pintervals：用于构建模型无关预测区间的 R 软件包

David Randahl战争研究系 瑞典国防大学

Anders HjortEiendomsverdi AS

Jonathan P.Williams统计系 北卡罗来纳州立大学

arXiv:2601.03994v1 [stat.AP] 7 Jan 2026

摘要

pintervals 软件包旨在提供一个统一框架，利用留出的校准数据，在模型无关的设置下构建预测区间并校准预测结果。它包含一系列程序，可基于任何能够输出点预测的模型构建保形预测区间、参数预测区间以及自助法预测区间。虽然已有多个 R 软件包和函数可用于构建预测区间，但它们通常侧重于特定的建模框架或预测类型，或要求针对不同模型或应用进行手动定制。通过为多种预测区间构造方法（均为模型无关方法）提供一致的接口，**pintervals** 使研究人员能够在不同建模框架和应用中运用并比较这些方法。

关键词：预测区间、保形预测、自助法、R。

1. 引言

近年来，预测已成为社会科学定量研究中日益重要的组成部分。从预测国内冲突和大规模暴行，到预测选举结果、经济衰退和政策采纳，社会科学家越来越多地致力于为各种经验现象构建预测模型（例如参见 Hegre, Metternich, Nyg˚ard, and Wucherpfennig 2017年; Nanlohy, Butcher, and Goldsmith 2017年; Lyu, Nie, and Yang 2021年; Hegre, Vesco, Colaresi, Vestby, Timlick,Kazmi, Lindqvist-McGowan, Becker, Binetti, Bodentien等2024年; Petropoulos, Apiletti,Assimakopoulos, Babai, Barrow, Taieb, Bergmeir, Bessa, Bijak, Boylan等2022）。这一转变体现出一种日益增强的认识：预测能够补充传统的因果推断，为政策制定者、公民社会和研究人员提供具有实践价值的前瞻性见解（Nanlohy等 2017年）。

这些预测模型大多只提供感兴趣结果的单一估计值，例如某个国家某个月份与冲突相关的预期死亡人数（例如Hegre、Allansson、Basedau、Colaresi、Croicu、Fjelde、Hoyles、Hultman、H¨ogbladh、Jansen 等人 2019; Hegre、Bell、Colaresi、Croicu、Hoyles、Jansen、Leis、Lindqvist-McGowan、Randahl、Rød等人2021年; Hegre 等人 2024年），或某次选举中发生选举暴力的概率（例如Randahl、Leis、G˚asste、Fjelde、Hegre、Lindberg 和 Wilson 2025a）。虽然此类点预测可以提供有价值的见解，但它们并未反映预测固有的不确定性，因此限制了其效用，并可能导致对 forecast 过度自信。在许多应用中，理解点预测的不确定性对于有效决策至关重要，

因为这能更全面地展现不同情景所涉及的风险和潜在结果（Raftery 2016）。例如，在冲突预测中，为了有效安排资源的优先顺序，利益相关者可能更需要了解冲突超过某一死亡人数阈值的可能性，而不只是预期死亡人数。同样，预测区间的宽度可以让决策者了解他们对某个点预测可以抱有多大信心，这对风险评估和规划至关重要。

这些局限使人们逐渐达成共识：仅有点预测不足以与利益相关者有效沟通，也不足以为决策者提供可采取行动的支持，在冲突预测等高风险领域尤其如此。当预测用于指导现实世界的干预时，仅仅知道最可能出现的结果还不够；决策者还必须了解最可能结果的不确定性有多大，以及他们对该结果可以有多大把握（Raftery，2016；Hegre 等 2024年；Brandt、Freeman 和 Schrodt，2014；Chadefaux，2017）。例如，预测显示暴力风险中等，但相关不确定性是高还是低，可能会导致政策应对大不相同。因此，透明地传达预测不确定性，有助于确保应对措施与风险水平相称，从而提高可靠性和问责性。

研究人员通常借助预测区间来量化预测的不确定性。构造此类区间的方法多种多样，包括参数方法（即假设预测误差服从特定分布）、自助法（从校准集重抽样误差）、分位数回归以及贝叶斯可信区间。每种方法各有优缺点。若能通过解析或近似方法求得，参数区间可能易于计算，但依赖较强的分布假设或渐近假设。自助法方法灵活，但假设数据中的误差分布保持不变，而且往往会低估方差（因为抽样依据的是经验分布函数，而非真实的累积分布函数）。分位数回归可提供非对称区间，但需要谨慎设定，且依赖于具体模型。贝叶斯方法可以给出完整的后验分布，但同样依赖于具体模型，并且需要谨慎进行先验设定（有关这些方法的综述，参见 Tian、Nordman 和 Meeker，2022；Zhang、Zimmerman、Nettleton 和 Nordman，2020；Hesterberg，2015）。其次，贝叶斯不确定性属于认识论的，而非随机性的，这意味着可信集未必能解释为 [频率学派的] 置信集；第三，贝叶斯后验推断受到虚假置信定理(Balch、Martin 和 Ferson，2019；Martin，2019；Carmichael 和 Williams，2018)的制约；该定理指出，始终存在一些不包含待预测真实值的集合，却会被赋予任意高的后验概率，而且这种情况会任意频繁地发生（即在对数据集进行重复抽样时）。这些标准方法的局限可能导致不可靠或误导性的不确定性量化，尤其是在社会科学应用中常见的复杂或异质数据情境下。因此，人们日益需要更加灵活、模型无关的不确定性量化方法，这类方法能够提供可验证的可靠性保证，并适用于各种预测模型和数据类型。为满足这一需求，保形预测（CP）正逐渐成为一种可行的不确定性量化方法。

在 t 可交换数据观测值的假设下，保形预测会生成预测集合，且 h

[1]可交换性 中，只要求数据的联合分布在排列下保持不变，这一假设 a 略微 弱于数据为独立同分布这一假设。

无论样本量固定为多少，也无论选择何种置信水平，都能保证覆盖率。保形预测的一大优势是其与模型无关：它可以应用于任何预测模型之上，从简单的线性回归模型到复杂的机器学习算法，而无需针对特定模型进行修改（Shafer 和 Vovk 2008；Vovk、Gammerman 和 Shafer 2022）。近期的扩展包括聚类保形预测（Ding、Angelopoulos、Bates、Jordan 和 Tibshirani 2023；Hjort、Williams 和 Pensar 2024）、加权保形预测（Hjort、Hermansen、Pensar 和 Williams 2025）以及分箱条件保形预测（Randahl、Williams 和 Hegre 2025b），这些方法通过解决结果分布偏斜或不平衡情形下的局部校准难题，进一步提升了保形预测的适用性。

本文旨在介绍 **pintervals** 软件包，并阐述其所支持的不确定性量化方法的理论与计算基础。我们还展示了如何使用该软件包分析2016年美国总统选举中各县的选举投票率数据。该软件包设计灵活且易于扩展，用户可以将其应用于社会科学及其他领域的广泛预测建模任务。

目前已有若干 R 软件包和函数可用于构建预测区间。例如，forecast 软件包（Hyndman、Athanasopoulos、Bergmeir、Caceres、Chhay、O'Hara-Wild、Petropoulos、Razbash、Wang 和 Yasmeen，2025）提供时间序列预测方法；tidymodels 和 mlr3 软件包（Kuhn 和 Wickham，2020；Lang、Binder、Richter、Schratz、Pfisterer、Coors、Au、Casalicchio、Kotthoff 和 Bischl，2019）则为机器学习工作流提供工具。可以使用 stats 软件包构建参数预测区间（适用于线性回归和广义线性回归模型；R Core Team，2025），也可以根据标准分布函数（例如，t 分布的 qt、正态分布的 qnorm 等）手动计算区间（若解析解可得）。可以使用 boot 软件包构建自助法预测区间（该软件包提供重抽样函数；Angelo Canty 和 B. D. Ripley，2024），也可以使用 stats 软件包手动实现自助抽样过程。还可以通过多个软件包获得保形预测区间，例如用于贝叶斯保形预测的 conformalbayes 软件包（McCartan，2025）、用于时间序列预测的 conformalForecast 软件包（Wang 和 Hyndman，2025），以及用于局部保形预测的 LCP 软件包（Guan，2021）。除这些软件包外，还有各种教程介绍如何使用 tidymodels（Kuhn 和 Wickham，2020）和 marginaleffects（Arel-Bundock、Greifer、McDermott、Bacher 和 Heiss，2025；Arel-Bundock，2025）等软件包中的函数手动获取保形预测区间。

