨older不等式

$$\left\{\int_T \mathbb{E}_{P_t}\left[|\hat{t}(\boldsymbol{X})-t|^p\right]\mu(t)\,dt\right\}^{1/p}$$
$$\geq \left\{\int_T\int_{\mathcal{X}} p(x,t)\mu(t)|\zeta(x,t)|^q\,d\nu(x)\,dt\right\}^{-1/q}\left|\int_T\int_{\mathcal{X}}(\hat{t}(x)-t)p(x,t)\mu(t)\zeta(x,t)\,d\nu(x)\,dt\right|.$$

Then

$$|\zeta(x,t)|^q=\frac{1}{\mu(t)^q}\left|\frac{\partial_t p(x,t)}{p(x,t)}\alpha(t)+\alpha'(t)\right|^q,$$

因此

$$\int_T\int_{\mathcal{X}} p(x,t)\mu(t)|\zeta(x,t)|^q\,d\nu(x)\,dt=\int_T \mu(t)^{1-q}\,\mathbb{E}_{P_t}\left[|\alpha'(t)+\rho_t(\boldsymbol{X})\alpha(t)|^q\right]dt.$$

此外，取

$$\mu(t):=\frac{\left\{\mathbb{E}_{P_t}\left[|\alpha'(t)+\rho_t(\boldsymbol{X})\alpha(t)|^q\right]\right\}^{1/q}}{\int_T\left\{\mathbb{E}_{P_\tau}\left[|\alpha'(\tau)+\rho_\tau(\boldsymbol{X})\alpha(\tau)|^q\right]\right\}^{1/q}d\tau}.$$

结合方程（17），我们得出

$$\sup_{t\in T}\left\{\mathbb{E}_{P_t}\left[|\hat{t}(\boldsymbol{X})-t|^p\right]\right\}^{1/p}\geq\frac{\left|\int_T\alpha(t)\,dt\right|}{\int_T\left\{\mathbb{E}_{P_t}\left[|\alpha'(t)+\rho_t(\boldsymbol{X})\alpha(t)|^q\right]\right\}^{1/q}dt}$$

□

A.6 定理6的证明

我们将定理6 的证明分别分为下界和上界两部分。

引理 9. 与定理 6的设定相同，

$$\liminf_{n\to\infty}\frac{\inf_{\hat{f}}\sup_{f\in\mathcal{H}(\beta,L)}\mathbb{E}_f\left[\left(\hat{f}(x_0)-f(x_0)\right)^2\right]}{\left(\frac{d^d(\beta+d)^{2\beta}\Gamma^{2\beta}(1+d/2)}{\pi^{\beta d}\beta^{2\beta}(2\beta+d)^d(1\vee\beta)^{2d}}\right)^{1/(2\beta+d)}\left(\frac{L^{d/\beta}\sigma^2}{p_X(x_0)n}\right)^{2\beta/(2\beta+d)}}\geq A_{\beta,d},$$

对于函数

$$A_{\beta,d}:=A\Big(\frac{2\beta}{2\beta+d}\Big),\qquad A(a):=\frac{1}{a^a(1-a)^{1-a}}\sup_{\substack{\lambda>0,\\ m>0}}\frac{\lambda^a}{\left\{(m+1)_2F_1\left(-\frac{1}{2},\frac{m}{2},\frac{m}{2}+1;-\frac{\lambda}{m}\right)\right\}^2}.$$

此外，

$$\inf_{\beta\in(0,2]}\inf_{d\in\mathbb{N}}A_{\beta,d}\geq\frac{1}{1.69},\quad \textit{and}\quad A_{2,1}\geq\frac{1}{1.37},$$

引理 9 的证明。 考虑由$P_t)_{t\in[-1,1]}$ 给出的分布族（

$$Y_i\,|\,X_i\overset{\text{iid}}{\sim}N\big(f_t(X_i),\,\sigma^2\big),\qquad f_t(x):=\frac{tLh_n^\beta}{1\vee\beta}K\Big(\frac{x-x_0}{h_n}\Big),$$

$$K(u)=\left(1-\|u\|_2^\beta\right)_+,\qquad R(K):=\int_{\mathbb{R}}K^2(u)\,du,\qquad h_n:=\left(\frac{(1\vee\beta)^2\sigma^2\lambda}{L^2p_X(x_0)R(K)n}\right)^{1/(2\beta+d)},$$

其中 $(\cdot)_+:=\max(\cdot,\,0)$。对于 $\beta\in(0,1]$,

$$\frac{|f_t(x)-f_t(x_0)|}{\|x-x_0\|_2^\beta}=\frac{tLh_n^\beta\left|K\left(\frac{x-x_0}{h_n}\right)-K(0)\right|}{\|x-x_0\|_2^\beta}\leq L,$$

并且对于 $\beta\in(1,2]$,

$$\frac{\|\nabla f_t(x)-\nabla f_t(x_0)\|_2}{\|x-x_0\|_2^{\beta-1}}=\frac{tLh_n^{\beta-1}\left\|\nabla K\left(\frac{x-x_0}{h_n}\right)-\nabla K(0)\right\|_2}{\beta\|x-x_0\|_2^{\beta-1}}\leq L.$$

