arXiv:2603.01971v1 [stat.ML] 2 Mar 2026

LOCUS：一种用于风险感知预测的无分布损失分位数评分

Matheus Barreto[1, 2], M´ario de Castro[2], Thiago R. Ramos[1], Denis Valle[3], 和 Rafael Izbicki[1]

[1]巴西圣保罗州圣卡洛斯市圣卡洛斯联邦大学统计系˜ao˜ao˜[2]巴西圣保罗州圣卡洛斯市圣保罗大学数学与计算机科学研究所˜ao˜ao˜[3]美利坚合众国佛罗里达州盖恩斯维尔佛罗里达大学森林、渔业与地理信息科学学院

2026年3月3日

摘要

现代机器学习模型的平均表现可能很准确，但仍会犯下主导部署成本的错误。我们提出 LOCUS，这是一种无分布假设包装器，可针对固定的预测函数，为每个输入生成损失尺度可靠性评分。它不量化标签的不确定性，而是使用任何能够根据输入输出损失预测分布的引擎，对预测函数的实际损失进行建模。一个简单的分割校准步骤可将该函数转化为一种分布无关且易于解释的评分，该评分可在不同输入之间进行比较，并可解读为损失上界。该评分本身可用于排序，也可选择设置阈值，以获得透明的标记规则，并对大损失事件进行分布无关控制。涵盖 13 个回归基准测试的实验表明，与标准启发式方法相比， LOCUS能够有效进行风险排序，并减少大损失事件的发生频率。

1 引言

现代机器学习模型平均而言可以非常准确 [LeCun等人，2015年，Lakshminarayanan等人，2017], 然而已部署的系统往往是按每一次决策来评判的。对于给定的输入 x，操作层面的问题不仅是模型做出何种预测，还包括该预测是否可以安全地据以行动，或者是否应已标记以供额外审查。这一需求出现在临床决策支持、信用评分和自主系统等领域，在这些领域中，罕见但巨大的误差主导着自动化的真实成本。

一个核心困难在于，预测误差本质上具有异质性。即使数据充足，不可约的（偶然）随机性也意味着结果 Y 并不是 X [Kendall和Gal， 2017]的确定性函数。在数据有限时，（认知）不确定性会进一步放大系统尚未被充分观测的区域中的风险 [H¨ullermeier和Waegeman， 2021]。因此，任何固定预测器 g（x）都可能在全局性能指标下看似良性的输入上偶尔招致灾难性损失。

标准性能汇总指标，如准确率、均方根误差、ROC曲线和曲线下面积，量化模型在部署分布 [Fawcett, 2006] 上的平均性能。它们对模型选择和基准测试很有用，但就其构造而言是全局性的：它们并不能告诉我们单次预测 g (x)的风险有多大。校准工具，如可靠性图、期望校准误差或共形预测集，通过提供更为局部的指标而构成重要改进，但它们仍然是对许多输入的聚合，并且往往无法转化为形如“将此 x 标记为对其的预测不可靠”的简单规则。

在实践中，单实例可靠性通常由某种不确定性概念来近似。从业者使用贝叶斯预测分布、深度集成、蒙特卡洛 dropout 、预测区间或预测集以及保形方法来估计条件分布 $Y \mid X = x$ （Lakshminarayanan 等，2017，Angelopoulos 和 Bates，2023；相关工作见第 1.2 节）。这些方法会产生预测方差、熵或区间宽度等量。一种常见的启发式方法是标记估计变异性较大或预测集较宽的点，并信任预测分布较紧的点。然而，这些不确定性代理指标与部署中通常真正重要的量仅有松散联系：实际所采用的预测所产生的损失。

形式上，我们固定一个已部署预测函数 $g : \mathcal{X} \to \mathcal{Y}$ 和一个任务损失 $L(\cdot, \cdot)$（平方误差、0–1 损失或领域特定成本）。对于一个新的样本对 (X, Y)，我们定义实际损失

$$Z = L\big(g(X), Y\big).$$

从业者通常会设定一个阈值 τ，用于界定不可接受的误差（例如，具有临床意义的偏差）。目标是设计一条标记规则，使我们选择信任的预测中事件 $\{Z > \tau\}$ 的发生率保持较低，并且该规则的保证在有限样本下仍然成立，且无需假设概率模型设定正确。

我们的方法：损失控制标记。我们提出 Locus（基于不确定性分数的损失控制），这是一种无分布假设包装器，可将对条件损失分布的任意估计转化为针对固定预测函数 $g(x)$、可用于部署的可靠性指标。具体而言，给定一个独立同分布校准样本，以及任何能够输出实际损失的预测累积分布函数的黑箱引擎，$\widehat{F}(\cdot \mid x) \approx \mathbb{P}(Z \le \cdot \mid X = x)$，我们会应用一个简单的分割校准步骤，为每个输入计算一个易于解读的分数 $U_\alpha(x)$。由此得到的 $U_\alpha(x)$在损失尺度上经过校准，并且可以在不同输入之间进行比较。指定损失容忍度 τ 后，只要 $U_\alpha(x) \le \tau$，我们就接受（不标记）该输入；否则就将其标记，从而形成一条透明规则，对大损失事件进行分布无关控制（定理 3）。

图 1 展示了为何预测方差（面板 b）可能无法准确反映部署风险。对于两个固定的预测函数 $g(x)$（左侧为线性模型；右侧为 k近邻模型），面板 (a) 绘制了数据和拟合的 $g(x)$，面板 (c) 显示实际损失 $Z = L(g(X)\ Y)$以及我们在损失尺度上校准的上界 $U_\alpha(x)$（一个估计的$(1-\alpha)$损失分位数），面板 (d) 则给出阈值 U_α，以便针对这一特定 g接受或标记输入。有趣的是，由于拟合不佳，线性模型在低方差区域产生了较大损失，而 $\sigma(x)$无法捕捉这一点，但 $U_\alpha(x)$可以。

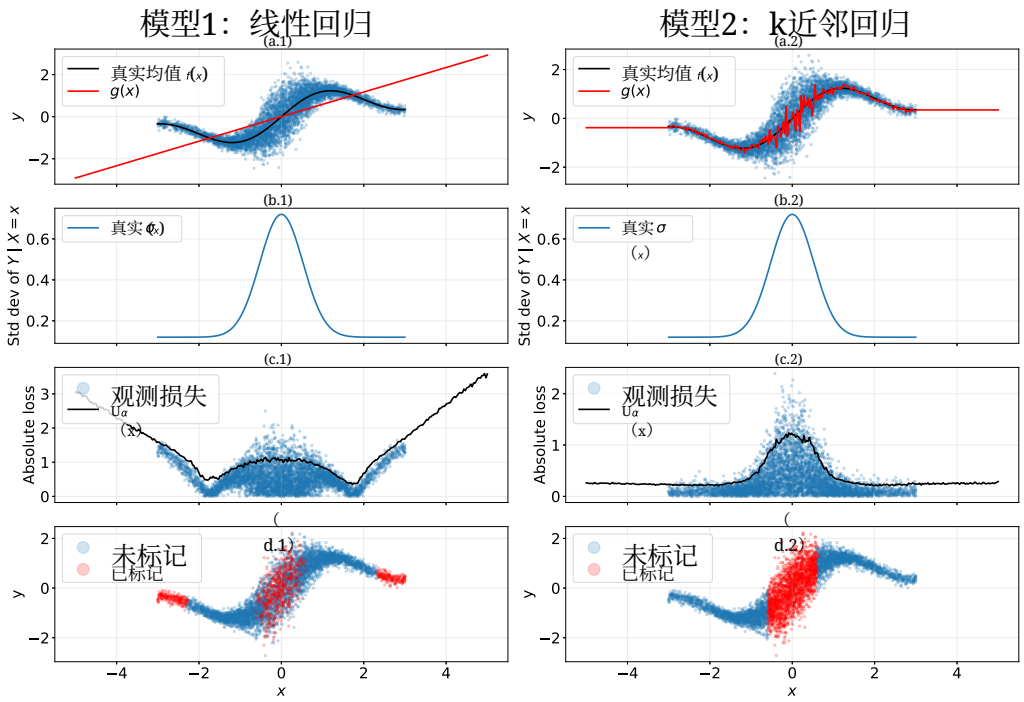

图 1：**以损失为中心的可靠性与方差。**我们对同一组数据拟合了两个模型：线性回归（左）和 kNN 回归（右）。(a) 数据及真实均值 $f(x)$（黑色）和拟合得到的预测函数 $g(x)$（红色）。(b) 真实条件标准差 $\sigma(x)$（偶然不确定性代理指标）。(c) 实际绝对损失 $Z = |g(X) - Y|$（点）和我们的 Locus 分数 $U_\alpha(x)$（曲线），这是对 $(1-\alpha)$ 损失分位数的校准估计。(d) 针对这一特定预测函数所导出的接受／标记规则：若 $U_\alpha(x) \le \tau$（蓝色）则接受，否则标记（红色）。U_α 以损失单位表示，易于解释，并能捕捉预测值中的问题（例如，低方差区域内的线性拟合不佳），而这些问题可能会被 $\sigma(x)$ 遗漏。

1.1 创新点与贡献

Locus 是针对固定预测函数 g 的一种以损失为中心的可靠性包装器，其核心是实际损失 $Z = L(g(X)\ Y)$，而非对 $Y \mid X = x$ 的不确定性。为了标记，我们假设用户指定了损失尺度上的容忍度 τ，以此界定何为不可接受的误差。我们的贡献如下：

- **由任意预测损失累积分布函数得到的校准损失分位数分数。**给定针对 $Z \mid X = x$ 的任意预测累积分布函数 $\widehat{F}(\cdot \mid x)$，我们构建具有有限样本、分布无关的边际有效性的局部评分 $U_\alpha(x)$：

$$\mathbb{P}\big(Z \le U_\alpha(X)\big) \ge 1 - \alpha \qquad \text{(Theorem 1)}.$$

在附加假设下， $U_\alpha(x)$ 还能实现渐近条件有效性（定理 2），并可解释为 x 处上 $1-\alpha$ 损失分位数的分布无关校准估计：若 $U_\alpha(x) = K$，则 x 处的损失大于 K的概率大约为 α 。

- **一种简单、易于理解且具有明确损失控制的标记规则**。对上界进行阈值化即可得到接受区域

$$A_{\lambda;\alpha} = \{x : U_\alpha(x) \le \lambda\}.$$

设置 $\lambda = \tau$ 可保证对“可信但不良”事件进行分布无关控制：

$$\mathbb{P}\big(Z > \tau,\ X \in A_{\tau;\alpha}\big) \le \alpha \qquad \text{(Theorem 3)}$$

这符合操作目标：限制我们选择信任的预测中不可接受损失的总体发生频率。此外，我们还有 $\mathbb{P}\big(Z > \tau \mid X \in A_{\tau;\alpha}\big) \lesssim \alpha$，也就是说，在接受的样本中出现大损失的概率约小于 α。

- **条件超越目标的实用调参**。虽然我们的核心保证控制了上述联合事件，但许多部署场景旨在针对接受点中的条件超越水平 η， $\mathbb{P}(Z > \tau \mid X \in A_{\lambda;\ \alpha}) \approx \eta$。我们提供了一种简单的启发式、基于验证的调参方案 Locus-Tuned，可在网格上调整λ 。我们还提供了对 λ 的完全分布无关的基于验证的选择，并以高概率保证 $\mathbb{P}(Z > \tau \mid X \in A_{\lambda^\star;\ \alpha}) \le \eta$ （附录 E，定理 5）。

- **不改变校准的认知不确定性感知膨胀**。我们提出了 F 的默认构造，使其在数据稀缺区域更为保守，从而提升外推情形下的稳健性，同时不影响无分布假设校准步骤及其保证。

1.2 与其他工作的关系

我们的方法与三个广泛的研究方向相关：保形预测与风险控制、局部或条件覆盖方法，以及选择性预测和拒绝选项的相关研究。下面重点介绍 Locus 与这些方法的不同之处。

选择性预测、拒绝选项与标记。选择性预测（带拒绝选项的分类）研究在预测与弃权之间进行选择的规则，其历史可追溯至 Chow 规则 [Chow， 1970]以及许多成本／覆盖率变体；综述请参见 [Zhang 等人，2023年，Hendrickx 等人， 2024]。现代研究通常会学习一个选择函数（有时与深度预测器联合学习），以权衡覆盖率与所选子集上的性能 [Geifman 和 El-Yaniv， 2017, 2019]。我们将 Locus 视为一种包装方法，它会诱导出一个接受区域 $A_{\lambda;\alpha} = \{x :\ U_\alpha\ (x)\ \le \lambda\}$，但我们并不学习选择器来优化覆盖率—准确率目标。相反，我们校准逐输入的实际损失上界，并通过设定阈值来控制被接受样本中的大损失事件。分数 U_α 也可单独用于风险排序和分诊。

正如 Hendrickx 等人所指出的，关于回归拒绝选项的文献比其分类对应领域更为有限。在这一背景下，Zaoui 等人研究了预算约束拒绝并获得了基于方差阈值化的规则，而 Locus 则以实际损失为目标，提供了分布无关的损失超越控制。

通过不确定性和分布外分数进行标记。 提供单实例可靠性的一种常见方式，是使用标量不确定性代理 （例如预测方差/熵、集成分歧）作为分诊信号，通常通过近似贝叶斯方法获得，例如蒙特卡洛 Dropout 或深度集成 [Gal and Ghahramani, 2016年, Lakshminarayanan 等人, 2017年, Zaoui 等人，2020]。然而，这类代理可能与部署风险不一致，并且在样本稀少的区域可能无法增大 [Ovadia 等人，2019]。与此密切相关，分布外（*OOD*）检测 和开放集识别 旨在利用基于置信度或距离的分数（例如最大 softmax 概率、ODIN、马氏特征距离）来标记异常输入 [Hendrycks and Gimpel, 2017年, Liang 等人, 2018年, Lee 等人， 2018], 相关综述将这些思想与异常检测和开放集识别联系起来 [Geng 等人, 2021年, Yang 等人, 2024]。尽管这些分数在检测异常输入方面有效，但它们通常未经过校准，无法直接控制固定部署预测器在接受集上的实际任务损失。 Locus转而对实际损失进行建模 $Z = L(g(X)\ Y)$，并对校准后的损失上界进行阈值化。

