使用条件分位间区间的快速保形预测

Naixin Guo[1*]、**Rui Luo**[2*]、**Zhixin Zhou** [3*]

[1]香港城市大学决策分析与运营系，中国香港特别行政区 [2]香港城市大学系统工程系，中国香港特别行政区[3]Alpha Benito Research，美国洛杉矶 naixinguo2-c@my.cityu.edu.hk、ruiluo@cityu.edu.hk、zzhou@alphabenito.com

arXiv:2601.02769v1 [stat.ML] 6 Jan 2026

摘要

我们提出了保形分位间回归（CIR），这是一种保形回归方法，能够高效构造接近最小且具有保证覆盖率的预测区间。CIR 利用黑盒机器学习模型，通过分位间距估计结果分布，并将这些估计值转换为紧凑的预测区间，同时实现近似条件覆盖率。我们进一步提出 CIR+（具有更多比较的条件分位间回归），通过纳入基于宽度的分位间区间选择规则来增强 CIR。这一改进能够在保持相当覆盖率的同时，生成更窄的预测区间，但计算时间略有增加。这两种方法解决了现有分布保形预测方法的关键局限：与保形化分位数回归相比，它们能更有效地处理偏斜分布；通过消除直方图构建的需求，它们的计算效率也显著高于保形直方图回归。在合成数据集和真实世界数据集上进行的大量实验表明，与现有方法相比，我们的方法能够最优地平衡预测准确度与计算效率。

代码— https://github.com/orince/CIR

1 引言

保形预测是一种强大的框架，可用于构造具有有限样本有效性保证的预测区间。借助数据可交换性，保形方法能够将任意机器学习算法的输出转化为集合值预测，并在不对底层数据施加分布假设的情况下达到预期的覆盖水平。

现有的保形回归方法主要分为两类：直接预测区间端点（Romano、Patterson 和 Cand`es，2019年；Kivaranovic、Johnson 和 Leeb，2020年；Sesia 和 Cand`es，2020年；Gupta、Kuchibhotla 和 Ramdas，2022年），或对估计得到的完整条件分布求逆（Izbicki、Shimizu 和 Stern，2020a；Chernozhukov、W¨uthrich 和 Zhu，2021年）。尽管这些方法在许多情况下效果良好，但当数据呈偏态时，它们可能产生次优区间。保形化分位数回归（CQR）（Romano、Patterson 和 Cand`es，2019年）在条件分布呈偏态时可能会产生不平衡区间，而基于密度的方法虽能适应偏度，却涉及复杂的调参和解释。保形直方图回归（CHR）方法（Sesia 和 Romano（2021年））利用直方图近似$Y|X$ 的条件分布，并寻求具有目标覆盖率的最短区间。与 CQR 等其他方法（Romano、Patterson 和 Cand`es，2019年）相比，该方法能从估计的条件分布中提取更多信息。CHR 的一项主要优势在于，它能自动适应数据分布可能存在的偏度；这一点使其有别于那些倾向于生成对称区间、且固定上下未覆盖率的方法。对于偏度未知的数据，这些方法可能并非最优（Romano、Patterson 和 Cand`es，2019年；Izbicki、Shimizu 和 Stern，2020b；Chernozhukov、W¨uthrich 和 Zhu，2021年）。

然而，CHR 在计算效率方面存在潜在局限，尤其是将响应空间离散化为分箱、构造直方图并在其中搜索区间边界的过程会带来开销。为解决这些效率问题，我们提出了保形分位间回归（CIR），这是一种利用条件分位数结构直接确定区间的新方法。CIR 使用具有等概率分位数的多输出分位数回归模型，直接估计新响应落入各分位数间隔的条件概率，从而避免显式地对响应空间进行分箱。这种设计使 CIR 能够适应数据偏斜，同时提高计算效率。

对于每个校准样本，CIR 从最窄的分位数间隔开始，逐步纳入下一个最短的相邻区间，直到覆盖观测到的响应变量 Y。所需区间的数量定义为符合度分数。直观来说，落在较长分位数间隔中的样本得分较高，而落在较短区间中的样本得分较低。遵循 (Luo and Zhou 2025b; Romano, Sesia, and Candes 2020) 中的阈值化原则，我们根据这些得分选取一个数据依赖截断值，使校准响应变量中包含在相应合并区间内的比例达到目标覆盖水平。随后，将该截断值应用于测试输入，以构造

*通讯作者 版权所有 © 2026，美国人工智能促进协会 (www.aaai.org)。保留所有权利。

最终预测区间。

为进一步优化预测区间，我们还提出了条件分位间回归扩展方法 CIR+，并对其进行了更多比较。CIR+ 引入了额外的决策机制，根据特定分位数间区间的长度来判断是否保留或舍弃这些区间。具体而言，如果包含 Y 的分位数间区间长度低于校准过程中确定的阈值，则将其舍弃。此步骤在保持覆盖率表现相当的同时，使预测集宽度略有减小。我们的分析表明，通过对 CIR 和 CIR+ 中分位数间区间的数量进行阈值筛选所获得的预测区间，能够保证边际覆盖率。此外，这些方法渐近地能够达到所需的条件覆盖率，并使预期预测区间长度最小化。在合成数据集和真实世界数据集上的实验结果展示了所提方法的优越性。CIR 和 CIR+ 的表现可与最先进的 CHR 方法相媲美，同时显著降低了计算开销。这种统计效率与计算经济性的结合，使我们的方法成为条件预测区间领域的有力替代方案。

本文其余部分的结构如下。第2节讨论相关工作，为我们的贡献提供背景。第 3 节详细介绍所提出的 CIR 和 CIR+方法。随后，我们在第4节对我们的方法进行简要的理论分析。第5节通过数值实验，将我们的方法与现有的保形回归技术在模拟数据集和真实数据集上的表现进行比较。最后，我们在第 6 节总结全文。

2 相关工作

保形预测（CP）已成功应用于分类（Luo and Zhou 2025f; Luo and Colombo 2024；Luo and Zhou 2025d；Wang 等，2025；Zhang 等，2025）和回归任务（Luo and Zhou 2025b,e；Bao 等，2025a），展现了其在分割（Luo and Zhou 2025a）、博弈（Luo 等，2024；Bao 等，2025b）、时间序列预测（Su、Zhou 和 Luo，2024）以及基于图的应用（Luo、Nettasinghe 和 Krishnamurthy，2023；Tang 等，2025；Luo and Zhou 2025c；Wang、Zhou 和 Luo 2025；Luo and Colombo 2025）等多种真实世界场景中的灵活性。

我们的工作聚焦于保形回归，旨在生成紧凑、可解释且具有保证覆盖率的预测区间。虽然 (Izbicki, Shimizu, and Stern 2022) 和 (Luo and Zhou 2025b) (CTI) 将基于轮廓距离的相似性与多输出分位数回归相结合，以近似条件密度，但我们优先考虑基于区间的不确定性量化；正如 (Sesia and Romano 2021) 所强调的，这种方法具有更清晰的可解释性，并可避免不规则且过度自信的区域。CTI 对分位数间隔段采用全局长度阈值，往往会产生高度碎片化的预测集；而我们的 CIR 和 CIR+ 方法在维持有限样本边际覆盖率和条件覆盖率的同时，力求尽可能生成单个连续区间。由于 CTI 明确以牺牲覆盖率来换取更窄的集合，因此其行为从根本上仍与我们的方法不同，即使在单峰条件分布且分位数估计完全准确的情况下也是如此。(Gao 等，2025) 构造了不依赖条件分布的无条件标签空间水平集，其目标是基于密度的最高后验密度集，而非回归区间。然而，他们的方法需要在所有由校准样本诱导的断点上反复计算经验覆盖率，并对这些候选集进行离散全局优化，因此计算开销较大。

3 提出的方法

问题设定

考虑一个一般的回归问题，其数据集为$\{(x_i,\ y_i)\}_{i=1}^{n}$，其中 $x_i \in \mathcal{X} \subseteq \mathbb{R}^d$ 是输入特征向量，$y_i \in \mathcal{Y} \subseteq \mathbb{R}$ 是相应的连续响应变量。该数据集分为三部分：训练集 $\mathcal{D}_{\text{train}}$、校准集 $\mathcal{D}_{\text{cal}}$和测试集 $\mathcal{D}_{\text{test}}$。相应的索引集分别记为 $\mathcal{I}_{\text{train}}$，$\mathcal{I}_{\text{cal}}$ 和$\mathcal{I}_{\text{test}}$。我们假设这些集合中的样本是可交换的。我们的目标是为每个测试输入 X 找到一个预测区间$C(X) \subseteq \mathcal{Y}$，使得真实响应值 Y 以至少$1-\alpha$的概率包含在 $C(X)$ 中，其中 $\alpha \in (0,\ 1)$是目标显著性水平。具体而言，$C(X)$ 应保证 $\mathcal{D}_{\text{cal}}$ 和 $\mathcal{D}_{\text{test}}$中所有样本的有限样本边际覆盖率:

$$\mathbb{P}[Y \in C(X)] \geq 1-\alpha$$

对于 X 和 Y的联合分布。

在实现边际覆盖率的同时，我们还希望$C(X)$能够以 $1-\alpha$为水平近似实现条件覆盖率，其中 $\alpha \in$ （0,1）：

$$\mathbb{P}[Y \in C(x) | X = x] \geq 1-\alpha$$

这意味着在实践中，该方法应近似实现这一目标，并在适当条件下，随着样本量趋于无穷大而渐近实现这一目标。最后，预测区间应尽可能窄。

设想有一个预言机可以获取 $P_{Y|X}$，即给定 X时 Y的分布，并利用此类信息构造最优预测区间，具体如下。为简便起见，假设 $P_{Y|X}$ 相对于勒贝格测度具有连续密度 $f(y \mid x)$，不过通过采用更复杂的记号，也可以放宽这一假设。于是，针对 $Y \mid X = x$ 的预言机区间为:

$$C^{\alpha}_{\text{oracle}}(x) = [\ell_{1-\alpha}(x), \upsilon_{1-\alpha}(x)]\,, \tag{1}$$

其中，对于任意 $\tau \in$ （0, 1], $\ell_\tau(x)$和 $\upsilon_\tau(x)$定义如下:

$$\underset{(\ell,\upsilon)\in\mathbb{R}^2:\upsilon\geq\ell}{\arg\min} \left\{ |\upsilon - \ell| : \int_{\ell}^{\upsilon} f(y \mid x)dy \geq \tau \right\}. \tag{2}$$

这是具有条件覆盖率的最短区间。回想一下，CHR（ Sesia 和 Romano（2021））将公式 (2) 中的 f 替换为直方图近似，但计算开销很大。具体而言，如果我们根据分位数将 Y 的定义域划分为 T 个相等部分，并得到相应的分位数 $q_0(x) \ldots, q_T(x)$，其中 $q_t(x)$ 是

算法 1：条件分位间回归（CIR）

输入：带标签数据 $\{(x_i, y_i)\}_{i\in\mathcal{I}}$、数据划分比例、无标签测试数据 $\{x_i\}_{i\in\mathcal{I}_{\text{test}}}$、黑盒学习算法 $\mathcal{B}$、分位数间区间的数量 T、水平 $\alpha \in (0, 1)$。