因此 $f_t \in \mathcal{H}(\beta,\ L)$。费舍尔信息为

$$\mathcal{I}(t) = \frac{L^2 n h_n^{2\beta+d}}{(1\vee\beta)^2\sigma^2}\int K^2(u)p_X(x_0+h_n u)\,du \to \frac{L^2 n h_n^{2\beta+d} R(K) p_X(x_0)}{(1\vee\beta)^2\sigma^2} = \lambda,$$

当 $n\to\infty$。对于任意Borel可测估计量$\hat{}\ f_n(x_0)$，定义$\hat{}\ t_n := \frac{1\vee\beta}{Lh_n^\beta}\hat{}\ f_n(x_0)$。然后由增广 van Trees 2不等式（6）（见例4）

$$\begin{aligned}
&\liminf_{n\to\infty}\sup_{f\in\mathcal{H}(\beta,L)} n^{2\beta/(2\beta+d)}\mathbb{E}_f\big[(\hat{f}_n(x_0)-f(x_0))^2\big]\\
&\geq \liminf_{n\to\infty}\sup_{t\in[-1,1]} n^{2\beta/(2\beta+d)}\mathbb{E}_{P_t}\big[(\hat{f}_n(x_0)-f_t(x_0))^2\big]\\
&\geq \liminf_{n\to\infty}\frac{L^2h_n^{2\beta}n^{2\beta/(2\beta+d)}}{(1\vee\beta)^2}\sup_{t\in[-1,1]}\mathbb{E}_{P_t}\big[(\hat{t}_n-t)^2\big]\\
&\geq \left(\frac{L^{d/\beta}\sigma^2}{(1\vee\beta)^{d/\beta}p_X(x_0)R(K)}\right)^{2\beta/(2\beta+d)}\sup_{\substack{\lambda>0,\\ m>0}}\left(\frac{\lambda^{\beta/(2\beta+d)}}{(m+1)_2F_1\big(-\frac12,\frac{m}{2},\frac{m}{2}+1;-\frac{\lambda}{m}\big)}\right)^2,
\end{aligned}$$

当 $n\to\infty$。现在，多元微积分计算给出

$$R(K) = \frac{2\pi^{d/2}\beta^2}{(\beta+d)(2\beta+d)\Gamma(1+\frac{d}{2})}. \tag{20}$$

因此

$$\begin{aligned}
&\liminf_{n\to\infty}\sup_{f\in\mathcal{H}(\beta,L)} n^{2\beta/(2\beta+d)}\mathbb{E}_f\big[(\hat{f}_n(x_0)-f(x_0))^2\big]\\
&\qquad\geq \left(\frac{d^d(\beta+d)^{2\beta}\Gamma^{2\beta}(1+d/2)}{\pi^{\beta d}\beta^{2\beta}(2\beta+d)^d(1\vee\beta)^{2d}}\right)^{1/(2\beta+d)}\left(\frac{L^{d/\beta}\sigma^2}{p_X(x_0)}\right)^{2\beta/(2\beta+1)}A\left(\frac{2\beta}{2\beta+d}\right).
\end{aligned}$$

数值计算得到 $\inf_{a\in[0,4/5]}A(a)\geq\frac{1}{1.69}$，其中特例 $(\beta,d)=(2,1)$ 通过计算 $A(4/5)\geq\frac{1}{1.37}$ 得到。

□

引理10。 如同定理6的设定，

$$\limsup_{n\to\infty}\frac{\inf_{\hat{f}}\sup_{f\in\mathcal{H}(\beta,L)}\mathbb{E}_f\Big[\big(\hat{f}(x_0)-f(x_0)\big)^2\Big]}{\Big(\frac{d^d(\beta+d)^{2\beta}\Gamma^{2\beta}(1+d/2)}{\pi^{\beta d}\beta^{2\beta}(2\beta+d)^d(1\vee\beta)^{2d}}\Big)^{1/(2\beta+d)}\Big(\frac{L^{d/\beta}\sigma^2}{p_X(x_0)n}\Big)^{2\beta/(2\beta+d)}}\leq 1.$$

证明。 我们考虑局部常数（Nadaraya–Watson）估计量

$$\hat{f}_{n,h}(x_0) := \frac{\frac1n\sum_{i=1}^n K_h(X_i-x_0)Y_i}{\frac1n\sum_{i=1}^n K_h(X_i-x_0)+n^{-r}},$$

其中带宽为核函数

$$h := \left(\frac{d(1\vee\beta)^2\sigma^2(x_0)R(K)}{2\beta L^2\mu_\beta^2(K)p_X(x_0)n}\right)^{1/(2\beta+d)},\qquad K(u) := \frac{(\beta+d)\Gamma(1+d/2)}{\pi^{d/2}\beta}\big(1-\|u\|_2^\beta\big)_+, \tag{21}$$

其中 $R(K) := \int_{\mathbb{R}}K^2(u)\,du$、$\mu_\beta(K) := \int_{\mathbb{R}}K(u)\|u\|_2^\beta\,du$、$K_h(\cdot) := K(\cdot/h)/h^d$，且 $r>0$ 是任意大的正数。注意，分母中额外的 n^{-r} 项确保 Nadaraya-Watson 估计量在事件 $\frac1n\sum_{i=1}^n K_h(X_i-x_0)=0$ 上有定义。