保形预测与风险控制。 经典保形预测在可交换性条件下构造具有有限样本边际覆盖率的集合值预测 [Vovk 等, 2005年， Angelopoulos 和 Bates， 2023年， Izbicki， 2025], 通常可保证 $\mathbb{P}(Y \in C(X)) \ge 1-\alpha$ ，且无需对预测模型作出假设。风险控制的扩展方法会校准预测集，从而界定未来样本点上的用户指定损失泛函的聚合值

（例如期望风险控制） [Bates 等，2021年，Angelopoulos 等， 2024]。从全局角度来看，这些方法界定了所有未来样本点上的平均损失。相比之下， LOCUS 通过 $\mathbb{P}(Z > \tau,\ \ X \in A_{\tau;\alpha}) \leq \alpha$设定损失上界，以控制接受集上的灾难性损失事件。

局部与条件保证。 越来越多的研究致力于突破纯粹的边际覆盖率，为能够适应预测任务局部难度的方法提供保证。例如，基于回归树和随机森林的局部保形方案 Cabezas 等， [2024, 2025a、b] ，以及其他用于局部保形预测的数据驱动分组方法 Lei 和 Wasserman [2014], Izbicki 等 [2020, 2022], Gibbs 等[2023]。我们构建 U_α（x）的方法在结构上与之相似，因为它旨在反映 x之间的异质性，但它针对的是损失 $Z = L$（g（X） Y）的分布，而非直接针对 Y。

认知不确定性感知保形方法。 近期研究开始在保形预测中明确结合偶然不确定性与认知不确定性。EPICSCORE 提出了一种通用机制，可通过贝叶斯估计为任意保形分数加入认知不确定性，从而生成在数据稀疏区域自动扩大的预测区域，同时保持有限样本边际覆盖率并实现渐近条件覆盖率[Cabezas 等，2025c]。类似地， LUCCA（局部不确定性保形校准）面向机器人动力学，在校准偶然不确定性估计的同时，考虑环境变化所引发的认知不确定性 Marques 和 Berenson [2024]。 LOCUS 与之互为补充：我们可以利用此类认知不确定性感知校准，为 $Z \mid X = x$构建更精确、更保守的预测累积分布函数；但本文的主要贡献是创建一种可解释的分数来衡量预测的可靠性，并提供一个标记层，将针对实际损失的任意预测累积分布函数转化为具有明确保证的接受规则。

2 方法

回顾部署预测器 g、损失 L 以及实际损失 $Z = L(g(X)\ Y)$（第 1 节）。令 $(X,\ Y)$ 表示一次独立的部署抽样，并令 $D = \{(X_i,\ Y_i)\}_{i=1}^n$ 表示来自同一部署分布、且独立于用于训练 g 的数据的独立同分布校准样本。定义观测到的校准损失 $Z_i := L(g(X_i)\ Y_i)$。利用 D，我们构造：(i) 针对 $Z \mid X = x$ 的经校准局部上界 $U_\alpha(x)$；以及 (ii) 基于对 $U_\alpha(x)$ 进行阈值化的标记规则。

我们的方法框架分为两个层次。首先，我们使用在部分校准数据上拟合的概率模型，为损失 $Z \mid X = x$ 构建基础预测分布。其次，我们在留出数据上应用无分布假设校准步骤，将基础损失预测分布转换为有效的局部上界 $U_\alpha(x)$，并满足有限样本边际保证，而无论基础模型是否设定正确。下一节将详细介绍这些步骤。

2.1 构建校准的局部损失界

步骤 1：数据划分

我们将校准样本划分为两个相互独立的部分：

$$D_1 = \{(X_i, Y_i)\}_{i\in I_1} \quad \text{and} \quad D_2 = \{(X_i, Y_i)\}_{i\in I_2},$$

集合 D_1 用于拟合一个预 测模型来预测 $Z \mid X$，而 D_2 则用于校准。我们构建由此产生的损失数 ta

$$\widetilde{D}_1 := \{(X_i, Z_i) : i \in I_1\} \text{ and } \widetilde{D}_2 := \{(X_i, Z_i) : i \in I_2\}.$$

步骤 2： $Z \mid X$ 的基础预测模型

~在此，我们为这类模型提出两种构造，不过步骤 3 与步骤 2 的具体执行方式无关。在 D_1上，我们拟合一个给定 X 时 Z 的贝叶斯模型。我们假设一族条件密度

$$\mathcal{F} = \{f(z \mid x, \theta) : \theta \in \Theta\},$$

对 $Z \mid X = x$ 的偶然变异进行建模，并为 θ (Θ 设定先验分布，该分布可以是无限维的；也就是说，这可以是一个非参数模型)。观察到 D_1 后，我们得到后验分布 $f(\theta \mid D_1)$ 以及相应的后验预测累积分布函数

$$\begin{aligned} F_Z(z \mid x, D_1) &:= \mathbb{P}(Z \leq z \mid x, D_1) \\ &= \int_\Theta F_Z(z \mid x, \theta)\, f(\theta \mid \widetilde{D}_1)\, d\theta, \end{aligned} \tag{1}$$

~其中， $F_Z(z \mid x, \theta)$ 是由 $f(z \mid x, \theta)$ 诱导的累积分布函数。最后，定义 $F(\cdot \mid x) \equiv F_Z(\cdot \mid x, D_1)$。

在实践中，后验分布可能无法通过闭式形式计算。此时，我们可以通过对后验抽样 $\theta^{(1)}$、.....、$\theta^{(S)} \sim f(\theta \mid \widetilde{D}_1)$ 进行蒙特卡洛聚合（例如，贝叶斯加性回归树（BART）、蒙特卡洛 dropout、贝叶斯神经网络），或采用变分近似（例如，变分高斯过程）来近似 (1)。

~**的另一种定义** F。(1) 中的后验平均可能无法反映数据稀缺或外推区域中损失概率较高的情况 [Ovadia 等人，2019年，Wang 等人， 2022]。为明确表征认知不确定性，我们还定义了预测累积分布函数的密度调制后验包络。给定后验抽样 $\{\theta^{(s)}\}_{s=1}^{S}$ 和一个截尾水平 $\gamma(x) \in (0,1)$（我们稍后会讨论如何选择该值），定义

$$\widetilde{F}(z \mid x) := \text{quantile}_{\gamma(x)}\Big(\big\{F(z \mid x, \theta^{(s)})\big\}_{s=1}^{S}\Big). \tag{2}$$

~对于每个固定的 x， $F(\cdot \mid x)$ 都是 z 上的有效累积分布函数。较小的 $\gamma(x)$ 对应于更保守（更低）的累积分布函数包络（右尾更重），因此上界更大。

步骤 3：通过概率积分变换值进行无分布假设校准

~ ~我们使用概率积分变换（PIT）值 [来校准 D_2 上的预测累积分布函数 F （Zhao 等人，2021年，Dey ~ ~等人， 2025]）。对于每个 $(X_i, Z_i) \in D_2$，定义 $W_i := F(Z_i \mid X_i)$。</style}对于目标尾部水平 $(0, 1)$，定义校准水平

$$t_{1-\alpha} := \text{quantile}_{1-\alpha}\big(\{W_i : (X_i, Z_i) \in \widetilde{D}_2\}\big).$$

最后，将局部性能分数定义为校准后的损失上界

$$U_\alpha(x) := \widetilde{F}^{-1}\big(t_{1-\alpha} \mid x\big).$$

对 $U_\alpha(x)$ 的解释：共形化损失分位数。 对于固定的 α，分数 $U_\alpha(x)$ 是一种局部风险度量：数值越大，表示模型在相应输入上可能遭受较大损失。它是对 x 处损失的 $1-\alpha$ 条件分位数的校准估计。

校准步骤可得出有限样本边际保证，无论基础预测模型是否正确。事实上，下面的定理表明，对于一个新的测试点，实际损失以至少 $1-\alpha$ 的概率低于校准后的上界，且不会低太多。

The定理 1 (边际有效性)。 在上述构造下，如果数据点是独立同分布的 .,

$$\mathbb{P}\big(Z \le U_\alpha(X)\big) \;\ge\; 1-\alpha.$$

此外，如果 *PIT* 分数之间出现平局的概率为0，则

$$1-\alpha+\frac{1}{|\tilde{D}_2|+1} \;>\; \mathbb{P}\big(Z \le U_\alpha(X)\big).$$

~此外，在附加假设下，即随着校准样本的增加，估计损失分布 $F(\cdot \mid x)$ 变得准确，并且真实的条件损失行为表现良好（见附录 F）， U_α 近似等于损失函数在 x 处的 $1-\alpha$ 分位数：

定理 2 (渐近条件覆盖率)。 在假设 *1* 下，我们有

$$\mathbb{P}\big(Z \le U_\alpha(x) \mid X = x,\ D\big) \xrightarrow{\mathbb{P}} 1-\alpha \quad as\ n \to \infty,$$

其中，左侧是一个随机变量（随机性来自 D），而 $\xrightarrow{\mathbb{P}}$ 表示相对于 D依概率收敛。

2.1.1 通过 kNN 半径选择 $\gamma(x)$

为了更好地编码认知不确定性，我们令截尾水平 $\gamma(x)$ 取决于 x 在 D_1中的设计点支持程度。直观而言，当 x 位于数据稀缺区域时，针对 $Z \mid X = x$ 的后验抽样可靠性较低，因此我们选择较小的 $\gamma(x)$，并在（2）中采用更保守的下包络，从而增大所得界限 $U_\alpha(x)$。下面我们将介绍一种简单且可扩展的方法，利用基于k近邻距离的局部密度代理，在表示空间中计算这样的 $\gamma(x)$。

设 $\phi : \mathcal{X} \to \mathbb{R}^d$ 是一种距离具有意义的表示。对于表格数据， $\phi(x)$ 可以是标准化后的协变量。对于高维输入，我们可以使用预测器 g 的嵌入（例如神经网络的倒数第二层），也可以使用用于预测 $Z \mid X$ 的模型所产生的嵌入。

首先，在 $\{\phi(X_i) : i \in I_1\}$ 上建立近邻索引。固定 k （例如 $k = 50$）。对于任意 x，定义 kNN 半径

$$r_k(x) := \|\phi(x) - \phi(X_{(k)}(x))\|,$$

其中， $X_{(k)}(x)$ 是 x 在 $\{X_i : i \in I_1\}$ 中的第 k 个近邻。较大的 $r_k(x)$ 表示较低的局部设计密度。接下来，通过 $q_{\mathrm{lo}} := $ 分位数 $_{0.50}(\{r_k(X_i) : i \in I_1\})$ 和 $q_{\mathrm{hi}} := $ 分位数 $_{0.90}(\{r_k(X_i) : i \in I_1\})$，在 D_1 上计算参考分位数。现在，定义标准化后的稀缺性评分

$$s(x) := \frac{r_k(x) - q_{\mathrm{lo}}}{q_{\mathrm{hi}} - q_{\mathrm{lo}} + \varepsilon},$$

对于一个较小的 ε 值（我们将其设为0.000001）。然后设定

$$\gamma(x) = \gamma_{\max} - (\gamma_{\max} - \gamma_{\min}) \cdot \sigma\left(\frac{s(x) - m}{s_\gamma}\right), \tag{3}$$

其中 σ 为逻辑函数。我们选取 $k = 50$、 $\gamma_{\min} = 0.15$、 $\gamma_{\max} = 0.9$、 $m = 0$和 $s_\gamma = 1$。在此选择下，$\gamma(x)$在低密度区域（较大的 $r_k(x)$）中更小，从而在(2)中得到更保守的包络以及更大的界$U_\alpha(x)$。

2.2 基于 $U_\alpha(x)$的标记规则

现在我们说明如何将局部评分 $U_\alpha(x)$转化为实用的标记规则，以在队列层面控制大损失。给定阈值 $\lambda \in \mathbb{R}$，我们定义未标记（接受）区域

$$A_{\lambda;\alpha} := \{x \in \mathcal{X} : U_\alpha(x) \le \lambda\},$$

及其补集，即已标记区域 $A^{\mathrm{c}}_{\lambda;\alpha} : = \{x : \ U_\alpha(x) > \lambda\}$. 我们假设用户在损失尺度上指定容差 τ，用以界定何为不可接受的误差。下文中，我们考虑选择 λ的两种方式，分别对应损失控制标记的两种变体。

2.2.1 使用 $\lambda = \tau$ 进行标记

最简单的选择是将接受阈值设为损失阈值 $\lambda = \tau$，此时接受区域为 $A_{\tau;\alpha} := \{x : U_\alpha(x) \le \tau\}$. 在这种情况下，我们得到以下保证：

定理 3. 在上述构造下，

$$\mathbb{P}\big(Z > \tau,\ X \in A_{\tau;\alpha}\big) \le \alpha.$$

联合界控制了“可信但表现不佳”的预测的总体比例：在所有未来案例中，未被标记且损失超过 τ 的案例所占比例至多为 α。

此外，下一个定理表明，对于较大的校准样本量，简单规则“当 $U_\alpha(x) \le \tau$ 时，信任 x ”近似等价于预言机规则；该规则仅在 x 处出现不可接受误差的概率至多为 α 时，才会信任 x 。也就是说，对 U_α 进行阈值化最终会在每个点上作出正确的标记决策。

