arXiv:2602.10018v1 [stat.ME] 10 Feb 2026

采用非对称选择规则的在线选择性保形预测：置换检验方法

郑明义和金莹*

摘要

选择性保形预测旨在构造预测集，使得对于由数据驱动机制选中的测试单元，其覆盖率有效。现有离线设置下的方法能够处理任何对带标签数据具有置换不变性的选择机制，但将其扩展到在线设置——数据按顺序到达，后续决策取决于先前决策——面临着一项挑战：选择机制天然具有非对称性。因此，现有方法只能处理有限范围的选择机制。

本文提出基于排列的 Mondrian 保形推断（PEMI），这是一个通用的基于排列的框架，适用于采用任意非对称选择规则的选择性保形预测。受全保形预测和孟德尔共形预测的启发，PEMI 找出所有会产生相同选择事件的观测数据排列（或其蒙特卡洛子集），并利用这一保持选择事件不变的参考集上的符合度分数校准预测集。在标准可交换性条件下，对于任何非对称选择机制和任何预测模型，我们的预测集都能实现有限样本精确选择条件覆盖率。PEMI 可自然纳入额外的离线标注数据，扩展至涉及多个测试样本的选择机制，并借助细粒度选择分类体系实现 FCR 控制。我们还针对常用的在线选择规则提出了若干高效的具体实现，包括基于协变量的规则、基于共形 p 值/e 值的程序，以及基于早期结果的选择。最后，我们在真实药物发现数据集上展示了这些方法在各种选择规则下的有效性，并通过模拟研究其性能。

1 引言

共形预测是一种用于量化黑箱预测模型不确定性的一般框架。给定一组带标签（校准）数据，在假设测试样本与带标签数据可交换的前提下，该框架会构建一个预测集，使其以预先设定的概率覆盖新（测试）样本的未知标签（Vovk 等人，2005）。然而，在许多场景中，从业者可能只关注部分测试点，而且是否发布预测集这一决定本身也是由数据驱动的。例如，科学家可能会在候选药物的预测活性较高时，查看其未知活性的预测集（Svensson 等人，2018）；当多个深度学习模型意见不一致时，临床医生可能会寻求医学诊断的不确定性量化（Loftus 等人，2022）。这种选择性发布会破坏标准共形预测的有效性，使其无法再保证被选实例的覆盖率（Jin and Ren，2024）。

这引出了选择性保形预测的问题：给定任何决定是否查询预测集的数据驱动过程，如何为被查询样本构造具有有效覆盖率的预测集？形式上，假设可以访问带标签的（校准）数据 $\{Z_i\}_{i=1}^{t-1}$，其中 $Z_i = (X_i, Y_i)$ 包含特征 $X_i \in \mathcal{X}$ 和标签 $Y_i \in \mathcal{Y}$，以及一个观测到特征 $X_t \in \mathcal{X}$ 但标签未知的新测试单元 $Y_t \in \mathcal{Y}$。选择机制是一个固定映射 $\mathcal{S}_t\colon (\mathcal{X} \times \mathcal{Y})^{t-1} \times \mathcal{X} \to \{0, 1\}$，它利用所有可用数据来

*宾夕法尼亚大学统计与数据科学系。电子邮件：yjinstat@wharton.upenn.edu。本文实验结果的复现代码可见于 https://github.com/mingyi811/PEMI。

通过指示函数 $S_t := \mathcal{S}_t(Z_1, \ldots, Z_{t-1}, X_t)$ 决定是否发布预测集。给定置信水平 $\alpha \in (0, 1)$，目标是构造一个满足以下条件的预测集 $\mathcal{C}_{\alpha,t} \subseteq \mathcal{Y}$

$$\mathbb{P}(Y_t \in \hat{\mathcal{C}}_{\alpha,t} \mid S_t = 1) \geq 1 - \alpha, \tag{1}$$

其中，（条件）概率是针对所有数据点的联合分布而言的。遵循相关文献（Jin and Ren，2024），我们将（1）称为选择条件覆盖率（SCC）。本文唯一的假设是可交换性：

假设 1。对于任意 $t \in \mathbb{N}^+$，序列 $(Z_1, \ldots, Z_t)$ 均为可交换的。

为清晰起见，我们首先关注在线设置，其中单个测试点在时刻 t到达；这一设置在选择性保形预测领域日益受到关注（Bao 等人，2024a；Sale and Ramdas，2025；Humbert 等人，2025）。随后我们将展示，我们的框架还可扩展至：(i) 纳入额外的离线数据（第 2.4节）；(ii) 在包含多个测试样本的标准离线设置中处理非对称选择规则（第 2.6节）；以及 (iii) 对某些类别的选择规则实现错误覆盖率（FCR）控制（第 2.5节）；上述结果均严格推广了文献中的现有结果。

选择性保形预测问题最初是在离线设置下研究的，其中选择机制会从多个测试点中挑选一个子集（Bao 等人，2024b；Jinand Ren，2024；Gazin 等人，2025）。已有研究指出，关键挑战在于，给定选择性发布（即以 $S_t = 1$ 为条件，如式（1）所示），带标签数据与未标记数据之间的可交换性不再成立，从而破坏了标准共形预测的有效性。为解决这一问题，现有方法通常会构造一个由带标签数据组成的“参考集”，使其在给定选择条件下仍与测试点可交换，并且仅使用该子集来校准预测集。这通常依赖某种“交换”技术，而 Jin and Ren （Jin and Ren （2024）提出的联合 Mondrian 保形推断（JOMI）方法对其进行了形式化。具体而言，参考集定义为这样一个带标签数据子集：将其中的数据与一个被选中的测试点交换后，会得到相同的选择事件。对于任何对1具有置换不变性的选择规则，JOMI 都能实现式（$\{Z_i\}_{i=1}^{t-1}$）。

在线设置中的一个根本难题是，选择规则 $\mathcal{S}_t$ 本质上具有非对称性：带标签的校准数据的顺序会影响选择轨迹，进而影响在时间 t是否进行选择。这打破了早期离线方法所需的唯一对称性假设。为应对这一挑战，现有将这种“交换”策略应用于在线非对称选择规则的解决方案，必须对选择规则以及“可交换”校准数据子集的构成方式施加强约束（Bao 等人，2024a；Sale and Ramdas，2025）。因此，这些方法（i）仅适用于一类范围狭窄的“决策驱动”规则，即时间 t 的决策仅取决于过去的决策 $\{S_i\}_{i=1}^{t-1}$ 和当前的X_t；（ii）可能产生无效或过于保守的预测集。

1.1 我们的方法：对排列进行选择性推断

本文提出基于排列的 Mondrian 保形推断（PEMI），这是一个一般框架，可在具有任意非对称选择规则的选择性保形预测中实现 SCC。我们的核心思路是对数据的排列进行选择性推断，而不是通过成对交换对单个校准数据点进行推断。具体而言，对于观测数据的一组 Π排列（Z_1, ……, Z_t），我们找出其中会导致在时间步 t发生相同选择事件的排列子集，并将其称为参考集。随后，我们按照共形预测的常规方法校准预测集 $C_{\alpha,t}$：仅当假设标签值 $y \in \mathcal{Y}$ 所产生的$C_{\alpha\text{—致性分数}t}$ 处于所选定排列下各分数所界定的常规范围内时，才将其纳入预测集。图1 对此作了可视化展示。