我们接着证明

$$\lim_{n\to\infty}\sup_{f\in\mathcal{H}(\beta,L)}\mathbb{E}_f\Big[\big(\hat{f}_{n,h}(x_0)-f(x_0)\big)^2\Big]\leq\left(\frac{d^d(\beta+d)^{2\beta}\Gamma^{2\beta}(1+d/2)}{\pi^{\beta d}\beta^{2\beta}(2\beta+d)^d(1\vee\beta)^{2d}}\right)^{1/(2\beta+d)}\left(\frac{L^{d/\beta}\sigma^2}{p_X(x_0)n}\right)^{2\beta/(2\beta+d)}.$$

定义 $e(x) := f(x) - f(x_0)$ 以及 $\sigma^2(x) := \mathbb{E}(\varepsilon^2 \mid X = x)$。则

$$\mathbb{E}_f\big[(\hat{f}_{n,h}(x_0) - f(x_0))^2\big] = \mathbb{E}\left[\left(\frac{\frac{1}{n}\sum_{i=1}^n K_h(X_i - x_0)e(X_i)}{\frac{1}{n}\sum_{i=1}^n K_h(X_i - x_0) + n^{-r}}\right)^2\right] + \frac{1}{n}\mathbb{E}\left[\frac{\frac{1}{n}\sum_{i=1}^n K_h^2(X_i - x_0)\sigma^2(X_i)}{\left(\frac{1}{n}\sum_{i=1}^n K_h(X_i - x_0) + n^{-r}\right)^2}\right].$$

定义 $D_n := \frac{1}{n p_X(x_0)}\sum_{i=1}^n K_h(X_i - x_0) + n^{-r}$ 以及 $D := \frac{1}{p_X(x_0)}\mathbb{E}[K_h(X - x_0)] + n^{-r} = \mathbb{E}(D_n)$。那么

$$\begin{aligned}
\frac{1}{n}\mathbb{E}\left[\frac{\frac{1}{n}\sum_{i=1}^n K_h^2(X_i - x_0)\sigma^2(X_i)}{D_n^2}\right] &= \frac{1}{nD^2}\mathbb{E}\left[\Big(\frac{1}{n}\sum_{i=1}^n K_h^2(X_i - x_0)\sigma^2(X_i)\Big)\right] \\
&\quad + \frac{1}{n}\mathbb{E}\left[\left(\frac{1}{D_n^2} - \frac{1}{D^2}\right)\Big(\frac{1}{n}\sum_{i=1}^n K_h^2(X_i - x_0)\sigma^2(X_i)\Big)\right] \\
&= \frac{1}{nD^2}\,\mathbb{E}\big[K_h^2(X - x_0)\sigma^2(X)\big] \\
&\quad - \frac{1}{D^2}\mathbb{E}\left[\left(\frac{D_n^2 - D^2}{D_n^2}\right)\Big(\frac{1}{n^2}\sum_{i=1}^n K_h^2(X_i - x_0)\sigma^2(X_i)\Big)\right].
\end{aligned}$$

对于第二项,

$$\begin{aligned}
\mathbb{E}\left[\left|\left(\frac{D_n^2 - D^2}{D^2 D_n^2}\right)\Big(\frac{1}{n^2}\sum_{i=1}^n K_h^2(X_i - x_0)\sigma^2(X_i)\Big)\right|\right] &\le \frac{1}{D^2}\mathbb{E}\left[\left(\frac{1}{n^2}\sum_{i=1}^n K_h^2(X_i - x_0)\sigma^2(X_i)\right)^2\right]^{1/2} \\
&\le \frac{\sigma_\infty^2}{D^2}\mathbb{E}\left[\left(\frac{D_n^2 - D^2}{D_n^2}\right)^2\right]^{1/2}\left\{\frac{n-1}{n^3}\big(\mathbb{E}[K_h^2(X - x_0)]\big)^2 + \frac{1}{n^3}\mathbb{E}[K_h^4(X - x_0)]\right\}^{1/2} \\
&\le \mathbb{E}\left[\left(\frac{D_n^2 - D^2}{D_n^2}\right)^2\right]^{1/2} O\Big(\frac{1}{nh^d}\Big).
\end{aligned}$$

最后，定义 $E_p := \mathbb{E}\big[|D_n - D|^p\big]$。那么，对于任意 $q > 0$,

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{D_n^2 - D^2}{D_n^2}\right)^2\right] &\le n^{4r}\mathbb{E}\big[(D_n^2 - D^2)^2 \mathbb{1}_{(|D_n - D| > D/2)}\big] + \frac{2^4}{D^4}\mathbb{E}\big[(D_n^2 - D^2)^2 \mathbb{1}_{(|D_n - D| \le D/2)}\big] \\
&\le \frac{2^q n^{4r}}{D^q}\mathbb{E}\big[(D_n + D)^2 |D_n - D|^{q+2}\big] + \frac{2^4}{D^4}\mathbb{E}\big[(D_n + D)^2 (D_n - D)^2\big] \\
&\le \frac{2^{q+1}}{D^q} n^{4r} E_{q+4} + \frac{2^{q+3}}{D^{q-2}} n^{4r} E_{q+2} + \frac{2^5}{D^4} E_4 + \frac{2^7}{D^2} E_2,
\end{aligned}$$