定理 *4*。固定 $\alpha \in (0, 1)$、阈值 $\tau \in \mathbb{R}$ 和点 $x \in \mathcal{X}$。假设 $F_{Z|X}(\cdot \mid x)$ 连续，且其 $(1-\alpha)$-分位数 $q_{1-\alpha}(x) := F^{-1}_{Z|X}(1-\alpha \mid x)$ 唯一。另假设当 $n \to \infty$ 时， $U_\alpha(x) \xrightarrow{\mathbb{P}} q_{1-\alpha}(x)$（这在定理 *2* 的条件下成立），并且 $F_{Z|X}(\cdot \mid x)$ 连续。那么，决策规则“当且仅当 $U_\alpha(x) \le \tau$ 时接受”等价于预言机规则“当且仅当 $\mathbb{P}(Z > \tau \mid X = x) \le \alpha$ 时接受”。

最后，定理 2 表明， $U_\alpha(x)$ 的表现类似于 $(1-\alpha)$ 条件损失分位数，而定理 6（附录 F）中的选择论证指出，以接受为条件应能保留这一尾部水平。因此，我们预计，在接受点中，超越概率应有 α 的上界：

$$\mathbb{P}\big(Z > \tau \mid X \in A_{\tau;\alpha}\big) \lesssim \alpha.$$

换言之，接受集应具有近似由 α控制的大损失率。

2.2.2 Locus-Tuned：调节 λ

人们会o 也希望接受集满足目标条件超越水平 η，即 y

$$q(\lambda) := \mathbb{P}(Z > \tau \mid X \in A_{\lambda;\alpha}) \approx \eta.$$

如果 $U_\alpha(x)$恰好是 $Z \mid X = x$的$(1-\alpha)$条件分位数，那么选择 $\alpha = \eta$ 并设定 $\lambda = \tau$便可近似实现 $q(\tau) \approx \eta$。然而，在有限样本中，边际校准并不能保证由此得到的接受集满足期望的条件超越率。当另有带标签验证集可用时，我们可以调节接受阈值 λ 以匹配目标 η。

具体而言，设 $D_{\text{val}} = \{(X_i, Y_i)\}_{i=1}^N$ 为独立同分布，且独立于用于构造 $U_\alpha(\cdot)$的数据，并定义验证损失 $Z_i := L(g(X_i)\ Y_i)$。一次性计算验证分数，

$$u_i := U_\alpha(X_i), \qquad i = 1, \dots, N,$$

并考虑一个有限 的阈值网格 $\Lambda \subset \mathbb{R}$。对于每个 $\lambda \in \Lambda$，定义接受索引集及其大小

$$I_\lambda := \{i : u_i \le \lambda\}, \qquad n_\lambda := |I_\lambda|,$$

并估计接受的验证点中的条件超越率，计算如下

$$\widehat{q}(\lambda) := \frac{1}{n_\lambda \vee 1} \sum_{i \in I_\lambda} \mathbf{1}\{Z_i > \tau\}.$$

一个简单的选择规则是

$$\widehat{\lambda} \in \arg\min_{\lambda \in \Lambda} |\widehat{q}(\lambda) - \eta|.$$

在实践中，通常还可以施加最小接受约束（以避免 n_λ 过小时估计不稳定），例如，将范围限制为 $\{n_\lambda / N \ge \rho_{\min}\}$。

这一步调优采用的是启发式方法。如果需要一种完全无分布假设、基于验证集来选择 λ 的方法，并保证 $\mathbb{P}($ $Z > \tau \mid X \in A_{\widehat{\lambda};\alpha})$ 在有限网格上以高概率 $\le \eta$，请参见附录 E。附录（D 节）还提供了一种互补的启发式方法：固定 $\lambda = \tau$ (Locus-α)，同时调优 α 。

3 实验

我们在 13 个数据集上测试我们的方法；完整描述请参见附录 A。每个数据集均划分为训练集（40%）、校准集（40%）、验证集（10%）和测试集（10%）。我们首先通过随机森林 [Breiman, 2001] 训练固定的预测函数 g 。利用校准集，我们为实际损失$Z = L(g(X)\ Y)$拟合一个概率模型，并应用第 2.1 节中的分割校准程序，以获得校准后的损失上界 ~bound $U_\alpha(x)$。我们使用以下预测损失分布 $F(\cdot \mid x)$ 来实例化 Locus：基于混合密度网络的蒙特卡洛 dropout [Bishop, 1994]，以及贝叶斯加性回归树 [Chipman et al., 2010]。我们还展示了这些方法考虑认知不确定性的 $\gamma(x)$ 膨胀变体的结果。为简单起见，对每个数据集，我们设置容忍水平 τ （用于定义不可接受的误差），使得验证集上 30% 的观测损失超过该水平，即满足 $\mathbb{P}(Z > \tau) \approx 0.3$。实现细节见附录 B。

3.1 U_α 与简单标记

首先，我们说明为什么经过损失校准的分数 U_α 比标准不确定性代理指标更能直观地反映单实例可靠性，并展示使用 $\lambda = \tau$ 的默认 Locus 标记器。我们聚焦于房屋数据集（金县房价）。在这项实验中，我们将 τ =$70，000 设为价格绝对误差不可接受的操作阈值，

并使用蒙特卡洛 Dropout 估计 F。作为不确定性基线，我们还报告 $\widehat{sd}(Y \mid x)$，即给定特征时标签的条件标准差估计。详情见第 3.2 节和附录 B。

表 1 将 $\widehat{sd}(Y \mid x)$ 与孤立森林（基线分布外检测器；Liu 等人 [2008]）给出的离群标记进行了对比。关键区别在于其在损失尺度上的可解释性。虽然 $sd(Y \mid x)$ 衡量了 $Y \mid x$ 的离散程度，但它并未说明该样本的实际损失 $Z = |g(x) - y|$ 可能有多大，因为它没有考虑到 g 的局部拟合质量。例如，第三行的实际绝对误差为 $Z \approx$\$187，830，尽管估计的标签变异性仅为 $sd(Y \mid x) \approx$\$17，150；这说明，即使 $Y \mid X = x$ 的离散程度相对较低，当 g 在局部出现偏差时，仍可能伴随非常大的损失。同一行中的损失分位数摘要明确展示了风险尺度：$U_{0.5}(x) \approx$\$38，640 表明典型误差可能达到数万美元，而 $U_{0.1}(x) \approx$\$188，920 则表明，对于该特定输入，上尾中出现六位数误差是可能的。

表 1：**房价预测示例**。来自房价预测的六个测试样本；所有金额均以千美元计。我们报告 $\hat{y} = g(x)$、y 和 $Z = |\hat{y} - y|$，以及 Locus 界限 $U_{0.5}(x)$ 和 $U_{0.1}(x)$（每个 x 处损失的近似条件中位数和第 90 百分位数）。Locus $\alpha = 10$%标记使用 $\tau =$ \$70，000（当且仅当 $U_{0.1}(x) > \tau$ 时标记）；IF 表示孤立森林离群标记；$sd(Y \mid X = x)$ 是标签不确定性基线。$U_\alpha(x)$ 以美元表示，易于解释，并可指出被方差或分布外风格分数遗漏的高损失案例。

被标记为不可靠 由 Locus 标记?	被标记为离群点 由 IF?	结果（千美元）			Locus（千美元）		其他
		估计价格 $\hat{y}$	真实价格 y	$Z = \lvert\hat{y} - y\rvert$	$U_{0.5}(x)$	$U_{0.1}(x)$	$\widehat{sd}(Y \mid X = x)$
yes	yes	2358.64	3567.00	1208.36	294.06	1276.52	104.75
yes	no	1648.52	2200.00	551.48	203.00	554.59	104.07
yes	yes	337.83	150.00	187.83	38.64	188.92	17.15
no	yes	217.85	215.00	2.85	17.94	67.93	76.95
no	yes	208.04	219.95	11.91	17.27	43.58	25.10
no	no	306.03	319.00	12.97	24.87	67.21	31.17

相比之下，第二行展示了 OOD 式筛查的另一种失效模式：孤立森林没有将该点标记为离群点，但实际误差为 $Z \approx$\$551,480（对售价为 \$2,200,000 的房屋，预测价格为 \$1,648,520）。在此，$U_{0.5}(x) \approx$ \$203,000 和 $U_{0.1}(x) \approx$\$554,590 再次以直观的方式传达信息 —— 其金额尺度与部署容忍度相同 —— 即对于此输入，预期误差远高于 $\tau =$\$70,000。最后，底部区域强调，Locus 不会标记那些大概率预期损失较小的样本。

接下来，我们通过实证评估默认标记规则能否控制已接受预测中大损失的发生率。附录表 6 报告显示，在 $\lambda = \tau$ 和 $\alpha = 10$% 的设定下，Locus 在几乎所有数据集上的条件超越概率都接近目标值：$\mathbb{P}\big(Z > \tau \mid X \in A_{\tau;\alpha}\big) \lesssim \alpha$. 此外，附录表 4 验证了边际校准性质：经验覆盖率 $\widehat{\mathbb{P}}\big(Z \le U_\alpha(X)\big)$ 始终不低于 $1 - \alpha$，这与定理 1 的预期一致。作为补充，我们还在附录表 5 中提供了 Locus 接受率。

3.2 Locus-Tuned

本节将 Locus-Tuned 与两种标记基线方法进行比较：Locus-Tuned 固定 α（我们使用 $\alpha = 10$%），并在验证集划分上调整接受阈值 λ。**(i) IFlag**：孤立森林异常分数 $O(x)$。**(ii) VARNet**：标签方差估计值，$V(Y \mid X = x)$。附录 C 介绍了 VARNet 的实现细节。

设 $S(x)$ 为分数（Locus 对应 $U_\alpha(x)$，VARNet 对应 $V(Y \mid X = x)$，IFlag 对应 $O(x)$），并定义 $A_\lambda := \{x : S(x) \le \lambda\}$. 对于每个数据集和方法，我们选择 λ，使得验证集上的 $\mathbb{P}(X \in A_\lambda)$ 为 ≈ 70%，然后报告测试集上的 $\mathbb{P}(Z > \tau \mid X \in A_\lambda)$。在接受率匹配的情况下，较低的 $\mathbb{P}(Z > \tau \mid X \in A_\lambda)$ 表示风险排序效果更好。

表 2 报告了在全部 13 个数据集上，$\mathbb{P}(Z > \tau \mid X \in A)$. 30 次运行结果的中位数（括号内为第 5% 和第 95% 分位数，用于衡量变异性）。表中显示，在匹配的接受率约为 70% 时，Locus-Tuned 分数在每个数据集上的条件大损失率都显著低于两种基线方法。孤立森林标记始终表现最差，标签方差基线虽有所改善，但

表 2：**匹配接受率下调优后的条件大损失率**。 接受点中的条件超越， $\mathbb{P}(Z > \tau \mid X \in A_\lambda)$ (%)。对于每个数据集和方法，在验证集划分上选择阈值 λ ，以达到 $\approx 70\%$ 的接受率；随后报告测试性能。 Locus-Tuned 方法使用Locus 分数 U_α ，校准水平固定为 $\alpha = 10\%$。条目为 30 次运行的中位数，括号内为第 5 和第 95 百分位数；数值越低越好。粗体表示每行中的最低值。

	基线方法（无 Locus）		Locus-Tuned（损失分位数分数）			
数据集	**IFlag** （IF 离群值）	**VARNet** （标签方差）	**Locus-BART**	**Locus-BART** （γ-膨胀）	**Locus-MC Dropout**	**Locus-MC Dropout** (γ-infl.)
翼型	31.0 (27.4; 35.3)	27.5（23.0；30.7）	24.7 (19.6；28.0)	**23.1 (18.4；28.2)**	25.0（18.6；28.1）	24.3（19.4；27.5）
bike	30.1（30.1；30.1）	28.4（26.1；29.2）	20.1 (19.4；21.6)	**20.0 (19.1; 20.8)**	22.6（21.8；23.7）	22.9（21.7；24.0）
混凝土	33.5（27.2；38.6）	27.7（23.0；31.2）	**25.3（18.7；29.3）**	26.2 (20.5；30.4)	27.7 (21.5；31.1)	27.4（22.7；32.4）
周期	28.9 (27.2; 30.0)	29.7 (28.5; 30.2)	**27.5（26.4；28.6）**	28.0 (26.6; 28.8)	28.7（26.8；30.4）	29.2（27.9；30.1）
电力	31.5 (30.1; 32.1)	29.4（28.1；30.7）	24.2 (22.3；25.4)	**24.0 (22.5；26.2)**	24.6 (22.3; 26.4)	24.6 (22.9; 25.8)
住宅	34.8 (33.7；36.0)	26.6（24.8；27.7）	**18.8（18.2；20.2）**	19.3（17.5；20.2）	19.2 (17.9; 20.3)	19.0（17.9；20.3）
meps19	36.4（35.4；37.4）	24.8 (21.5；26.6)	15.7（14.0；17.4）	15.7（13.2；17.4）	14.3 (12.6；15.2)	**14.0 (12.5；15.1)**
蛋白质	28.0（27.4；29.0）	27.4 (25.8；29.4)	23.9（22.8；24.9）	23.9（23.2；25.0）	23.6（22.5；24.3）	**23.5（22.5；24.4）**
star	29.1（26.3；31.9）	29.0 (26.9; 33.8)	29.0（26.3；31.7）	28.8 (26.4; 32.2)	**28.6（24.3；32.4）**	29.3（26.8；32.5）
超导性	24.0 (23.1；25.3)	25.8（21.3；28.1）	18.1 (17.0; 19.7)	**18.0 (16.5；19.5)**	18.8（17.7；20.9）	18.8 (17.8；20.3)
wec	34.8 (33.5; 35.3)	23.0（21.4；24.6）	15.2（14.1；16.6）	14.9（13.3；16.0）	8.9 (7.9；10.3)	**8.8（8.3；9.8）**
红葡萄酒	30.2 (26.6; 32.9)	28.8 (24.9; 32.9)	26.7（23.3；30.8）	**26.4 (23.1；29.7)**	27.4（22.8；31.5）	27.2（22.0；31.1）
白葡萄酒	29.9（28.2；31.5）	30.3 (28.3; 31.7)	27.4（25.8；29.5）	**26.8（24.5；29.1）**	28.2 (25.5；29.0)	27.8（24.9；29.7）