1：随机划分索引 $\mathcal{I}$ 为 $\mathcal{I}_{\text{train}}$ 和 $\mathcal{I}_{\text{cal}}$。2：使用 $\mathcal{I}_{\text{train}}$中的样本训练 $\mathcal{B}$，并获得针对 $t = 0$、1、. . .、T的分位数估计函数 $\widehat{q}_t$。3：对于每个 $i \in \mathcal{I}_{\text{cal}} \cup \mathcal{I}_{\text{test}}$，计算 $\widehat{q}_t(x_i)$，其中 t =取0、1、. . .、T。4：对于任意给定的 k 和 $i \in \mathcal{I}_{\text{cal}}$，将 l_k 定义为包含 k 个连续分位数间区间的最短区间的下端点，并将相应区间定义为 $C_k^\alpha(x_i) = (\widehat{q}_{l_k}(x_i)\ \widehat{q}_{l_k+k}(x_i)]$。5：对于 $i \in \mathcal{I}_{\text{cal}}$，令 k_i 为满足 $y_i \in C_k^\alpha(x_i)$ 的最小 k （等价于公式 9 中的 $s(x_i, y_i)$ 。6：将 k 计算为 r_α 最小的 k_i，其中 $r_\alpha = \lceil(1-\alpha)(1+|\mathcal{I}_{\text{cal}}|)\rceil$ 。

输出：预测区间 $C_{\widehat{k}}^\alpha(x_i)$。

$Y|X = x$ 的条件分布的 t/T 分位数，当 T 足够大时，根据以下表达式得出的区间近似于公式（2）中的区间：

$$\underset{\substack{l=0,\ldots,T-1\\ l<u\le T}}{\arg\min}\Big\{|q_u(x)-q_l(x)|: \sum_{t=l}^{u-1}\mathbb{P}\big[Y\in(q_t(x),q_{t+1}(x))\big]\ge\tau\Big\}. \tag{3}$$

它对应的区间为：

$$C^\alpha(x) = (q_{l_\tau}(x),\ q_{u_\tau}(x)]\,. \tag{4}$$

这种针对连续基于分位数的分箱进行的优化，启发了我们的分位间区间方法，该方法避免了显式的响应空间分箱。

条件分位间回归

我们的方法直接从最小的分位数间隔开始，选择连续的分位数间隔，因此得名保形分位数回归（*CIR*）。具体而言，我们在训练集 $\mathcal{D}_{\text{train}}$ 上应用分位数回归，以预测每个 $x \in \mathcal{X}$对应的条件分布 $Y|X = x$ 的第t/T分位数，其中 t 的取值从 0 到 T，步长为 1。t/T的估计分位数记为 $\widehat{q}_t(x)$，分位数间隔随后定义为：

$$I_t(x) = (\widehat{q}_{t-1}(x),\ \widehat{q}_t(x)] \quad \text{for} \quad t = 1,\ldots,T. \tag{5}$$

然后，公式 3 可以近似写为：

$$\underset{\substack{l=0,\ldots,T-1,\\ 1\le k\le T-l-1}}{\arg\min}\left\{\sum_{t=l}^{l+k}|I_{t+1}(x)| : \sum_{t=l}^{l+k}\mathbb{P}[Y\in I_{t+1}(x)]\ge\tau\right\}, \tag{6}$$

其中，l 表示区间的起始索引，k 表示要纳入的连续分位数间隔数量，二者均由 l 确定。

相应的区间可定义为

$$C_{k(\tau)}^\alpha(x) = \left(\widehat{q}_{l(\tau)}(x),\ \widehat{q}_{l(\tau)+k(\tau)}(x)\right],$$

具体而言，当 $P_{Y|X}$ 为单峰时，我们估计的区间可由以下公式得出

$$C_{\widehat{k}}^\alpha(x) = \left(\widehat{q}_{l_{\widehat{k}}}(x),\ \widehat{q}_{l_{\widehat{k}}+\widehat{k}}(x)\right], \tag{7}$$

其中

$$\widehat{k} := k(\widehat{\tau});\ l_{\widehat{k}} := l(\widehat{\tau}) = \underset{l=0,\ldots,T-1}{\arg\min}\left\{\sum_{t=l}^{l+\widehat{k}}|I_{t+1}(x)|\right\}. \tag{8}$$

请注意，$l_{\widehat{k}}$ 表示下端点索引，其中$\widehat{\ }$下标 k 体现了单峰性质：所包含的分位数间区间数量唯一决定了最窄区间及其对应的下端点。$\widehat{\tau}$ 将通过在留出数据上计算适当的符合度分数来确定；如果 f 的模型不够准确，为确保满足这一条件，该值通常会大于真实的目标覆盖率 τ。然而，如果分位数估计准确，确保每个区间覆盖真实标签Y的概率均约为 $1/T$，并且分位数间区间的数量趋于无穷大，那么公式 (7) 将与公式 (1) 中预言机的结果非常接近。

$\widehat{\ }$决定我们预测区间的关键量是 k，即要纳入的分位数间区间数量。为确-$\widehat{\ }$定 k，我们首先引入一个一致性分数函数：

$$s(x,y) = \min\Big\{k\in 1,\ldots,T : y\in C_k^\alpha(x)\Big\}. \tag{9}$$

此函数计算包含 y 所需的最小数量的分位间区间。该实现遵循迭代过程。从最短的分位间区间开始，对于每个 $i \in \mathcal{I}_{\text{cal}}$，如果 y_i 落在此区间之外，我们便将搜索范围扩展到相邻的分位间区间（在任一侧扩展，边界处则仅向一侧扩展）。随后，我们选择较窄的相邻区间，并将其与原始区间合并。此过程持续进行，直到 y_i 落入扩展后的区间。最终区间中分位间区间的总数即为 k_i。

为提高计算效率，我们将符合度分数实现为 $s(x_i, y_i) = k_i$，适用于 $i \in \mathcal{I}_{\text{cal}}$。随后，我们定义$r_\alpha = \lceil(1-\alpha)(1+|\mathcal{I}_{\text{cal}}|)\rceil$ 并选择

$$\widehat{k} \leftarrow \text{the } r_\alpha\text{-th smallest } k_i, \tag{10}$$

确保 $y_i \in C_{\widehat{k}}^\alpha(x_i)$ 对校准集中的至少 r_α 个实例成立。最后，我们在公式 (7) 中使用此 k，为所有 $x_i, i \in \mathcal{I}_{\text{test}}$ 获取预测集。具体步骤见算法 1。

具有更多比较的条件分位间回归（CIR+）

在 CIR 中，符合度分数定义为包含响应 y 所需的最小分位数间区间数量。接着，我们选取其中第 r_α 小的数值

算法 2：采用更多比较的 CIR（CIR+） 输入：带标签数据 $\{(x_i, y_i)\}_{i\in\mathcal{I}}$、无标签测试数据 $\{x_i\}_{i\in\mathcal{I}_{\text{test}}}$、数据划分比例、黑盒学习算法 $\mathcal{B}$、分位数间区间数量 T。 1：随机划分索引 $\mathcal{I}$，将其分为 $\mathcal{I}_{\text{train}}$ 和 $\mathcal{I}_{\text{cal}}$ 2：使用 $\mathcal{I}_{\text{train}}$ 中的样本训练 $\mathcal{B}$，并为 $t = 0$、1、.....、T 计算分位数估计函数 $\widehat{q}_t$。 3：对于每个 $i \in \mathcal{I}_{\text{cal}} \cup \mathcal{I}_{\text{test}}$，计算 $\widehat{q}_t(x_i)$，其中 $t =$ 为 0、1、.....、T。 4：对于任意给定的 k 和 $i \in \mathcal{I}_{\text{cal}}$，将 l_k 定义为包含 k 个连续分位数间区间的最短区间的下端点，并将相应区间定义为 $C_k^{\alpha}(x_i) = (\widehat{q}_{l_k}(x_i)\ \widehat{q}_{l_k+k}(x_i)]$。 5：定义 $s(x_i, y_i) = k_i - 1 + e_{k_i}(x_i)$，其中 k_i 是满足以下条件的最小 k：对于 $i \in \mathcal{I}_{\text{cal}}$， $y_i \in C_k^{\alpha}(x_i)$，且 $e_{k_i}(x_i) \in (0, 1)$ 是 x_i 的第 k_i 个分位数间区间的重缩放长度，如公式（12）所定义。 6：定义 $r_\alpha = \lceil (1-\alpha)(1+|\mathcal{I}_{\text{cal}}|)\rceil$，然后将阈值 $\widehat{s}$ 定义为对于 $i \in \mathcal{I}_{\text{cal}}$， $s(x_i, y_i)$ 中第 r_α 小的值。 7：对于 $i \in \mathcal{I}_{\text{test}}$， $k_i = \lfloor \widehat{s} \rfloor + \mathbf{1}\{e_{\lfloor \widehat{s} \rfloor}(x_i) > \widehat{s} - \lfloor \widehat{s} \rfloor\}$。 输出： C^{α}

k_i（x_i）对于 y_i，其中 $i \in \mathcal{I}_{\text{test}}$。

校准集作为预测区间中分位数间区间最小数量的阈值。然而，当校准集中多个区间在第 r_α 位具有相同宽度时，仅使用区间数量可能会降低准确性。例如，假设在 CIR 校准过程中有 5 个样本的排名为 r_α，则所选的 k 将覆盖从 r_α 到 $r_\alpha + 4$ 的排名。为了获得更精确的排名，我们引入额外信息，并将 CIR 修改为具有更多比较的条件分位回归（CIR+）。符合度分数定义如下：

$$s(x, y) = \min\Big\{k - 1 + e_k(x) : y \in C_k^{\alpha}(x)\Big\}, \tag{11}$$

其中

$$e_k(x) = \frac{1}{c}\min\{\widehat{q}_{l_{k-1}}(x) - \widehat{q}_{l_{k-1}-1}(x), \widehat{q}_{l_{k-1}+k}(x) - \widehat{q}_{l_{k-1}+k-1}(x)\}, \tag{12}$$

并且， c 可以是任意使 $e_k(x) \in (0, 1)$ 的固定常数。换句话说，当我们从最短的分位数间隔开始扩展搜索时，我们细化了递增 k 的过程。与其在找到一个包含 y 的分位数间区间时简单地将 $k-1$ 增加到 k，不如加上新纳入区间的重缩放长度——具体而言，即相邻分位数间区间中较短的那个。通常，由于不同样本的预测分位数不同，分位数间区间的长度也往往不同。这使我们能够准确地从这些符合度分数中选出第 r_α 小的分数。在校准步骤中，我们仍然首先按照 CIR 找到 k_i。然后，我们只需关注包含 y_i 且对应于第 r_α 小位置并列排名的分位数间区间的长度。接着，我们根据公式（11）对这些并列分数重新排名，并将阈值定义为：

$$\widehat{s} \leftarrow \text{the } r_\alpha\text{-th smallest } k_i - 1 + e_{k_i}(x_i). \tag{13}$$

在预测阶段，我们为测试集中的每个样本识别 $C^{\alpha}_{\lfloor \widehat{s} \rfloor}(x_i)$。随后，根据长度确定是否纳入第 $\lfloor \widehat{s} \rfloor$ 个合并后的分位数间区间。具体而言，我们将缩放后的长度与 $\widehat{s} - \lfloor \widehat{s} \rfloor$ 进行比较。若更大，则保留该部分，预测集变为 $C^{\alpha}_{\lceil \widehat{s} \rceil}(x_i)$。若更小，则将其舍弃，集合变为 $C^{\alpha}_{\lfloor \widehat{s} \rfloor}(x_i)$。该过程如算法 2 所示。

4 理论结果

本节主要介绍 CIR 的理论结果，并在附录中提供更多细节，包括对 CIR+ 的分析。首先，我们证明，若假设分位数回归模型足够准确，那么在考虑边际分布时，CIR 有望实现预测区间的最优规模。CIR 的边际覆盖概率证明遵循标准保形预测一般框架中所使用的常规论证思路。