其中，最后一个不等式成立是因为对于所有 $x, y \in \mathbb{R}$，$(x + y)^2 \le 2(x - y)^2 + 8y^2$。由 Rosenthal不等式，对于任意 $p \ge 2$，存在一个仅依赖于 p 的常数 C_p，使得

$$E_p \le C_p\left(\frac{\mathbb{E}(|K_h(X - x_0)|^p)}{n^{p-1}} + \frac{\{\mathrm{Var}(K_h(X - x_0))\}^{p/2}}{n^{p/2}}\right) = O\left(\frac{1}{(nh^d)^{p/2}}\right).$$

因此，对于任意 $q > 0$,

$$\mathbb{E}\left[\left(\frac{D_n^2 - D^2}{D_n^2}\right)^2\right] = O\left(\frac{1}{nh^d} + \frac{n^{4r}}{(nh^d)^{1+q/2}}\right) = O\left(n^{-2\beta/(2\beta+d)} + n^{4r - \beta(q+2)/(2\beta+d)}\right).$$

因此取 $q > 4(2 + d/\beta)r - 2$，我们得到

$$\mathbb{E}\left[\left(\frac{D_n^2 - D^2}{D_n^2}\right)^2\right] = o(1),$$

thus

$$\mathbb{E}\left[\left(\frac{D_n^2 - D^2}{D^2 D_n^2}\right)\Big(\frac{1}{n^2}\sum_{i=1}^n K_h^2(X_i - x_0)\sigma^2(X_i)\Big)\right] = o\left(\frac{1}{nh^d}\right).$$

通过类似的论证,

$$\mathbb{E}\left[\frac{\big(\frac{1}{n}\sum_{i=1}^n K_h(X_i - x_0)e(X_i)\big)^2}{D_n^2}\right] = o\big(h^{2\beta}\big).$$

此外，

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}K_h(X_i-x_0)e(X_i)\right)^2\right]=\frac{n-1}{n}\left\{\mathbb{E}\left[K_h(X-x_0)e(X)\right]\right\}^2+\frac{1}{n}\mathbb{E}\left[K_h^2(X-x_0)e^2(X)\right]$$
$$=\left\{\mathbb{E}\left[K_h(X-x_0)e(X)\right]\right\}^2+o\left(h^{2\beta}+\frac{1}{nh^d}\right).$$

因此

$$\mathbb{E}_f\left[(\hat{f}_{n,h}(x_0)-f(x_0))^2\right]=\frac{1}{nD^2}\mathbb{E}\left[K_h^2(X-x_0)\sigma^2(X)\right]+\frac{1}{D^2}\left\{\mathbb{E}\left[K_h(X-x_0)e(X)\right]\right\}^2+o\left(h^{2\beta}+\frac{1}{nh^d}\right). \quad (22)$$

现在，根据 p_X 的绝对连续性，

$$D=\frac{1}{p_X(x_0)}\mathbb{E}[K_h(X-x_0)]=\frac{1}{p_X(x_0)}\int_{\mathbb{R}}K(u)p_X(x_0+hu)\,du=1+o(1).$$

此外，根据 σ^2 的绝对连续性，

$$\mathbb{E}\left[K_h^2(X-x_0)\sigma^2(X)\right]=\frac{1}{h^d}\int_{\mathbb{R}}K^2(u)\sigma^2(x_0+hu)p_X(x_0+hu)\,du=\frac{R(K)p_X(x_0)\sigma^2(x_0)}{h^d}(1+o(1)).$$

当 $\beta\in(0,1]$，

$$\left|\mathbb{E}\left[K_h(X-x_0)e(X)\right]\right|\le\int_{\mathbb{R}^d}K(u)\left|f(x_0+hu)-f(x_0)\right|p_X(x_0+hu)du$$
$$\le Lh^\beta\int_{\mathbb{R}^d}K(u)\|u\|_2^\beta p_X(x_0+hu)\,du=p_X(x_0)Lh^\beta\mu_\beta(K)(1+o(1)).$$

另一方面，当 $\beta\in(1,2]$，

$$\left|\mathbb{E}\left[K_h(X-x_0)e(X)\right]\right|=\left|\int_{\mathbb{R}^d}K(u)\big(f(x_0+hu)-f(x_0)\big)p_X(x_0+hu)\,du\right|$$
$$=h\left|\int_{\mathbb{R}^d}\int_0^1K(u)u^\top\big(\nabla f(x_0+thu)-\nabla f(x_0)\big)\,dt\,p_X(x_0+hu)\,du\right|$$
$$\le h\int_{\mathbb{R}^d}\int_0^1t\,K(u)\|u\|_2\|\nabla f(x_0+thu)-\nabla f(x_0)\|_2\,dt\,p_X(x_0+hu)\,du$$
$$\le Lh^\beta\int_0^1t^{\beta-1}dt\int_{\mathbb{R}^d}K(u)\|u\|_2^\beta p_X(x_0+hu)\,du=\frac{p_X(x_0)Lh^\beta\mu_\beta(K)}{\beta}(1+o(1)),$$