仍远逊于最佳的损失分位数得分。在 Locus-Tuned 变体中，表现最佳者因数据集而异：BART（或其经 γ 膨胀的版本）在若干表格基准上表现最强，而采用 $\gamma(x)$ 膨胀的蒙特卡洛 Dropout 往往表现最佳或并列最佳。

表7展示了测试集划分中的经验接受率。我们注意到，70% 的接受率限制基本得到了遵守。

或者，也可以调整 α 参数，其结果与 Locus-Tuned相似。附录 D 描述了这一流程。指标 $P(X \in A)$ 和 $P(Z > \tau | X \in A)$ 分别见附录表8和表9。

4 最后总结

Locus 针对每个输入输出一个单一数值，即校准损失得分 $U_\alpha(x)$，其尺度与部署损失$Z = L(g(X)\ Y)$相同。 U_α 本身就很有参考价值，：它提供了一种可解释且与决策相关的风险信号，可用于排序、分诊和资源分配，而且通常无需预先 确定固定的接受/拒绝操作点。

根据构造， U_α 满足分布无关保证 $\mathbb{P}\big(Z \le U_\alpha(X)\big) \ge 1 - \alpha$；在基础损失模型满足温和条件时， $U_\alpha(x)$ 收敛于 $Z \mid X = x$的条件（$1 - \alpha$）分位数。

需要进行二值决策时，一个直观透明的默认做法是：当 $U_\alpha(x) \le \tau$时予以接受，并明确控制“可信但不良”事件： $\mathbb{P}\big(Z > \tau,\ U_\alpha(X) \le \tau\big) \le \alpha$ ，且近似满足 $\mathbb{P}\big(Z > \tau \mid X \in A_{\tau;\alpha}\big) \lesssim \alpha$ （实验结果也证实了这一点）。或者，也可以调整 α 或 λ ，以控制接受集内的错误。

我们的框架具有模块化特性：任何能够输出实际损失 $Z \mid X = x$ 的预测累积分布函数的模型，都可用于步骤 2（例如，归一化流 [Papamakarios等人， 2021], FlexCode [Izbicki和Lee， 2017], 或混合密度网络 [Bishop, 1994]或其他方法 [Izbicki和Lee，2016年，Dalmasso等人， 2020]）；而步骤 3 则通过共形化将其转化为有效的上界。可选的 $\gamma(x)$膨胀包络可提高数据稀缺区域的保守性，同时不改变校准步骤或有限样本保证。在计算方面，一旦拟合损失模型 $\sim F(\cdot \mid x)$并计算出校准PIT值，改变 α 只会影响标量阈值 $t_{1-\alpha}$；因此，只需极少的额外开销即可得到相应的函数族 $\{U_\alpha(x)\}_\alpha$ ，而在新的输入上评估 $U_\alpha(x)$ 只需进行一次CDF求逆。

尽管本文的介绍侧重于回归， Locus只要求已知一个标量损失。因此，只要任务由单一的数值误差度量来定义，它便可自然地扩展到分类及其他代价敏感问题。

致谢

M´ario de Castro 的研究部分由巴西 CNPq 资助（资助编号 301596/2025-5）。Rafael Izbicki 感谢 CNPq（422705/2021-7、305065/2023-8 和 403458/2025-0）及 FAPESP（资助编号 2023/07068-1）提供的经费支持。Thiago R. Ramos 感谢 CNPq（403458/2025-0）提供的经费支持。Denis Valle

获得了美国农业部国家食品与农业研究所 McIntire–Stennis 项目 1005163 以及美国国家科学基金会资助项目 2040819 的部分资助。

参考文献

AHRQ。《医疗支出小组调查》（MEPS）2019 年全年合并数据文件（HC-181）。https://meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-181，2019。访问日期：2026年2月。Anastasios N. Angelopoulos 和 Stephen Bates。共形预测：简明入门。机器学习基础与趋势，16(4):494–591，2023。Anastasios N. Angelopoulos、Stephen Bates、Adam Fisch、Lihua Lei 和 Tal Schuster。共形风险控制。载于国际学习表征会议 *(ICLR)*，2024。arXiv:2208.02814。Stephen Bates、Anastasios N. Angelopoulos、Lihua Lei、Jitendra Malik 和 Michael I. Jordan。无分布假设的风险控制预测集。*ACM*会刊，68(6)，2021。doi: 10.1145/3478535。文章 43。Christopher M. Bishop。混合密度网络。技术报告 NCRG/94/004，神经计算研究组，阿斯顿大学，英国伯明翰，1994年2月。URL https://publications.aston.ac.uk/id/ eprint/373/1/NCRG_94_004.pdf。此前以 NCRG/94/4288 发布。Leo Breiman。随机森林。机器学习，45(1):5–32，2001。Luben Cabezas、Guilherme P. Soares、Thiago R. Ramos、Rafael B. Stern 和 Rafael Izbicki。统计置信集的无分布假设校准。arXiv 预印本 arXiv:2411.19368，2024。Luben Cabezas、Mateus P. Otto、Rafael Izbicki 和 Rafael B. Stern。用于快速自适应预测区间的回归树。信息科学，686，2025a。文章 121369。Luben Cabezas、Vagner S Santos、Thiago R Ramos、Pedro LC Rodrigues 和 Rafael Izbicki。Cp4sbi：基于模拟的推断中可信集的局部共形校准。*arXiv* 预印本 *arXiv:2508.17077*，2025b。Luben Cabezas、Vagner Silva Santos、Thiago Ramos 和 Rafael Izbicki。共形分数中的认知不确定性：统一方法。载于 Silvia Chiappa 和 Sara Magliacane（编），第四十一届人工智能不确定性会议论文集，第 286 卷，机器学习研究论文集，第 443–470 页。PMLR，2025年7月21–25日。URL https://proceedings.mlr.press/v286/cruz-cabezas25a.html。Hugh A. Chipman、Edward I. George 和 Robert E. McCulloch。BART：贝叶斯加性回归树。应用统计年鉴，4(1):266–298，2010。doi: 10.1214/09-AOAS285。C. K. Chow。论最优识别错误率与拒绝权衡。*IEEE*信息论汇刊，16(1): 41–46，1970。doi: 10.1109/TIT.1970.1054406。Paulo Cortez、A. Cerdeira、F. Almeida、T. Matos 和 J. Reis。葡萄酒质量（红葡萄酒）。UCI机器学习资料库，2009。Niccol`o Dalmasso、Taylor Pospisil、Ann B Lee、Rafael Izbicki、Peter E Freeman 和 Alex I Malz。使用 Python 和 R 进行条件密度估计的工具及其在光度红移和无似然宇宙学推断中的应用。天文学与计算，30:100362，2020。L. Devroye 和 G. Lugosi。密度估计中的组合方法。施普林格统计学丛书。纽约施普林格，2001。ISBN 9780387951171。URL https://books.google.com.br/books?id=jvT-sUt1HZYC。Biprateep Dey、David Zhao、Brett H Andrews、Jeffrey A Newman、Rafael Izbicki 和 Ann B Lee。迈向实例级校准：条件密度的局部摊销诊断与重塑（LADAR）。机器学习：科学与技术，6(4):045058，2025。Dheeru Dua、Casey Graff 等。UCI机器学习资料库。http://archive.ics.uci.edu/ml，2017。加利福尼亚州尔湾：加利福尼亚大学信息与计算机科学学院。Tom Fawcett。ROC分析简介。模式识别快报，27(8):861–874，2006。Yarin Gal 和 Zoubin Ghahramani。随机失活作为贝叶斯近似：在深度学习中表示模型不确定性。载于第三十三届国际机器学习会议 *(ICML)* 论文集，2016年。

Yonatan Geifman 和 Ran El-Yaniv。深度神经网络的选择性分类。载于神经信息处理系统进展，第 30 卷，2017年。Yonatan Geifman 和 Ran El-Yaniv。Selectivenet：集成拒绝选项的深度神经网络。载于第三十六届国际机器学习会议论文集，第 97 卷，机器学习研究论文集，第 2151–2159 页。PMLR，2019年。Chuanxing Geng、Sheng-Jun Huang 和 Songcan Chen。开放集识别的最新进展：综述。*IEEE*模式分析与机器智能汇刊，43(10):3614–3631，2021年。Isaac Gibbs、Chenyi Cherian 和 Emmanuel J. Cand`es。具有条件保证的共形预测。arXiv 预印本 arXiv:2305.12616，2023年。Kam Hamidieh。超导数据。UCI机器学习资料库，2018年。Kilian Hendrickx、Lorenzo Perini、Dries Van der Plas、Wannes Meert 和 Jesse Davis。带拒绝选项的机器学习：综述。机器学习，113(5):3073–3110，2024年。Dan Hendrycks 和 Kevin Gimpel。用于检测神经网络中误分类样本和分布外样本的基线方法。载于国际学习表征会议 *(ICLR)*，2017年。arXiv:1610.02136。Eyke H¨ullermeier 和 Willem Waegeman。机器学习中的偶然不确定性与认知不确定性：概念与方法简介。机器学习，110(3):457–506，2021年。Rafael Izbicki。超越点预测的机器学习：不确定性量化。第 1 版，2025年。ISBN 978-65-01-20272-3。Rafael Izbicki 和 Ann B Lee。高维回归情境下的非参数条件密度估计。计算与图形统计学杂志，25(4):1297–1316，2016年。Rafael Izbicki 和 Ann B. Lee。将高维回归转化为高维条件密度估计。电子统计学杂志，11(2):2800–2831，2017年。doi: 10.1214/17-EJS1302。URL https://projecteuclid.org/journals/electronic-journal-of-statistics/volume-11/issue-2/Converting-high-dimensional-regression-to-high-dimensional-conditional-density-estimation/10.1214/17-EJS1302.full。Rafael Izbicki、Gilson Shimizu 和 Rafael Stern。使用密度估计器构建灵活的无分布假设条件预测带。载于国际人工智能与统计学会议，第 3068–3077 页。PMLR，2020年。Rafael Izbicki、Gilson Shimizu 和 Rafael B Stern。Cd-split 和 hpd-split：高维空间中的高效共形区域。机器学习研究杂志，23(87):1–32，2022年。Kaggle。美国金县房屋销售。https://www.kaggle.com/harlfoxem/housesalesprediction/metadata，2016年。访问日期：2019年8月。Alex Kendall 和 Yarin Gal。贝叶斯深度学习中需要哪些不确定性来处理计算机视觉？载于神经信息处理系统进展，第 30 卷，2017年。Balaji Lakshminarayanan、Alexander Pritzel 和 Charles Blundell。使用深度集成进行简单且可扩展的预测不确定性估计。载于神经信息处理系统进展，第 30 卷，2017年。Yann LeCun、Yoshua Bengio 和 Geoffrey Hinton。深度学习。《自然》，521(7553):436–444，2015年。Kimin Lee、Kibok Lee、Honglak Lee 和 Jinwoo Shin。用于检测分布外样本和对抗攻击的简单统一框架。载于神经信息处理系统进展 *(NeurIPS)*，2018年。Jing Lei 和 Larry Wasserman。非参数回归的无分布假设预测带。英国皇家统计学会杂志：*B*辑（统计方法学），76(1):71–96，2014年。Shiyu Liang、Yixuan Li 和 R. Srikant。提高神经网络中分布外图像检测的可靠性。载于国际学习表征会议 *(ICLR)*，2018年。Fei Tony Liu、Kai Ming Ting 和 Zhi-Hua Zhou。孤立森林。载于*2008*年 *IEEE*国际数据挖掘会议论文集，第 413–422 页。IEEE，2008年。doi: 10.1109/ICDM.2008.17。

Lu´ıs Marques 和 Dmitry Berenson。通过局部共形校准量化偶然动态不确定性与认知动态不确定性。arXiv 预印本 arXiv:2409.08249，2024年。另曾发表于 WAFR 2024。Raj Mehra。共享单车需求——均方根对数误差：0.3194。https://www.kaggle.com/code/rajmehra03/ bike-sharing-demand-rmsle-0-3194/input?select=train.csv，2023年。Kaggle。访问时间：2026年2月。Mehdi Neshat、Bradley Alexander、Nataliia Y. Sergiienko 和 Markus Wagner。采用多策略进化框架优化大型波浪能农场。载于遗传与进化计算大会（*GECCO*）论文集，第1150–1158页，2020年。Yaniv Ovadia、Emily Fertig、Jie Ren、Zachary Nado、David Sculley、Sebastian Nowozin、Joshua Dillon、Balaji Lakshminarayanan 和 Jasper Snoek。你能相信模型的不确定性吗？评估数据集偏移下的预测不确定性。神经信息处理系统进展，第32卷，2019年。George Papamakarios、Eric Nalisnick、Danilo Jimenez Rezende、Shakir Mohamed 和 Balaji Lakshminarayanan。用于概率建模与推断的归一化流。机器学习研究杂志，22：1–64，2021年。网址：https://jmlr.org/papers/v22/19-1028.html。Vladimir Vovk、Alex Gammerman 和 Glenn Shafer。随机世界中的算法学习。Springer，2005年。Meijiang Wang、Jingyu He 和 P. Richard Hahn。用于 BART模型的局部高斯过程外推及其在因果推断中的应用。arXiv 预印本 arXiv:2204.10963，2022年。第2版（2023年2月24日修订）。Jingkang Yang、Kaiyang Zhou、Yixuan Li 和 Ziwei Liu。广义分布外检测：综述。国际计算机视觉杂志，132：5635–5662，2024年。Ahmed Zaoui、Christophe Denis 和 Mohamed Hebiri。带拒绝选项的回归及其在 K 近邻中的应用。载于神经信息处理系统进展，2020 年。Xu-Yao Zhang 、 Guo-Sen Xie 、 Xiuli Li 、 Tao Mei 和 Cheng-Lin Liu。拒绝学习综述。*IEEE*会刊，111(2)：185–215，2023年。doi：10.1109/JPROC.2023.3238024。David Zhao、Niccol`o Dalmasso、Rafael Izbicki 和 Ann B Lee。条件密度模型与贝叶斯推断算法的诊断。载于人工智能中的不确定性，第1830–1840页。PMLR，2021年。