尽管现有软件包为构建预测区间提供了有价值的工具，但它们通常缺乏统一的模型无关接口。大多数软件包都依赖特定的建模框架，或需要大量手动调整才能适配不同模型或应用，因此研究人员难以在不同数据集和预测任务中一致地应用和比较这些方法。**pintervals**软件包通过提供一个连贯且模型无关的框架来生成预测区间，弥补了这一不足。它实现了多种广泛使用的现代和经典方法，包括保形预测（CP，包括加权、聚类和分箱条件保形预测等子类别）、基于自助法的区间以及参数方法，并且兼容任何能够生成点预测的模型。**pintervals** 采用一种结构，使用独立的校准数据来估计预测的不确定性，从而实现灵活而稳健的不确定性量化。

2. 预测区间

预测区间通过给出一个范围来量化模型预测的不确定性，并声称真实结果以给定的置信水平落在该范围内，例如 90% 的预测区间。与刻画固定但未知模型参数不确定性的置信区间不同，预测区间反映的是后续数据实现中的不确定性，因此是评估和传达预测模型可靠性的重要工具。在应用研究中，尤其是社会科学领域，这一区别至关重要，因为利益相关者和政策制定者往往更关注预测结果的不确定性，而非参数推断。

pintervals 软件包实现了三类模型无关预测区间：保形区间、自助法区间和参数区间。这些方法的基本假设和估计策略各不相同，但共享统一的接口和工作流程。保形方法和自助法都依赖于一个单独的校准集，其中包含观测结果值，用于根据预测误差的经验分布校准预测区间。相比之下，参数区间假设预测误差服从某个指定分布，其参数既可以根据校准数据估计，也可以直接提供。该软件包共同提供了一个灵活且连贯的框架，可用于在广泛的模型和应用中量化预测不确定性。

2.1. 保形预测

保形预测是一种通用的无分布假设方法，在对底层预测模型仅作最少假设的情况下，能够生成具有保证边际覆盖率的预测区间。该方法依赖于构造非一致性分数，以量化给定数据点与一组数据样本之间的差异程度。假设非一致性分数可交换，则可利用其经验分布构造一个p值，用于检验以下原假设：测试点的非一致性分数与观测数据的分数可交换。例如，这可以判断是否应排除给定测试点作为合理预测结果的可能性。保形p值是一种常见的频率学派p值，其随机性上不小于均匀分布(0,1)随机变量，可对第一类错误提供有限样本控制，并且对应一个预测区间。具体而言，置信水平为 $1-\alpha$ 的保形预测区间包含所有其保形p值大于α的结果值。关于保形预测的全面背景介绍，请参见Vovk等人(2022)；关于其与其他统计范式的联系，请参见Williams (2023年)；Williams 和 Liu (2024年)。

保形预测算法通过与由特征-响应对（x_1， y_1）.....（x_N， y_N）组成的训练集进行比较，为给定的相应特征x_{N+1}构造 y_{N+1} 的预测区间。非一致性分数 s_i 可以衡量特征-响应对（x_i， y_i）与其余数据点（x_1， y_1）.....（x_{i-1}， y_{i-1}）、（x_{i+1}， y_{i+1}）.....（x_{N+1}， y_{N+1}）之间的相异程度。尽管数据的分布和非一致性分数的分布都未知，但假设数据具有可交换性意味着 s_{N+1} 在 $N+1$ 个分数中取任意秩的可能性相同，即 $\text{rank}(s_{N+1}) \sim \text{uniform}\{1,\ldots,N+1\}$。因此，如果 $\hat{}q_{1-\alpha}$ 表示校准分数的 $1-\alpha$ 经验分位数，那么在假设具有可交换性的情况下，$\text{P}(s_{N+1} \leq \hat{}q_{1-\alpha}) \geq 1-\alpha$。换言之，$\text{rank}(s_{N+1})/(N+1)$ 是原假设的p值，即 s_{N+1} 确实与 s_1、.....、s_N可交换。在任意水平 $1-\alpha \in [0,1]$下，都可以通过考虑 y_{N+1} 可能取的所有值，并且仅纳入满足 $\text{rank}(s_{N+1})/(N+1) > \alpha$ 的值，来构造一个CP集合。

ˆ ˆ 在监督设定中，我们通常会创建一个预测模型 f，使得 $f(x_{N+1})$表示模型对 y_{N+1}的最佳估计。在回归设定中，典型的非一致性分数ˆ可能是绝对残差 $s_i := |y_i - f(x_i)|$；而在分类中，标准选择是*1*减去分配给真实类别的*softmax* 概率。无论哪种情况，共形预测程序都会产生一个边际预测集

$$C_{1-\alpha}(x_{N+1}) := \{y \in \mathcal{Y} : s_{N+1}(y) \leq \hat{q}_{1-\alpha}\},$$

其中， $s_{N+1}(y)$ 表示与 $y_{N+1} = y$ 对应的非一致性分数。这种构造确保了以下有限样本边际覆盖保证，

$$\mathsf{P}\{Y_{N+1} \in C_{1-\alpha}(X_{N+1})\} \geq 1 - \alpha,$$

这一保证几乎不需要对底层预测模型或非一致性分数的选择作任何假设。

保形预测算法主要有两种变体：传导式（或完整共形预测）算法和归纳式（或拆分式共形预测）算法。传导式算法使用全部训练数据，因此统计效率最高，但通常需要针对每个测试值 y重新训练预测模型，计算成本较高。归纳式算法则牺牲部分统计效率以提高计算效率：它留出一部分数据作为校准集，该部分数据不参与训练，但只需训练预测模型一次。构造出的预测区间通常根据测试集上的经验覆盖率进行评估，并以名义 $1-\alpha$ 置信水平为参照，针对每个 $\alpha \in [0,1]$进行评估；如果覆盖率差距 超过通常的蒙特卡洛模拟波动范围，则意味着非一致性分数不可交换。在后一种情况下，采用其他形式的非一致性度量和/或预测模型，可能更好地反映数据的结构成分，从而得到可交换分数。在实际应用中，预测区间的大小也很重要，因为“更窄”的区间能向用户提供更多信息。

Mondrian 共形预测

虽然标准保形预测在边际上提供有效的覆盖率保证，但它没有考虑数据中可能存在的异质性，因此可能导致数据的某些区域出现覆盖过度或覆盖不足。为解决这一问题，Mondrian 共形预测（MCP），也称为标签条件共形预测，通过根据协变量或其他特征将校准集划分为若干组，扩展了共形框架。这些互不相交的组使得构建组条件预测区间成为可能，即为每个组生成一个 CP 集合。假设特征空间 $\mathcal{X}$可以划分为互不相交的子集 $\mathcal{X}_1, \ldots, \mathcal{X}_K$，则 MCP 保证，对于每个 $k \in \{1, \ldots, K\}$和 $\alpha \in [0,1]$,

$$\mathsf{P}\{Y_{N+1} \in C_{1-\alpha}(x_{N+1}) \mid x_{N+1} \in \mathcal{X}_k\} \geq 1 - \alpha.$$

在 MCP 中，校准步骤按组分别进行，而非针对全部校准数据进行；对每个 α，需计算 K、$1-\alpha$ 级分位数 $\hat{q}_{1-\alpha}^{(1)}$，……，$\hat{q}_{1-\alpha}^{(K)}$。虽然 MCP 可能具有组条件有效性，但其效率（例如区间的窄度）受限于每组都必须有足够数量的校准点。最后，已知（Lei 和 Wasserman 2013），精确的条件

覆盖率保证，即在给定精确值 x_{N+1} 而非 Mondrian组成员身份的条件下实现有效覆盖率，在保形预测的背景下是不可能的。

聚类保序预测

聚类保序预测（CCP）是 MCP 的扩展，适用于组数 K 较多，和/或某些组中的校准点太少，无法有效估计所需分位数的情况。该方法最初由Ding 等人(2023年）提出，其做法是将 K 初始组聚为 $M << K$ 个簇，并在簇级别而非组级别进行校准。通过汇总相似组的信息，CCP 提高了统计效率，也有助于避免每组样本过少引发的数值问题。