在第 2.3 节中，我们证明，Π 有两种取值方式可以使 PEMI 预测集具有有限样本 SCC：一种是 Π 为 $\{1, \ldots, t\}$上的全体排列，另一种是 Π 为蒙特卡罗样本，即从所有排列中随机抽取的子集。在这两种情况下，PEMI 的基本原理都与选择性推断中的条件推断方法相关（Lee 等人，2016；Taylor 和 Tibshirani，2016；Markovic

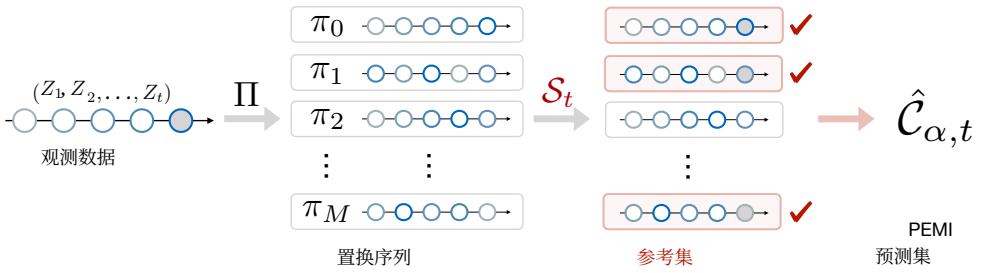

图 1：所提出的 PEMI 方法示意图。从观测数据的一组排列 Π 出发，我们找出经过排列后会产生相同选择事件的子集。随后，根据该排列子集对最终预测集进行校准。

等人，2017；Reid 等人，2017；Tibshirani 等人，2018），其中推断基于给定选择事件条件下检验统计量的分布。这里的基本随机对象是导致观测数据的随机排列。形式上，令无序数据的“数据袋” $[z_1, \ldots, z_t]$ 表示由实现的数据值构成的无序集，其中每个 $z_i \in \mathcal{X} \times \mathcal{Y}$ 都是一个固定值。在可交换性假设下，排列$\hat{}\pi$ 满足 $(Z_1, \ldots, Z_t) = (z_{\hat{\pi}(1)}, \ldots, z_{\hat{\pi}(t)})$ 时，在所有排列构成的集合上服从均匀分布。当Π 是全体排列构成的集合时，我们利用这样一个事实：给定选择事件后，$\hat{}\pi$ 在参考集上服从均匀分布。当 Π 是随机抽取的子集时，覆盖保证来自$\hat{}\pi$ 与其他随机抽取的排列之间的可交换性。

因此，我们的关键技术可以视为将 JOMI 中的交换思想（Jin and Ren，2024）推广到排列层面。具体而言，JOMI 方法会寻找所有这样的带标签数据 $Z_i = (X_i, Y_i)$：若将其视为测试点（即与 Z_t交换），便会产生相同的选择事件。事实上，交换（Z_i，Z_t）是（Z_1，.....，Z_t）的一种特殊排列；当选择规则具有对称性时，交换便是唯一需要考虑的因素（备注 1）。然而，当带标签数据的顺序会产生影响时，我们必须在排列层面进行推理，才能恢复可交换性。这一思路建立在将全共形预测（FCP）视为置换检验的视角之上；详见第2.1 节。

1.2 结果预览

PEMI 框架详见第 2 节，我们将在该节中为涉及单个未标记数据点的在线选择规则构建主要方法与理论。随后，我们将讨论一些自然扩展：(i) 纳入离线数据；(ii) 处理适用于多个未标记数据点的非对称选择规则；以及 (iii) 针对某些选择规则实现错误覆盖率控制。所有这些结果都严格推广了现有文献中的结论。

我们的一般框架要求针对每个插补标签计算置换参考集$y \in \mathcal{Y}$，以处理第 t 个数据点；对于连续值响应变量，这在计算上可能难以处理。在第 3 节中，我们将针对各种在线或非对称选择规则，构建 PEMI 的计算高效实例化方案，包括：

(i) 基于协变量的规则，其中 $\mathcal{S}_t$ 仅涉及协变量 $\{X_i\}_{i=1}^t$。这涵盖了早期研究中探讨的由决策驱动的选择规则（Bao 等人，2024a；Sale and Ramdas，2025），以及基于协变量评分的加权分位数/平均值进行的选择，和任何黑箱优化程序。(ii) 基于共形 p/e 值的在线多重检验程序，例如 Xu and Ramdas (2023)。(iii) 基于先前标签加权分位数的选择。

在第4节中，我们通过一个药物发现数据集，展示了 PEMI 对多种实用在线选择规则的有效性和效率。仅需少量带标签数据，PEMI 即可生成有限样本预测集，并在所有设置下实现有效的选择条件覆盖率。在第5节中，我们还将

通过模拟研究 PEMI 的行为，并分析选择规则和符合性分数中的若干因素对其性能的影响，同时将其与原始共形预测进行比较。

1.3 相关工作

本文丰富了选择性共形预测领域的相关研究（Bao 等人，2024b；Jinand Ren，2024；Gazin 等人，2025；Bao 等人，2024a；Sale andRamdas，2025）。在离线设置中，与我们工作最接近的是 JinandRen （2024），其方法实现了选择条件覆盖率，但仅适用于对称选择规则。我们的方法将 JOMI 严格推广到非对称选择规则，将参考集的构造对象从数据点扩展到排列，并利用可交换性实现选择条件覆盖率。我们还表明，在对称选择规则下，PEMI 可简化为 JOMI（备注 1）。

许多现有的在线设置下的选择性共形预测方法，包括 Bao 等人（2024a）和 Sale and Ramdas（2025），都依赖基于交换的机制来寻找“参考集”，其概念与 Jin and Ren（2024）的方法相似。具体而言，Bao 等人（2024a）提出了 CAP 程序，根据自适应选取的带标签数据子集构造预测集，但其方法仅适用于所谓的决策驱动型（即 $\mathcal{S}_t$ 仅涉及 $\{S_i\}_{i<t}$ 和 X_t）和对称的协变量依赖型选择规则。近期提出的 EXPRESS 方法（Sale and Ramdas，2025）改进了 CAP，实现了精确 SCC，但仍仅适用于同一类决策驱动型选择规则。此外，为确保有效性，EXPRESS 对参考集（由数据点构成）的选取施加了严格要求，这可能导致预测集空洞且长度无限。相比之下，我们的实证结果表明，PEMI 能降低空洞预测集出现的频率，从而提升实用性，同时扩大所涵盖的选择规则范围。最后，另一种方法是将自适应保形推断（ACI）的思路（Gibbs 和 Candes，2021）扩展至选定样本：该方法根据过去在选择时间点的表现更新符合度分数阈值，以提供渐近 SCC 和错误发现率控制，或更强的瞬时错误率控制（Bao 等人，2024a；Humbert 等人，2025）。尽管这些方法提供了渐近 SCC 和 FCR 控制，我们的目标是实现有限样本精确选择条件覆盖率，并利用数据中的可交换性结构。我们的实验（第 4 节）表明，这些方法的表现可能对超参数选择敏感，也容易受到近似误差的影响；而在可交换性假设下，我们的方法展现出稳健的性能。