由 b 得到 通过带积分余项的泰勒展开。将（22）与上述内容结合 e,

$$\mathbb{E}_f\left[(\hat{f}_{n,h}(x_0)-f(x_0))^2\right]=\frac{L^2h^{2\beta}\mu_\beta^2(K)}{(1\vee\beta)^2}+\frac{R(K)\sigma^2(x_0)}{p_X(x_0)\,nh^d}+\left(h^{2\beta}+\frac{1}{nh^d}\right).$$

回顾（21），并注意到 $R(K)=\frac{2}{(\beta+d)\Gamma(1+d/2)\,\pi^{d/2}(2\beta+d)}$，以及 $\mu_\beta(K)=\frac{d}{2\beta+d}$，即可得到所需结果；

上确界 $_{f\in\mathcal{H}(\beta,\ L)}\mathbb{E}_f\left[(\hat{}f_{n,\ h}(x_0)\ -f\right.$
$\left.(x_0))^2\right]$

$$=\left(\frac{d^d(\beta+d)^{2\beta}\Gamma^{2\beta}(1+d/2)}{\pi^{\beta d}\beta^{2\beta}(2\beta+d)^d(1\vee\beta)^{2d}}\right)^{1/(2\beta+d)}\left(\frac{L^{d/\beta}\sigma^2(x_0)}{p_X(x_0)n}\right)^{2\beta/(2\beta+d)}(1+o(1))。 \quad (23)$$

□

A.7 定理7的证明

我们首先证明逐点均方误差的有限样本下界。

引理 11（Hölder 函数估计的有限样本下界）。考虑模型 (10)，其中$p_X(\cdot)/p_X(x_0) \in \mathcal{H}(\alpha, L_X)$，且某个 $\alpha \in (0,1]$, $L_X > 0$。那么，对于任意 $n \in \mathbb{N}, n \geq 3$, $d \in \mathbb{N}$, $L > 0, \sigma^2 > 0$, $x_0 \in \mathbb{R}^d$,

$$\sup_{f \in \mathcal{H}(\beta, L)} \mathbb{E}_f\left[(\hat{f}_n(x_0) - f(x_0))^2\right] \geq \underline{A}\Big(\frac{2\beta}{2\beta+d}\Big)\Big(\frac{1}{1+c_{n,d}}\Big)\left(\frac{d^d(\beta+d)^{2\beta}\Gamma^{2\beta}(1+d/2)}{\pi^{\beta d}\beta^{2\beta}(2\beta+d)^d(1\vee\beta)^{2d}}\right)^{1/(2\beta+d)}\left(\frac{L^{d/\beta}\sigma^2}{p_X(x_0)n}\right)^{2\beta/(2\beta+1)}, \quad (24)$$

其中

$$c_{n,d} := L_X\left(\frac{(\beta+d)(2\beta+d)\Gamma(1+d/2)\sigma^2}{2\pi^{d/2}\beta^2 p_X(x_0)L^2 n}\right)^{\alpha/(2\beta+d)},$$

and

$$\underline{A}(a) := \frac{1}{a^a(1-a)^{1-a}} \sup_{\substack{0<\lambda\leq 1 \\ m>0}} \frac{\lambda^a}{\left\{(m+1)_2F_1\left(-\frac{1}{2}, \frac{m}{2}, \frac{m}{2}+1; -\frac{\lambda}{m^2}\right)\right\}^2}.$$

引理 *11* 的证明。 证明方法与引理 9 的证明类似。考虑如下分布族 $(P_t)^{t\in[-1,1]}$ ：

$$Y_i \,|\, X_i \overset{\text{iid}}{\sim} N\big(f_t(X_i), \sigma^2\big), \qquad f_t(x) := \frac{tLh_n^\beta}{1\vee\beta}K\Big(\frac{x-x_0}{h_n}\Big), \qquad K(u) = \big(1-\|u\|_2^\beta\big)_+,$$
$$R_{\beta,d} := \frac{2\pi^{d/2}\beta^2}{(\beta+d)(2\beta+d)\Gamma(1+d/2)}, \qquad h_n := \left(\frac{(1\vee\beta)^2\sigma^2\lambda}{L^2 p_X(x_0)R_{\beta,d}(1+\phi_n)n}\right)^{1/(2\beta+d)},$$
$$\phi_n := L_X\left(\frac{(1\vee\beta)^2\sigma^2\lambda}{L^2 p_X(x_0)R_{\beta,d}n}\right)^{\alpha/(2\beta+d)},$$

对任意 $\lambda \in (0,1]$。与引理 9 的证明一样， $f_t \in \mathcal{H}(\beta,\ L)$。费舍尔信息为

$$\mathcal{I}(t) \leq \frac{L^2 n h_n^{2\beta+d} R_{\beta,d} p_X(x_0)}{(1\vee\beta)^2\sigma^2}\big(1+h_n^\alpha L_X\big) = \frac{1+h_n^\alpha L_X}{1+\phi_n}\lambda \leq \lambda.$$