CCP 的核心思想是将非保形分数经验分布相似的组聚在一起。如Ding 等人(2023年）所示，任意给定组的预期覆盖率差距直接取决于被聚在一起的各组之间的相似程度（请注意，CCP 保证簇级别的有效性）。被聚在一起的经验累积分布函数越相似，覆盖率差距就越小。实践中，可以通过 k均值聚类，或根据经验累积分布函数之间的 Kolmogorov-Smirnov 距离进行聚类（Ding 等人2023年；Hjort等人2024年）。

距离加权保序预测

对保形预测程序的另一种修改是一类称为加权保形预测的方法。加权保形预测并非只计算一次感兴趣的分位数（或针对每个组或集群各计算一次），而是将 $\hat{}q_{1-\alpha}$ 计算为加权分位数，其中分数 s_1、...、s_N 根据某些权重 w_1、...、w_N进行加权。通过计算加权分位数，该方法更加重视那些与测试实例更接近可交换的数据所确定的非一致性分数，例如参考时间或空间时。举例来说，加权保序预测框架可以采用 $w_i \propto \exp\{-\text{dist}(x_i, x_{N+1})\}$形式的权重，其中 $\text{dist}(x_i, x_{N+1})$ 是一个特定于应用的距离函数，用于量化校准特征与测试特征之间的距离。随后，针对 y_{N+1} 的预测区间使用这些分数的加权 $1-\alpha$ 分位数进行校准。相关文献包含该方法的若干变体，包括用于应对协变量偏移的数据驱动权重（Tibshirani、Foygel Barber、Candes 和 Ramdas，2019）、用于处理不可交换数据的固定权重（Foygel Barber、Cand`es、Ramdas 和 Tibshirani，2023）、空间权重（Mao、Martin 和 Reich，2023）以及特征距离权重（Guan，2023）。该方法的覆盖率保证取决于加权函数的选择；有关加权保形预测方法统计性质的综述，请参见 Barber 和 Tibshirani（2025）。

分箱条件共形预测

众所周知，标准保形预测方法在结果空间密度较低的区域往往覆盖不足 $\mathcal{Y}$，而在密度较高的区域则会过度覆盖 $\mathcal{Y}$ (Guan 2023）。虽然MCP和CCP有助于解决数据中不同组的覆盖不足和过度覆盖问题，但如果结果空间不存在自然或定义明确的分组结构，这些方法就难以应用。一个例子是来自偏态分布的数据。分箱条件保形预测（BCCP）通过将校准数据划分为结果空间中由用户定义的分箱（例如低值、中值和高值）来解决这一问题（Randahl等

2025b）。

对于BCCP，在每个由用户定义的分箱内，使用校准数据集计算非一致性分数。由于新测试点所属的真实分箱未知，对于新的测试点，BCCP算法会在每个分箱内计算其分箱特定的非一致性分数。随后构造一个CP集合，将测试点的分箱特定的保形p值超过 α的所有分箱都纳入其中。对于单变量结果，最终预测集合可以通过以下任一方式构成：取CP集合中每个分箱所对应的不相交区间的并集，或使用CP集合中各分箱的最左端点和最右端点将这些区间连续化。这两种构成方式都能达到至少所需的覆盖水平，但连续化后的区间可能会出现过度覆盖。

当结果空间呈偏态或不平衡时，BCCP 尤其有用，因为它能够构建更灵活、适应性更强的CP集合，从而比标准保形预测更准确地反映结果变量的分布。它还允许用户根据对数据的了解和具体应用来定义分箱，从而以更有针对性的方式进行不确定性量化。与 MCP 和 CCP 类似，BCCP 需要仔细权衡分箱的数量和设置，以及每个分箱中的数据量，以最大限度地提高统计效率。

2.2. 自助法预测区间

在保形预测之外，构建预测区间的另一种广泛使用的方法是自助法。自助法是一种重抽样技术，通过从观测数据中反复有放回地抽取样本，来近似任何感兴趣量（例如点预测）的抽样分布。当预测误差的分布未知或复杂时，这种方法尤其有吸引力，因为它无需参数假设，即可对预测不确定性进行经验估计（Beran 1990年；Tian等 2022）。

自助法预测区间预期在以下假设下达到名义覆盖率：校准集能够代表数据生成过程，且预测误差近似独立同分布。该算法首先计算校准集上的预测误差，然后通过有放回地反复重抽样这些误差，构建经验误差分布（Tian 等 2022；Hesterberg，2015）。对于每个新的测试观测，将大量重抽样得到的误差加到模型的点预测上，从而生成可能结果的模拟分布。然后通过取 $\hat{}q_{\alpha/2}$ 和 $\hat{}q_{1-\alpha/2}$，得到预测区间的上下界。

自助法有多种变体，包括残差自助法、野生自助法和参数自助法。这些变体在重抽样的执行方式，以及对底层模型或误差结构所需的假设方面均有所不同。

2.3. 参数预测区间

参数预测区间是在预测误差遵循特定概率分布这一假设下构建的。这种方法以解析方式而非经验方式对不确定性进行建模，从而可以直接根据所假设的概率分布及其估计参数推导出区间界限。常见的选择包括正态分布、对数正态分布、泊松分布或负二项分布，具体取决于结果变量

以及模型的残差。
例如，若假设预测误差服从中心正态分布，则可根据校准集估计其标准差，并通过在点预测周围加上和减去正态分布的相应分位数，得到新观测的预测区间。更一般地，任何具有可用分位数函数的分布都可用于构造预测区间，包括用户指定的分布或混合分布。

参数预测区间的主要优势在于计算效率高，且其表现平滑、易于解释。然而，其有效性在很大程度上取决于所假设的误差分布是否恰当：一旦这一假设不成立，得到的区间可能会出现覆盖率偏差或不可靠。

3. pintervals 软件包及其主要函数

该**pintervals**软件包提供了统一的模型无关接口，使用上述方法构造预测区间。软件包中的所有函数都采用一致的语法和工作流程，便于用户在不同方法之间切换，并将其应用于多种预测模型。每个函数都接收测试集的预测值向量，以及校准数据中的预测结果和真实结果作为输入[2]，，这些输入可以是双列 tibble，也可以是分开的向量。函数返回一个 tibble，其中包含预测区间的上下界，以及相应的已校准或未校准预测值。基于 MCP 和 CCP 的函数还会返回每个观测的组或簇分配。

对于依赖数据分区的方法（如 MCP、CCP 或 BCCP），必须指定一个额外的分组变量。该变量可以作为独立向量提供，也可以作为校准 tibble 中的一列提供。

使用 pinterval_*() 函数系列构造预测区间，其中包括以下主要函数：

- pinterval_conformal()，用于构造标准保形预测区间
- pinterval_mondrian()，用于构造 Mondrian 保形预测区间
- pinterval_ccp()，用于构造聚类保形预测区间
- pinterval_bccp()，用于构造分箱条件保形预测区间
- pinterval_bootstrap() 用于计算自助法预测区间
- pinterval_parametric() 用于计算参数预测区间

所有函数都遵循相同的通用结构，并支持通过额外参数控制区间构造的关键方面，例如所需的覆盖水平、自助法重复次数，或参数区间的分布参数。这种统一的设计便于在共同的分析框架内灵活开展实验，并直接比较不同的区间估计方法。每个函数的输出都是一个

[2]校准 参数预测区间的校准数据为可选项，因为可以直接提供参数

tibble，其中包含点预测、预测区间的下界和上界，以及其他相关信息（例如 MCP 和 CCP 的组分配或聚类分配）。

3.1. 保形预测区间

pinterval_conformal() 函数实现标准保形预测区间。它以测试集中的预测值和校准数据作为输入，并返回一个 tibble，其中包含预测区间的上下界。校准数据应为一个包含两列的 tibble，其中第一列为校准集中的预测值，第二列为真实值；或者，也可以将预测值作为向量传递给 calib 参数，并将真实值作为向量传递给 calib_true 参数。

此外，用户可以指定所需的覆盖水平以及要使用的非一致性评分函数。默认使用绝对误差，也可以选择相对误差、零调整相对误差、异质误差和原始误差。用户还可以指定一个自定义的实值函数，该函数以预测值和真实值作为输入，并返回一个非一致性分数向量。这样，用户便可以根据具体应用或数据特征灵活定义非一致性分数。