除 SCC 外，选择性保形预测的另一个目标是控制错误覆盖率（FCR）（Weinstein 和 Ramdas，2019；Xu and Ramdas，2023；Humbert 等人，2025）。Jin and Ren（2024）通过更细致的选择分类体系，针对对称选择规则在离线设置下讨论了 FCR 控制。在在线设置下，Bao 等人（2024a）和 Humbert 等人（2025）建立了某种渐近 FCR 控制，而 Sale and Ramdas（2025）则表明，EXPRESS 对决策驱动的选择规则实现了 FCR 控制。正如我们稍后将展示的，我们的框架可以推广到在类似条件下实现 FCR 控制。与经典选择性推断中控制 FCR 的方法（Benjamini 和 Yekutieli，2005；Weinstein 和 Ramdas，2019；Xu 等人，2024）相比，由于我们处理的是预测推断问题，设置和方法有显著不同。

我们的工作也与经典选择后推断（POSI）中关于条件推断的文献密切相关（Lee 等人，2016；Taylor 和 Tibshirani，2016；Markovic 等人，2017；Reid 等人，2017；Tibshirani 等人，2018）。具体而言，许多现有方法着重于构造在给定选择事件条件下所选参数的置信区间。我们的方法关注的是构造预测集，而非模型参数的置信集；它利用数据中固有的可交换性结构来实现选择条件覆盖率，而不是依赖检验统计量的分布。因此，我们的研究在设定和方法论上都与 POSI 有根本区别。

2 一般方法

2.1 预热：作为置换检验的全共形预测

我们首先回顾全共形预测（FCP）程序（Vovk 等人，2005），并将其解释为置换检验。这一视角并非全新，但有助于引出我们的方法（Angelopoulos 等人，2025）。它还将 FCP 扩展到非对称一致性分数。

为与在线记号保持一致，我们令 $n = t-1$，并将 $\mathcal{D}_n = \{(X_i, Y_i)\}_{i=1}^n$ 记为时刻 t的带标签数据。增广数据记作 $\mathcal{D}_{n+1}^y = (Z_1, \ldots, Z_n, Z_{n+1}^y)$，其中 $Z_i = (X_i, Y_i)$，且对于假设值 $y \in \mathcal{Y}$，$Z_{n+1}^y = (X_{n+1}, y)$。一致性分数是一个函数 $\mathcal{V}$，它将数据点 (x, y) 和数据集 $\mathcal{D}$ 映射为 $\mathcal{V}((x, y); \mathcal{D}) \in \mathbb{R}$。FCP 假设 $\mathcal{V}$是对称的，也就是说，对于 $\mathcal{D}$ 中元素的任意排列 π （将 $\mathcal{D}$ 视为有序序列），都有 $\mathcal{V}((x, y); \mathcal{D}) = \mathcal{V}((x, y); \mathcal{D}_\pi)$。

对于每个 $i \in [n]$，我们定义 $V_i^y = \mathcal{V}\big((X_i, Y_i); \mathcal{D}_{n+1}^y\}\big)$，以及 $V_{n+1}^y = \mathcal{V}\big((X_{n+1}, y); \mathcal{D}_{n+1}^y\big)$. 然后，FCP 集可通过对保序 p 值求逆得到：

$$\hat{\mathcal{C}}_{\alpha,n+1} = \{y \in \mathcal{Y} : p_{\text{FCP}}^y > \alpha\}, \quad \text{where} \quad p_{\text{FCP}}^y := \frac{1 + \sum_{i=1}^n \mathbb{1}\{V_i^y \geq V_{n+1}^y\}}{n+1}.$$

直观而言，FCP 可以解释为对假设值进行插补 y，用于 Y_{n+1}，并评估（X_{n+1}, y）是否与观测数据 $\{(X_i, Y_i)\}_{i=1}^n$相符。这一视角与置换检验背后的核心思想密切契合（Angelopoulos 等人，2025），下面我们将对此进行阐述。

给定检验统计量 $T\colon \mathcal{Z}^{n+1} \to \mathbb{R}$ 以及（有序）数据 $(Z_1, \ldots, Z_{n+1})$，置换检验的 p 值为

$$p_{\text{perm}} = \frac{\sum_{\pi \in \Pi_{n+1}} \mathbb{1}\{T(Z_{\pi(1)}, \ldots, Z_{\pi(n+1)}) \geq T(Z_1, \ldots, Z_{n+1})\}}{(n+1)!},$$

其中，Π_{n+1}是 $\{1, \ldots, n+1\}$ 上所有排列的集合。现在，考虑检验统计量 $T(z_1, \ldots, z_{n+1}) = \mathcal{V}(z_{n+1}; (z_1, \ldots, z_{n+1}))$，其中任意 $(z_1, \ldots, z_{n+1}) \in \mathcal{Z}^{n+1}$。由于 $\mathcal{V}$ 的对称性，对于任意 $\pi \in \Pi$，

$$T(Z_{\pi(1)}, \ldots, Z_{\pi(n+1)}) = \mathcal{V}\big(Z_{\pi(n+1)}; (Z_{\pi(1)}, \ldots, Z_{\pi(n+1)})\big) = \mathcal{V}\big(Z_{\pi(n+1)}; \mathcal{D}_{n+1}\big) = V_{\pi(n+1)}^{Y_{n+1}}.$$

使用这一特定统计量 T 得到的置换 p 值可简化为

$$\begin{aligned} p_{\text{perm}} = \frac{\sum_{\pi \in \Pi_{n+1}} \mathbb{1}\{V_{\pi(n+1)}^{Y_{n+1}} \geq V_{n+1}^{Y_{n+1}}\}}{(n+1)!} &= \frac{\sum_{i=1}^{n+1} n! \cdot \mathbb{1}\{V_i^{Y_{n+1}} \geq V_{n+1}^{Y_{n+1}}\}}{(n+1)!} \\ &= \frac{1 + \sum_{i=1}^n \mathbb{1}\{V_i^{Y_{n+1}} \geq V_{n+1}^{Y_{n+1}}\}}{n+1} = p_{\text{FCP}}^{Y_{n+1}}. \end{aligned}$$

换言之，在 $T(\cdot)$ 下的置换 p 值等价于填入真实值 Y_{n+1} 后的 FCP p 值。因此，p^y FCP 可视为一种针对 $(Z_1, \ldots, Z_n, (Z_{n+1}, y))$ 可交换性的置换检验，用于确定每个 $y \in \mathcal{Y}$是否应纳入预测集。