A 实验细节

A.1 数据集

我们使用保形预测文献中常见的13个标准回归数据集：翼型 [Dua 等人，2017], 自行车 [Mehra， 2023], 混凝土 [Dua 等人， 2017], 周期 [Dua 等人， 2017], 住宅 [Kaggle， 2016], 电力[Dua 等人， 2017], meps19 [AHRQ, 2019], 蛋白质 [Dua 等人， 2017], 星形 [Mehra， 2023], 超导性[Hamidieh， 2018], wec [Neshat 等人, 2020], 红葡萄酒 [Cortez 等人， 2009], 以及白葡萄酒 [Cortez 等人， 2009]。表3总结了样本数和特征数。

数据集	n	p
翼型	1503	5
Bike	10885	12
混凝土	1030	8
周期	9568	4
住宅	21613	17
电力	10000	12
meps19	15781	141
蛋白质	45730	8
Star	2161	48
超导性	21263	81
WEC	54000	49
红葡萄酒	4898	11
白葡萄酒	1599	11

表3：数据集摘要（规模和特征数量）。

A.2 划分与预处理

对于每个数据集，我们先随机打乱整个数据集，再将其划分为**40% 训练集、40% 校准集、10% 验证集**和**10% 测试集**。训练集仅 用于拟合部署预测器 g。校准集用于通过内部划分 D_1/D_2 （第 2.1 节）构建 $U_\alpha(\cdot)$。验证集用于调优后的变体（附录 3.2）。测试集仅用于最终报告。

我们使用在训练集划分上拟合的标准化器对特征和目标进行标准化，然后将相同的变换应用于校准集/验证集/测试集。所有报告的损失均在此标准化尺度上计算。

自行车特征工程。 对于自行车，我们根据时间戳构建时间派生协变量（小时、日、月、年），并删除原始的 dateti me、casual 和 registered 列，以与所提供的脚本保持一致。

A.3 部署预测器 g

我们在所有数据集上使用随机森林回归器作为固定的部署预测器 g ，其中 n_estimators 为 =300，其余采用默认设置。该预测器仅在训练划分上训练。

A.4 损失与阈值 τ

我们使用标准化响应上的绝对误差来评估实际损失： $Z = |g(X) - Y|$。每次运行将不可接受损失阈值 τ （代码中记为 T）定义为 g 下测试损失的经验 0.7 分位数。这一选择在各数据集之间产生可比的尾部事件（按构造大致为 $\mathbb{P}(Z > \tau) \approx 0.3$ ），且仅用于基准测试；在部署中， τ 将由领域要求设定。

A.5 评估指标

给定拟合上界 $U_\alpha(\cdot)$ 与接受阈值 λ，定义接受区域 $A_{\lambda;\alpha} = \{x : U_\alpha(x) \le \lambda\}$。我们报告：

- **接受率：** $p_A = \mathbb{P}\left(X \in A_{\lambda;\ \alpha}\right)$。
- **边际尾部率：** $p_{\text{big }Z} = \mathbb{P}(Z > \tau)$，在所有示例中固定为30%。
- **已接受样本中的条件尾部率：** $p_{\text{big }Z|A} = \mathbb{P}\left(Z > \tau \mid X \in A_{\lambda;\ \alpha}\right)$。
- **边际覆盖率检验：** $p_{\text{conf}} = \mathbb{P}\left(Z \le U_\alpha(X)\right)$（经验估计）。

A.6 重复运行次数与不确定性报告

对于每个数据集，我们使用不同的随机种子运行 30 次。表格报告各次运行的平均指标，并在括号中通过百分位区间总结各次运行的变异性。

B 实现细节：$\widetilde{F}(\cdot \mid x)$

本附录记录了在已发布代码中实例化预测损失累积分布函数 $F(\cdot \mid x)$ 时所采用的具体设置。

B.1 SharedLocus 构造

给定校准数据 (X_i, Y_i)，我们计算 $Z_i = |g(X_i) - Y_i|$，并将校准集随机划分为两个大小相等的子集，即 D_1 和 D_2。在 (X, Z) 上拟合概率损失模型，并利用 D_1 定义预测损失分布；随后使用 D_2 计算概率积分变换值和校准水平 $t_{1-\alpha}$，从而得到最终界限 $U_\alpha(x) = \widetilde{F}^{-1}(t_{1-\alpha} \mid x)$。

B.2 通过混合密度网络实现蒙特卡洛 Dropout

我们使用混合密度网络（MDN），将 $Z \mid X = x$ 建模为具有 $K = 5$的 K分量高斯混合。该 MDN 包含一个宽度为 64 的隐藏层，丢弃率为 0.4。为获得近似的后验预测分布，我们进行 500 次随机前向传播，采用蒙特卡洛 Dropout。对于每个 x，我们从所得混合分布中抽样，以估计分位数和概率积分变换值，从而实现 $\widetilde{F}^{-1}(\cdot \mid x)$ 和校准。

B.3 BART

我们使用通过 pymc-bart / pymc 实现的异方差BART模型。具体而言，我们使用两个 BART组件分别对（i）均值函数和（ii）对数尺度函数进行建模，每个组件包含 $m = 100$ 棵树，并根据这些潜在函数采用高斯似然对Z 进行建模。后验推断使用四条链，每条链进行 2000 次 MCMC抽样。使用 $Z \mid X = x$ 的后验预测样本计算分位数和概率积分变换值。

B.4 具有认知不确定性感知的 $\gamma(x)$膨胀

对于 BART 和 MC Dropout 模型，我们通过基于标准化特征空间中 D_1 个 $k = 50$ 最近邻计算得到的 k近邻密度代理来调节保守程度。随后，通过逻辑变换将所得的 kNN 半径映射到 $\gamma(x) \in (0,\ 1)$；在稀疏区域中，较小的 $\gamma(x)$ 会产生更保守的预测累积分布函数包络。

具体而言，对于每个测试输入 x，我们会在后验抽样中形成一组候选累积分布函数值，并取其下包络（一个较小的分位数），从而得到膨胀后的累积分布函数 $\widetilde{F}(\cdot \mid x)$。通过二分法进行数值求逆，以求得 $\widetilde{F}^{-1}$。

C VARNet 的实现细节

为了估计 $V(Y|X)$，我们拟合两个神经网络，每个网络都包含两个各有 64 个神经元的层和一个单神经元输出，分别用于预测 $E(Y^2 \mid X = x)$ 和 $E^2(Y \mid X = x)$。然后我们取

$$\widehat{V}(Y \mid X = x) = \widehat{E}(Y^2 \mid X = x) - \widehat{E}^2(Y \mid X = x),$$

作为 $V(Y|X = x)$ 的矩估计。实现中使用了 PyTorch 库。选择的优化损失函数是平均绝对误差 (MAE)。

D 使用 $\lambda = \tau$ (Locus-α) 进行 Alpha调优

本附录介绍一种简单的启发式方法，用于调优校准水平 α，以便在希望将操作阈值固定为 $\lambda = \tau$ 时使用，即当 $U_\alpha(x) \leq \tau$ 时接受。其目的是使接受点中的条件超越水平 η 达到期望值：

$$\mathbb{P}(Z > \tau \mid X \in A_{\tau;\alpha}) \approx \eta, \qquad A_{\tau;\alpha} = \{x : U_\alpha(x) \leq \tau\}.$$

设 $D_{\text{val}} = \{(X_i, Y_i)\}_{i=1}^N$ 是一个额外的带标签验证集，该集合独立同分布，且独立于用于构造 $U_\alpha(\cdot)$ 的校准数据。定义验证损失 $Z_i := L(g(X_i) Y_i)$。固定一个候选水平的有限网格 $\Gamma \subset (0, 1)$。对于每个 $\alpha \in \Gamma$，计算相应的界 $U_\alpha(\cdot)$ 和验证分数

$$u_i(\alpha) := U_\alpha(X_i), \qquad i = 1, \ldots, N,$$

然后定义接受索引及其数量

$$I_\alpha := \{i : u_i(\alpha) \leq \tau\}, \qquad n_\alpha := |I_\alpha|.$$

估计被接受的验证点中的条件超越率：

$$\widehat{q}(\alpha) := \frac{1}{n_\alpha \vee 1} \sum_{i \in I_\alpha} \mathbf{1}\{Z_i > \tau\}.$$

一种基本的选择规则是

$$\widehat{\alpha} \in \arg\min_{\alpha \in \Gamma} |\widehat{q}(\alpha) - \eta|.$$

与 λ 调优一样，通常适合施加最小接受约束，例如，仅限于 $\{\alpha \in \Gamma\colon \ n_\alpha/N \geq \rho_{\min}\}$。

~ ~**计算说明**。在 D_1 上拟合好 $F(\cdot \mid x)$ 后，就可以一次性计算 D_2 上的概率积分变换值 $W_i = F(Z_i \mid X_i)$；改变 α 只会改变经验分位数水平 $t_{1-\alpha}$，进而改变最终界限 $U_\alpha(x) = \widetilde{F}^{-1}(t_{1-\alpha} \mid x)$。因此，扫描 α 主要只需重复利用相同的概率积分变换值并重新计算阈值 $t_{1-\alpha}$。

保证。对于任意固定的（预先指定的） α， Locus-τ 在满足 $\lambda = \tau$ 时仍保有无分布联合尾部保证$\mathbb{P}(Z > \tau, X \in A_{\tau;\alpha}) \leq \alpha$ （定理 3）。上述数据驱动选择 $\widehat{\alpha}$ 是一种实用启发式方法，报告时也会明确说明这一点。

E Locus-Tuned 的分布无关替代方法

为补充 Locus-Tuned 中对 λ 的启发式调优，我们提供一种完全分布无关的方法，利用验证样本选择 λ。我们的目标是保证：以高概率（针对校准数据和验证数据的随机性），接受的预测中的条件超越率不超过目标 η：

$$\mathbb{P}\big(Z > \tau \mid X \in A_{\lambda;\alpha}\big) \leq \eta,$$

其中 $A_{\lambda;\alpha} := \{x : U_\alpha(x) \leq \lambda\}$. 我们通过扫描有限网格 Λ并利用针对该比率的同时上置信证书施加约束，来确定数据驱动的选择 $\lambda^\star$

$$q(\lambda) = \mathbb{P}(Z > \tau \mid X \in A_{\lambda;\alpha}) = \frac{H(\lambda)}{G(\lambda)},$$

其中

$$G(\lambda) := \mathbb{P}(X \in A_{\lambda;\alpha}), \qquad H(\lambda) := \mathbb{P}(Z > \tau,\ X \in A_{\lambda;\alpha}).$$

使用一致有限样本 偏差界来控制 Λ 上经验分子和分母，我们构造 t
一个经验上界证书 ，其形式为

$$\overline{q}(\lambda) := \frac{\widehat{H}_N(\lambda) + \varepsilon_H}{\widehat{G}_N(\lambda) - \varepsilon_G},$$

^并将 $\lambda^\star$ 定义为 Λ 中满足 $q(\lambda) \le \eta$ 的最大阈值（另外要求 $G_N(\lambda) > \varepsilon_G$，以确保分母为正）。如果不存在这样的阈值，则按照惯例令 $A_{\lambda^\star;\ \alpha} = \varnothing$。

下一个定理表明，这样选择 $\lambda^\star$ 可以得到一个分布无关的有限样本保证：以至少 $1-\delta$ 的概率，在 $(D,\ D_{\mathrm{val}})$ 下，接受点中的条件超越概率至多为 η。

定理 5。 固定 $\alpha \in (0,1)$，并令 U_α 由独立校准数据集 D构造而成。令 $D_{\mathrm{val}} = \{(X_i, Y_i)\}_{i=1}^N$ 为额外的独立同分布验证样本，且独立于 D，并记 $Z_i = L(g(X_i)\ Y_i)$ 和 $u_i = U_\alpha(X_i)$。固定 $\tau \in \mathbb{R}$ 和一个有限的候选阈值集 Λ。

对于 $\lambda \in \Lambda$，令

$$G(\lambda) := \mathbb{P}(X \in A_{\lambda;\alpha}) = \mathbb{P}(u \le \lambda), \qquad H(\lambda) := \mathbb{P}(Z > \tau,\ X \in A_{\lambda;\alpha}) = \mathbb{P}(Z > \tau,\ u \le \lambda),$$

并且，当 $G(\lambda) > 0$时，

$$q(\lambda) := \mathbb{P}(Z > \tau \mid X \in A_{\lambda;\alpha}) = \frac{H(\lambda)}{G(\lambda)}.$$

另令

$$\widehat{G}_N(\lambda) := \frac{1}{N}\sum_{i=1}^N \mathbf{1}\{u_i \le \lambda\}, \qquad \widehat{H}_N(\lambda) := \frac{1}{N}\sum_{i=1}^N \mathbf{1}\{u_i \le \lambda\}\mathbf{1}\{Z_i > \tau\}.$$

固定 $\delta \in (0,\ 1)$ 以及 $\eta \in (0,1)$，并设

$$\varepsilon_H := 2\sqrt{\frac{\log(2(N+1))}{N}} + \sqrt{\frac{\log(4/\delta)}{2N}}, \qquad \varepsilon_G := \sqrt{\frac{\log(4/\delta)}{2N}}.$$