Mondrian 和聚类保形预测区间

函数 pinterval_mondrian() 和 pinterval_ccp 实现了 MCP 和 CCP，可根据分组变量将校准集划分为互不相交的组。MCP 和 CCP 的语法基于 pinterval_conformal()，但要求在校准 tibble 中为分组变量额外添加一列（第三列）。分组变量也可以作为向量传入参数 calib_group。

对于 CCP，用户还可以指定要使用的簇数，或选择通过最小化所提议聚类结构的 Calinski-Harabasz 指数来优化簇数（Caliński 和 Harabasz，1974年）。为避免簇出现过拟合，还可以进一步将校准数据划分为两个子集：一个用于对类别进行聚类，另一个用于计算相应的预测区间。[3]

分箱条件保形预测区间

函数 pinterval_bccp() 实现了 BCCP区间，支持将校准集划分为结果空间中由用户定义的分箱。其语法与 pinterval_mondrian() 和 pinterval_ccp() 类似：它使用校准 tibble 中的额外列 bins 指定分箱，或者通过参数 calib_bins 传入一个向量。用户还可以通过参数 contiguize 指定是否将区间连续化，该参数默认为 FALSE。如果设为 TRUE，函数会取各分箱中最左侧和最右侧的端点，返回连续区间；否则会将相邻分箱对应的区间连续化。当 contiguize 为 FALSE 时，函数输出一个 tibble，其中包含预测区间及其对应的下界和上界，这些值存放在名为 intervals 的列表列中；否则，函数输出标准 tibble，其中包含点预测以及连续化后区间的下界和上界。用户还可以像在 pinterval_conformal() 中一样，指定所需的覆盖率和要使用的非一致性评分函数。

[3]有关如何划分数据的指南详情，请参见 Ding *et al.*(2023年)。

距离加权保形预测区间

在**pintervals** 中，所有保形预测方法也支持距离加权保形预测（DWCP），该方法会提高校准集中取值接近新测试观测的观测所对应的非一致性分数的权重。

将 distance_weighted_cp = TRUE 设为启用 DWCP。DWCP 还需要校准协变量和测试协变量，分别通过参数 distance_features_calib 和 distance_features_pred 提供。校准数据与测试数据中的协变量行之间的距离，使用马氏距离（默认）或欧几里得距离计算。还可以在计算后对距离进行归一化（欧几里得距离通常更适合进行归一化），方法包括最小-最大归一化，或将距离除以其标准差（Z分数归一化）。使用距离加权保形预测时，在计算预测区间的分位数之前，会根据校准集中的非一致性分数与新测试观测之间的距离对这些分数进行加权。

此外，用户可以指定一个权重函数，将距离转换为非一致性分数的权重。默认情况下，该函数使用高斯核：$w = \exp(-d^2)$，其中 d 表示新测试观测的预测值与校准集中的预测值之间的欧几里得距离。内置的其他选项包括柯西核：$w = \frac{1}{1+d^2}$、逻辑斯蒂核：$w = \frac{1}{1+\exp(d)}$，以及倒数线性核：$w = \frac{1}{1+d}$。用户还可以使用参数 distance_weighting_function 指定自定义距离加权函数。该函数应以距离向量为输入，并返回一个非负权重向量，用于对非一致性分数加权。这样，用户便可以根据具体应用或数据特征，灵活定义距离加权函数。

3.2. 自助法预测区间

pinterval_bootstrap() 函数用于实现自助法预测区间。它通过重抽样，利用校准集中的预测误差来模拟误差分布。该函数的输入包括测试集的预测值和校准数据，校准数据可以是一个包含两列的 tibble（第一列为预测值，第二列为真实值），也可以分别作为向量传入 calib 和 calib_truth 参数。

用户还可以通过其他参数指定所需的覆盖水平（默认为 0.9）、自助样本数量（默认为 1000），以及用于自助抽样的误差类型。参数 error_type 可设为 "raw"，以使用原始的有符号误差；也可设为 "absolute"，以使用带随机符号的绝对误差。用户还可以将 distance_weighted 参数设为 TRUE，指定使用距离加权重抽样。这样会提高校准集中预测值接近新测试观测的样本所对应误差的权重，从而有助于提高预测区间的准确性。距离加权自助法遵循与上文所述距离加权保形预测相同的原理，由 distance_features_calib、distance_features_pred、distance_type、normalize_distance 和 weight_function 参数控制。

3.3. 参数预测区间

pinterval_parametric() 函数基于用户指定的概率分布实现参数预测区间。用户提供测试

集的预测值、所需的分布（例如正态分布、对数正态分布、泊松分布、负二项分布），以及校准集中的预测值和真实值或 该分布的参数。
原生支持的分布包括常见的连续分布和计数分布，例如正态分布、对数正态分布、泊松分布和负二项分布。用户通过参数 dist 指定分布。对于原生支持的分布，相关分布参数（例如正态分布的标准差和负二项分布的离散参数）将根据校准集进行估计。或者，用户也可以通过参数 pars 传入一个包含各参数的命名列表，直接提供这些参数。与其他 pinterval_*() 函数不同，如果直接提供了分布参数，pinterval_parametric() 不需要校准数据。同样，对于单参数分布（例如泊松分布和卡方分布），由于可以直接根据预测值估计唯一参数，因此无需校准数据。

除了原生支持的分布外，用户还可以通过向参数 `dist` 提供自定义分位数函数，使用任意分布。这使得使用自定义分布或混合分布具有很大的灵活性，例如使用 **mistr** 包（Sablica 和 Hornik 2023年，2020）。指定自定义分布时，用户需要将分布参数作为命名列表传给参数 `pars`。理想情况下，这些参数应在校准数据集上估计。随后，函数将使用指定的分位数函数计算预测区间的上下界。

参数区间简单、高效且易于解释，但依赖于对基础分布的正确设定。建议用户在应用此方法前，使用校准误差检验分布假设。

4. 使用 pintervals：预测县级投票率的示例

为说明**pintervals**软件包的实际用途，我们演示如何将其应用于生成2016年美国总统选举的县级选民投票率预测区间。此示例以具体的现实应用展示了上文介绍的方法，采用典型的回归问题工作流程：先从拟合的模型中获得点预测，再使用**pintervals**提供的方法构建预测区间，以量化这些预测的不确定性。

数据包含在**pintervals**软件包中，涵盖2016年美国总统选举的县级选民投票率信息，以及可能影响投票率的各种人口统计和社会经济协变量，数据集来自MIT选举数据与科学实验室（2018年）发布的数据集。该数据集包含人口规模、收入中位数、教育水平，以及各县的种族、年龄和性别构成等变量。我们关注的结果变量是选民投票率，定义为在选举中投票的符合资格的选民所占比例。数据集还包含使用所有可用协变量（州和纬度/经度坐标等地理特征除外），通过随机森林回归模型得到的投票率点预测。为确保所有预测均为样本外预测，这些点预测是通过留一交叉验证生成的。[4]

本分析旨在使用

[4]有关该数据集的更多详情，请参阅Randahl 和 Williams (2025) 的文档。

pintervals**pintervals** 中实现的不同方法，为选民投票率的点预测构建预测区间，并评估其覆盖率和宽度。我们将比较不同保形预测方法（标准、Mondrian、聚类和分箱条件方法）的性能，以及假设误差服从正态分布和贝塔分布的自助法预测区间与参数预测区间。

我们首先加载必要的软件包和县级投票率数据集，并将数据拆分为校准集和测试集。在此示例中，我们将数据随机按 50/50 的比例拆分为校准集和测试集，以说明该过程。在实际应用中，通常会使用更大的校准集，以确保能够准确估计预测区间。需要特别注意的是，用于校准的点预测必须由模型基于样本外数据生成，以确保有效的覆盖率保证。[5]

```
R> library(pintervals)
R> library(dplyr)
R> library(tidyr)
R> data("county_turnout", package = "pintervals")
R> # Split data into calibration and test sets
R> set.seed(20260106)
R> nobs <- nrow(county_turnout)
R> calib_indices <- sample(nobs, size = 0.5 * nobs)
R> calib_data <- county_turnout[calib_indices, ]
R> test_data <- county_turnout[-calib_indices, ]
```