采用非对称评分的置换 FCP。 对于非对称的一致性分数，同样的置换检验构造仍然合理。形式上，考虑任意对数据排序敏感的一致性分数 $\mathcal{V}((x,\ y); \mathcal{D})$，其中 $\mathcal{D}$。对于任意假设值 $y \in \mathcal{Y}$，我们将 $\mathcal{D}_{n+1}^y = (Z_1,\ \ldots\ldots,\ Z_n,\ Z_{n+1}^y)$ 及其第 i 个元素记为 Z_i^y，并将 $V_i^y = \mathcal{V}(Z_i^y; \mathcal{D}_{n+1}^y)$ 记为 $V_i^y = \mathcal{V}(Z_i^y; \mathcal{D}_{n+1}^y)$。对于任意排列 $\pi \in \Pi_{n+1}$，我们记作$V_{\pi,\ n+1}^y := \mathcal{V}(Z_{\pi(n+1)}^y; \mathcal{D}_\pi^y)$，其中 $\mathcal{D}_\pi^y = (Z_{\pi(1)}^y \cdots,\ Z_{\pi(n+1)}^y)$。那么，上述论证意味着

$$\hat{\mathcal{C}}_{\alpha,n+1}^{\text{perm}} := \{y \in \mathcal{Y}\colon p_{\text{perm}}^y > \alpha\}, \quad \text{where} \quad p_{\text{perm}}^y = \frac{\sum_{\pi \in \Pi_{n+1}} \mathbb{1}\{V_{\pi,n+1}^y \geq V_{n+1}^y\}}{(n+1)!} \tag{2}$$

可在可交换性条件下实现有限样本覆盖率： $\mathbb{P}(Y_{n+1} \in \hat{\mathcal{C}}_{\alpha}^{\mathrm{perm}}(Z_{1:n}, X_{n+1})) \geq 1-\alpha$。然而，与对称情形不同，$p_{\mathrm{perm}}^{y}$ 不一定能简化为基于秩的表达式，因此最好将这一表述视为一种概念性工具。这一观察启发了 PEMI：我们将比较单位从单个数据“提升”到排列，但现在仅限于那些能够保留选择决策的排列。

2.2 PEMI：一般流程

现在我们可以介绍 PEMI 框架了，它可以被视为一种具有受限参考类的置换检验。从现在起，我们恢复使用时间点 $t \in \mathbb{N}^{+}$ 的记号。在时间 t，我们考虑 $\{1, \ldots, t\}$ 上的一组排列 $\Pi_t^{\star}$。对于候选标签 $y \in \mathcal{Y}$，我们在 $\Pi_t^{\star}$ 中会导向相同选择决策的排列子集上计算置换检验的 p 值，并对其求逆以获得预测集。

选择 $\Pi_t^{\star}$。遵循置换检验的思路，一个自然的想法是令 $\Pi_t^{\star} = \Pi_t$ 为 $\{1, \ldots, t\}$ 上所有排列的集合。为确保计算可行性，我们在此关注随机化版本，其中 $\Pi_t^{\star}$ 是从均匀分布 $\mathrm{Unif}(\Pi_t)$ 中抽取的 $M \in \mathbb{N}^{+}$ 个排列所构成的独立同分布样本，记为

$$\Pi_t^{(M)} = \{\pi^{(1)}, \ldots, \pi^{(M)}\},$$

且 $M \in \mathbb{N}^{+}$ 是任意预先固定的正整数。重要的是，PEMI 的保证对任意固定的 M 均成立。

在第 2.2 节和第 2.3 节中，选择规则是任意映射 $\mathcal{S}_t\colon \mathcal{Z}^{t-1} \times \mathcal{X} \to \{0,1\}$，预测集则依赖于用户指定的一致性分数 $\mathcal{V}_t\colon \mathcal{Z}^{t} \to \mathbb{R}$；二者都对参数中的数据排序敏感。例如，对于预训练预测模型 $\hat{\mu}\colon \mathcal{X} \to \mathcal{Y}$ 和 $\mathcal{Y} = \mathbb{R}$，可以直接设定 $\mathcal{V}_t(Z_1, \ldots, Z_t) = |Y_t - \hat{\mu}(X_t)|$。对于任意 $y \in \mathcal{Y}$，令 $Z_t^{y} = (X_t, y)$ 表示增广数据。

PEMI 的第一步是找出那些能够保持选择决策不变的 $\pi \in \Pi_t^{\star}$，也就是说，在 $\mathcal{S}_t$ 下对数据进行排列会得到相同的选择事件。形式上，对于任意 $\pi \in \Pi^{\star}$，我们将排列后的数据定义为

$$\mathcal{D}_{\pi,t}^{y} = \left(Z_{\pi(1)}^{y}, Z_{\pi(2)}^{y}, \ldots, Z_{\pi(t-1)}^{y}, X_{\pi(t)}\right), \tag{3}$$

其中，对于 $j \in [t-1]$，我们定义排列后的插补数据点

$$Z_{\pi(j)}^{y} = \begin{cases} Z_t^{y}, & \pi(j) = t, \\ Z_{\pi(j)}, & \pi(j) \neq t. \end{cases}$$

这里我们对增广数据集的先前记法稍作调整，使得现在 $\mathcal{D}_{\pi,t}^{y}$ 记录用于在时间点 t（排列后）做出选择性发布决策的信息。

对于任意排列 $\pi \in \Pi_t^{\star}$，我们将排列后数据的选择决策定义为 $S_t^{y}(\pi) = \mathcal{S}_t(\mathcal{D}_{\pi,t}^{y})$。于是，置换参考集定义为

$$\hat{R}_t(y; \Pi_t^{\star}) = \{\pi \in \Pi_t^{\star} : S_t^{y}(\pi) = 1\} \cup \{\pi_0\}, \tag{4}$$

其中 π_0 表示恒等映射。此外，回顾排列后数据的一致性分数 $V_t^{y}(\pi) = \mathcal{V}_t(Z_{\pi(1)}^{y}, Z_{\pi(2)}^{y}, \cdots, Z_{\pi(t)}^{y})$。最后，PEMI 预测集构造如下

$$\hat{\mathcal{C}}_{\alpha,t}(\Pi_t^{\star}) = \left\{y \in \mathcal{Y} : p_t(y; \Pi_t^{\star}) > \alpha\right\}, \quad \text{where} \quad p_t(y; \Pi_t^{\star}) = \frac{\sum_{\pi \in \hat{R}_t(y;\Pi_t^{\star})} \mathbb{1}\{V_t^{y}(\pi_0) \leq V_t^{y}(\pi)\}}{|\hat{R}_t(y; \Pi_t^{\star})|}. \tag{5}$$

式（5）中的 p 值可以视为基于所选置换参考集的置换检验（2）。正如我们将在第 2.3 节的理论中看到的，恰当地选择 $\Pi_t^{\star}$ 后，这种调整能够恢复置换检验在给定选择事件条件下的有效性，从而实现选择条件有效性：即使一致性分数 $\mathcal{V}_t$ 和选择规则 $\mathcal{S}_t$ 都对数据排序敏感，这一结论仍然成立。