定理 4.1。 *(有限样本)* 若 (X_i, Y_i) 对于 $i \in \mathcal{I}_{cal} \cup \mathcal{I}_{test}$ 可交换，则对于 $i \in \mathcal{I}_{test}$，算法 *1* 的输出满足：

$$\mathbb{P}\left[Y_i \in C^{\alpha}_{\widehat{k}}(X_i)\right] \geq 1 - \alpha. \tag{14}$$

证明。 注意，评分函数 $s(X_i, Y_i) = \min\Big\{k \in \mathbf{1}, \ldots, T : Y_i \in C_k^{\alpha}(X_i)\Big\}$ 对于 $i \in \mathcal{I}_{\text{cal}} \cup \mathcal{I}_{\text{test}}$ 也可交换。对于测试集中的任意 (X_i, Y_i)， $s(X_i, Y_i)$ 的秩小于 k，而算法 1 中这一情况发生的概率为 $\frac{\lceil (1+|\mathcal{I}_{\text{cal}}|)(1-\alpha) \rceil}{1+|\mathcal{I}_{\text{cal}}|} \geq 1 - \alpha$。 □

这一理论结果对 CIR+ （算法 2）也成立，详细讨论见附录。如第 3 节所述，如果分区 T 足够大，则式 (3)–(4) 的输出近似于式 (1)–(2)。以下结果表明，如果分位数估计准确且样本独立同分布，则随着样本量 $|\mathcal{I}| \to \infty$ 增大（也表示 $|\mathcal{I}_{cal}| \to \infty$），由 CIR（算法 1）计算出的 k 渐近等于最优值。接下来的理论依赖于以下假设：

1.（独立同分布）样本独立同分布，这比可交换性更强。

2.（一致性）当黑盒模型能够一致地估计 $P_{Y|X}$ 时，每个分位数间区间都应以概率 $1/T$ 包含真实标签 Y，而包含 Y 的最优分位数间区间数应为 $\lceil T(1-\alpha) \rceil$，以达到校准集中真实标签的 $(1-\alpha)$ 覆盖率。该假设至关重要，与（Romano、Patterson 和 Cand`es，2019年；Sesia 和 Romano（2021年））中的假设类似。尽管在实践中难以满足，但已有理论结果证明某些模型符合该假设，例如（Meinshausen，2006年）中的模型。

3.（单峰性）真实的 $P_{Y|X}$ 是单峰的。

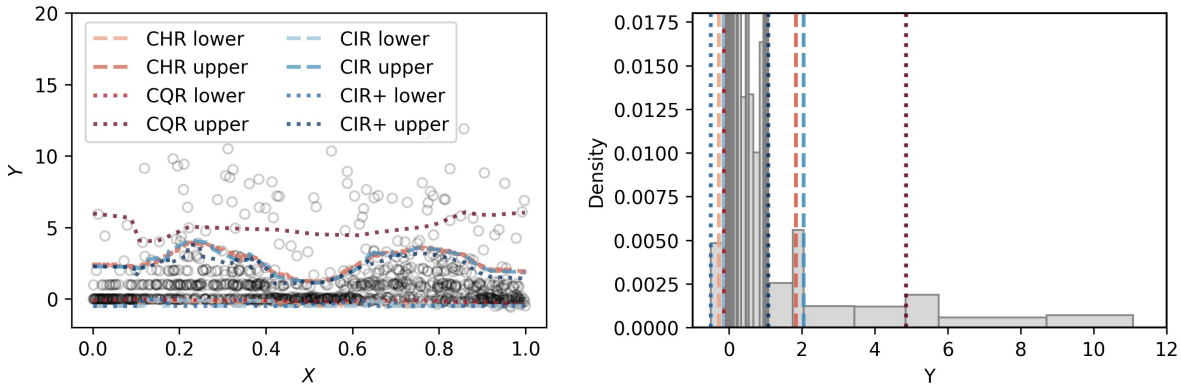

图1：在单变量示例中比较 CIR、CIR+、CQR（Romano、Patterson 和 Cand`es，2019年）与 CHR（Sesia 和 Romano，2021年）。所有方法均使用相同的深度分位数模型，并保证 90%的边际覆盖率。左：预测带随 X 变化的情况。所有方法的经验边际覆盖率和估计条件覆盖率均为 0.9（CQR 的条件覆盖率除外：0.7）。平均长度：CIR（3.0）、CIR+（2.7）、CHR（3.0）、CQR（5.2）。右：在 $X \approx 0.4$ 处的分位数间区间，使用 50 个分位数，并截断高密度区域。区间越窄，表示密度越高。CIR 和 CIR+ 选择了更窄的区间；CIR+ 选择的区间少一个，因为其缩放后的长度小于 $\widehat{s} - s$。

定理 4.2。 （无限样本）在假设 *1–3* 下，对于 $(X, Y) \in \mathcal{D}^{test}$，通过算法 *1* 获得的 $\widehat{k}$ 具有以下性质：

$$\mathbb{P}\left[|\widehat{k} - \lceil T(1-\alpha)\rceil| > \epsilon_n\right] \leq \eta_n, \tag{15}$$

$$\mathbb{P}\left[\mathbb{P}\left[Y \in C^{\alpha}_{\widehat{k}}(X) \mid X\right] \geq 1 - \alpha - \gamma_n\right] \geq 1 - \zeta_n \tag{16}$$

其中我们假设校准样本量为 $|\mathcal{I}_{cal}| = n \to \infty$，且 $\epsilon_n \to 0$、$\eta_n \to 0$、$\gamma_n \to 0$、$\zeta_n \to 0$。

直观而言，当 T 足够大时，算法 2 中额外的选择步骤会变得可以忽略，因此从理论上说，CIR+ 会收敛到 CIR。证明细节见附录。

备注： 单峰性设定对于实现我们的方法或在定理 4.1 中实现边际覆盖率并非必要，只有定理 4.2 中关于最优区间长度和条件覆盖率的性质需要该设定。在非单峰情形下，最优预测集可能由多个互不相连的区间组成，而非单个连通区间（单峰性设定可保证后者）。为此，更有效的方法是按长度对分位间区间排序，并从这一有序集合中进行选择。尽管如此，后续章节和附录中的实证分析表明，即使面对非单峰分布，两种方法仍能保持稳健的性能，体现了它们在实际应用中的灵活性。

5 实验

我们的实验实现基于 Sesia 和 Romano（2021）的公开可用代码 Conformal HistogramRegression（CHR），该代码可访问：https://github.com/msesia/chr。我们对他们的工作深表感谢，它为我们的研究提供了重要的灵感。我们的数值实验在计算集群中的 Intel Xeon 2.10GHz CPU 上进行，并且每个数据集均使用单核进行分析。

$\lvert\mathcal{I}\rvert$		CHR	CQR	CIR	CIR+
500	边际	0.902 (0.001)	0.901 (0.001)	0.900 (0.002)	0.897 (0.002)
	条件	0.881 (0.003)	0.879 (0.004)	0.878 (0.004)	0.871 (0.004)
	宽度	3.670 (0.039)	5.165（0.043）	3.930（0.057）	3.790（0.053）
	Time	36.597 (0.396)	0.008 (0.000)	0.072 (0.001)	0.097 (0.001)
3000	边际	0.901 (0.001)	0.900（0.001）	0.905（0.001）	0.900（0.001）
	条件	0.887 (0.003)	0.875 (0.003)	0.895 (0.003)	0.886 (0.003)
	宽度	3.425 (0.017)	5.320 (0.019)	3.764 (0.028)	3.632 (0.028)
	Time	73.735 (0.785)	0.014 (0.000)	0.404 (0.005)	0.526 (0.007)
5000	边际	0.901 (0.001)	0.900 (0.001)	0.905 (0.001)	0.900 (0.001)
	条件	0.889 (0.003)	0.875（0.003）	0.895（0.003）	0.889（0.003）
	宽度	3.389 (0.013)	5.345（0.016）	3.776（0.022）	3.643（0.022）
	Time	103.637 (1.105)	0.019 (0.000)	0.663 (0.008)	0.862 (0.010)

表1：我们的方法（CIR 和 CIR+）与基准方法在不同样本量（$|\mathcal{I}|$）的合成数据上的性能比较，偏度为 2.7。相应的标准差列于括号中。

合成数据分析

我们模拟了一个合成数据集，其中包含一维特征 X 和连续响应变量 Y，数据生成过程遵循（Romano、Patterson 和 Cand`es 2019年；Sesia 和 Romano（2021年）），详情见附录。图 1展示了我们的方法在这一示例数据上的应用，并将 CIR 和 CIR+与保形化直方图回归（CHR）（Sesia 和 Romano（2021年））及保形化分位数回归（CQR）（Romano、Patterson 和 Cand`es 2019年）进行了比较。

Skew		CHR	CQR	CIR	CIR+
0.66	边际	0.901 (0.001)	0.900 (0.001)	0.906 (0.001)	0.902 (0.001)
	条件	0.891 (0.005)	0.883（0.005）	0.898（0.004）	0.889（0.005）
	宽度	5.627 (0.046)	5.950 (0.031)	5.900 (0.049)	5.715 (0.054)
	Time	106.995 (2.049)	0.018 (0.000)	0.710(0.015)	0.911(0.019)
1.71	边际	0.901 (0.001)	0.900（0.001）	0.907（0.001）	0.902（0.001）
	条件	0.891 (0.003)	0.888 (0.003)	0.897 (0.003)	0.891 (0.003)
	宽度	4.785 (0.025)	5.436 (0.018)	5.110 (0.026)	4.870 (0.030)
	Time	103.026（1.150）	0.020（0.000）	0.687（0.009）	0.887（0.011）
2.70	边际	0.901 (0.001)	0.900 (0.001)	0.905 (0.001)	0.900 (0.001)
	条件	0.889 (0.003)	0.875 (0.003)	0.895 (0.003)	0.889 (0.003)
	宽度	3.389 (0.013)	5.345 (0.016)	3.776 (0.022)	3.643 (0.022)
	Time	103.637 (1.105)	0.019 (0.000)	0.663(0.008)	0.862(0.010)

表 2：我们的方法（CIR 和 CIR+）与基准方法在不同偏度的合成数据上的性能比较，偏度指响应变量条件分布的偏度（Skew），样本量为 5000。

左侧图展示了独立测试数据上的预测带，并比较了所有方法的相应输出。CIR 和 CIR+均基于同一个深度神经网络，并保证 90% 的边际覆盖率。CHR 能够利用基模型估计的所有条件分位数中的信息，自动适应估计的数据分布，并生成相对较窄的区间。尽管我们的方法省略了直方图步骤，但这并未显著影响结果。CIR 和 CIR+ 生成的区间与 CHR 一样窄。相比之下，CQR 只能利用预先指定的下分位数和上分位数（例如 5% 和 95%），因此无法适应偏度。右侧图展示了 $X \approx 0.4$ 的分位数间区间及其对应的概率。为了更清晰地呈现 CIR 和 CIR+算法的内部过程，我们只设置了 50 个分位数。此外，我们截去了密度较高的区域，并聚焦于信息集中的部分。请注意，较窄的分位数间区间具有较高的密度。可以看到，CIR 和 CIR+在每一步都会选择相对较窄的分位数间区间。在这个样本中，CIR+ 比 CIR 少选择了一个区间，因为该区间的缩放长度小于 $\widehat{s}-s$，因此被排除。我们模拟了一个合成数据集，其中包含一维特征 X 和连续响应变量 Y，它们取自图 1所示的分布。我们的方法应用于来自该分布的 2,000 个独立观测值，其中前 1,000 个用于训练深度分位数回归模型，其余用于校准。与（Sesia 和 Romano（2021年））一致，我们为 CHR 设置了 100 个分箱。