现在，我们可以使用**pintervals**中提供的不同方法来构建预测区间。对于每种方法，我们将使用随机森林模型生成的点预测，以及校准集中的真实投票率，为测试集生成区间。随后，我们将使用该软件包中的 interval_coverage() 函数评估经验覆盖率和平均宽度。

```
R> # Standard CP intervals
R> conformal_intervals <- pinterval_conformal(
+   pred = test_data$predicted_turnout,
+   calib = calib_data$predicted_turnout,
+   calib_truth = calib_data$turnout,
+   alpha = 0.1+ )R> # Evaluate coverage and width
R> conformal_coverage <- interval_coverage(
+   truth = test_data$turnout,
+   lower_bound = conformal_intervals$lower_bound,
+   upper_bound = conformal_intervals$upper_bound)
R> conformal_coverage

[1] 0.9086229
```

[5]The 数据集中的点预测通过留一交叉验证生成，以确保它们 e
样本外 示例。

我们可以对参数预测区间和自助法预测区间重复此过程，并根据各方法的具体要求调整函数调用。

```
R> # Parametric prediction intervals assuming normal distribution
R> norm_intervals <- pinterval_parametric(
 +   pred = test_data$predicted_turnout,
 +   calib = calib_data$predicted_turnout,
 +   calib_truth = calib_data$turnout,
 +   dist = "norm",+   alpha = 0.1+ )+
```

```
R> # Evaluate coverage
R> norm_coverage <- interval_coverage(
+   truth = test_data$turnout,
+   lower_bound = norm_intervals$lower_bound,
+   upper_bound = norm_intervals$upper_bound)+
```

```
R> # Parametric prediction intervals assuming logistic distribution
R> logis_intervals <- pinterval_parametric(
 +   pred = test_data$predicted_turnout,
 +   calib = calib_data$predicted_turnout,
 +   calib_truth = calib_data$turnout,
 +   dist = "logis",+   alpha = 0.1+ )+
```

```
R> # Evaluate coverage
R> logis_coverage <- interval_coverage(
+   truth = test_data$turnout,
+   lower_bound = logis_intervals$lower_bound,
+   upper_bound = logis_intervals$upper_bound)+
```

```
R> # Bootstrapped prediction intervals
R> bootstrap_intervals <- pinterval_bootstrap(
 +   pred = test_data$predicted_turnout,
 +   calib = calib_data$predicted_turnout,
 +   calib_truth = calib_data$turnout,
 +   alpha = 0.1,+   n_bootstrap = 1000+ )+
```

```
R> # Evaluate coverage
R> bootstrap_coverage <- interval_coverage(
 +   truth = test_data$turnout,
 +   lower_bound = bootstrap_intervals$lower_bound,
 +   upper_bound = bootstrap_intervals$upper_bound)
```

```
 +
R> c(norm_coverage, logis_coverage, bootstrap_coverage)
```

```
[1] 0.9105534 0.9079794 0.9066924
```

我们可以看到，所有方法的覆盖率都接近名义90%水平。附录中，我们展示了一项更全面的模拟研究，比较这些方法在多种校准-测试集划分下的表现。结果表明，在不同模拟中，SCP方法的覆盖率最稳定且接近名义水平，而参数方法的平均覆盖率略高，自助法区间的覆盖率则略低。

4.1. 使用 Mondrian 和聚类保形预测实现子组覆盖率

接下来，我们将演示如何使用 Mondrian 和聚类保形预测来提高数据子集内的覆盖率。在此示例中，我们将县的地理分组划分用作两种方法的分组变量。我们将使用 Mondrian 和聚类保形预测，以及基于县与县之间地理距离的距离加权保形预测，生成预测区间。随后，我们将评估总体以及美国人口普查四个地区各自的经验覆盖率。对于 CCP，我们使用美国人口普查分区作为分组变量，并根据 Cali´nski-Harabasz 指数优化簇数，将簇数上限设为 5。

```
R> # MCP intervals
R> mondrian_intervals <- pinterval_mondrian(
+   pred = test_data$predicted_turnout,
+        pred_class = test_data$region,
+   calib = calib_data$predicted_turnout,
+   calib_truth = calib_data$turnout,
+   calib_class = calib_data$region+   alpha = 0.1
+ )R> # Evaluate coverage
R> mondrian_coverage <- interval_coverage(
+   truth = test_data$turnout,
+   lower_bound = mondrian_intervals$lower_bound,
+   upper_bound = mondrian_intervals$upper_bound)+
R> # Clustered CP intervals
R> clustered_conformal_intervals <- pinterval_ccp(
 +   pred = test_data$predicted_turnout,
 +         pred_class = test_data$division,
 +   calib = calib_data$predicted_turnout,
 +   calib_truth = calib_data$turnout,
 +   calib_class = calib_data$division,
 +   alpha = 0.1,+   optimize_n_clusters = TRUE,
```

```
    +   max_n_clusters = 5,+ )R> # Evaluate coverage
R> clustered_conformal_coverage <- interval_coverage(
+   truth = test_data$turnout,
+   lower_bound = clustered_conformal_intervals$lower_bound,
+   upper_bound = clustered_conformal_intervals$upper_bound)+
R> # Distance-weighted CP intervals using geographic distances
R> dw_cp_intervals <- pinterval_conformal(
+   pred = test_data$predicted_turnout,
+   calib = calib_data$predicted_turnout,
+   calib_truth = calib_data$turnout,+   alpha = 0.1,
+   distance_weighted_cp = TRUE,
+   distance_features_calib = calib %>%
+     select(latitude, longitude),
+   distance_features_pred = test_data %>%
+     select(latitude, longitude),
+         normalize_distance = "sd"+ )R> # Evaluate coverage
R> dw_cp_coverage <- interval_coverage(
+   truth = test_data$turnout,
+   lower_bound = dw_cp_intervals$lower_bound,
+   upper_bound = dw_cp_intervals$upper_bound))
R> c(mondrian_coverage, clustered_conformal_coverage, dw_cp_coverag
                                                                    e)
```

```
[1] 0.8931789 0.9021879 0.9054054
```

我们再次看到，这些方法总体上实现了接近名义90%水平的覆盖率。为了评估每个美国人口普查区内的分组覆盖率，我们合并所有保形预测方法的结果，并计算每个组内的经验覆盖率。

```
R> # Adding grouping variable and true turnout values
R> conformal_intervals <- conformal_intervals %>%
+   mutate(region = test_data$region,+          method = "SCP",
+          turnout = test_data$turnout)
R> mondrian_intervals <- mondrian_intervals %>%
+   mutate(region = test_data$region,
+                                      method = "MCP",
+          turnout = test_data$turnout)
R> clustered_conformal_intervals <- clustered_conformal_intervals %>%
+   mutate(region = test_data$region,
+                                      method = "CCP",
+          turnout = test_data$turnout)
```

```
R> dw_cp_intervals <- dw_cp_coverage %>%
+   mutate(region = test_data$region,
+                                       method = "DWCP",
+          turnout = test_data$turnout)
R> bootstrap_intervals <- bootstrap_intervals %>%
+   mutate(region = test_data$region,
+                                       method = "Bootstrap",
+          turnout = test_data$turnout)
R> norm_intervals <- norm_intervals %>%
+   mutate(region = test_data$region,
+                                       method = "Normal",
+          turnout = test_data$turnout)
R> logis_intervals <- logis_intervals %>%
+   mutate(region = test_data$region,
+                                       method = "Logistic",
+          turnout = test_data$turnout)+
R> # Combine all intervals for group-wise coverage evaluation
R> all_conformal_intervals <- bind_rows(
+   conformal_intervals,+   mondrian_intervals,
+   clustered_conformal_intervals,+   dw_cp_intervals,
+   bootstrap_intervals,+   norm_intervals,+   logis_intervals
+ )R> # Evaluate group-wise coverage
R> group_wise_coverage <- all_conformal_intervals %>%
+   group_by(method, region) %>%+   summarise(
+     coverage = interval_coverage(+       truth = turnout,
+       lower_bound = lower_bound,
+       upper_bound = upper_bound+     ),+     .groups = "drop"
+   )R> group_wise_coverage %>%+    pivot_wider(
+     names_from = region,+     values_from = coverage+   )一个
tibble: 4 × 8re
```

```
gion Bootstra p CCP DWCP Lo gistic MCP 正态法 SCP
<字符> <双精度> <双精度> <双精度> <双精度> <双精度> <双精度> <双精度>
```