再次取 $\hat{} \ t_n := \frac{1\vee\beta}{Lh_n^\beta}\hat{} f_n(x_0)$,

$$\sup_{f\in\mathcal{H}(\beta,L)} \mathbb{E}_f\left[(\hat{f}_n(x_0) - f(x_0))^2\right] \geq \frac{1}{(1\vee\beta)^2}\left(\frac{(1\vee\beta)^2 L^{d/\beta}\sigma^2}{p_X(x_0)R_{\beta,d}n}\cdot\frac{1}{1+c_{n,d}}\right)^{2\beta/(2\beta+d)}\left(\frac{(2\beta)^{2\beta}d^d}{(2\beta+d)^{2\beta+d}}\right)^{1/(2\beta+d)}\underline{A}\Big(\frac{2\beta}{2\beta+d}\Big).$$

结论由此得出。

□

定理 7 由引理 11 得出，概述如下。

定理的证明 7。 在假设的渐近条件下，

$$p_X(x_0)L^2 R_{\beta,d} \geq \frac{L^2 R_{\beta,d}}{C\,\mathrm{vol}\,\mathcal{B}_1(0)} = \frac{2\beta^2 L^2}{C(\beta+d)(2\beta+d)} \sim \frac{2\beta^2 L^2}{Cd^2},$$

因此

$$c_{n,d} \sim L_X L^{-\frac{2\alpha}{d}} n^{-\frac{\alpha}{d}} = L_X e^{-2\alpha\frac{\log L}{d}} e^{-\alpha\frac{\log n}{d}} \to 0. \quad (25)$$

此外，对于 $a \in (0, 1/2)$,

$$\underline{A}(a) = \frac{1}{a^a(1-a)^{1-a}} \sup_{\substack{0<\lambda\leq 1 \\ m>0}} \frac{4\lambda^a}{(m+1)^2\big(\int_0^1 t^{\frac{m}{2}-1}\sqrt{m^2+\lambda t}\,dt\big)^2} \geq \frac{1}{a^a(1-a)^{1-a}} \sup_{\substack{0<\lambda\leq 1 \\ m>0}} \frac{m^2\lambda^a}{(m+1)^2(m^2+\lambda)} = a^{2a}(1-a)^{2(1-a)},$$

上确界在 $(\lambda, m) = \left(\frac{a^3}{1-a}()^3, \frac{a}{1-a} \right)$ 处取得。因此，对于 $\beta \in (0, 2]$, $d \geq 2$,

$$\underline{A}\left(\frac{2\beta}{2\beta+d}\right) \geq \left(\frac{(2\beta)^{2\beta} d^d}{(2\beta+d)^{2\beta+d}}\right)^{2/(2\beta+d)} \sim 1,$$

为 $d \to \infty$。与引理 10 的证明类似，渐近论证在高维情形下同样成立，因此（23）仍然成立。因此，

$$\lim_{n\to\infty} \frac{\inf_{\hat{f}} \sup_{f\in\mathcal{H}(\beta,L)} \mathbb{E}_f\left[\left(\hat{f}(x_0) - f(x_0)\right)^2\right]}{\left(\frac{d^d(\beta+d)^{2\beta}\Gamma^{2\beta}(1+d/2)}{\pi^{\beta d}\beta^{2\beta}(2\beta+d)^d(1\vee\beta)^{2d}}\right)^{1/(2\beta+d)} \left(\frac{L^{d/\beta}\sigma^2}{p_X(x_0)n}\right)^{2\beta/(2\beta+d)}} = 1,$$

应用斯特林近似,

$$\Gamma^{2\beta/(2\beta+d)}(1+d/2) \sim \left\{\sqrt{\pi d}\left(\frac{d}{2e}\right)^{d/2}\right\}^{2\beta/d} \sim \pi^{\beta/d}\left(d^{1/d}\right)^{\beta}\left(\frac{d}{2e}\right)^{\beta} \sim \left(\frac{d}{2e}\right)^{\beta}.$$

因此

$$\left(\frac{d^d(\beta+d)^{2\beta}\Gamma^{2\beta}(1+d/2)}{\pi^{\beta d}\beta^{2\beta}(2\beta+d)^d(1\vee\beta)^{2d}}\right)^{1/(2\beta+d)} \sim \frac{d^{2\beta/d}}{\pi^{\beta}(1\vee\beta)^2}\left(\frac{d}{2e}\right)^{\beta} \sim \frac{d^{\beta}}{(2\pi e)^{\beta}(1\vee\beta)^2},$$

由此可得所需结果。

□

A.8 定理 8 的证明

定理 *8* 的证明。 我们推广定理 1 的证明。定义

$$\zeta(x,t) := \frac{\partial_t\big(p(x,t)\alpha(t)\big)}{p(x,t)\mu(t)}.$$

考虑矩阵

$$\begin{pmatrix} \int_T\int_{\mathcal{X}} \big(\hat{\psi}(x)-\phi(t)\big)\big(\hat{\psi}(x)-\phi(t)\big)^\top p(x,t)\mu(t)\,d\nu(x)\,dt & \int_T\int_{\mathcal{X}} \big(\hat{\psi}(x)-\phi(t)\big)\zeta(x,t)^\top p(x,t)\mu(t)\,d\nu(x)\,dt \\ \int_T\int_{\mathcal{X}} \zeta(x,t)\big(\hat{\psi}(x)-\phi(t)\big)^\top p(x,t)\mu(t)\,d\nu(x)\,dt & \int_T\int_{\mathcal{X}} \zeta(x,t)\zeta(x,t)^\top p(x,t)\mu(t)\,d\nu(x)\,dt \end{pmatrix}$$

$$= \int_T\int_{\mathcal{X}} \begin{pmatrix} \hat{\psi}(x)-\phi(t) \\ \zeta(x,t) \end{pmatrix}\begin{pmatrix} \hat{\psi}(x)-\phi(t) \\ \zeta(x,t) \end{pmatrix}^\top p(x,t)\mu(t)\,d\nu(x)\,dt \succcurlyeq 0,$$