表1展示了在1000次独立实验中，CIR和CIR+在不同样本量（$|\mathcal{I}|$）下的表现，其中 $|\mathcal{I}|$在训练集和校准集之间均分，分位数间区间的数量固定为 $T = 100$。我们评估边际覆盖率、最差切片条件覆盖率（定义见 Cauchois、Gupta 和 Duchi（2021；Romano, Sesia, and Candes 2020））、平均区间宽度，以及校准和预测的单样本计算成本，所有方法均使用相同的基模型。我们的方法和CHR都比CQR产生更窄的区间，同时保持边际覆盖率和条件覆盖率。在样本量较小时，由于数据有限导致分位数估计不够精确，我们的方法表现出一定波动；但随着样本增加，分位数回归会更加准确，结果也趋于稳定。由于符合度分数并列，CIR的边际覆盖率高于目标水平。在理想情况下，当 N较大时，CIR+应接近覆盖率水平，而CIR的覆盖率则应额外高出约$1/(2T) = 0.005$ 于覆盖率水平 $1-\alpha = 0.9$（由于每个分位数间区间承载 $1/T$ 的概率，阈值可能出现在区间内任意位置，但平均而言会落在中点）。值得注意的是，CHR的计算量很大，其运行时间会随着样本量增加而增长，这主要是因为直方图构建过程所需时间不断增加。由于我们的置信区域直接根据分位数间区间计算得出，不涉及分箱过程，因此校准和预测每个样本所需的计算时间可以忽略不计。

表 2 展示了样本量为 5000 时，不同分布偏度下的结果。遵循 Sesia 和 Romano（2021），我们对每个数据点进行有偏硬币投掷，以概率 Y 将其反转为 $-Y$，并以硬币偏向程度作为控制变量。这会生成从偏斜（表1）到对称 $P_{Y|X}$的分布。结果按期望偏度 $\mathbb{E}[(Y-\mu(X))^3/\sigma^3(X)]$,展示，其中 $\mu(X)$和 $\sigma(X)$分别是 $Y \mid X$的均值和标准差。对于对称的 $P_{Y|X}$（偏度接近 0），所有方法产生的区间长度相近。随着偏度增加，我们的方法在不同不对称程度下均能达到与 CHR 相当的性能。尽管 CHR 始终产生最窄的区间，但其较高的计算成本依然存在。无论偏度如何，CIR 和 CIR+都能保持高效，在性能与实用性之间取得平衡。附录提供了扩展分析，包括与其他基准方法（如分布式保形预测（DCP）（Chernozhukov 等（2019））和 DistSplit（Izbicki 等（2020））））的比较，以及对更广泛的样本量和偏度水平的评估，并展示了使用随机森林模型的结果。

真实数据分析

我们在七个公共领域数据集上评估 CIR 和 CIR+。这些数据集此前已在（Romano、Patterson 和 Cand`es 2019）中进行分析：蛋白质三级结构的物理化学性质（bio）（Rana 2019）、博客反馈（blog）（Buza 2014）、Facebook 评论量变体一（fb1）和二（fb2）（Singh 2016），以及医疗支出小组调查数据集 19（meps19）、20（meps20）和 21（meps21）（Af HRaQ 2021）。前四个数据集来自 UCI 机器学习仓库（Dua 和 Graff 2017），而 MEPS 数据集来自（Cohen、Cohen 和 Banthin（2009））。关于这些数据集的详细描述，读者可参阅（Romano、Patterson 和 Cand`es 2019）。我们将所提出的方法与 CHR、CQR、DistSplit 和 DCP-CQR 进行比较。后者由（Sesia 和 Romano（2021））提出，旨在

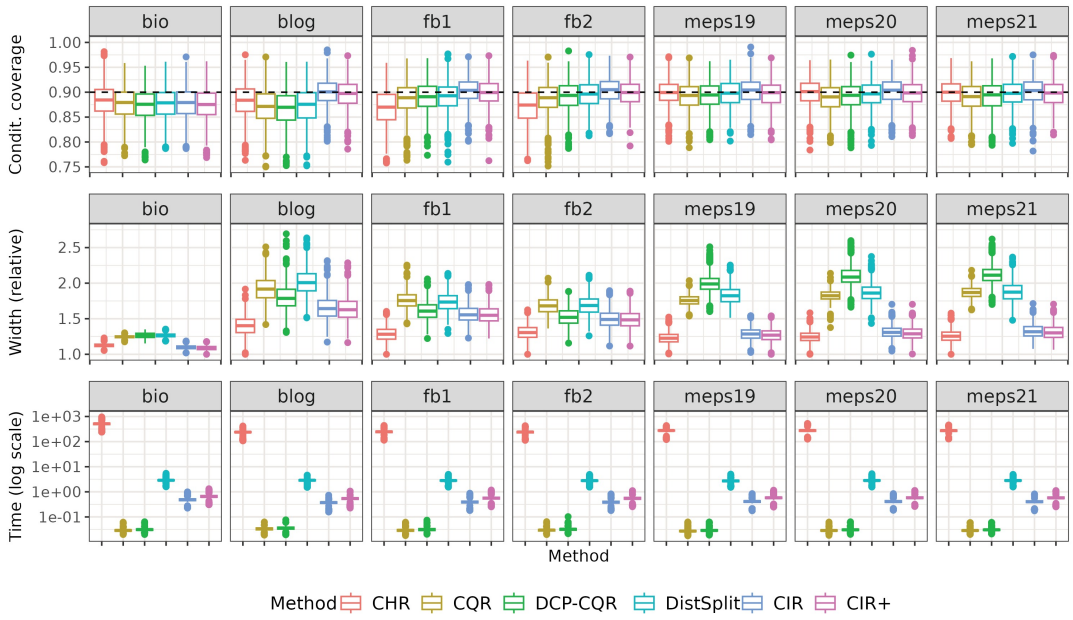

图 2：我们的方法（CIR）与基于相同深度神经网络回归模型的朴素的未校准预测区间的性能比较。请注意，该图上半部分展示的是边际覆盖率。

通过将 DCP（Chernozhukov、W¨uthrich 和 Zhu 2019）与 CQR（Romano、Pat- terson 和 Cand`es 2019）结合来增强 DCP 的稳定性。所有方法均采用相同的深度分位数回归模型进行预测。作为比较，我们还在附录中展示了使用随机森林方法得到的结果。森林方法（Meinshausen 2006）在一组决策树之间聚合分位数估计，而神经网络方法（Taylor 2000）则通过单个网络中的共享参数，以向量形式输出多个分位数。这两种方法在设计上都能避免分位数交叉问题。按照（Sesia 和 Romano（2021））的方法，在本部分中我们为 CHR 设置了 1000 个分箱。我们遵循上一节所述的流程评估其性能，并对每个数据集进行 100 次独立实验后取平均结果。每次实验使用 2000 个样本进行训练，使用 2000 个样本进行校准，其余样本用于测试。所有特征均标准化为均值为零、方差为一。名义覆盖率设为 90%。

图 2 比较了不同方法和数据集下的条件覆盖率与区间宽度。为了有意义地比较不同数据集，预测区间宽度在各数据集内进行归一化，并将最小宽度设为 1。计算时间经过类似归一化后，以对数尺度呈现。结果表明，所有方法都实现了令人满意的条件覆盖率。尽管 CHR 和 CIR+ 在不同数据集中交替生成最有效（最窄）的预测区间，但与 CHR 相比，CIR+ 和 CIR 的计算开销显著降低。因此，CIR+ 成为最佳选择，能够有效平衡计算效率与预测性能。CQR 和 DistSplit 的表现相近，而 DCP-CQR 在某些情况下倾向于生成更宽的预测区间。附录中提供了全面分析，包括边际覆盖率（理论上所有方法均能保证）。

6 结论

本文介绍了两种新颖的保形回归方法：保形分位间回归（CIR）及其增强变体 CIR+。这些方法旨在构建兼具覆盖率保证和计算效率的预测区间，从而有效解决现有方法的关键局限。具体而言，我们的方法克服了偏态数据上性能欠佳的挑战，同时避免了高昂的计算负担。理论分析表明，在适当条件下，CIR 和 CIR+不仅能够保证边际覆盖率，还能实现所需的条件覆盖率，同时将预期预测区间长度降至最低。通过在合成数据集和真实数据集上进行广泛测试，我们证明这些方法在显著降低计算成本的同时，能够达到领先替代方法的性能。这两种方法为覆盖率与效率的平衡提供了不同的解决方案：CIR 生成更宽、更保守的区间，而 CIR+ 通过分析端点区间分位数宽度，以略高的计算成本生成更窄的区间。由于难以有效训练分位数回归模型，CIR 和 CIR+的性能可能会受到小数据集的影响。未来，我们将专注于把这些框架扩展到时间序列数据，在保留其理论保证的同时拓展其应用范围。

致谢

本研究得到国家自然科学基金委员会（项目编号：62506315）和香港城市大学（项目编号：9610639、7020161）的资助。

参考文献

AfHRaQ。2021年。医疗支出小组调查。https: // meps.ahrq.gov/mepsweb/data stats/data overview.jsp。Bao, J.; Colombo, N.; Manokhin, V.; Cao, S.; 和 Luo, R. 2025a。《单变量保形回归方法综述与比较分析》。载于第十四届保形与概率预测及其应用研讨会（*COPA 2025*），282–304。PMLR。Bao, J.; Dang, C.; Luo, R.; Zhang, H.; 和 Zhou, Z. 2025b。《通过保形预测增强对抗鲁棒性：保证模型可靠性的框架》。载于第四十二届国际机器学习会议（*ICML*）论文集。Buza, K. 2014年。BlogFeedback。UCI 机器学习仓库。数字对象标识符：https://doi.org/10.24432/C58S3F。Cauchois, M.; Gupta, S.; 和 Duchi, J. C. 2021年。知其所知：多类别和多标签预测中的有效且经过验证的置信集。机器学习研究期刊，22(81)：1–42。Chernozhukov, V.; W¨ uthrich, K.; 和 Zhu, Y. 2019年。分布保形预测。*arXiv*预印本：*1909.07889*。Chernozhukov, V.; W¨ uthrich, K.; 和 Zhu, Y. 2021年。分布保形预测。美国国家科学院院刊，118(48)：e2107794118。Cohen, J. W.; Cohen, S. B.; 和 Banthin, J. S. 2009年。医疗支出小组调查：为医疗成本研究以及政策与实践提供信息的全国性信息资源。*Medical Care*，S44–S50。Dua, D.; 和 Graff, C. 2017年。UCI 机器学习仓库。Gao, C.; Shan, L.; Srinivas, V.; 和 Vijayaraghavan, A. 2025年。结构化预测集中的保形预测体积最优性。*arXiv*预印本 *arXiv:2502.16658*。Gupta, C.; Kuchibhotla, A. K.; 和 Ramdas, A. 2022年。嵌套保形预测与分位数袋外集成方法。模式识别，127：108496。Izbicki, R.; Shimizu, G.; 和 Stern, R. 2020a。使用密度估计器的灵活无分布条件预测带。载于国际人工智能与统计学会议，3068–3077。PMLR。Izbicki, R.; Shimizu, G.; 和 Stern, R. 2020b。使用密度估计器的灵活无分布条件预测带。载于 Chiappa, S.; 和 Calandra, R. 编，第二十三届国际人工智能与统计学会议论文集，第108卷，机器学习研究论文集，3068–3077。PMLR。Izbicki, R.; Shimizu, G.; 和 Stern, R. B. 2022年。Cd-split 和 hpd-split：高维中的高效保形区域。机器学习研究期刊，23(87)：1–32。