为实现精确覆盖率，可以计算随机化 PEMI 预测集。给定用于打破并列值的独立同分布随机变量 $U_1, \ldots, U_t \sim$ 均匀分布 $([0,1])$，我们设 $\hat{\mathcal{C}}^{\mathrm{rand}}_{\alpha\, t}(\Pi^\star_t) = \{y \in \mathcal{Y}\colon\ p^{\mathrm{rand}}_t(y;\ \Pi^\star_t)\ > \alpha\}$，其中

$$p^{\mathrm{rand}}_t(y;\Pi^\star_t) = \frac{\sum_{\pi\in\hat{R}_t(y;\Pi^\star_t)} \mathbb{1}\{V^y_t(\pi_0) < V^y_t(\pi)\} + U_t \cdot \sum_{\pi\in\hat{R}_t(y;\Pi^\star_t)} \mathbb{1}\{V^y_t(\pi_0) = V^y_t(\pi)\}}{|\hat{R}_t(y;\Pi^\star_t)|}. \tag{6}$$

备注 *1.* 我们的方法，尤其是带有 $\Pi^\star_t = \Pi_t$ 的“完整参考集”版本，是对 *JOMI*（*Jin* 和 *Ren*，*2024*）的严格推广。为说明这一点，我们固定一个时间点 t，并令 $\mathcal{S}(\cdot)$ 为任意对称选择规则，用以决定第 t 个数据点的选择，该规则对前 $t-1$ 个数据点置换不变，以及一个仅涉及测试点的一致性分数 $V\colon\ \mathcal{Z} \to \mathbb{R}$。我们再次如 *(3)* 中那样表示排列后的数据。*JOMI* 框架对每个 $y \in \mathcal{Y}$ 识别出一个子集 $R_{JOMI}(y) \subseteq \{1, \ldots, t-1\}$，该子集由数据点 $i \in [t-1]$ 组成，使得将 Z_i 与 (X_{n+1}, y) 交换时会导致相同的选择事件。预测集随后由 $\hat{\mathcal{C}}^{JOMI}_{\alpha,t} = \{y\colon\ V(X_{n+1}, y) \le$ 分位数$(1-\alpha;\ \{V(X_i, Y_i)\}_{i\in \hat{R}_{JOMI}(y)} \cup \{+\infty\})\}$ 定义。事实上，在对称选择函数 $\mathcal{S}_t = \mathcal{S}$ 下，可以证明 *(4)* 中的 *PEMI* 参考集由 $\hat{R}_t(y) = \cup_{i\in \hat{R}_{JOMI}(y)\cup\{t\}}\{\pi \in \Pi\colon\ \pi(i) = t\}$ 给出。在一致性分数 $\mathcal{V}_t(z_1, \ldots, z_t) := V(z_t)$ 下，*PEMI* 预测集因此与 *JOMI* 预测集一致。在完全一般的情形下，*JOMI* 框架处理的是离线设置，其中选出多个测试点的一个子集；我们在第 *2.6* 节讨论 *PEMI* 向多个测试点的扩展，该扩展同样以类似方式推广了 *JOMI*。

2.3 理论保证

定理 1 确立了上述预测集的有限样本选择条件有效性，且不对选择机制作任何假设。

定理 1. 假设数据 $\{Z_i\}_{i=1}^t$ 可交换。对于任意选择规则 $\mathcal{S}_t\colon\ \mathcal{Z}^{t-1} \times \mathcal{X} \to \{0,1\}$ 和任意一致性分数 $\mathcal{V}_t\colon\ \mathcal{Z}^t \to \mathbb{R}$，以及上述所述的 $\Pi^\star_t \in \{\Pi_t, \Pi^{(M)}_t\}$，有

$$\mathbb{P}\big(Y_t \in \hat{\mathcal{C}}_{\alpha,t}(\Pi^\star_t) \,\big|\, S_t = 1\big) \ge 1 - \alpha, \quad \forall\, t \ge 1. \tag{7}$$

此外，

$$\mathbb{P}\big(Y_t \in \hat{\mathcal{C}}^{\mathrm{rand}}_{\alpha,t}(\Pi^\star_t) \,\big|\, S_t = 1\big) = 1 - \alpha, \quad \forall\, t \ge 1. \tag{8}$$

值得注意的是，在 $\Pi^\star_t = \Pi^{(M)}_t$ 的情况下，对于任意有限的 $M \in \mathbb{N}^+$，覆盖保证都能精确成立，无需依赖渐近蒙特卡罗近似。较大的 M 可提高 p 值的分辨率，但需要更多计算。

我们定义详细证明见附录B.1，并在此提供一些直观解释。认识到 t

$$Y_t \in \hat{\mathcal{C}}_{\alpha,t}(\Pi^\star_t) \quad \Leftrightarrow \quad p_t(Y_t;\Pi^\star_t) > \alpha,$$

关键思路是证明，使用 $\hat{}\ R_t(Y_t;\ \Pi^\star_t)$ 的置换检验在给定选择事件的条件下具有有效性。针对 $\Pi^\star_t$ 的两种选择，具体技术略有不同，我们将分别讨论。

为了证明全参考变体 $\hat{\mathcal{C}}_{\alpha,\ t}(\Pi_t)$ 的有效性，我们证明一个更强的结果

$$\mathbb{P}\big(p_t(Y_t;\Pi_t) \le \alpha \mid [\mathcal{D}_t], S_t = 1\big) \le \alpha,$$

其中，我们定义完整观测值的无序数据袋 $[\mathcal{D}_t] = [Z_1, \ldots, Z_t]$。对于无序集 $[d_t] = [z_1, \ldots, z_{t-1}, z_t]$ 的任意实现值，在给定 $[\mathcal{D}_t] = [d_t]$ 的条件下，唯一的随机性来自 $(Z_1$、……、$Z_t)$ 相对于 $(z_1$、……、$z_t)$ 的排列。由可交换性可知，$\hat{}\pi$——使得 $(Z_1$、……、$Z_t) = (z_{\hat{\pi}(1)}$、……、$z_{\hat{\pi}(t)})$ 的随机排列——在 Π_t 上服从均匀分布。接着，我们证明，在给定 $[\mathcal{D}_t] = [d_t]$ 和 $S_t = 1$ 的条件下，$\hat{}\pi \sim$ 服从均匀分布 Unif$(R_t(Y_t;\Pi_t))$（事实上，我们的构造确保

$\hat{} R_t(Y_t;\ \Pi_t)$ 仅由 $[\mathcal{D}_t]$ 决定。借助贝叶斯法则和塔式性质，即可得到定理 1 中的 (a)。

$\hat{}\mathcal{C}_{\alpha,t}(\Pi_t^{(M)})$ 和 $\hat{}\mathcal{C}_{\alpha\ t}^{\text{rand}},(\Pi_t^{(M)})$ 的有效性依赖于一种略有不同的结构：$\hat{}\pi$（使得 $(Z_1,\ldots,Z_t) = (^{z}_{\hat{\pi}(1)},\ldots,{}^{z}_{\hat{\pi}(t)})$ 的排列）与随机抽取的 $\{\pi^{(m)}\}_{m=1}^{M}$ 之间的可交换性。由于 $\{\pi^{(m)}\}_{m=1}^{M}$ 是从 Π_t 中独立且均匀抽取的，我们能够证明排列 $\{\hat{}\pi,\ \pi^{(1)}\circ\hat{}\pi,\ldots,\ \pi^{(M)}\circ\hat{}\pi\}$ 是可交换的。随后，我们利用这种可交换性，并依赖于 $\hat{}\pi$ 在相应的排列参考集 $\{\hat{}\pi,\ \pi^{(1)}\circ\hat{}\pi,\ldots,\ \pi^{(M)}\circ\hat{}\pi\}$ 中的选择条件分布，在定理 1 中证明它们的有效性。