1 中西部 0.9520.896 0.925 0.950 0.876 0.954 0.950 2 东北部 0.9290.902 0.902 0.946 0.902 0.946 0.946 3 南部 0.8840.904 0.890 0.884 0.899 0.886 0.885 4 美国西部 0.8620.913 0.913 0.867 0.913 0.867 0.867

结果表明，与标准保形预测、自助法和参数预测区间相比，Mondrian、聚类和距离加权保形预测算法在所有美国人口普查地区都能实现更均匀且接近名义水平的覆盖率。

为进一步分析结果，我们在下文计算了各方法在不同地区的覆盖率平均绝对误差（MAE），以量化覆盖率的均匀性。分析结果显示，Mondrian、聚类和距离加权保形预测的平均绝对误差约为其他方法的一半。附录中的模拟研究进一步支持了这些发现，表明考虑群组的保形预测方法和距离加权保形预测方法在美国人口普查地区内始终能实现更接近名义水平的覆盖率。

```
R> group_wise_coverage %>%+         group_by(method) %>%
+         summarize(mae_coverage = mean(abs(coverage - 0.9))) %>%
+     pivot_wider(+                 names_from = method,
+                 values_from = mae_coverage+         )
```

一个 tibble: 1 × 7自助法 CCP DWCP 逻辑分布 MCP 正态法 SCP<dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>1 0.03380.00579 0.0122 0.03650.00994 0.03680.0361

4.2. 使用分箱条件保形预测考察目标变量不同范围内的覆盖率

最后，我们将展示如何使用 BCCP 提高结果变量不同范围内的覆盖率。在本例中，我们将为低于 50%、介于 50% 和 60% 之间、介于 60% 和 65% 之间以及高于 65% 的投票率划分分箱[6] 并使用 BCCP 生成预测区间。随后，我们将评估总体及各分箱内的经验覆盖率。

```
R> # Create bins of turnout rates in the calibration set
R> calib_data <- calib_data %>%
 +   mutate(turnout_bin = case_when(
 +     turnout < 0.5 ~ 1,
 +     turnout < 0.6 ~ 2,
 +     turnout < 0.65 ~ 3,
```

[6]这些分箱的划分大致对应于校准集投票率的四分位数。实际应用中，这些分箱应根据领域知识和具体应用场景来确定。

e,

```
 +     TRUE ~ 4
 +   ))
 +
R> bccp_contiguized_intervals <- pinterval_bccp(
 +   pred = test_data$predicted_turnout,
 +   calib = calib_data$predicted_turnout,
 +   calib_truth = calib_data$turnout,
 +   calib_bins = calib_data$turnout_bin,
 +   alpha = 0.1,+   contiguize = TRUE+ )+


R> # Evaluate coverage
R> bccp_contiguized_coverage <- interval_coverage(
+   truth = test_data$turnout,
+   lower_bound = bccp_contiguized_intervals$lower_bound,
+   upper_bound = bccp_contiguized_intervals$upper_bound)+R
 > bccp_discontigous_intervals <- pinterval_bccp(
 +   pred = test_data$predicted_turnout,
 +   calib = calib_data$predicted_turnout,
 +   calib_truth = calib_data$turnout,
 +   calib_bins = calib_data$turnout_bin,
 +   alpha = 0.1,+   contiguize = FALSE+ )
 +
R> # Evaluate coverage
R> bccp_discontigous_coverage <- interval_coverage(
+   truth = test_data$turnout,
+   intervals = bccp_discontigous_intervals$intervals
+ )+
R> c(bccp_contiguized_coverage, bccp_discontigous_coverage)
```

```
[1] 0.9026227 0.9018326
```

连续化 BCCP区间和非连续化 BCCP区间的总体覆盖率均接近名义90%水平。为了评估每个投票率分箱内的分箱覆盖率，我们合并两种方法的结果，并计算各分箱内的经验覆盖率。

```
R> # Adding bin variable to the intervals
R> test_data <- test_data %>%
+   mutate(turnout_bin = case_when(
+     turnout < 0.5 ~ 1,+     turnout < 0.6 ~ 2,
+     turnout < 0.65 ~ 3,
```

```
    +     TRUE ~ 4+   ))+
R> bccp_contiguized_intervals <- bccp_contiguized_intervals %>%
+   mutate(turnout_bin = test_data$turnout_bin,
+          method = "BCCP(c)",
+          turnout = test_data$turnout)+
R> bccp_discontigous_intervals <- bccp_discontigous_intervals %>%
+   mutate(turnout_bin = test_data$turnout_bin,
+          method = "BCCP(d)",
+          turnout = test_data$turnout)+
R> conformal_intervals <- conformal_intervals %>%
+   mutate(turnout_bin = test_data$turnout_bin)+
R> bootstrap_intervals <- bootstrap_intervals %>%
+   mutate(turnout_bin = test_data$turnout_bin)+
R> norm_intervals <- norm_intervals %>%
+   mutate(turnout_bin = test_data$turnout_bin)+
R> logis_intervals <- logis_intervals %>%
+   mutate(turnout_bin = test_data$turnout_bin)+
R> # Combine all intervals for bin-wise coverage evaluation
R> all_intervals_bins <- bind_rows(+   bccp_contiguized_intervals,
+   bccp_discontigous_intervals,+   conformal_intervals,
+   bootstrap_intervals,+   norm_intervals,+   logis_intervals+ )+
R> # Evaluate bin-wise coverage
R> bin_wise_coverage <- all_intervals_bins %>%
+   group_by(method, turnout_bin) %>%+   summarise(
+     coverage = interval_coverage(+       truth = turnout,
+       lower_bound = lower_bound,
+       upper_bound = upper_bound,
+             intervals = intervals+     ),+     .groups = "drop"
+   )
```

```
    +R> bin_wise_coverage %>%
+     pivot_wider(
+       names_from = turnout_bin,
+       values_from = coverage
+     )
```

一个 tibble: 6 × 5
方法`1` `2` `3` `4`<字符> <
双精度数> <双精度数> <双精度数> <双精度数>
1 BCCP(c) 0.909 0.9270.921 0.904 2 BCCP(d)
0.909 0.9150.918 0.904 3 自助法 0.806
0.9690.944 0.854 4 逻辑分布 0.814 0.9690.941
0.856 5 正态法 0.814 0.9710.944 0.861 6 SCP
0.814 0.9690.941 0.859

在这种情况下，我们看到，与标准保形预测、自助法和参数预测区间相比，连续化和非连续化 BCCP 在每个投票率分箱中的覆盖率都更接近名义水平。

为了进一步分析结果，我们在下文计算了各方法在不同分箱中的覆盖率平均绝对误差，以量化其表现。分析表明，连续化 BCCP 方法在各分箱中的覆盖率平均绝对误差最低，非连续化版本紧随其后。附录中的模拟研究进一步支持了这些发现，显示 BCCP 在结果变量的不同取值范围内始终能达到更接近名义水平的覆盖率。

```
R> bin_wise_coverage %>%+         group_by(method) %>%
+         summarize(mae_coverage = mean(abs(coverage - 0.9))) %>%
+         pivot_wider(+                 names_from = method,
+                 values_from = mae_coverage+ )
```

一个 tibble: 1 × 6
`BCCP(c)` `BCCP(d)` 自助法 逻辑分布 正态法 SCP
<双精度> <双精度> <双精度> <双精度> <双精度> <双精度>
1 0.0155 0.0116 0.0631 0.0598 0.0598 0.0592

5. 结论

本文 per 已为 R 推出了 **pintervals** 软件包，这是一个统一且可扩展的框架 k

用于通过保形方法、自助法和参数方法构造预测区间。该软件包旨在降低应用研究人员和从业者将正式的不确定性量化纳入预测分析的门槛。通过提供一致的语法和灵活的接口，**pintervals**使用户能够在不同的区间构造方法之间无缝切换，同时保持工作流程透明且可复现。