^对于满足 $G_N(\lambda) > \varepsilon_G$的 $\lambda \in \Lambda$，令

$$\overline{q}(\lambda) := \frac{\widehat{H}_N(\lambda) + \varepsilon_H}{\widehat{G}_N(\lambda) - \varepsilon_G}.$$

Let

$$\mathcal{F} := \Big\{\lambda \in \Lambda :\ \widehat{G}_N(\lambda) > \varepsilon_G \ \textit{and}\ \overline{q}(\lambda) \le \eta\Big\},$$

并选择 $\lambda^\star := \max \mathcal{F}$，约定若 $\mathcal{F} = \varnothing$，则选择 $\lambda^\star$，使得 $A_{\lambda^\star;\alpha} = \varnothing$。

则在 (D, D_{val}) 的随机性下，以至少 $1-\delta$ 的概率，

$$\mathbb{P}\big(Z > \tau \mid X \in A_{\lambda^\star;\alpha}\big) \le \eta,$$

其中内部概率是相对于从部署分布中独立抽取的一次测试 (X, Y) 而言，并且我们约定：若 $\mathbb{P}(X \in A_{\lambda^\star;\alpha}) = 0$，则左端等于 0。

F 证明与附加定理

定理 *1* 的证明。 设 $n_2 := |I_2|$，并对满足 $Z = L(g(X)\ Y)$ 的独立测试对 (X, Y) 定义，

$$W_{n_2+1} := \widetilde{F}(Z \mid X), \qquad W_i := \widetilde{F}(Z_i \mid X_i)\ \ (i \in I_2).$$

在给定 D_1的条件下，变量 W_1、.....、W_{n_2}、W_{n_2+1} 是独立同分布的（因此具有可交换性）。令 $W_{(1)} \le \cdots \le W_{(n_2)}$ 表示 $\{W_i$ 的次序统计量：$i \in I_2\}$ 并令

$$k := \lceil (1-\alpha)(n_2+1) \rceil, \qquad t_{1-\alpha} := \begin{cases} W_{(k)}, & k \le n_2, \\ 1, & k = n_2+1, \end{cases}$$

这就是通常的分割保形经验（$1-\alpha$）分位数。根据可交换性，

$$\mathbb{P}\big(W_{n_2+1} \le t_{1-\alpha} \mid D_1\big) \ge \frac{k}{n_2+1} \ge 1-\alpha.$$

~由于对于每个 x，映射 $z \mapsto F(z \mid x)$ 都是非递减的，因此 $W_{n_2+1} \le t_{1-\alpha}$ 等价于

$$Z \le \widetilde{F}^{-1}(t_{1-\alpha} \mid X) = U_\alpha(X).$$

对 D_1 取期望可得 $\mathbb{P}(Z \le U_\alpha(X)) \ge 1-\alpha$。

现在，为证明定理的第二部分，设 W_{n_2+1} 在 $W_1, \dots, W_{n_2}, W_{n_2+1}$ 中的秩为 $R := 1 + \sum_{i \in I_2} \mathbf{1}\{W_i < W_{n_2+1}\}$. 根据可交换性和无并列假设，$R$ 在 $\{1, \dots, n_2+1\}$ 上服从均匀分布，因此

$$\mathbb{P}(R \le k \mid D_1) = \frac{k}{n_2+1}.$$

此外，当 $k \le n_2$ 时，我们有 $t_{1-\alpha} = W_{(k)}$，且无并列意味着

$$\{W_{n_2+1} \le t_{1-\alpha}\} = \{W_{n_2+1} \le W_{(k)}\} = \{R \le k\},$$

whil也就是说，如果 $k = n_2+1$，则 $t_{1-\alpha} = 1$，并且同一恒等式显然成立。因此，在所有情况下 ，

$$\mathbb{P}(W_{n_2+1} \le t_{1-\alpha} \mid D_1) = \frac{k}{n_2+1}.$$

~再次利用 $z \mapsto F(z \mid x)$ 的单调性，可得

$$\mathbb{P}\big(Z \le U_\alpha(X) \mid D_1\big) = \frac{k}{n_2+1}.$$

最后，由于 $k = \lceil (1-\alpha)(n_2+1) \rceil$，我们有 $k < (1-\alpha)(n_2+1) + 1$，因此

$$\mathbb{P}\big(Z \le U_\alpha(X) \mid D_1\big) = \frac{k}{n_2+1} < 1 - \alpha + \frac{1}{n_2+1}.$$

对 D_1 取期望，得到 $\mathbb{P}(Z \le U_\alpha(X)) < 1 - \alpha + \frac{1}{n_2+1}$，从而证明了所述上界。 □

定理 3 的证明。</style}在 $\{X \in A_{\tau;\alpha}\}$ 上，我们有 $U_\alpha(X) \le \tau$，因此

$$\{Z > \tau,\ X \in A_{\tau;\alpha}\} \subseteq \{Z > U_\alpha(X)\}.$$

根据定理 1，$\mathbb{P}(Z > U_\alpha(X)) \le \alpha$，从而证明 $\mathbb{P}(Z > \tau, \quad X \in A_{\tau;\alpha}) \le \alpha$。 □

假设 1 (渐近条件覆盖率的条件)。 **</style}考虑一个由总校准样本量 索引的序列，其中划分后的样本量为：** $= |D_1| \to \infty$ 和 $n_2 := |D_2| \to \infty$，当 $n \to \infty$ 时。令 $F_{n_1}(\cdot \mid x)$ </style}表示由 构造的预测累积分布函数（在正文中记为 $(\cdot \mid x)$ *</style}*，并显式标出其对 n_1 的依赖）。假设：

（*A1*）随机设计一致性（依概率）。对于一次独立的部署抽样（X、Z）（与 D_1独立），定义

$$\Delta_{n_1} := \sup_{z \in \mathbb{R}} \big| \widetilde{F}_{n_1}(z \mid X) - F_{Z|X}(z \mid X) \big|.$$

则，

$$\Delta_{n_1} \xrightarrow{\mathbb{P}} 0 \qquad (n_1 \to \infty),$$

其中，$\xrightarrow{\mathbb{P}}$ 表示关于 D_1 中的随机性以及独立抽样X的依概率收敛。

（*A2*）在目标 x 处的逐点一致性（依概率）。对于所考虑的固定 $x \in \mathcal{X}$，

$$\sup_{z \in \mathbb{R}} \big| \widetilde{F}_{n_1}(z \mid x) - F_{Z|X}(z \mid x) \big| \xrightarrow{\mathbb{P}} 0 \qquad (n_1 \to \infty),$$

其中，概率是针对 D_1 中的随机性而言的。

(A3) 真实条件分布的正则性及目标 x。*(i)* 对于 P_X-几乎所有的 x'，真实条件累积分布函数 $F_{Z|X}(\cdot \mid x')$ 都是连续的。*(ii)* 对于所考虑的固定 $x \in \mathcal{X}$， $F_{Z|X}(\cdot \mid x)$ 是连续的，且其$(1-\alpha)$分位数 $q_{1-\alpha}(x) := F^{-1}_{Z|X}(1-\alpha \mid x)$ 是唯一的。

定理 *2* 的证明。 按定理所述固定 $x \in \mathcal{X}$。我们分四步证明该结果 .

步骤 1：概率积分变换值渐近服从均匀分布，且其条件累积分布函数收敛于恒等函数。设 (X, Z) 为一次独立的部署抽样，且独立于 D_1。定义理想概率积分变换值

$$V := F_{Z|X}(Z \mid X).$$

在假设 1(A3)(i) 下，对 P_X-几乎所有的 x'，条件累积分布函数 $F_{Z|X}(\cdot \mid x')$ 是连续的，因此概率积分变换给出

$$V \sim \mathrm{Unif}(0,1).$$

定义估计的概率积分变换值

$$W := \widetilde{F}_{n_1}(Z \mid X).$$

根据假设 1(A1),

$$|W - V| = \left|\widetilde{F}_{n_1}(Z \mid X) - F_{Z|X}(Z \mid X)\right| \le \sup_{z \in \mathbb{R}} \left|\widetilde{F}_{n_1}(z \mid X) - F_{Z|X}(z \mid X)\right| = \Delta_{n_1} \xrightarrow{\mathbb{P}} 0.$$

因此 $W - V \xrightarrow{\mathbb{P}} 0$，且由于 $V \sim$ 服从 Unif(0,1)，斯卢茨基定理表明

$$W \xrightarrow{Dist} \mathrm{Unif}(0,1).$$

我们还需要对给定 D_1 时 W 的条件累积分布函数进行统一控制。令

$$H_{n_1}(w) := \mathbb{P}(W \le w \mid D_1), \qquad w \in [0,1].$$

固定 $\delta \in (0,\ 1)$。利用以下包含关系

$$\{V \le w - \delta\} \subseteq \{W \le w\} \cup \{|W - V| > \delta\}, \qquad \{W \le w\} \subseteq \{V \le w + \delta\} \cup \{|W - V| > \delta\},$$

我们得到，对所有 $w \in [0,1]$,

$$w - \delta - \mathbb{P}(|W - V| > \delta \mid D_1) \le H_{n_1}(w) \le w + \delta + \mathbb{P}(|W - V| > \delta \mid D_1).$$

因此,

$$\sup_{w \in [0,1]} |H_{n_1}(w) - w| \le \delta + \mathbb{P}(|W - V| > \delta \mid D_1). \tag{4}$$

剩下需要证明 $\mathbb{P}(|W - V| > \delta \mid D_1) \xrightarrow{\mathbb{P}} 0$。由于 $|W - V| \le \Delta_{n_1}$，我们有

$$\mathbb{P}(|W - V| > \delta \mid D_1) \le \mathbb{P}(\Delta_{n_1} > \delta \mid D_1).$$

此外，假设 1(A1) 意味着 $\mathbb{P}(\Delta_{n_1} > \delta) \to 0$。于是，对任意 $\varepsilon > 0$，由马尔可夫不等式可得

$$\mathbb{P}(\mathbb{P}(\Delta_{n_1} > \delta \mid D_1) > \varepsilon) \le \frac{\mathbb{E}\left[\mathbb{P}(\Delta_{n_1} > \delta \mid D_1)\right]}{\varepsilon} = \frac{\mathbb{P}(\Delta_{n_1} > \delta)}{\varepsilon} \to 0,$$

因此 $\mathbb{P}(\Delta_{n_1} > \delta \mid D_1) \xrightarrow{\mathbb{P}} 0$，从而 $\mathbb{P}(|W - V| > \delta \mid D_1) \xrightarrow{\mathbb{P}} 0$ 也成立。

将其代入 (4)，然后令 $\delta \downarrow 0$ ，得到

$$\sup_{w \in [0,1]} |H_{n_1}(w) - w| \xrightarrow{\mathbb{P}} 0. \tag{5}$$

步骤 2：经验PIT分位数收敛至 $1 - \alpha$。 在给定 D_1 的条件下， D_2 中的点是独立同分布的，且与 D_1 独立，因此 $\{W_{i,n}\}_{i \in I_2}$ 是具有累积分布函数 H_{n_1} 的独立同分布样本。根据格利文科 – 坎特利定理（在给定 D_1 的条件下应用），

$$\sup_{w \in [0,1]} |\widehat{H}_{n_2}(w) - H_{n_1}(w)| \xrightarrow{\text{a.s.}} 0 \qquad (n_2 \to \infty),$$

在给定 D_1 的条件下，关于 D_2 的随机性几乎必然成立。结合式 (5) 和三角不等式可得

$$\sup_{w\in[0,1]} |\widehat{H}_{n_2}(w) - w| \xrightarrow{\mathbb{P}} 0.$$

现在利用分位数泛函在均匀累积分布函数处的连续性。由于极限累积分布函数 $w \mapsto w$ 连续且严格递增，累积分布函数的一致收敛意味着广义逆收敛，因此

$$t_{1-\alpha,n} = \widehat{H}_{n_2}^{-1}(1-\alpha) \xrightarrow{\mathbb{P}} 1-\alpha.$$

步骤 3：$U_{\alpha,\ n}(x)$ 收敛到预言机分位数。 令 $q_{1-\alpha}(x) := F_{Z|X}^{-1}(1-\alpha \mid x)$，由假设 1(A3)可知其唯一。根据假设 1(A2)，

$$\sup_{z\in\mathbb{R}} \big|\widetilde{F}_{n_1}(z \mid x) - F_{Z|X}(z \mid x)\big| \xrightarrow{\mathbb{P}} 0,$$

并且步骤 2 给出 $t_{1-\alpha,\ n} \xrightarrow{\mathbb{P}} 1-\alpha$。根据步骤 2 中使用的相同分位数映射连续性论证（此处利用假设 1(A3)中目标分位数处的严格递增性和连续性），可得

$$U_{\alpha,n}(x) = \widetilde{F}_{n_1}^{-1}(t_{1-\alpha,n} \mid x) \xrightarrow{\mathbb{P}} F_{Z|X}^{-1}(1-\alpha \mid x) = q_{1-\alpha}(x).$$

步骤 4：条件覆盖率收敛。 由于 D 与测试点独立，在给定 D 和 $X = x$ 的条件下可得

$$\mathbb{P}\big(Z \le U_{\alpha,n}(x) \mid X = x,\ D\big) = F_{Z|X}\big(U_{\alpha,n}(x) \mid x\big).$$

根据步骤 3，$U_{\alpha,\ n}(x) \xrightarrow{\mathbb{P}} q_{1-\alpha}(x)$。由于 $u \mapsto F_{Z|X}(u \mid x)$ 在 $q_{1-\alpha}(x)$ 处连续（假设 1(A3)），由连续映射定理可得

$$F_{Z|X}\big(U_{\alpha,n}(x) \mid x\big) \xrightarrow{\mathbb{P}} F_{Z|X}\big(q_{1-\alpha}(x) \mid x\big) = 1-\alpha.$$