完成核心一般框架的介绍后，接下来的三个小节将探讨 PEMI 的若干自然推广。这些结果有助于与现有方法建立联系，并表明 PEMI 严格推广了先前的方法。若读者想了解如何使用具体的非对称选择规则实现 PEMI，可直接阅读第 3节，不会错过关键概念。

2.4 纳入离线数据

在线设置中的若干现有研究纳入了某些离线数据，即未参与任何选择决策的带标签数据（Bao 等人，2024a；Sale and Ramdas，2025）。虽然 PEMI 无需离线数据即可运行，但只需将排列扩展至整个数据序列，便可处理这些设置。

沿用相关文献的记法，我们将离线数据记为 $\mathcal{D}_{\text{off}} = \{Z_{-n_{\text{off}}+1},\ldots,Z_0\}$，其中 $n_{\text{off}} \in \mathbb{N}^+$是离线样本的数量。它们在在线选择过程开始之前收集。时刻 t 之前的在线数据记为 $\mathcal{D}_{t,\ \text{on}} = \{Z_1,\ldots,X_t\}$ 并将完整数据记为 $\mathcal{D}_t = \{Z_{-n_{\text{off}}+1},\ldots,Z_{t-1},X_t\}$。我们假设 $\mathcal{D}_t$ 中的数据是可交换的。接着，我们定义选择指示变量 $S_t = \mathcal{S}_t(\mathcal{D}_t)$，其中 $\mathcal{S}_t$ 是一个固定映射 $\mathcal{S}_t\colon (\mathcal{X}\times\mathcal{Y})^{n_{\text{off}}+t-1}\times\mathcal{X}\to\{0,1\}$。请注意，这一一般设置涵盖了我们的默认设置。因此，在本节中，一致性分数 $\mathcal{V}_t\colon (\mathcal{X}\times\mathcal{Y})^{n_{\text{off}}+t}\to\mathbb{R}$ 是一个预先指定的函数，其参数为整个数据序列。

在这里，我们通过对全部在线和离线数据进行排列来扩展 PEMI。对于假设的 $y\in\mathcal{Y}$以及 $\{-n_{\text{off}}+1,\ldots,t\}$ 的一个排列 π，我们将排列后的数据集表示为 $\mathcal{D}^{y}_{\pi,\ t} = (Z_{\pi(-n_{\text{off}}+1)}(y)\ldots,\ X_{\pi(t)})$。与之前类似，$Z_i(y) = Z_i$ 适用于 $i<t$，而 $Z_t(y) = (X_t,\ y)$。排列下的选择决策为$S_t^y(\pi) = \mathcal{S}_t(\mathcal{D}^{y}_{\pi,\ t})$，一致性分数为 $V_t^y(\pi) = \mathcal{V}_t(Z_{\pi(-n_{\text{off}}+1)}(y)\ldots,\ Z_{\pi(t)}(y))$。然后，我们定义

$$p_t^{\text{off}}(y) = \frac{\sum_{\pi\in\hat{R}_t^{\text{off}}(y)}\mathbb{1}\{V_t^y(\pi_0)\le V_t^y(\pi)\}}{|\hat{R}_t^{\text{off}}(y)|},\quad \text{where}\quad \hat{R}_t^{\text{off}}(y) = \left\{\pi\in\Pi_{t,\text{off}}^{(M)} : S_t^y(\pi)=1\right\}\cup\{\pi_0\},$$

以及 $\Pi_t^{(M)}{}_{,\ \text{离线}} = \{\pi^{(1)},\ldots,\pi^{(M)}\}$ 分别从 $\{-n_{\text{off}}+1,\ldots,t\}$ 上的所有排列中独立且均匀地抽取。最后，我们将预测集定义为

$$\hat{\mathcal{C}}_{\alpha,t}^{\text{off}} = \{y\in\mathcal{Y} : p_t^{\text{off}}(y) > \alpha\}. \tag{9}$$

定理 2。 假设 $\{Z_i\}_{i=-n_{\text{off}}+1}^{0}\cup\{Z_j\}_{j=1}^{t}$ 是可交换的。那么 $\hat{}\mathcal{C}_{\alpha\ t}^{off}{}_{,}$ 定义见 (9) 满足

$$\mathbb{P}(Y_t\in\hat{\mathcal{C}}_{\alpha,t}^{off}\,|\,S_t=1)\ge 1-\alpha,\quad \forall\ t\ge 1. \tag{10}$$

定理 2 的详细证明见附录 B.2。这一扩展很直观：当有更多数据可用时，我们只需将第 2 节中的相同思路应用于扩大的数据集。在可交换性成立的条件下，相同的理论论证仍然适用。应用式（6）中的随机化技巧也能得到精确覆盖率，此处为简洁起见不再赘述。

纳入离线数据的实际好处是，大量排列可能被纳入参考集，从而得到更稳定、更高效的预测集。例如，考虑所有

保持在线数据 $\{1,\ldots,t-1\}$ 不变、只对 $\{-n_{\text{off}}+1,\ldots,-1,t\}$进行排列的排列。在这种排列下，时间 t 之前的所有选择决策都保持不变；只要将离线数据换到位置 t 后选择决策仍然相同，该排列就会被纳入 $\hat{}$ $R_t^{\text{off}}(y)$。最后需要指出的是，我们的框架并不一定需要离线数据才能运行；提出这一扩展主要是为了使论述更加完整，并与现有文献衔接。

2.5 一般选择分类体系与 FCR 控制

除了 SCC 之外，许多现有研究还考虑在选取多个测试点时的错误覆盖率（FCR）。虽然我们目前的讨论一直聚焦于 SCC，但在本部分中，我们引入选择分类体系的概念，并扩展 PEMI，以便在特定情况下实现 FCR 控制。

类似于 Jin and Ren (2024)，我们将选择分类体系定义为 $\mathfrak{S}\subseteq\{0,1\}^t$ 长度为 t的任意预先指定的二进制序列集合。直观地说，分类体系是一组满足预先设定目标属性的选择轨迹。给定一个分类体系 $\mathfrak{S}$，选择条件覆盖率可以推广为

$$\mathbb{P}(Y_t\in\hat{\mathcal{C}}_{\alpha,t}\mid S_t=1,(S_1,\ldots,S_t)\in\mathfrak{S})\geq 1-\alpha, \tag{11}$$

其中，$S_i=\mathcal{S}_i(Z_1,\ldots,Z_{i-1},X_i)\in\{0,1\}$ 表示时间 i的选择决策。

为实现（11），我们构造参考集，使其保留分类体系中的选择决策：

$$\hat{R}_t(y)=\{\pi\in\Pi_t^{(M)}:S_t^y(\pi)=1,\ (S_1^y(\pi),\ldots,S_t^y(\pi))\in\mathfrak{S}\}\cup\{\pi_0\}.$$