通过一个预测2016年美国总统选举县级选民投票率的实证示例，我们展示了如何将该软件包应用于真实世界的数据，以及不同方法在覆盖率表现和区间宽度方面的差异。该示例还说明，Mondrian、聚类、分箱条件和距离加权保形预测等扩展方法，如何应对数据或预测目标在不同子群体维度上的异质性和局部校准挑战。结果证实，保形预测方法不仅在总体上，而且在有意义的子群体内，都能实现接近名义水平的经验覆盖率；在分布假设可能不成立或子群体平衡不完善的应用研究中，这是一项重要特性。附录中的模拟研究进一步验证了这些结果，表明**pintervals**中实现的保形预测方法具有稳健性和可靠性。

除了本文介绍的示例，**pintervals** 还可以轻松集成到各种预测建模工作流中，从经典回归到复杂的机器学习流程皆可。其设计注重可解释性和可扩展性，因此能够轻松评估新方法，或根据特定领域的需求调整现有方法。

我们希望 **pintervals** 能为研究人员和分析师提供一种易于使用、严谨且可复现的方法，以量化预测不确定性。该软件包将易用性与严谨的统计学基础相结合，旨在弥合方法创新与实际应用之间的差距，从而使应用预测建模中的推断和决策更加可信。

参考文献

Angelo Canty, B D Ripley（2024年）。*boot*：*R*（*S-Plus*）自助法函数。R软件包版本 1.3-31。Arel-Bundock V（2025）。从模型到意义：如何使用*R*和*Python*解读统计模型。CRC出版社。Arel-Bundock V、Greifer N、McDermott G、Bacher E、Heiss A（2025）。*marginaleffects*：预测、比较、斜率、边际均值和假设检验。doi:10.32614/CRAN.package.marginaleffects。R软件包版本 0.31.0，网址 https://CRAN.R-project.org/package=marginaleffects。Balch MS、Martin R、Ferson S（2019）。“卫星交会分析与虚假置信定理。”《皇家学会会刊*A*辑》，**475**(20180565)。Barber RF、Tibshirani RJ（2025）。“统一保形预测的不同理论。”*arXiv*预印本 *arXiv:2504.02292*。

Beran R（1990年）。“校准预测区域。”《美国统计学会杂志》，85(411)，715–723。Brandt PT、Freeman JR、Schrodt PA（2014年）。“评估政治冲突动态预测。”《国际预测学杂志》，30(4)，944–962。Cali´nski T、Harabasz J（1974年）。“一种用于聚类分析的树突法。”《统计学理论与方法通讯》，3(1)，1–27。Carmichael I、Williams JP（2018年）。“虚假置信定理阐述。”《Stat》，7(1)，e201。Chadefaux T（2017年）。“冲突预测及其局限。”《数据科学》，1(1-2)，7–17。Ding T、Angelopoulos A、Bates S、Jordan M、Tibshirani RJ（2023年）。“多类别的类条件保形预测。”《神经信息处理系统进展》，36，64555–64576。Foygel Barber R、Cand`es E、Ramdas A、Tibshirani RJ（2023年）。“超越可交换性的共形预测。”《统计年刊》，51(2)，816–845。Guan L（2021年）。LCP：局部化共形预测。R软件包版本 1.0，提交版本 64e5166864379a5e53d55ca22f5fe01986cde631，网址 https://github.com/LeyingGuan/LCP。Guan L（2023年）。“局部化共形预测：保形预测的广义推断框架。”《生物计量学》，110(1)，33–50。Hegre H、Allansson M、Basedau M、Colaresi M、Croicu M、Fjelde H、Hoyles F、Hultman L、H¨ogbladh S、Jansen R 等（2019）。“ViEWS：政治暴力预警系统。”《和平研究杂志》，56(2)，155–174。Hegre H、Bell C、Colaresi M、Croicu M、Hoyles F、Jansen R、Leis MR、Lindqvist-McGowan A、Randahl D、Rød EG 等（2021年）。“ViEWS2020：修订并评估ViEWS政治暴力预警系统。”《和平研究杂志》，58(3)，599–611。Hegre H、Metternich NW、Nyg˚ard HM、Wucherpfennig J（2017年）。“引言：和平研究中的预测。”《和平研究杂志》，54(2)，113–124。Hegre H、Vesco P、Colaresi M、Vestby J、Timlick A、Kazmi NS、Lindqvist-McGowan A、Becker F、Binetti M、Bodentien T 等（2024年）。“2023/24 VIEWS预测挑战赛：预测武装冲突死亡人数及其不确定性。”《和平研究杂志》，第00223433241300862页。Hesterberg TC（2015年）。“教师应了解的自助法：本科统计学课程中的重抽样。”《美国统计学家》，69(4)，371–386。Hjort A、Hermansen GH、Pensar J、Williams JP（2025）。“使用空间加权共形预测量化自动估值模型中的不确定性。”《国际数据科学与分析杂志》，第1–18页。Hjort A、Williams JP、Pensar J（2024年）。“住房市场的聚类共形预测。”《机器学习研究论文集》，230，1–21。

Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropou- los F, Razbash S, Wang E, Yasmeen F (2025). forecast：用于时间序列和线性模型的预测函数。R软件包版本 8.24.0，网址 https://pkg.robjhyndman.com/ forecast/. Kuhn M, Wickham H (2020). Tidymodels：基于 tidyverse 原则用于建模和机器学习的软件包集合。网址 https://www.tidymodels.org. Lang M, Binder M, Richter J, Schratz P, Pfisterer F, Coors S, Au Q, Casalicchio G, Kotthoff L, Bischl B (2019). "mlr3：R 中现代的面向对象机器学习框架。"《开放源代码软件期刊》。doi:10.21105/joss.01903. 网址 https://joss.theoj. org/papers/10.21105/joss.01903. Lei J, Wasserman L (2013). "非参数回归的无分布假设预测带。"《皇家统计学会杂志》B辑：统计方法论，76(1)，71–96。ISSN 1369-7412。doi:10.1111/rssb.12021. https://academic. oup.com/jrsssb/article-pdf/76/1/71/49514328/jrsssb_76_1_71.pdf，网址 https: // doi.org/10.1111/rssb.12021. Lyu Y, Nie J, Yang SKX (2021). "利用跨国数据预测美国经济衰退期的经济增长。"《经济学快报》，198，109668。Mao H, Martin R, Reich BJ (2023). "有效的无模型空间预测。"《美国统计学会杂志》，119(546)，904–914。Martin R (2019). "虚假置信、非可加信念与有效统计推断。"《国际近似推理杂志》，113，39–73。McCartan C (2025). conformalbayes：贝叶斯模型的 Jackknife(+) 预测区间。doi:10.32614/CRAN.package.conformalbayes. R软件包版本 0.1.4，网址 https: // CRAN.R-project.org/package=conformalbayes. MIT选举数据与科学实验室 (2018). "美国 2018 年大选——分析数据集。" https://github.com/MEDSL/2018-elections-unoffical/blob/master/ election-context-2018.csv. 访问日期：2025-11-03。Nanlohy S, Butcher C, Goldsmith BE (2017). "暴行定量预测模型的政策价值。"《皇家联合军种研究所期刊》，162(2)，24–32。Petropoulos F, Apiletti D, Assimakopoulos V, Babai MZ, Barrow DK, Taieb SB, Bergmeir C, Bessa RJ, Bijak J, Boylan JE, et al. (2022). "预测：理论与实践。"《国际预测学杂志》，38(3)，705–871。R Core Team (2025). R：统计计算语言与环境。R Foundation for Statistical Computing，维也纳，奥地利。网址 https://www.R-project.org/. Raftery AE (2016). "概率预测的使用与交流。"《统计分析与数据挖掘：美国统计协会数据科学期刊》，9(6)，397–410。Randahl D, Leis M, G˚asste T, Fjelde H, Hegre H, Lindberg SI, Wilson S (2025a). "预测选举暴力。"《国际预测学杂志》。ISSN 0169-2070。doi:https:

//doi.org/10.1016/j.ijforecast.2025.09.003. 网址 https://www.sciencedirect.com/science/article/pii/S0169207025000871. Randahl D, Williams JP (2025). pintervals：模型无关预测区间。R软件包版本 0.9.2。Randahl D, Williams JP, Hegre H (2025b). “武装冲突伤亡人数的分箱条件保形预测。”《政治分析》，第 1–13 页。doi:10.1017/pan.2025.10010. Sablica L, Hornik K (2020). “mistr：混合分布和复合分布的计算框架。”《R 期刊》，12(1)，283–299。doi:10.32614/RJ-2020-003. 网址 https://journal.r-project.org/archive/2020/RJ-2020-003/index.html. Sablica L, Hornik K (2023). mistr：混合分布和复合分布。doi:10.32614/ CRAN.package.mistr. R软件包版本 0.0.6，网址 https://CRAN.R-project.org/ package=mistr. Shafer G, Vovk V (2008). “保形预测教程。”《机器学习研究杂志》，9(3)。Tian Q, Nordman DJ, Meeker WQ (2022). “计算预测区间的方法：综述与新成果。”《统计科学》，37(4)，580–597。Tibshirani RJ, Foygel Barber R, Candes E, Ramdas A (2019). “协变量偏移下的保形预测。”载于《神经信息处理系统进展》，第 32 卷。Vovk V, Gammerman A, Shafer G (2022). 《随机世界中的算法学习》。Springer Nature。Wang X, Hyndman R (2025). conformalForecast：用于多步超前时间序列预测的保形预测方法。doi:10.32614/CRAN.package.conformalForecast. R软件包版本 0.1.0，网址 https://CRAN.R-project.org/ package=conformalForecast. Williams JP (2023). “无模型广义信念推断。”arXiv 预印本 arXiv:2307.12472。Williams JP, Liu Y (2024). “通过无模型广义信念推断实现决策理论。”载于《信念函数：理论与应用》，第 14909 卷，第 131–139 页。Springer。Zhang H, Zimmerman J, Nettleton D, Nordman DJ (2020). “随机森林预测区间。”《美国统计学家》。

附录：预测区间方法的基于模拟的评估

本附录介绍一项模拟研究，旨在系统评估 **pintervals** 软件包中实现的预测区间方法的经验覆盖率特性。正文通过单次校准-测试集划分展示了这些方法，而模拟框架则评估它们在对数据进行多次随机划分时的表现。研究重点关注有效性的三个互补维度：总体覆盖率、子组覆盖率，以及结果变量不同取值范围内的覆盖率。对于每种设定，我们总共进行 1000 次模拟迭代，并在每次迭代中生成新的校准集和测试集。

在所有模拟设定中，名义覆盖水平固定为 $1 - \alpha = 0.9$。每次模拟运行均采用正文所述的相同流程构建预测区间，并使用以下指标总结表现：(i) 经验覆盖率，即真实结果落在相应预测区间内的测试观测所占比例，并对各次模拟运行取平均；(ii) 覆盖率相对于名义水平的平均绝对误差（MAE）。MAE 指标定义如下：

$$\text{MAE} = \frac{1}{G} \sum_{g=1}^{G} |\hat{c}_g - (1 - \alpha)|$$

其中，$\hat{c}_g$ 表示组 g 中的经验覆盖率，G 表示所考虑的组或分箱数量。MAE 值越低，表明覆盖率越均匀且局部有效。每次模拟迭代都会计算 MAE，然后对所有运行结果取平均，以汇总性能。

总体覆盖率模拟

第一组模拟用于评估总体覆盖率，即对测试集中的所有观测取平均，不按子组或结果范围进行条件划分。每次模拟迭代中，数据都会随机分为校准集和测试集，并使用标准保形预测、参数预测区间和基于自助法的区间构造预测区间。在此，我们将标准保形预测方法与假设服从正态分布和逻辑斯蒂分布的参数区间，以及自助法区间进行比较。

结果显示，所有四种方法的覆盖率都非常接近 0.9 的名义水平，平均绝对误差均低于 0.01。两种参数方法都略有过度覆盖，而自助法的覆盖不足程度极小。[7] 这些发现证实，保形预测方法无需依赖分布假设，即可稳健地实现有效覆盖率。

方法	平均覆盖率	MAE 覆盖率
逻辑分布	0.903	0.008
SCP	0.899	0.009
自助法	0.898	0.009
正态法	0.907	0.009

[7]通过使用不同的随机种子重新运行分析，证实了这些模式：自助法的覆盖率为 89.7–89.9%，逻辑斯蒂方法的覆盖率为 90.3–90.5%，而正态方法的覆盖率略高。

分组覆盖率模拟

第二组模拟评估了数据预定义子组内的覆盖率。这些检验旨在探究这样一种担忧：总体覆盖率正确的方法，在具有实质意义的组内仍可能出现系统性的覆盖不足或过度覆盖。

在每次模拟运行中，分别使用标准保形预测（SCP）、Mondrian 共形预测（MCP）、聚类共形预测（CCP）和距离加权保形预测（DWCP）构建预测区间，同时还构建自助法区间和参数区间（逻辑分布和正态法）。

组成员资格根据地理分类确定，对应于美国人口普查地区和分区，具体如正文所述。在每个组内分别计算覆盖率，同时跨组计算覆盖率平均绝对误差（MAE of coverage），以量化覆盖率的均匀性。

结果表明，尽管所有方法在总体上都实现了接近名义水平的经验覆盖率，但与标准保形预测、自助法区间和参数区间相比，分组感知保形预测方法（MCP、CCP、DWCP）的分组覆盖率更接近名义水平。具体而言，MCP、CCP 和 DWCP。此外，与标准保形预测、自助法区间和参数区间相比，这些方法在各地区的覆盖率平均绝对误差显著更低。这表明，分组感知方法能够在子群体内提供更加均匀且可靠的覆盖率，从而应对数据中潜在的异质性。

	按区域划分的经验覆盖率				总体指标	
方法	中西部	东北部	南部	West	平均覆盖率	MAE 覆盖率
MCP	0.898	0.890	0.898	0.894	0.897	0.022
CCP	0.904	0.884	0.894	0.880	0.895	0.024
DWCP	0.942	0.901	0.877	0.883	0.901	0.028
自助法	0.948	0.927	0.873	0.844	0.898	0.040
正态法	0.956	0.937	0.882	0.851	0.907	0.040
SCP	0.949	0.929	0.873	0.847	0.899	0.040
逻辑分布	0.952	0.934	0.877	0.849	0.903	0.040

目标模拟不同范围内的覆盖率

第三组模拟聚焦于结果变量不同范围内的覆盖率。当预测误差在结果空间中呈系统性变化时，例如在分布尾部，这一情景尤其具有相关性。

在每次模拟迭代中，根据观测到的结果将校准数据划分为若干分箱，并选择阈值，使四个分箱中的观测数量大致均衡。随后，采用分箱条件保形预测（BCCP）构造预测区间，包括连续化和非连续化两种变体；同时还构造标准保形预测、参数法和基于自助法的区间。

分别计算每个结果分箱内的覆盖率。与分组模拟一样，使用平均总体覆盖率以及分箱覆盖率相对于名义水平的平均绝对误差来汇总性能。

结果表明，与标准保形预测相比，连续化 BCCP 和非连续化 BCCP 在每个结果分箱内的经验覆盖率都更接近名义水平，boot-

strap 和参数预测区间相比，连续化和非连续化 BCCP 在每个结果分箱内实现的经验覆盖率都更接近名义水平。各分箱的覆盖率平均绝对误差以连续化 BCCP 方法最低，非连续化版本紧随其后。这些发现凸显了 BCCP 在处理结果空间中预测误差异质性方面的有效性，从而实现更可靠的不确定性量化。正如预期，由于连续化的影响，连续化 BCCP 的总体覆盖率略高于名义水平。

	按结果分箱的经验覆盖率				总体指标	
方法	1	2	3	4	平均覆盖率	平均绝对误差覆盖率
BCCP(d)	0.895	0.898	0.899	0.894	0.897	0.019
BCCP(c)	0.895	0.909	0.912	0.894	0.906	0.020
正态法	0.804	0.960	0.937	0.779	0.907	0.079
逻辑分布	0.796	0.957	0.934	0.773	0.903	0.081
自助法	0.796	0.954	0.929	0.760	0.898	0.082
SCP	0.788	0.955	0.932	0.768	0.899	0.083

所属机构：
David Randahl 瑞典国防大学战争研究系 Drottning Kristinas V¨ag 47 114 28 斯德哥尔摩，瑞典
电子邮箱：david.randahl@fhs.se网址：http://randahl.org/david