Kingma, D. P.; and Ba, J. 2014年。Adam：一种随机优化方法。*arXiv* 预印本 *arXiv:1412.6980*。Kivaranovic, D.; Johnson, K. D.; and Leeb, H. 2020年。适用于深度网络的自适应、分布无关预测区间。载于人工智能与统计国际会议，4346–4356。PMLR。Luo, R.; Bao, J.; Zhou, Z.; and Dang, C. 2024年。针对医学影像中对抗攻击的稳健保形预测博弈论防御。*arXiv* 预印本 *arXiv:2411.04376*。Luo, R.; and Colombo, N. 2024年。熵重加权保形分类。载于第*13* 届保形与概率预测及其应用研讨会，264–276。PMLR。Luo, R.; and Colombo, N. 2025年。使用传导式图自编码器进行保形负荷预测。机器学习，114(3)：1–22。Luo, R.; Nettasinghe, B.; and Krishnamurthy, V. 2023年。边可交换社交网络模型中的异常边检测。载于保形与概率预测及其应用，287–310。PMLR。Luo, R.; and Zhou, Z. 2025a. 条件保形风险自适应。*arXiv* 预印本 *arXiv:2504.07611*。Luo, R.; and Zhou, Z. 2025b. 用于高效回归的保形阈值区间。*AAAI* 人工智能会议论文集，39(18)：19216–19223。Luo, R.; and Zhou, Z. 2025c. 使用对称校准的保形区间算术。*AAAI* 人工智能会议论文集，39(18)：19207–19215。Luo, R.; and Zhou, Z. 2025d. 用于分类的一致性分数平均。载于第四十二届国际机器学习会议（*ICML*）论文集。Luo, R.; and Zhou, Z. 2025e. 密度排序预测集：面向多目标回归的高效保形预测。模式识别，112513。Luo, R.; and Zhou, Z. 2025f. 通过基于秩的保形预测集实现可靠分类。模式识别，112330。Meinshausen, N. 2006年。分位数回归森林。机器学习研究期刊，7：983–999。Rana, P. 2019年。蛋白质三级结构的物理化学性质数据集。https://archive.ics.uci.edu/ml/datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure。访问日期：2019年7月。Romano, Y.; Patterson, E.; and Cand`es, E. 2019年。保形化分位数回归。载于神经信息处理系统进展，3538–3548。Romano, Y.; Sesia, M.; and Candes, E. 2020年。具有有效且自适应覆盖率的分类。神经信息处理系统进展，33：3581–3591。Sesia, M.; and Cand`es, E. 2020年。若干保形分位数回归方法的比较。*Stat*，9(1)。Sesia, M.; and Romano, Y. 2021年。使用条件直方图进行保形预测。神经信息处理系统进展，34：6304–6315。

Singh, K. 2016年。Facebook 评论量数据集。UCI 机器学习仓库。数字对象标识符：10.24432/C5Q886。

Su,X.; Zhou,Z.;和 Luo, R. 2024年. 基于隐马尔可夫模型的粒子滤波自适应保形推断。*arXiv* 预印本 *arXiv:2411.01558*.

Tang, L.; Luo, R.; Zhou, Z.;和 Colombo, N. 2025年. 利用校准不确定性集增强路线规划。机器学习, 114(5):1–16.

Taylor, J. W. 2000年. 一种用于估计多期收益条件密度的分位数回归神经网络方法。《预测学杂志》, 19(4): 299–311.

Wang, T.; Zhang, K.; Luo, R.; Zhou, Z.;和 Li, W. J. 2025年. 基于加权聚合保形预测的可信药物相互作用预测模型。载于第*23*届亚太生物信息学会议.

Wang, T.;Zhou, Z.;和 Luo, R. 2025年. 通过基于排名的保形训练提升图神经网络的可信度。*AAAI* 人工智能会议论文集, 39(20): 21261–21268.

Zhang,Z.; Bao,J.; Zhou, Z.; Cheng, L.; Luo, R.; 等. 2025. 面向图神经网络的残差重加权保形预测. 载于 第 *41* 届人工智能不确定性会议。

A 附加理论结果

CIR 的边际覆盖概率证明遵循与一般保形预测框架中标准论证相同的逻辑。

定理 A.1. （有限样本）若(X_i, Y_i)（对于 $i \in \mathcal{I}_{cal} \cup \mathcal{I}_{test}$）是可交换的，则对于 $i \in \mathcal{I}_{test}$，算法 *2* 的输出满足：

$$\mathbb{P}\left[Y_i \in C^{\alpha}_{\widehat{k}_i}(X_i)\right] \geq 1-\alpha. \tag{17}$$

证明。 注意，评分函数

$$s(X_i, Y_i) = \min\Big\{k - 1 + e_k(X_i) : Y_i \in C^{\alpha}_k(X_i)\Big\}$$

其中

$$e_k(x) = \frac{1}{c}\min\{\widehat{q}_{l_{k-1}}(x) - \widehat{q}_{l_{k-1}-1}(x), \\ \widehat{q}_{l_{k-1}+k}(x) - \widehat{q}_{l_{k-1}+k-1}(x)\},$$

对于 $i \in \mathcal{I}_{\text{cal}} \cup \mathcal{I}_{\text{test}}$ 也具有可交换性。对于测试集中的任意 (X_i, Y_i)， $s(X_i, Y_i)$ 的秩小于 $\widehat{s}$ （算法 2 中的）$\lceil (1+|\mathcal{I}_{\text{cal}}|)(1-\alpha) \rceil / |\mathcal{I}_{\text{cal}}|$ 的概率。 $\geq 1-\alpha$ □

定理4.2的证明 假设满足以下条件：

1.（独立同分布）样本是独立同分布的，这比可交换性更强。

2.（一致性）黑盒模型能够一致地估计 $P_{Y|X}$，这意味着分位数得到了正确估计。满足这一条件时，每个分位数间区间覆盖真实标签的概率都应为 $1/T$ Y，而为了覆盖校准集中真实标签的 $(1-\alpha)$ 比例，包含 Y 的分位数间区间的最优数量应为 $\lceil T\ (1-\alpha)\ \rceil$。

3.（单峰性）真实的 $P_{Y|X}$ 是单峰的。

定义

$$\widehat{F}(y \mid x) := \widehat{t}(y)/T,$$

其中 $\widehat{t}(y) = \max\{t \in \{1,\dots,T\} :\ y \leq \widehat{q}_t\}$。注意， $\widehat{\ } F(\widehat{q}_t \mid x) = t/T$。假设对于所有 $t \in \{1,\dots,T\}$， $|\mathcal{I}_{cal}| = n$，

$$\mathbb{P}\Big[\mathbb{E}\Big[(t/T - F(\widehat{q}_t \mid X))^2 \mid \mathcal{D}^{\text{train}}\Big] \leq \rho_n^2\Big] \geq 1 - \rho_n^2. \tag{18}$$

将事件 A_n 定义为

$$A_n := \left\{x : \sup_{t\in\{1,\dots,T\}} |\widehat{F}(\widehat{q}_t \mid x) - F(\widehat{q}_t \mid x)| > \rho_n^{1/3}\right\}.$$

对于任意固定的 $t \in \{1,\dots,T\}$,

$$\begin{aligned}
&\mathbb{P}[X \in A_n] \\
&= \mathbb{P}\left[\sup_{t'\in\{1,\dots,T\}} |\widehat{F}(\widehat{q}_{t'} \mid x) - F(\widehat{q}_{t'} \mid x)|^2 > \rho_n^{2/3}\right] \\
&\leq T\,\mathbb{P}\Big[|\widehat{F}(\widehat{q}_t \mid x) - F(\widehat{q}_t \mid x)|^2 > \rho_n^{2/3}\Big] \\
&\leq T\left(\rho_n^{-2/3}\mathbb{E}\Big[\mathbb{E}\Big[|\widehat{F}(\widehat{q}_t \mid x) - F(\widehat{q}_t \mid x)|^2 \mid \mathcal{D}^{\text{train}}\Big]\Big]\right) \\
&\leq T\left(\rho_n^{-2/3}(4\rho_n^2 + \rho_n^2)\right) \\
&\leq 5T\rho_n^{4/3}
\end{aligned}$$

上面的第二个不等式是马尔可夫不等式。令$T = \rho_n^{-1}$，我们有 $\mathbb{P}[X \in A_n^c] \geq 1 - 5\rho_n^{1/3}$。将校准数据点划分为

$$\mathcal{D}^{\text{cal},a} := \{i \in \mathcal{D}^{\text{cal}} : X_i \in A_n\},$$
$$\mathcal{D}^{\text{cal},b} := \{i \in \mathcal{D}^{\text{cal}} : X_i \in A_n^{\text{c}}\},$$

对于任意 $\epsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\Big[|\mathcal{D}^{\text{cal},a}| \geq 5nT\rho_n^{4/3} + \epsilon\Big] \\
&\leq \mathbb{P}\big[|\mathcal{D}^{\text{cal},a}| \geq n\mathbb{P}\{X \in A_n\} + \epsilon\big] \\
&\leq \mathbb{P}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{cal}} \mathbf{1}\{X_i \in A_n\}\right] \geq \mathbb{P}\Big[X_i \in A_n\} + \frac{\epsilon}{n}\Big] \\
&\leq \exp\left(-\frac{2\epsilon^2}{n}\right).
\end{aligned}$$

因此，令 $\epsilon = c\sqrt{n\log\ n}$，其中某个常数 $c > 0$，可得

$$\mathbb{P}\Big[|\mathcal{D}^{\text{cal},a}| \geq 5nT\rho_n^{4/3} + c\sqrt{n\log n}\Big] \leq n^{-2c^2}.$$

对于任意固定的 $t \in \{0,\dots,T\}$, $k_i \geq t$，假设集合 $C_{k_i} = (\widehat{q}_{l_{k_i}-1},\ \widehat{q}_{l_{k_i}+k_i})$。在我们的设定中，包含更多分位数间区间的区间严格长于包含较少区间的区间——这一性质在单峰

设定中自然成立。形式上，我们假设对每个 k， $C_k^\alpha(x) \subset C_{k+1}^\alpha(x)$。然后，注意到

$$\begin{aligned}\mathbb{P}[k_i \le t] &= \mathbb{P}[Y_i \in C_{k_i}] \\ &= F(\widehat{q}_{\widehat{l}_{k_i}+k_i}) - F(\widehat{q}_{\widehat{l}_{k_i}-1}) \\ &\ge \widehat{F}(\widehat{q}_{\widehat{l}_{k_i}+k_i}) - \widehat{F}(\widehat{q}_{\widehat{l}_{k_i}-1}) - 2\rho_n^{1/3} \\ &\ge t/T - 2\rho_n^{1/3}.\end{aligned}$$

上文中的第一个不等式由 $\mathcal{D}^{\mathrm{cal},b}$ 的定义得出。等价地，我们可以将其改写为

$$\mathbb{P}\Big[k_i > t + 2\rho_n^{1/3} + \delta_n\Big] \le 1 - t/T - \delta_n,$$

对任意 $\delta_n > 0$。现在，将 $\mathcal{D}^{\mathrm{cal},b}$ 划分为以下两个互不相交的子集：

$$\begin{aligned}\mathcal{D}^{\mathrm{cal},b1} &:= \{i \in \mathcal{D}^{\mathrm{cal},b} : k_i \le t + 2\rho_n^{1/3} + \delta_n\}, \\ \mathcal{D}^{\mathrm{cal},b2} &:= \{i \in \mathcal{D}^{\mathrm{cal},b} : k_i > t + 2\rho_n^{1/3} + \delta_n\}.\end{aligned}$$