PEMI 预测集，记作 $\hat{}\mathcal{C}_{\alpha,\,t}(\mathfrak{S})$，随后可用 $\hat{}\,R_t(y)$ 按照 (5) 或 (6) 的类似方式构造。式 (11) 的证明是定理 1证明的直接推广。

FCR 控制。 现在我们展示，通过设定适当的分类体系，可以实现 FCR 控制。FCR 衡量所选预测集中未能覆盖真实结果的预测集所占比例的期望值：

$$\text{FCR}=\mathbb{E}\left[\frac{\sum_{t=1}^T S_t\,\mathbb{1}\{Y_t\notin\hat{\mathcal{C}}_{\alpha,t}\}}{1\vee\sum_{t=1}^T S_t}\right]. \tag{12}$$

这里 $T\in\mathbb{N}^+$ 是时间点的总数。

现有方法可以针对所谓的决策驱动的选择规则实现 FCR 控制，在这些规则中，选择规则是过去选择决策的确定性函数（Sale and Ramdas, 2025；Bao et al., 2024a）。通过引入适当的选择分类体系，PEMI 在类似情形下实现了 FCR 控制。在每个时间 t，我们将观测到的选择轨迹记录为 $s_1=\mathcal{S}_1(X_1)$、$s_2=\mathcal{S}_2(Z_1,X_2)$、……、$s_t=\mathcal{S}_t(Z_1,\ldots,Z_{t-1},X_t)$，并将分类体系定义为 $\mathfrak{S}:=\{(s_1,\ldots,s_t)\}$。

下述定理表明，在对未来选择决策作出条件独立假设的情况下，这种 PEMI 预测集的构造能够实现 FCR 控制。

定理 3。 假设对于每个 $t\in\mathbb{N}^+$，在以 $(S_1,\ldots,S_t)$为条件时，选择决策 $\{S_{t'}\}_{t'>t}$ 与 $\mathbb{1}\{Y_t\in\hat{}\mathcal{C}_{\alpha,\,t}(\mathfrak{S})\}$ 相互独立。那么，对于 $\mathfrak{S}:=\{(s_1,\ldots,s_t)\}$，上述 $\{\hat{}\mathcal{C}_{\alpha,\,t}(\mathfrak{S})\}_{t=1}^T$ 满足 $FCR\leq\alpha$。

定理 3 的证明见附录B.3。此处的核心思路遵循 Sale and Ramdas (2025，命题 5.2)。通过将选择事件限制在分类体系内，预测集 $\{\hat{}\mathcal{C}_{\alpha,\,t}(\mathfrak{S})\}_{t=1}^T$ 可在给定整个选择轨迹的条件下达到覆盖率，从而实现 FCR 控制。决策驱动的选择规则是满足条件独立假设的一个自然场景。然而，一般而言，我们指出，在不过度保守的情况下实现精确的 FCR 控制颇具挑战。对于任意选择规则，多个时间点上的决策与标签不确定性之间存在复杂的依赖关系。因此，我们将这一问题留作未来研究方向。

2.6 推广到多个测试样本

在本小节中，我们将 PEMI 推广至一种类似离线的设定，其中可以选择多个测试点。为便于清晰表述，我们采用与 JOMI 框架相似的记号（Jin 和 Ren，2024）。

假设可以获得校准数据 $\mathcal{D}_{\text{calib}} = (Z_1, \ldots\ldots, Z_n)$ 和测试数据 $\mathcal{D}_{\text{test}} = (X_{n+1}, \ldots\ldots, X_{n+m})$。我们考虑针对多个测试样本 $\mathcal{S}$ 的一般选择机制：$(\mathcal{X} \times \mathcal{Y})^n \times \mathcal{X}^m \to 2^{\{1,\ldots,m\}}$，该机制可能对数据排序敏感。我们还依赖一个预先指定的一致性分数 $\mathcal{V}$：$(\mathcal{X} \times \mathcal{Y})^{n+1} \to \mathbb{R}$，它对数据排序敏感（例如，仅作为最后一个数据点的函数）。我们的目标是仅为选定样本构造预测集，并使其具有选择条件覆盖率。当假设 $\mathcal{S}$ 对前 n 个参数（校准数据）置换不变时，此设置也可简化为 JOMI 的设置。

我们的方法遵循第 2.2 节中的相同原理。考虑选择集 $\hat{S} = \mathcal{S}(\mathcal{D}_{\text{calib}}, \mathcal{D}_{\text{test}}) \subseteq \{1, \ldots, m\}$。令 $\Pi_{n+j}^{(M)} = (\pi^{(1)}, \ldots\ldots, \pi^{(M)})$ 为从关于 $\{1, \ldots, n, n+j\}$ 的所有 $\hat{}$ 排列集合中随机且均匀抽取的排列。对于每个 $j \in S$ 及任意假设值 y，我们以类似方式定义排列后的校准数据 $\mathcal{D}_{\text{calib}}^{\pi}(y)$ 和排列后的测试数据 $\mathcal{D}_{\text{test}}^{\pi}(y)$：

$$\mathcal{D}_{\text{calib}}^{\pi}(y) = \left(Z_{\pi(1)}(y), Z_{\pi(2)}(y), \ldots, Z_{\pi(n-1)}(y), Z_{\pi(n)}(y)\right),$$
$$\mathcal{D}_{\text{test}}^{\pi}(y) = \left(X_{n+1}, \ldots, X_{n+j-1}, X_{\pi(n+j)}, X_{n+j+1}, \ldots, X_{n+m}\right),$$

其中 $Z_i(y) = (X_i, Y_i)$ 对应 $i \le n$，而 $Z_{n+j}(y) = (X_{n+j}, y)$ 对应 $i \le n$。因此，排列后的选择集为 $\hat{S}^{\pi}(y) = \mathcal{S}(\mathcal{D}_{\text{calib}}^{\pi}(y), \mathcal{D}_{\text{test}}^{\pi}(y))$，一致性分数为 $V_{n+j}^{y}(\pi) = \mathcal{V}(Z_{\pi(1)}(y)\ldots\ldots, Z_{\pi(n)}(y), Z_{\pi(n+j)}(y))$。我们定义参考集

$$\hat{R}_{n+j}^{M}(y) = \{\pi \in \Pi_{n+j}^{(M)} : j \in \hat{S}^{\pi}(y)\} \cup \{\pi_0\},$$

并构造 PEMI 预测集

$$\hat{\mathcal{C}}_{\alpha,n+j} = \{y \in \mathcal{Y} : p_{n+j}(y) > \alpha\}, \quad \text{where} \quad p_{n+j}(y) = \frac{\sum_{\pi \in \hat{R}_{n+j}^{M}(y)} \mathbb{1}\{V_{n+j}^{y}(\pi_0) \le V_{n+j}^{y}(\pi)\}}{|\hat{R}_{n+j}^{M}(y)|} \tag{13}$$