这就证明了定理中显示的第一个收敛关系。

□

最后，我们证明定理 5。

定理 *5* 的证明。令

$$G(\lambda) := \mathbb{P}(X \in A_{\lambda;\alpha}) = \mathbb{P}(u \le \lambda), \qquad H(\lambda) := \mathbb{P}(Z > \tau,\ X \in A_{\lambda;\alpha}) = \mathbb{P}(Z > \tau,\ u \le \lambda),$$

使得对于 $G(\lambda) > 0$,

$$q(\lambda) := \mathbb{P}(Z > \tau \mid X \in A_{\lambda;\alpha}) = \frac{H(\lambda)}{G(\lambda)}.$$

步骤 1：分子项的 VC 型一致偏差。定义

$$\Delta_H := \sup_{\lambda\in\Lambda} \big|\widehat{H}_N(\lambda) - H(\lambda)\big|.$$

由于改变一个观测值（$X_i,\ Z_i$）会使每个加数 $A_i^{\lambda}B_i \in \{0,1\}$ 至多变化 1，因此对于任意 λ，它会使 $\widehat{H}_N(\lambda)$ 至多变化 $1/N$，进而使 Δ_H 至多变化 $1/N$。根据 McDiarmid 不等式 [Devroye 和 Lugosi，2001，定理 2。2], 对于所有 $\varepsilon > 0$,

$$\mathbb{P}\left(\big|\Delta_H - \mathbb{E}\,\Delta_H\big| > \varepsilon\right) \le 2\exp(-2N\varepsilon^2).$$

为界定 $\mathbb{E}\,\Delta_H$，考虑集合类

$$\mathcal{C} := \Big\{(x,z) : U_\alpha(x) \le \lambda,\ z > \tau \ \Big|\ \lambda \in \Lambda\Big\}.$$

则 Δ_H =上确界 $_{C\in\mathcal{C}}\,|P_N(C) - P(C)|$，并根据极大不等式 [Devroye 和 Lugosi，2001，定理3.1],

$$\mathbb{E}\,\Delta_H \le 2\sqrt{\frac{\log\big(2\mathcal{S}(N)\big)}{N}},$$

其中 $\mathcal{S}(N)$ 是 $\mathcal{C}$ 的碎裂系数。固定任意样本 $\{(x_i,\ z_i)\}_{i=1}^N$，并记 $u_i := U_\alpha(x_i)$。由于 τ 是固定的，

$$\mathbf{1}\{(x_i, z_i) \in \mathcal{C}_\lambda\} = \mathbf{1}\{u_i \le \lambda\}\mathbf{1}\{z_i > \tau\}.$$

具有 $z_i \le \tau$ 的点始终标记为 0，而在具有 $z_i > \tau$ 的点中，标签由关于 $\{u_i\}$ 的单个阈值 λ 决定。因此，随着 λ 变化，只有当 λ越过至多 N 个值$\{u_i\}_{i=1}^N$之一时，标签才会改变，因此最多可能有 $N+1$ 种不同的标签标记方式。因此 $\mathcal{S}(N) \le N+1$，并且

$$\mathbb{E}\,\Delta_H \le 2\sqrt{\frac{\log\big(2(N+1)\big)}{N}}.$$

结合 McDiarmid 不等式并取 $\delta_H \in (0,\ 1)$，可得以至少 $1-\delta_H$ 的概率，

$$\sup_{\lambda\in\Lambda}\big|\widehat{H}_N(\lambda) - H(\lambda)\big| \le \varepsilon_H := 2\sqrt{\frac{\log\big(2(N+1)\big)}{N}} + \sqrt{\frac{\log(2/\delta_H)}{2N}}. \tag{6}$$

备注 1 (更精细的一致偏差界)。 利用更精细的经验过程技术（例如链式方法/熵界），可以得到阶为以下形式的 *DKW* 型一致偏差界

$$\sup_{\lambda\in\Lambda}\big|\widehat{H}_N(\lambda) - H(\lambda)\big| \ \lesssim\ \sqrt{\frac{\log(1/\delta_H)}{N}}$$

（相差一个普适常数因子），从而去除额外的 $\sqrt{\log N}$ 因子；例如，参见 *[Devroye 和 Lugosi，2001，第 3.2] 节*。为简洁起见，我们保留(6)。

步骤 2：分母的 DKW 一致偏差界。 定义 $\Delta_G := \sup_{\lambda\in\Lambda}|G_N(\lambda) - G(\lambda)|$。由于 $G_N(\lambda)$ 是标量变量 $u_i = U_\alpha(X_i)$ 的经验累积分布函数，Dvoretzky–Kiefer–Wolfowitz [Devroye 和 Lugosi，2001，定理 3.3] 不等式意味着，对于任意 $\delta_G \in (0,\ 1)$，以至少 $1-\delta_G$ 的概率，

$$\sup_{\lambda\in\Lambda}\big|\widehat{G}_N(\lambda) - G(\lambda)\big| \le \varepsilon_G := \sqrt{\frac{\log(2/\delta_G)}{2N}}. \tag{7}$$

步骤 3：一致比率界与选择 $\lambda^\star$。 令 $\mathcal{E}_H$ 表示式（6）中的事件，令 $\mathcal{E}_G$ 表示式（7）中的事件，即

$$\mathcal{E}_H := \Big\{\sup_{\lambda\in\Lambda}\big|\widehat{H}_N(\lambda) - H(\lambda)\big| \le \varepsilon_H\Big\}, \qquad \mathcal{E}_G := \Big\{\sup_{\lambda\in\Lambda}\big|\widehat{G}_N(\lambda) - G(\lambda)\big| \le \varepsilon_G\Big\}.$$

定义它们的交集

$$\mathcal{E} := \mathcal{E}_H \cap \mathcal{E}_G = \Big\{\sup_{\lambda\in\Lambda}\big|\widehat{H}_N(\lambda) - H(\lambda)\big| \le \varepsilon_H,\ \sup_{\lambda\in\Lambda}\big|\widehat{G}_N(\lambda) - G(\lambda)\big| \le \varepsilon_G\Big\}.$$

根据式（6）和式（7）以及并集界（取 $\delta_H = \delta_G = \delta/2$），

$$\mathbb{P}(\mathcal{E})\ \ge\ 1-\delta.$$

对于满足 $G_N(\lambda) > \varepsilon_G$ 的 $\lambda \in \Lambda$，定义经验比率

$$\overline{q}(\lambda) := \frac{\widehat{H}_N(\lambda) + \varepsilon_H}{\widehat{G}_N(\lambda) - \varepsilon_G}.$$

现在通过施加 $q(\lambda) \le \eta$ 来选择 λ。定义阈值的可行集

$$\mathcal{F} := \Big\{\lambda \in \Lambda :\ \widehat{G}_N(\lambda) > \varepsilon_G \ \text{ and } \ \overline{q}(\lambda) \le \eta\Big\},$$

并令 $\lambda^\star := \max\mathcal{F}$（约定若 $\mathcal{F} = \varnothing$，则 $A_{\lambda^\star;\alpha} = \varnothing$）。根据构造，每当 $\mathcal{F} \ne \varnothing$ 时，所选的 $\lambda^\star$ 都满足 $G_N(\lambda^\star) > \varepsilon_G$。

$\widehat{}$ $\widehat{}$在事件 $\mathcal{E} = \mathcal{E}_H \cap \mathcal{E}_G$ 上，对所有 $\lambda \in \Lambda$，我们都有 $H(\lambda) \le H_N(\lambda) + \varepsilon_H$，且 $G(\lambda) \ge G_N(\lambda) - \varepsilon_G$。将$\widehat{}$应用于 $\lambda = \lambda^\star$，并利用 $G_N(\lambda^\star) > \varepsilon_G$，可得

$$G(\lambda^\star) \;\ge\; \widehat{G}_N(\lambda^\star) - \varepsilon_G \;>\; 0,$$

因此 $q(\lambda^\star) = H(\lambda^\star)/G(\lambda^\star)$ 定义良好，而且

$$q(\lambda^\star) = \frac{H(\lambda^\star)}{G(\lambda^\star)} \le \frac{\widehat{H}_N(\lambda^\star) + \varepsilon_H}{\widehat{G}_N(\lambda^\star) - \varepsilon_G} = \overline{q}(\lambda^\star) \le \eta,$$

其中，最后一个不等式根据 $\lambda^\star \in \mathcal{F}$ 的定义成立。

最后，由于 $\mathcal{E} \subseteq \{q(\lambda^\star) \le \eta\}$，我们得出

$$\mathbb{P}\Big(q(\lambda^\star) \le \eta\Big) \;\ge\; 1 - \delta.$$

□

定理 6. 设 (X, Z) 为随机变量，且 $U_\alpha : \mathcal{X} \to \mathbb{R}$ 为可测函数。固定 $\tau \in \mathbb{R}$，并定义选择事件 $A = \{U_\alpha(X) < \tau\}$，并假设 $\mathbb{P}(A) > 0$。若 U_α 在水平 α 上条件有效，即

$$\mathbb{P}\big(Z > U_\alpha(X) \mid X\big) \le \alpha \quad a.s., \tag{8}$$

则 $\mathbb{P}\left(Z > \tau \mid A\right) \le \alpha.$

证明。 在事件 A 上，我们有 $U_\alpha(X) < \tau$，因此 $\{Z > \tau\} \subseteq \{Z > U_\alpha(X)\}$。所以，

$$\mathbb{P}(Z > \tau \mid X)\mathbf{1}_A \;\le\; \mathbb{P}(Z > U_\alpha(X) \mid X)\mathbf{1}_A \;\le\; \alpha \mathbf{1}_A,$$

其中，最后一个不等式用到了 (8)。对给定的 A 取条件期望，得到

$$\mathbb{P}(Z > \tau \mid A) = \mathbb{E}\big[\mathbb{P}(Z > \tau \mid X) \mid A\big] \le \alpha.$$

□

G 其他表格

这里我们展示 Locus 其余变体的模拟结果。

表4：**Locus 的边际覆盖率检验**。报告覆盖概率 $\mathbb{P}(Z < U_\alpha(x))$，其预期约为 $1-\alpha$。

数据集	**BART** (%)	**BART (γ)**(%)	**蒙特卡洛 Dropout**(%)	**蒙特卡洛 Dropout(γ)**(%)
翼型	89.4（80.4；93.1）	90.7（84.1；95.4）	88.7（81.1；93.4）	91.1 (80.9; 94.8)
bike	90.6（89.6；91.4）	91.3 (90.0; 91.9)	91.1 (90.5；92.2)	92.8（92.0；93.6）
混凝土	90.8（84.5；94.7）	91.3 (84.4; 96.1)	89.3（84.9；95.2）	91.7 (83.0；97.1)
周期	90.3（88.5；91.6）	90.9（89.4；92.1）	90.0 (88.5; 91.8)	91.1 (89.3; 92.3)
电力	90.3（88.8；92.0）	90.5（89.2；92.4）	90.0（88.1；91.3）	91.3 (89.8；92.7)
住宅	90.5 (89.0；91.3)	90.9 (90.0；91.9)	90.3 (89.2；91.1)	92.4 (91.1；93.5)
meps19	89.7 (88.4; 91.1)	92.0 (90.3; 93.0)	89.9 (88.2；91.3)	90.9 (89.8；92.8)
蛋白质	89.9（89.1；90.6）	90.6（89.9；91.6）	89.7（89.0；90.7）	91.6（90.8；92.2）
star	90.1（85.9；92.4）	91.0（86.4；95.0）	90.8（86.1；93.1）	90.3 (87.1; 94.1)
超导性	89.8 (88.3; 90.8)	90.6（88.8；91.5）	89.9 (87.8; 90.9)	91.6 (90.1；92.6)
wec	89.9 (89.3; 90.5)	90.8（90.0；91.2）	89.8 (89.4; 90.3)	92.3 (91.6；93.1)
红葡萄酒	90.3（86.5；95.1）	92.2（86.5；96.0）	90.0（84.9；95.6）	91.2 (87.2；96.0)
白葡萄酒	90.2 (88.2；92.8)	91.1 (89.5; 93.5)	90.3 (87.5；92.5)	91.4（89.2；93.9）

表 5：**（Locus）的接受率**。测试集划分上的接受概率，$\mathbb{P}(X \in A_\lambda)$ **(%)**。 Locus 使用分数 U_α，校准水平固定为 $\alpha = 10$**%**。条目为 30 次运行的中位数，括号中为第 5 和第 95 百分位数。

数据集	**BART** (%)	**BART (γ)**(%)	**蒙特卡洛 Dropout**(%)	**蒙特卡洛 Dropout(γ)**(%)
翼型	6.3 (0.0; 26.2)	**9.9（0.7；23.8）**	2.6 (0.0; 13.1)	1.0（0.0；23.2）
bike	**28.0（25.3；29.0）**	26.5（25.2；28.2）	15.5（12.5；19.2）	13.8（11.8；16.8）
混凝土	**0.5（0.0；5.8）**	0.0 (0.0；1.9)	0.0 (0.0; 0.0)	0.0（0.0；1.0）
周期	**0.0 (0.0；0.5)**	**0.0 (0.0；0.1)**	**0.0 (0.0；0.4)**	**0.0 (0.0；0.0)**
电力	1.8 (0.3；3.2)	**2.8 (1.0；4.4)**	0.7 (0.1；2.5)	0.3 (0.0；2.7)
住宅	15.8 (12.1；18.6)	**16.1 (12.7；18.2)**	14.0 (9.8; 18.0)	11.7（9.3；14.6）
meps19	4.7 (0.3；8.4)	3.9（0.3；19.0）	**33.1（25.0；39.1）**	31.7 (27.2; 37.5)
蛋白质	0.0 (0.0; 0.0)	0.0 (0.0; 0.0)	**11.2（9.4；12.7）**	9.1（7.4；11.5）
star	**0.0（0.0；1.4）**	**0.0（0.0；0.5）**	**0.0 (0.0；3.1)**	**0.0 (0.0; 0.5)**
超导性	28.3（25.0；32.7）	26.5（22.6；30.3）	**34.6 (31.3; 37.4)**	29.2（26.0；33.6）
wec	5.3 (3.0；9.8)	6.7 (2.9；9.3)	**52.9 (50.6; 54.6)**	49.8（47.1；51.6）
红葡萄酒	**0.0（0.0；5.7）**	**0.0 (0.0；4.2)**	**0.0 (0.0；3.9)**	**0.0（0.0；2.6）**
白葡萄酒	**0.0 (0.0；0.3)**	**0.0 (0.0；1.1)**	**0.0（0.0；0.4）**	**0.0 (0.0；0.2)**