其中 $\succcurlyeq$ 表示勒夫纳序。于是,

$$\begin{aligned} &\int_T\int_{\mathcal{X}} \big(\hat{\psi}(x)-\phi(t)\big)\big(\hat{\psi}(x)-\phi(t)\big)^\top p(x,t)\mu(t)\,d\nu(x)\,dt \\ &\succcurlyeq \left(\int_T\int_{\mathcal{X}} \zeta(x,t)\big(\hat{\psi}(x)-\phi(t)\big)^\top p(x,t)\mu(t)\,d\nu(x)\,dt\right)^\top \left(\int_T\int_{\mathcal{X}} \zeta(x,t)\zeta(x,t)^\top p(x,t)\mu(t)\,d\nu(x)\,dt\right)^{-1} \\ &\quad\cdot\left(\int_T\int_{\mathcal{X}} \zeta(x,t)\big(\hat{\psi}(x)-\phi(t)\big)^\top p(x,t)\mu(t)\,d\nu(x)\,dt\right). \end{aligned}$$

首先，由于 $\alpha(t) = 0$ 位于 ∂T T的边界上,

$$\int_T \zeta(x,t)p(x,t)\mu(t)\,dt = \int_T \partial_t\big(p(x,t)\alpha(t)\big)\,dt = 0$$

对于 ν-几乎处处的 x，因此

$$\int_T\int_{\mathcal{X}} \hat{\psi}(x)p(x,t)\mu(t)\,d\nu(x)\,dt = 0. \tag{26}$$

通过分部积分

$$\int_T \phi(t)\partial_t\big(p(x,t)\alpha(t)\big)\,dt = -\int_T p(x,t)\alpha(t)\nabla\phi(t)\,dt,$$

因此

$$\int_T \int_{\mathcal{X}} \phi(t)\zeta(x,t)p(x,t)\mu(t)\,d\nu(x)\,dt = -\int_T \alpha(t)\nabla\phi(t)\,dt. \tag{27}$$

Now,

$$\begin{aligned}
&\int_T \int_{\mathcal{X}} \zeta(x,t)\zeta(x,t)^\top p(x,t)\mu(t)\,d\nu(x)\,dt \\
&= \int_T \int_{\mathcal{X}} \frac{\big(\partial_t p(x,t)\alpha(t) + p(x,t)\nabla\alpha(t)\big)\big(\partial_t p(x,t)\alpha(t) + p(x,t)\nabla\alpha(t)\big)^\top}{p(x,t)\mu(t)}\,d\nu(x)\,dt \\
&= \int_T \int_{\mathcal{X}} \left(\frac{\partial_t p(x,t)\,\partial_t p(x,t)^\top}{p(x,t)}\frac{\alpha^2(t)}{\mu(t)} + \frac{\nabla\alpha(t)\nabla\alpha(t)^\top}{\mu(t)}p(x,t)\right) d\nu(x)\,dt \\
&= \int_T \int_{\mathcal{X}} \left(\mathcal{I}(t)\frac{\alpha^2(t)}{\mu(t)} + \frac{\nabla\alpha(t)\nabla\alpha(t)^\top}{\mu(t)}\right) dt,
\end{aligned} \tag{28}$$

其中第二个等式中的交叉项为零。应用詹森不等式

$$\int_{\mathcal{X}} \|\partial_t p(x,t)\|_2\,d\nu(x) \le \|\mathcal{I}(t)\|_{\mathrm{op}}^{1/2} < \infty,$$

于是应用控制收敛

$$\begin{aligned}
\int_{\mathcal{X}} \frac{\big(\partial_t p(x,t)\,\alpha(t)\big)\big(p(x,t)\nabla\alpha(t)^\top\big)}{p(x,t)\mu(t)}\,d\nu(x) &= \frac{\alpha(t)}{\mu(t)}\int_{\mathcal{X}} \partial_t p(x,t)\,d\nu(x)\nabla\alpha(t)^\top \\
&= \frac{\alpha(t)}{\mu(t)}\left(\partial_t \int_{\mathcal{X}} p(x,t)\,d\nu(x)\right)\nabla\alpha(t)^\top = 0.
\end{aligned}$$

结合（26）、（27）和（28）可得

$$\begin{aligned}
&\int_T \int_{\mathcal{X}} \big(\hat{\psi}(x) - \phi(t)\big)\big(\hat{\psi}(x) - \phi(t)\big)^\top p(x,t)\mu(t)\,d\nu(x)\,dt \\
&\qquad \succcurlyeq \left(\int_T \nabla\phi(t)\alpha(t)\,dt\right)^\top \left(\int_T \frac{\mathcal{I}(t)\alpha^2(t) + \nabla\alpha(t)\nabla\alpha(t)^\top}{\mu(t)}\,dt\right)^{-1} \left(\int_T \nabla\phi(t)\alpha(t)\,dt\right).
\end{aligned}$$