接着，我们利用 Hoeffding不等式 对 $|\mathcal{D}^{\mathrm{cal},b2}|$ 进行界定。对任意 $i \in \mathcal{D}^{\mathrm{cal}}$，若 $i \in \mathcal{D}^{\mathrm{cal},b}$，则定义 $\tilde{k}_i = k_i$；否则定义为其他值。对任意 $\epsilon > 0$,

$$\begin{aligned}&\mathbb{P}\Big[|\mathcal{D}^{\mathrm{cal},b2}| \ge n(1 - t/T - \delta_n) + \epsilon\Big] \\ &\le \mathbb{P}\Big[\frac{1}{n}\sum_{\mathcal{D}^{\mathrm{cal},b}} \mathbf{1}\{\tilde{k}_i > t + 2\rho_n^{1/3} + \delta_n\} \\ &\qquad \ge \mathbb{P}\Big[k_i > t + 2\rho_n^{1/3} + \delta_n\Big] + \frac{\epsilon}{n}\Big] \\ &\le \mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{\tilde{k}_i > t + 2\rho_n^{1/3} + \delta_n\} \ge \\ &\qquad \mathbb{P}\Big[\tilde{k}_i > t + 2\rho_n^{1/3} + \delta_n\Big] + \frac{\epsilon}{n}\Big] \\ &\le \exp\left(-\frac{2\epsilon^2}{n}\right).\end{aligned}$$

因此，令 $\epsilon = c\sqrt{n\log\, n}$, 其中 $c > 0$ 为某个常数，可得

$$\mathbb{P}\Big[|\mathcal{D}^{\mathrm{cal},b2}| \ge n(1 - t/T - \delta_n) + c\sqrt{n\log n}\Big] \le n^{-2c^2}.$$

由于 $|\mathcal{D}^{\mathrm{cal},b1}| = n - |\mathcal{D}^{\mathrm{cal},a}| - |\mathcal{D}^{\mathrm{cal},b2}|$，结合上述结果可得：

$$\begin{aligned}&\mathbb{P}\Big[|\mathcal{D}^{\mathrm{cal},b1}| \ge nt/T + n\delta_n - 5nT\rho_n^{4/3} - 2c\sqrt{n\log n}\Big] \\ &\ge 1 - 2n^{-2c^2}.\end{aligned}$$

如果我们选择

$$\delta_n = t/(nT) + T\left(\rho_n^{4/3} + \rho_n^2\right) + 2c\sqrt{(\log n)/n},$$

这就变成了

$$\mathbb{P}\big[|\mathcal{D}^{\mathrm{cal},b1}| \ge t(n+1)/T\big] \ge 1 - 2n^{-2c^2},$$

这意味着

$$\begin{aligned}&\mathbb{P}[k(t) \le t + t/(nT) + 2\rho_n^{1/3} + 5T\rho_n^{4/3} + 2c\sqrt{(\log n)/n}] \\ &\ge 1 - 2n^{-2c^2}\end{aligned}$$

$\widehat{\ }$正如我们将第 k_i 个 $\lceil(1-\alpha)T\rceil$ 记为 k,

$$\begin{aligned}\mathbb{P}[\widehat{k} \le \lceil(1-\alpha)T\rceil + \lceil(1-\alpha)T\rceil/nT + 2\rho_n^{1/3} + 5T\rho_n^{4/3} \\ + 2c\sqrt{(\log n)/n}] \ge 1 - 2n^{-2c^2}\end{aligned}$$

由于 $n \to \infty$，不失一般性，令 $T = \rho_n^{-1}$,

$$\epsilon_n = 1/n + 7\rho_n^{1/3} + 2c\sqrt{(\log n)/n}$$

以及 $\eta_n = 2n^{-2c^2}$，从而得到我们的结果。与前面的证明类似，

$$\begin{aligned}\mathbb{P}[k_i \le t] &= \mathbb{P}[Y_i \in C_{k_i}] \\ &= F(\widehat{q}_{\widehat{l}_{k_i}+k_i}) - F(\widehat{q}_{\widehat{l}_{k_i}-1}) \\ &\le F(\widehat{q}_{\widehat{l}_{k_i}+k_i}) - F(\widehat{q}_{\widehat{l}_{k_i}-1}) + 2\rho_n^{1/3} \\ &\le t/T + 2\rho_n^{1/3}.\end{aligned}$$

上文中，第一个不等式由 $\mathcal{D}^{\mathrm{cal},b}$ 的定义得出。第二个不等式则基于以下观察：如果 $F(\widehat{q}_{\widehat{l}_{k_i}+k_i}) - F(\widehat{q}_{\widehat{l}_{k_i}-1}) \ge t/T + 2K\rho_n$，那么 C_{k_i} 就不可能是最优的，因为移除最左侧或最右侧的分位数间隔，就可以得到一个更短的可行区间。

对于任意 $\delta_n > 0$，将 $\mathcal{D}^{\mathrm{cal},b}$ 划分为以下两个互不相交的子集：

$$\begin{aligned}\mathcal{D}^{\mathrm{cal},b1} &:= \{i \in \mathcal{D}^{\mathrm{cal},b} : k_i \le t - 2\rho_n^{1/3} - \delta_n - 2K\rho_n\}, \\ \mathcal{D}^{\mathrm{cal},b2} &:= \{i \in \mathcal{D}^{\mathrm{cal},b} : k_i > t - 2\rho_n^{1/3} - \delta_n - 2K\rho_n\}.\end{aligned}$$

然后，我们也用 Hoeffding不等式 对 $|\mathcal{D}^{\mathrm{cal},b2}|$ 进行界定。对于任意 $i \in \mathcal{D}^{\mathrm{cal}}$，若 $i \in \mathcal{D}^{\mathrm{cal},b}$，则定义 $\tilde{k}_i = k_i$；否则定义为 $\tilde{k}_i = t$。对于任意 $\epsilon > 0$,

$$\begin{aligned}&\mathbb{P}\Big[|\mathcal{D}^{\mathrm{cal},b2}| \ge n(1 - t/T - \delta_n) + \epsilon\Big] \\ &\le \mathbb{P}\Big[\frac{1}{n}\sum_{\mathcal{D}^{\mathrm{cal},b}} \mathbf{1}\{\tilde{k}_i \le t - 2\rho_n^{1/3} - \delta_n - 2K\rho_n\} \\ &\qquad \ge \mathbb{P}\Big[k_i \le t - 2\rho_n^{1/3} - \delta_n - 2K\rho_n\Big] + \frac{\epsilon}{n}\Big] \\ &= \mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{\tilde{k}_i \le t - 2\rho_n^{1/3} - \delta_n - 2K\rho_n\} \\ &\qquad \ge \mathbb{P}\Big[k_i \le t - 2\rho_n^{1/3} - \delta_n - 2K\rho_n\Big] + \frac{\epsilon}{n}\Big] \\ &\le \mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{\tilde{k}_i \le t - 2\rho_n^{1/3} - \delta_n - 2K\rho_n\} \\ &\qquad \ge \mathbb{P}\Big[\tilde{k}_i \le t - 2\rho_n^{1/3} - \delta_n - 2K\rho_n\Big] + \frac{\epsilon}{n}\Big] \\ &\le \exp\left(-\frac{2\epsilon^2}{n}\right).\end{aligned}$$

因此，令 $\epsilon = c\sqrt{n\log\, n}$, 其中 $c > 0$ 为某个常数，可得

$$\mathbb{P}\Big[|\mathcal{D}^{\mathrm{cal},b2}| \ge n(1 - t/T - \delta_n) + c\sqrt{n\log n}\Big] \le n^{-2c^2}.$$

由于 $|\mathcal{D}^{\mathrm{cal},b1}| = n - |\mathcal{D}^{\mathrm{cal},a}| - |\mathcal{D}^{\mathrm{cal},b2}|$，结合上述结果可得：

$$\mathbb{P}\Big[|\mathcal{D}^{\mathrm{cal},b1}| \geq nt/T + n\delta_n - 5nT\rho_n^{4/3} - 2c\sqrt{n\log n}\Big] \geq 1 - 2n^{-2c^2}.$$

如果我们选择

$$\delta_n = t/(nT) + T\left(\rho_n^{4/3} + \rho_n^2\right) + 2c\sqrt{(\log n)/n},$$

这就变成了

$$\mathbb{P}\big[|\mathcal{D}^{\mathrm{cal},b1}| \geq t(n+1)/T\big] \geq 1 - 2n^{-2c^2},$$

这意味着

$$\mathbb{P}\Big[k(t) \geq t - 2\rho_n^{1/3} - 2K\rho_n - t/(nT) - T\left(\rho_n^{4/3} - \rho_n^2\right) - 2c\sqrt{(\log n)/n}\Big] \geq 1 - 2n^{-2c^2}.$$

$\widehat{\ }$正如我们注意到的，$\lceil(1-\alpha)T\rceil$个 k_i 为 k，

$$\mathbb{P}[\widehat{k} \geq \lceil(1-\alpha)T\rceil - 2K\rho_n - \lceil(1-\alpha)T\rceil/nT - 2\rho_n^{1/3} - 5T\rho_n^{4/3} - 2c\sqrt{(\log n)/n}] \geq 1 - 2n^{-2c^2}.$$

请记住，我们之前证明了 $\mathbb{P}[X \in A_n^c] \geq 1 - 5\rho_n^{1/3}$以及 $\widehat{k}$ 和 $C_{\widehat{k}}^{\alpha}$的定义。不失一般性，令 $T = \rho_n^{-1}$，

$$\begin{aligned}\mathbb{P}[\mathbb{P}[Y \in C_{\widehat{k}}^{\alpha}(x)|X = x] &\geq 1 - \alpha - 1/n - 7\rho_n^{1/3} \\ &\quad - 2K\rho_n - 2c\sqrt{(\log n)/n} - 2\rho_n^{1/3}] \\ &\geq 1 - 5\rho_n^{1/3} - 2n^{-2c^2}\end{aligned}$$

因此，$\gamma_n = 1/n - 9\rho_n^{1/3} - 2K\rho_n - 2c\sqrt{(\log n)/n}$和 $\zeta_n = 2n^{-2c^2} + 5\rho_n^{1/3}$。

B 实证结果

本节补充介绍合成数据和真实数据实验的更多细节。在实现算法 1 和算法 2 时，我们可以根据理论最优值 $k = \lceil(1-\alpha)T\rceil$，使用 $k\ \lceil rate \times (1-\alpha)T\rceil\widehat{\ }$区间进行初始化。这一策略在保持结果完整性的同时降低了计算开销，并且 $rate$ 参数可针对较大的数据集进行调整。为确保公平比较，我们报告未进行速率优化（$rate = 0$）时的结果。即使在这种未经优化的设置下，与基线方法相比，CIR 和 CIR+仍展现出更高的效率和更优的区间质量。

实验设置

我们使用以下分位数回归模型以及与（Sesia 和 Romano, 2021年）相似的超参数集来估计 $Y \mid X$ 的分布。

- **深度神经网络（NNet）：**
  – 架构：三个全连接层，隐藏维度为 64– 激活函数：ReLU 函数– 损失函数：用于条件分位数估计的 Pinball 损失（Taylor 2000）
  – 正则化：Dropout，丢弃率为 0.1– 优化器：Adam（Kingma 和 Ba 2014），学习率为 0.0005– 训练：* 最大训练轮数：2000 年 * 通过交叉验证确定最佳训练轮数 * 标准：最小化留出数据上的损失函数