表 6：**条件大损失率**。测试集划分中接受点的条件超越率，$\mathbb{P}(Z > \tau \mid X \in A_\lambda)$（%）。 Locus 使用分数 U_α，校准水平固定为 $\alpha = 10\%$。条目为 30 次运行的中位数，括号内为第 5 和第 95 百分位数；数值越低越好。粗体表示每行中的最低值。

数据集	**BART** (%)	**BART (γ)**(%)	**蒙特卡洛 Dropout**(%)	**蒙特卡洛 Dropout(γ)**(%)
翼型	**0.0（0.0；11.5）**	**0.0 (0.0；18.7)**	7.1（0.0；30.7）	6.9 (0.0; 15.2)
bike	1.6 (0.7；2.4)	**1.0（0.4；1.7）**	2.2（0.7；4.4）	1.9（0.7；4.1）
混凝土	**0.0（0.0；65.0）**	**0.0（0.0；50.0）**	100.0 (100.0；100.0)	**0.0 (0.0；90.0)**
周期	33.3 (2.5；86.7)	100.0 (10.0；100.0)	**0.0 (0.0；0.0)**	**– (–; –)**
电力	4.2 (0.0；20.5)	4.8 (0.0; 15.9)	16.7 (0.0; 37.3)	**0.0 (0.0; 31.2)**
住宅	5.0（3.5；7.0）	5.6（3.3；6.6）	4.9（2.9；6.9）	**4.0 (2.0；5.5)**
meps19	2.6 (0.0; 6.7)	**2.4（0.0；7.7）**	5.3 (4.0; 6.8)	4.5（2.9；6.0）
蛋白质	**0.0 (0.0; 0.0)**	**– (–; –)**	3.9（2.3；5.2）	2.7 (1.2；4.2)
star	**0.0（0.0；80.0）**	**0.0 (0.0; 0.0)**	**0.0（0.0；60.0）**	**0.0（0.0；100.0）**
超导性	2.5 (1.4；3.2)	**2.3（1.2；3.2）**	3.0 (2.0; 4.1)	**2.3 (1.0; 3.0)**
wec	4.3（1.7；8.1）	2.9（1.6；7.0）	2.3（2.0；3.1）	**2.1 (1.7; 2.7)**
红葡萄酒	10.0 (0.0; 80.0)	**0.0（0.0；77.8）**	**0.0 (0.0；69.4)**	**0.0 (0.0; 0.0)**
白葡萄酒	**0.0（0.0；28.3）**	**0.0 (0.0；14.6)**	**0.0 (0.0；68.1)**	**0.0 (0.0; 90.0)**

表7：**目标覆盖率匹配时的调优接受率（Locus-Tuned）**。接受概率，$\mathbb{P}(X \in A_\lambda)$（%）。对于每个数据集和方法，在验证集划分上选择阈值 λ，以达到 $\approx 70\%$ 的目标接受率；随后报告测试性能。 Locus-Tuned 方法采用 Locus 分数 U_α，校准水平固定为 $\alpha = 10\%$。条目为 30 次运行的中位数，括号内为第 5 和第 95 百分位数。

	基线方法（无 Locus）		**Locus 调优（损失分位数分数）**			
数据集	**IFlag**	**VARNet**	**Locus-BART**	**Locus-BART (γ)**	**Locus-MC**	**Locus-MC (γ)**
翼型	73.2（63.8；80.8）	**82.8（69.1；95.9）**	72.8（65.1；78.8）	71.2（61.5；78.9）	70.5（61.2；77.2）	71.5 (62.5；77.2)
bike	73.0（73.0；73.0）	**88.2（74.8；94.3）**	72.2 (70.9；74.6)	71.9（70.5；74.3）	73.3 (71.9; 75.1)	72.5（71.0；73.7）
混凝土	72.3 (66.0; 80.1)	**80.6 (73.8; 89.9)**	72.3 (61.6; 83.2)	72.3 (61.0; 85.9)	72.8 (64.5；86.1)	74.3（64.1；80.7）
周期	72.4 (67.9; 76.2)	**92.2（74.8；98.9）**	73.8（69.2；78.6）	73.1 (68.4；78.9)	75.0（70.8；81.9）	74.0（68.8；80.5）
电力	71.5 (67.5; 75.9)	**79.0 (71.5; 83.9)**	71.7 (68.6; 74.9)	71.5 (69.1；75.4)	71.0 (68.2；74.7)	71.4 (69.1; 75.1)
住宅	71.9 (69.6; 74.3)	**80.8（72.4；88.1）**	71.3（69.3；73.1）	71.3 (69.4；73.5)	71.3（69.2；73.7）	70.8 (68.5；73.1)
meps19	71.7（69.5；74.5）	**79.4（68.5；84.8）**	73.1 (69.3；76.1)	72.3（69.7；75.5）	71.3（68.8；73.4）	71.3 (68.8；74.0)
蛋白质	72.3 (70.0；74.5)	**73.1 (71.3；79.2)**	71.1 (69.0; 73.7)	71.1 (69.6；74.0)	71.2 (68.7；72.7)	71.2（69.9；73.2）
star	72.8 (63.9; 79.3)	73.0（65.2；80.5）	**76.3（67.0；84.4）**	75.1（66.4；83.5）	73.5 (66.4；77.3)	74.7（65.7；81.4）
超导性	71.5 (69.7; 74.0)	**82.4（71.6；88.8）**	70.9（69.3；73.6）	71.2 (69.2；72.5)	71.4 (69.2；75.0)	71.4 (68.6; 73.9)
wec	71.6 (69.7; 74.0)	**88.1 (85.2；89.7)**	71.2（69.6；72.8）	71.1（69.4；72.2）	70.6（69.0；72.5）	70.3（69.1；72.1）
红葡萄酒	70.9（62.1；77.6）	**73.4（65.0；80.6）**	70.9（63.1；79.4）	67.5（59.0；78.9）	73.1 (62.6；76.9)	71.6 (61.7; 81.2)
白葡萄酒	72.9（68.1；76.4）	72.7（66.2；79.4）	**80.2 (69.9; 85.2)**	77.1（68.6；81.9）	76.5 (69.7; 83.8)	76.3（69.8；82.6）

表8：**在目标覆盖率匹配时调整接受率（Locus-α）**。接受概率，$\mathbb{P}(X \in A_\lambda)$（%）。对于每个数据集和方法，在验证集划分上选择阈值 λ，以达到$\approx 70\%$的目标接受率；随后报告测试性能。 Locus-α 方法使用校准水平固定为 $\alpha = 10\%$的 Locus 分数 U_α。条目为30次运行的中位数，括号内为第5和第95百分位数。

	基线方法（无 Locus）		**Locus-α （损失分位数分数）**			
数据集	**IFlag**	**VARNet**	**Locus-BART**	**Locus-BART（γ）**	**Locus-MC**	**Locus-MC (γ)**
翼型	73.2（63.8；80.8）	**82.8（69.1；95.9）**	71.5 (64.5；78.1)	71.5（62.5；79.2）	70.5 (60.3; 77.5)	71.5（63.5；82.3）
bike	73.0（73.0；73.0）	**88.2（74.8；94.3）**	71.4（69.5；73.3）	71.7（70.3；73.6）	72.3 (70.9; 74.0)	70.9（69.2；72.8）
混凝土	72.3（66.0；80.1）	**80.6 (73.8; 89.9)**	71.8 (62.4; 83.7)	71.8（62.5；84.2）	73.8 (62.0; 83.6)	72.8（64.5；79.2）
周期	72.4 (67.9; 76.2)	**92.2 (74.8; 98.9)**	71.6 (68.5; 75.1)	71.6 (67.4; 75.8)	72.6（68.2；79.1）	75.8（68.8；85.3）
电力	71.5 (67.5; 75.9)	**79.0 (71.5; 83.9)**	71.6（68.2；74.2）	70.8（67.8；73.4）	71.0 (67.7; 74.0)	71.1（68.5；75.2）
住宅	71.9 (69.6; 74.3)	**80.8（72.4；88.1）**	71.6 (68.6; 72.9)	70.5 (68.8; 73.5)	70.9 (68.5; 73.5)	70.7 (68.9；72.8)
meps19	71.7（69.5；74.5）	**79.4（68.5；84.8）**	71.1 (67.2；72.3)	70.4 (68.5; 72.3)	70.0（68.0；72.0）	70.2（67.5；71.9）
蛋白质	72.3 (70.0；74.5)	**73.1 (71.3；79.2)**	**73.1（69.9；76.5）**	72.1 (70.5; 74.4)	71.2（69.5；72.7）	71.6 (69.2; 74.0)
star	72.8 (63.9; 79.3)	73.0（65.2；80.5）	**75.1 (66.6; 83.3)**	73.5 (64.2；80.6)	73.5 (66.7; 81.2)	72.6（67.3；79.6）
超导性	71.5 (69.7; 74.0)	**82.4（71.6；88.8）**	70.9 (68.9；72.8)	70.8 (68.6；72.8)	71.9（69.2；73.3）	71.1（69.3；73.3）
wec	71.6 (69.7; 74.0)	**88.1 (85.2；89.7)**	70.6 (69.2; 72.0)	70.1 (68.9; 72.0)	70.1 (68.5; 71.5)	70.3 (69.0；71.3)
红葡萄酒	70.9（62.1；77.6）	73.4 (65.0；80.6)	**74.4 (62.8; 80.1)**	71.9（61.3；80.7）	74.1 (60.8; 80.3)	71.9（62.8；81.7）
白葡萄酒	72.9（68.1；76.4）	72.7（66.2；79.4）	**75.1（67.1；78.0）**	73.7（67.4；77.9）	72.3（65.5；79.4）	73.0（67.6；77.2）

表9：**匹配接受率下调优后的条件大损失率（Locus-α）**。接受点中的条件超越， $\mathbb{P}(Z > \tau \mid X \in A_\lambda)$（%）。对于每个数据集和方法，均在验证集划分上选择阈值 λ，以达到 $\approx 70\%$ 的接受率；随后报告测试性能。 Locus-α 方法使用具有固定校准水平 $\alpha = 10\%$ 的 Locus 分数 U_α。条目为30次运行的中位数，括号内为第5和第95百分位数；数值越低越好。每行中的最低值以粗体标出。

	基线方法（无 Locus）		Locus-α （损失分位数分数）			
数据集	**IFlag**	**VARNet**	**Locus-BART**	**Locus-BART (γ)**	**Locus-MC**	**Locus-MC (γ)**
翼型	31.0（27.4；35.3）	27.5（23.0；30.7）	23.3（19.6；26.7）	**22.2（18.4；28.3）**	24.5 (19.2; 29.3)	24.7 (19.1; 27.4)
bike	30.1 (30.1; 30.1)	28.4 (26.1；29.2)	**19.0（18.3；20.8）**	19.4（18.3；20.6）	22.1 (21.3; 22.7)	21.8 (20.6；23.1)
混凝土	33.5 (27.2; 38.6)	27.7（23.0；31.2）	**23.7 (19.0; 28.0)**	25.2 (20.1; 29.1)	27.6（20.0；31.9）	27.4 (21.8; 32.9)
周期	28.9 (27.2; 30.0)	29.7 (28.5; 30.2)	**26.7 (25.2; 27.8)**	27.1 (25.5; 28.2)	28.4 (26.7；29.6)	28.9 (27.5；30.1)
电力	31.5（30.1；32.1）	29.4 (28.1；30.7)	**24.1 (22.2; 25.4)**	**24.1（22.4；25.8）**	24.7（22.9；26.0）	24.6（23.2；26.4）
住宅	34.8 (33.7；36.0)	26.6 (24.8; 27.7)	19.2（17.9；20.1）	19.3 (17.7; 20.6)	19.0 (17.6；20.2)	**18.9 (18.1；20.0)**
meps19	36.4 (35.4; 37.4)	24.8 (21.5；26.6)	14.0 (12.3; 15.0)	13.8（11.9；14.5）	**13.0 (11.7; 14.3)**	**13.0（11.9；14.4）**
蛋白质	28.0（27.4；29.0）	27.4 (25.8; 29.4)	24.6（23.5；25.2）	23.8（22.9；25.2）	22.8 (21.4；23.5)	**22.6 (21.9; 23.5)**
star	29.1 (26.3；31.9)	29.0（26.9；33.8）	29.1 (26.1; 31.9)	29.3 (26.1; 32.4)	**28.5（25.6；31.5）**	28.9 (26.5; 33.1)
超导性	24.0 (23.1；25.3)	25.8（21.3；28.1）	**17.9（17.0；19.0）**	18.1 (16.4；19.2)	18.3 (16.7；19.7)	**17.9（16.7；19.3）**
wec	34.8 (33.5；35.3)	23.0 (21.4; 24.6)	15.1 (12.2; 18.4)	14.6（12.9；18.7）	**8.2 (7.0；9.2)**	**8.2 (7.4; 8.9)**
红葡萄酒	30.2 (26.6; 32.9)	28.8 (24.9；32.9)	27.5（24.0；30.7）	**26.9（23.3；29.3）**	27.5（23.4；30.5）	27.5 (23.1; 31.5)
白葡萄酒	29.9（28.2；31.5）	30.3（28.3；31.7）	27.1（24.6；29.3）	**26.5 (24.3; 29.3)**	27.8（24.1；29.4）	27.4（25.0；29.2）