然后，应用 Takatsuand Kuchibhotla （2024年，引理 4）（即应用反三角不等式）即可得到该结果。

□

A.9 与 Takatsuand Kuchibhotla（2024年）的联系

在近期的研究中，Takatsu and Kuchibhotla （2024年）通过引入额外的逼近函数 ϕ: $T \to \mathbb{R}$，提出了广义 van Trees 不等式，它扩展了 van Trees 不等式。这一推广使得我们能够为不可微泛函以及非正则统计模型中的估计问题导出极小极大下界。在第2节的单变量情形下，他们的广义 van Trees 不等式简化为

$$\sup_{\phi} \left(\sqrt{\frac{\big(\int_T \phi'(t)\mu(t)\,dt\big)^2}{\int_T \mathcal{I}(t)\mu(t)\,dt + \mathcal{J}(\mu)}} - \sqrt{\int_T \big(\phi(t) - t\big)^2\mu(t)\,dt}\right)_+^2.$$

由于这并非人们预期会用到广义 van Trees 逼近函数的情形（该函数旨在加以扩展，以允许不可微泛函），我们预期在这一特定情形下 van Trees 不等式将被恢复，即

$$\sup_{\phi} \left(\sqrt{\frac{\big(\int_T \phi'(t)\mu(t)\,dt\big)^2}{\int_T \mathcal{I}(t)\mu(t)\,dt + \mathcal{J}(\mu)}} - \sqrt{\int_T \big(\phi(t) - t\big)^2\mu(t)\,dt}\right)_+^2 = \frac{1}{\int_T \mathcal{I}(t)\mu(t)\,dt + \mathcal{J}(\mu)},$$

且上确界 通过取 $\phi = \mathrm{id}$ 所达到的 mum。为看出这一点，对某个 φ，取 $\phi(t) = t + \varphi(t)$: $T \to \mathbb{R}$。然后

$$\begin{aligned}
&\sqrt{\frac{\left(\int_T \phi'(t)\mu(t)\,dt\right)^2}{\int_T \mathcal{I}(t)\mu(t)\,dt + \mathcal{J}(\mu)}} - \sqrt{\int_T \big(\phi(t)-t\big)^2\mu(t)\,dt} \\
&= \frac{\left|\int_T \phi'(t)\mu(t)\,dt\right|}{\sqrt{\int_T \mathcal{I}(t)\mu(t)\,dt + \mathcal{J}(\mu)}} - \sqrt{\int_T \big(\phi(t)-t\big)^2\mu(t)\,dt} \\
&= \frac{\left|1 - \int_T \varphi(t)\mu'(t)\,dt\right|}{\sqrt{\int_T \mathcal{I}(t)\mu(t)\,dt + \mathcal{J}(\mu)}} - \sqrt{\int_T \big(\varphi(t)\big)^2\mu(t)\,dt}.
\end{aligned}$$

现在，由于 $|1-x| \le 1 + |x|$，并利用柯西–施瓦茨不等式，

$$\left|1 - \int_T \varphi\mu'\right| \le 1 + \left|\int_T \varphi\mu'\right| \le 1 + \left(\int_T \varphi^2\mu\right)^{1/2}\left(\int_T \frac{(\mu')^2}{\mu}\right)^{1/2} = 1 + \sqrt{\int_T \varphi^2\mu}\,\sqrt{\mathcal{J}(\mu)}.$$

Thus

$$\begin{aligned}
\sqrt{\frac{\left(\int_T \phi'(t)\mu(t)\,dt\right)^2}{\int_T \mathcal{I}(t)\mu(t)\,dt + \mathcal{J}(\mu)}} - \sqrt{\int_T \big(\phi(t)-t\big)^2\mu(t)\,dt} &\le \frac{1}{\sqrt{\int_T \mathcal{I}\mu + \mathcal{J}(\mu)}} + \frac{\sqrt{\int_T \varphi^2\mu}\,\sqrt{\mathcal{J}(\mu)}}{\sqrt{\int_T \mathcal{I}\mu + \mathcal{J}(\mu)}} - \sqrt{\int_T \varphi^2\mu} \\
&\le \frac{1}{\sqrt{\int_T \mathcal{I}\mu + \mathcal{J}(\mu)}},
\end{aligned}$$

因此

$$\sup_{\phi}\left(\sqrt{\frac{\left(\int_T \phi'(t)\mu(t)\,dt\right)^2}{\int_T \mathcal{I}(t)\mu(t)\,dt + \mathcal{J}(\mu)}} - \sqrt{\int_T \big(\phi(t)-t\big)^2\mu(t)\,dt}\right)_+^2 = \frac{1}{\int_T \mathcal{I}(t)\mu(t)\,dt + \mathcal{J}(\mu)}.$$

这一观察自然地区分了逼近函数的Takatsu 和 Kuchib- hotla (2024年) 与我们的增广函数各自的用途。正如我们在第4节所见，我们还可以将逼近函数和增广函数一并纳入增广广义 *van Trees* 不等式。