- **随机森林（RF）：**
  – 实现：用于分位数回归森林的 Python Scikit-garden 软件包（Meinshausen，2006年）– 超参数：
  * 划分内部节点所需的最小样本数：50 * 树的数量：100 * 其他参数：默认设置

合成数据集

我们构建了一个合成回归数据集，旨在捕捉非线性、异方差和非对称噪声模式（Romano、Patterson 和 Cand`es，2019年；Sesia 和 Romano（2021年））。令 $X \in [0,1]$ 为均匀抽样的输入。对于每个输入 x，响应变量 Y 按以下方式生成：

$$\begin{aligned}Y_0 = &\mathrm{Poisson}\big(\sin^2(2\pi x) + 0.1\big) + 0.2\, x\, \varepsilon_1 \\ &+ \mathbf{1}\{\varepsilon_2 < 0.09\}\,(5 + 2\varepsilon_3)\,,\end{aligned} \tag{19}$$

其中 ε_1，ε_2，ε_3 是从 $\mathcal{N}(0,\ 1)$ 或 Unif(0，1) 中独立抽样得到的，具体取决于相应情况。第一项引入了幅度变化的平滑周期结构；第二项产生输入相关方差；第三项则注入了罕见但幅度较大的正向跳跃，以模拟现实世界中突发且剧烈的行为。为明确控制观测数据的偏度，我们引入一个对称性参数 $symmetry \in [0,1]$。最终的响应变量为：

$$Y = \begin{cases} -Y_0, & \text{with probability } symmetry, \\ Y_0, & \text{otherwise.} \end{cases}$$

当 $symmetry = 0$ 时，由于跳跃成分的影响，分布呈现出明显的右偏；当 $symmetry$ =0.5 时，分布变得近似对称；而对于 $symmetry > 0.5$，分布则变为左偏。

C 非单峰模型情形

在非单峰设置下，对于每个输入，我们按照分位数间区间的长度从短到长进行排序，并记录每个区间在该排序中的排名。在校准集上，每个观测值恰好落入一个区间；我们将该区间的排名作为其符合度分数。随后，在目标失覆盖率水平下，使用标准保形校准规则选择一个截止排名。对于新输入，预测集是所有排名小于或等于该校准截止排名的区间的并集。

其他实验

由于篇幅限制，表格中仅展示了基于神经网络的合成数据结果的一部分。更全面的结果请参见图3，其中展示了更多结果；图4则提供了使用随机森林基模型进行的比较。在这两项实验中，我们的所提方法和 CHR 始终能够生成更短的预测区间。

我们还根据这些方法的最差切片条件覆盖率（Cauchois、Gupta 和 Duchi，2021年）对其进行评估，并按照（Romano, Sesia, and Candes2020）中的方法进行估算。值得注意的是，所有方法都达到了理论保证的90%的边际覆盖率。

我们的分析表明，尽管 CHR 效果良好，但与其他方法相比，它在校准和预测阶段的计算开销都很大。相比之下，CIR 和 CIR+整体表现最佳，兼顾了计算效率与预测准确性。这种均衡的表现使 CIR 和 CIR+ 成为同时考虑时间效率和预测能力时最具优势的方法。

我们还通过真实数据上的额外结果进一步扩展了分析。图5展示了使用随机森林基模型进行的更多比较。该图的结构如下：

- 底部面板展示了各方法在校准和预测步骤中消耗的总时间。值得注意的是，与其他方法相比，CHR 的计算需求显著更高。
- 中部面板展示了各方法在不同数据集上的平均区间长度。
- 顶部面板比较了这些替代方法的最差切片条件覆盖率（Cauchois、Gupta 和 Duchi 2021），该指标按照（Romano, Sesia, and Candes 2020）中的方法进行估算。

对于 blog、fb1 和 fb2 数据集，虽然 CIR 和 CIR + 产生的预测区间略长于 CHR，但其条件覆盖率表现更优。区间长度与覆盖准确性之间的这种平衡，再加上它们的计算效率，使 CIR 和 CIR+ 在综合考虑时间效率和预测能力时，成为整体表现最佳的方法。

重要的是，所有方法都达到了理论保证的90%的边际覆盖率。如需全面了解性能指标，请参阅表3。

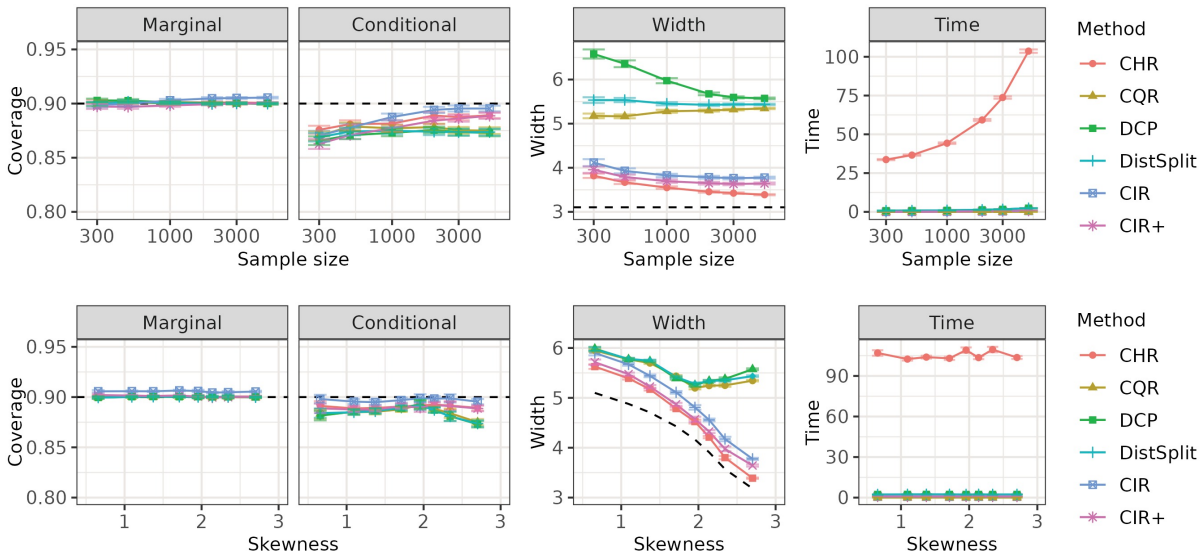

图3：基于神经网络的合成数据上，我们的方法（CIR 和 CIR+）与基准方法的性能比较。上图：样本量的影响。虚线及曲线对应全知预言机。垂直误差条表示均值上下两个标准误的范围。下图：样本量为 5000 时，响应变量条件分布的分布偏度所产生的影响。最大偏度（接近 3）对应上图中的数据。

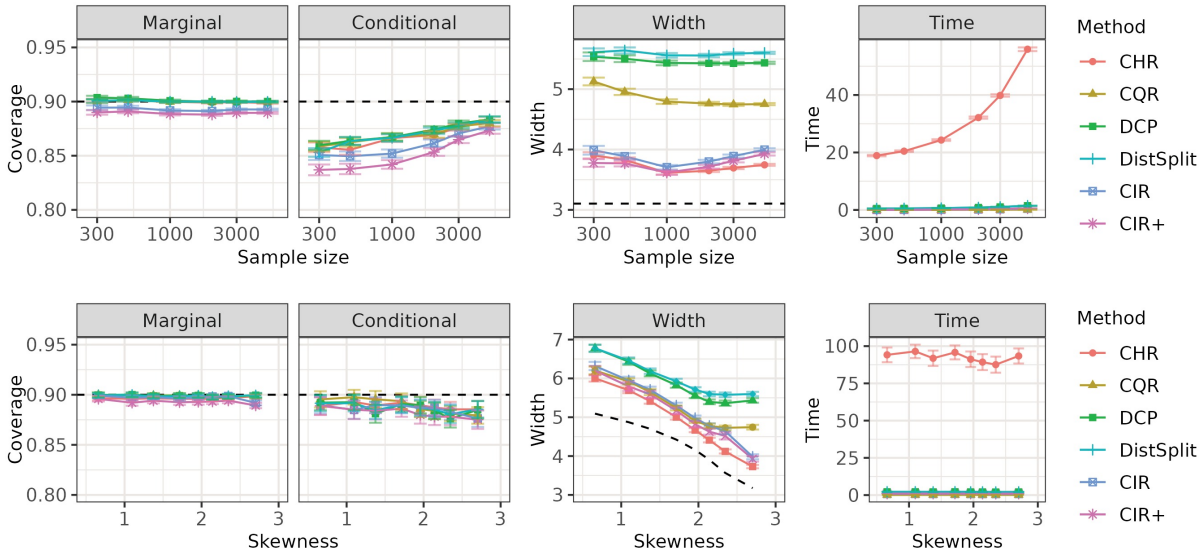

图4：基于随机森林的合成数据上，我们的方法（CIR 和 CIR+）与基准方法的性能比较。上图：样本量的影响。虚线及曲线对应全知预言机。垂直误差条表示均值上下两个标准误的范围。下图：样本量为 5000 时，响应变量条件分布的分布偏度所产生的影响。最大偏度（接近 3）对应上图中的数据。

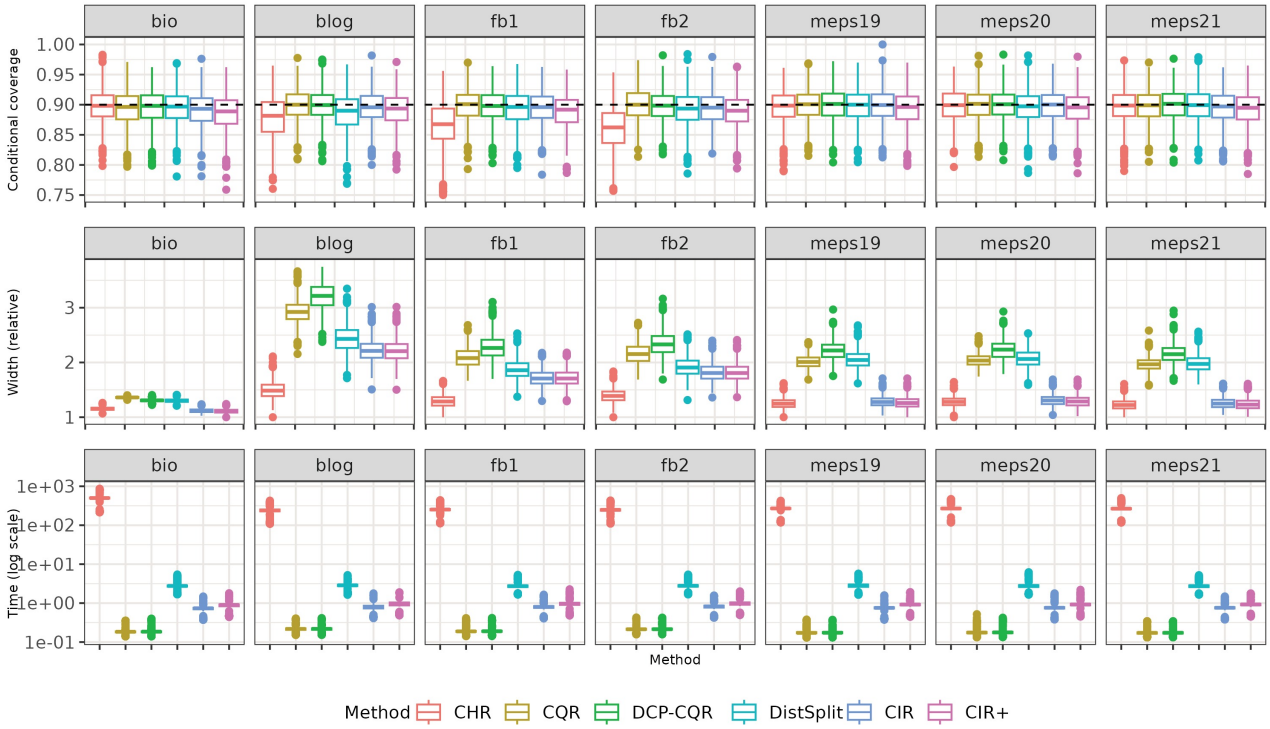

图5：我们的方法与基于随机森林回归模型的朴素未校准预测区间的性能比较。请注意，该图上半部分显示的是边际覆盖率。

		神经网络				随机森林			
		覆盖率				覆盖率			
Data	方法	边际	条件	宽度	Time	边际	条件	宽度	Time
bio	CHR	0.90 (0.01)	0.88 (0.03)	13.1 (0.3)	505.99 (97.50)	0.90 (0.01)	0.90 (0.03)	10.7 (0.3)	478.47 (106.51)
	CQR	0.90 (0.01)	0.88 (0.03)	14.5 (0.2)	0.03 (0.00)	0.90 (0.01)	0.89 (0.03)	12.6 (0.1)	0.18 (0.03)
	DCP	0.90 (0.01)	0.88 (0.03)	14.6 (0.3)	2.81 (0.42)	0.90 (0.01)	0.90 (0.03)	11.9 (0.3)	2.69 (0.47)
	DCP-CQR	0.90 (0.01)	0.87 (0.03)	14.7 (0.4)	0.03 (0.00)	0.90 (0.01)	0.90 (0.03)	12.1 (0.3)	0.18 (0.03)
	DistSplit	0.90 (0.01)	0.88 (0.03)	14.7 (0.3)	2.83 (0.43)	0.90 (0.01)	0.90 (0.03)	12.0 (0.3)	2.70 (0.48)
	CIR	0.90 (0.01)	0.88（0.03）	12.7（0.3）	0.48（0.09）	0.90（0.01）	0.89（0.03）	10.3（0.3）	0.70 (0.14)
	CIR+	0.89 (0.01)	0.87 (0.03)	12.6 (0.4)	0.65 (0.12)	0.89 (0.01)	0.89 (0.03)	10.3 (0.4)	0.85 (0.17)
blog	CHR	0.90 (0.01)	0.88 (0.03)	11.1 (1.1)	229.03 (45.34)	0.90 (0.01)	0.88 (0.03)	10.6 (1.1)	231.77 (48.09)
	CQR	0.90 (0.01)	0.87 (0.04)	15.2 (1.5)	0.03 (0.01)	0.90 (0.01)	0.90 (0.03)	20.8 (1.5)	0.21 (0.03)
	DCP	0.90 (0.01)	0.88 (0.03)	1422.3 (0.1)	2.74 (0.47)	0.90 (0.01)	0.90 (0.03)	1421.3 (0.1)	2.76 (0.47)
	DCP-CQR	0.90 (0.01)	0.87 (0.04)	14.2 (1.5)	0.04 (0.01)	0.90 (0.01)	0.90 (0.03)	23.0 (2.0)	0.21 (0.03)
	DistSplit	0.90 (0.01)	0.87 (0.04)	16.0 (1.6)	2.74 (0.47)	0.90 (0.01)	0.89（0.03）	17.3（1.7）	2.78（0.47）
	CIR	0.91（0.01）	0.90（0.03）	13.1（1.3）	0.36 (0.08)	0.90 (0.01)	0.90 (0.03)	15.8 (1.4)	0.77 (0.16)
	CIR+	0.90 (0.01)	0.90 (0.03)	12.9 (1.3)	0.52 (0.11)	0.89 (0.01)	0.89 (0.03)	15.7 (1.5)	0.93 (0.19)
fb1	CHR	0.90 (0.01)	0.87 (0.04)	10.6 (0.8)	242.20 (45.94)	0.90 (0.01)	0.87 (0.04)	11.5 (1.0)	244.09 (52.70)
	CQR	0.90 (0.01)	0.89 (0.03)	14.6 (1.0)	0.03 (0.00)	0.90 (0.01)	0.90 (0.03)	18.7 (1.6)	0.19 (0.03)
	DCP	0.90 (0.01)	0.89 (0.03)	1303.3 (0.1)	2.77 (0.42)	0.90 (0.01)	0.90 (0.03)	1302.5 (0.1)	2.71 (0.49)
	DCP-CQR	0.90 (0.01)	0.89 (0.03)	13.3 (1.1)	0.03 (0.00)	0.90 (0.01)	0.90 (0.03)	20.3 (1.9)	0.19 (0.03)
	DistSplit	0.90 (0.01)	0.89 (0.03)	14.4 (1.1)	2.77 (0.42)	0.90 (0.01)	0.89 (0.03)	16.7 (1.6)	2.70 (0.49)
	CIR	0.90 (0.01)	0.90 (0.02)	12.9 (1.0)	0.39 (0.08)	0.90 (0.01)	0.89 (0.03)	15.3 (1.3)	0.78 (0.16)
	CIR+	0.90 (0.01)	0.90 (0.03)	12.9 (1.0)	0.55 (0.11)	0.89 (0.01)	0.89 (0.03)	15.4 (1.4)	0.94 (0.20)
fb2	CHR	0.90 (0.01)	0.87 (0.03)	11.0 (0.8)	236.96 (43.77)	0.90 (0.01)	0.86 (0.04)	11.1 (0.9)	239.79 (51.39)
	CQR	0.90 (0.01)	0.89 (0.03)	14.2 (1.0)	0.03 (0.00)	0.90 (0.01)	0.90 (0.03)	17.1 (1.4)	0.21 (0.03)
	DCP	0.90 (0.01)	0.90 (0.03)	1964.0 (0.1)	2.76 (0.40)	0.90 (0.01)	0.90 (0.03)	1963.3 (0.1)	2.75 (0.47)
	DCP-CQR	0.90 (0.01)	0.89 (0.03)	12.8 (1.0)	0.03 (0.01)	0.90 (0.01)	0.90 (0.03)	18.6 (1.7)	0.21 (0.03)
	DistSplit	0.90 (0.01)	0.90 (0.03)	14.2 (1.1)	2.75 (0.39)	0.90 (0.01)	0.89 (0.03)	15.2 (1.3)	2.75 (0.46)
	CIR	0.90 (0.01)	0.90 (0.03)	12.5 (1.0)	0.38 (0.07)	0.90 (0.01)	0.89 (0.03)	14.4 (1.3)	0.80 (0.15)
	CIR+	0.90 (0.01)	0.90 (0.03)	12.5 (1.0)	0.54 (0.10)	0.89 (0.01)	0.89 (0.03)	14.4 (1.3)	0.95 (0.19)
meps19	CHR	0.90 (0.01)	0.90 (0.03)	20.3 (1.3)	270.10 (51.03)	0.90 (0.01)	0.90 (0.03)	19.4 (1.5)	258.25 (54.26)
	CQR	0.90 (0.01)	0.89 (0.03)	29.0 (1.2)	0.03 (0.00)	0.90 (0.01)	0.90 (0.03)	31.2 (1.8)	0.17 (0.03)
	DCP	0.90 (0.01)	0.89 (0.03)	559.3 (0.0)	2.71 (0.48)	0.90 (0.01)	0.89 (0.03)	559.0 (0.0)	2.79 (0.54)
	DCP-CQR	0.90 (0.01)	0.89 (0.03)	33.0 (2.3)	0.03 (0.00)	0.90 (0.01)	0.90 (0.03)	34.3 (2.6)	0.17 (0.03)
	DistSplit	0.90 (0.01)	0.90 (0.03)	30.2 (2.1)	2.70 (0.47)	0.90 (0.01)	0.90 (0.03)	31.8 (2.4)	2.77 (0.52)
	CIR	0.90 (0.01)	0.90 (0.03)	21.2 (1.5)	0.41 (0.09)	0.90 (0.01)	0.90 (0.03)	19.9 (1.5)	0.73 (0.16)
	CIR+	0.90 (0.01)	0.90 (0.03)	21.0 (1.6)	0.57 (0.12)	0.90 (0.01)	0.89 (0.03)	19.6 (1.6)	0.89 (0.19)
meps20	CHR	0.90 (0.01)	0.90 (0.03)	19.2 (1.2)	272.60 (58.48)	0.90 (0.01)	0.90 (0.03)	18.6 (1.3)	258.49 (57.63)
	CQR	0.90 (0.01)	0.89 (0.03)	28.1 (1.2)	0.03 (0.00)	0.90 (0.01)	0.90 (0.03)	29.6 (1.6)	0.18 (0.03)
	DCP	0.90 (0.01)	0.89 (0.03)	520.3 (0.0)	2.78 (0.47)	0.90 (0.01)	0.89 (0.03)	520.1 (0.0)	2.75 (0.59)
	DCP-CQR	0.90 (0.01)	0.89 (0.03)	32.4 (2.3)	0.03 (0.00)	0.90 (0.01)	0.90 (0.02)	32.3 (2.5)	0.18 (0.03)
	DistSplit	0.90 (0.01)	0.90 (0.03)	28.7 (2.0)	2.78 (0.45)	0.90 (0.01)	0.90 (0.03)	30.0 (2.3)	2.74 (0.60)
	CIR	0.91 (0.01)	0.90 (0.03)	20.2 (1.3)	0.41 (0.09)	0.90 (0.01)	0.90 (0.03)	18.9 (1.4)	0.74 (0.18)
	CIR+	0.90 (0.01)	0.90 (0.02)	19.9 (1.4)	0.57 (0.12)	0.90 (0.01)	0.89 (0.03)	18.7 (1.5)	0.90 (0.21)
meps21	CHR	0.90 (0.01)	0.90 (0.03)	20.4 (1.4)	266.75 (50.50)	0.90 (0.01)	0.90 (0.03)	20.0 (1.5)	255.58 (56.01)
	CQR	0.90 (0.01)	0.89 (0.03)	30.3 (1.3)	0.03 (0.00)	0.90 (0.01)	0.90 (0.03)	32.1 (1.8)	0.17 (0.03)
	DCP	0.90 (0.01)	0.89 (0.03)	531.3 (0.0)	2.75 (0.43)	0.90 (0.01)	0.89 (0.03)	531.0 (0.0)	2.69 (0.46)
	DCP-CQR	0.90 (0.01)	0.89 (0.03)	34.3 (2.3)	0.03 (0.00)	0.90 (0.01)	0.90 (0.03)	35.1 (2.7)	0.17 (0.03)
	DistSplit	0.90 (0.01)	0.90 (0.03)	30.4 (2.2)	2.76 (0.43)	0.90 (0.01)	0.90 (0.03)	32.3 (2.5)	2.69 (0.47)
	CIR	0.91 (0.01)	0.90 (0.03)	21.5 (1.6)	0.40 (0.08)	0.90 (0.01)	0.90 (0.03)	20.4 (1.6)	0.73 (0.14)
	CIR+	0.90 (0.01)	0.90 (0.03)	21.2 (1.7)	0.56 (0.11)	0.89 (0.01)	0.89 (0.03)	20.1 (1.7)	0.89 (0.18)

表3：在多个真实数据集上，以深度神经网络或随机森林作为基模型时，所提方法与既有基准的性能指标。数值为 1000 个随机测试集上的平均值，括号内为标准差。