定理 4。假设 $\{Z_i\}_{i=1}^{n} \cup \{Z_{n+j}\}$ 在给定 $\{X_{n+l}\}_{l \in [m] \setminus \{j\}}$ 的条件下可交换，对任意 $j \in [m]$均成立。那么 $\hat{\mathcal{C}}_{\alpha,\, n+j}$ 定义于(13) 满足

$$\mathbb{P}(Y_{n+j} \in \hat{\mathcal{C}}_{\alpha,n+j} \mid j \in \hat{S}) \ge 1 - \alpha. \tag{14}$$

定理 4 的详细证明推迟至附录 B.4。$\{Z_i\}_{i=1}^{n+m}$ 的标准联合可交换性蕴含条件可交换性条件。与定理 1 的证明类似，我们证明一个更强的结果，即 $\mathbb{P}\big(p_{n+j}(Y_{n+j}) \le \alpha \mid j \in \hat{S}, [\mathcal{D}_j], \mathcal{D}_j^{c}, [\hat{\pi}^{M}]\big) \le \alpha$，其中我们将完整观测值的无序集定义为 $[\mathcal{D}_j] = [Z_1, \ldots, Z_n, Z_{n+j}]$，将剩余测试数据定义为 $\mathcal{D}_j^{c} = \mathcal{D}_{\text{test}} \setminus \{X_{n+j}\}$，并将无序排列集定义为 $[\hat{\pi}^{M}] = [\hat{\pi}, \pi^{(1)} \circ \hat{\pi}, \ldots, \pi^{(M)} \circ \hat{\pi}]$。总体论证仍沿用相同思路，但现在我们要对所有其他测试点进行条件化。

3 计算效率高的具体实现

到目前为止，我们已经为任意选择规则建立了一个具有选择条件覆盖率的预测推断一般框架。在分类问题中，可以通过枚举所有类别 $y \in \mathcal{Y}$ 来求解 PEMI 预测集。然而，当 $|\mathcal{Y}|$ 是无限集时，这种枚举在计算上难以处理。

本节介绍针对具有特殊结构的多种常见选择规则而设计的高效算法。我们重点讨论三大类选择规则：协变量依赖规则、保形选择，以及基于早期结果的选择。为提高计算效率，全文将范围限定在仅涉及最后一个数据点的一致性分数函数，即 $\mathcal{V}_t(z_1, \ldots\ldots, z_t) = v(x_t, y_t)$，其中 $v : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ 是预先指定的函数。我们也只关注蒙特卡洛变体 $\hat{\mathcal{C}}_{\alpha,\, t}(\Pi_t^{(M)})$ 和 $\hat{\mathcal{C}}_{\alpha\, t}^{\text{rand}}(\Pi_t^{(M)})$，因为枚举整个排列集合的计算开销很大。

3.1 协变量依赖型选择规则

我们首先考虑一大类称为协变量依赖规则（Jin and Ren，2024）的选择规则。在这种设定下，时刻 t 的选择决策仅取决于协变量 $\{X_i\}_{i=1}^{t}$，而不取决于响应值 $\{Y_i\}_{i=1}^{t-1}$。这涵盖了许多常用的选择规则，包括现有研究中的以下规则：

1. 决策驱动型选择。这是 Bao 等人（2024a）；Sale and Ramdas（2025）研究的一类规则。在此，选择 S_t 仅是 X_t 和所有过去选择决策 $\{S_i\}_{i=1}^{t-1}$ 的函数。由于这类规则不涉及响应变量 $\{Y_i\}_{i=1}^{t-1}$，因此属于协变量依赖型。2. 基于加权分位数的选择。如果某个单元 t的预测值高于过去 $t-1$ 数据中预测值（或任何分数）在 q水平上的加权分位数，则选择该单元。这类思路已被用于提高对分布偏移的稳健性（Barber 等，2023）。在线设置中，权重$\{w_i\}_{i=1}^{t-1}$ 可能随时间变化，且选择规则对带标签数据是不对称的。3. 基于加权平均数的选择。如果某个单元 t 的预测值大于过去 $t-1$ 数据中预测值的加权平均数，则选择该单元。与分位数的情况类似，选择规则可以自然地是不对称的。4. 基于黑箱在线优化的选择。许多在线算法根据流式数据生成决策；如果优化问题仅涉及协变量，则这些算法属于此类，例如资源分配中的在线背包问题（Chakrabarty 等人，2008）。

现在，我们给出适用于任意协变量依赖型选择规则的程序。简化的关键在于，对于任意假设值 $y \in \mathcal{Y}$，参考集保持不变。换言之，对于协变量依赖型选择规则，参考集 $\hat{R}_t(y;\Pi_t^{(M)})$ 可独立于 y 计算。对于末点分数 $v(x_t, y_t)$，PEMI p 值（5）会在这个单一参考集中，将 $v(X_t, y)$ 与 $v(X_{\pi(t)}, Y_{\pi(t)})$ 的分数进行比较，而预测集只依赖于一个阈值。由于选择规则不涉及标签，我们将任意插补响应下的选择决策记为 $S_t^y(\pi) \equiv S_t(\pi)$。

以下命题给出了 PEMI 预测集的显式形式，证明见附录 B.5。算法 1概述了该程序；为简洁起见，我们仅给出 $\hat{\mathcal{C}}_{\alpha,t}(\Pi_t^{(M)})$。

命题 1。假设 $S_t = 1$。令 $\Pi_t^{(M)}$ 为来自 Π_t 上均匀分布的独立同分布样本。定义

$$\hat{R}_t = \{\pi \in \Pi_t^{(M)} : S_t(\pi) = 1\} \cup \{\pi_0\}, \qquad B_t = \{\pi \in \Pi_t^{(M)} : S_t(\pi) = 1,\ \pi(t) \neq t\}. \tag{15}$$

PEMI 预测集为 $\hat{\mathcal{C}}_{R_t^{\alpha},\,t}(\Pi_t^{(M)}) = \{y \in \mathcal{Y} :\ v(X_t,\ y) \le \text{分位数}\left(\beta;\ \{v(X_{\pi(t)},\ Y_{\pi(t)})\}_{\pi\in B_t} \cup \{+\infty\}\right)\}$ 对于 $\beta = \frac{\lceil (1-\alpha)\cdot(|B_t|+1)\rceil}{|B_t|}$.

此处，集合 B_t 收集了将一个带标签点移至最后位置并产生 $v(X_{\pi(t)},\ Y_{\pi(t)}) \neq v(X_t,\ y)$ 的信息性置换；该表达式出现在 PEMI p 值的约化形式中。随机化集合 $\hat{\mathcal{C}}_{\alpha\,t}^{\mathrm{rand}}(\Pi_t^{(M)})$ 也具有闭式表达，为简洁起见，我们将其推迟至附录 4 中的命题 A.1 。

3.2 保形选择

第二类选择规则称为保形选择，旨在选出结果满足特定条件的单位，同时控制某种第一类错误（Jinand Cand`es，2023；Jin andCand`es，2025；Xuand Ramdas，2023）。这类决策由假设检验程序作出，所依据的是为每个测试点构造的保序 p 值或 e 值。对于在线检验程序，这类选择规则通常是不对称的，并涉及先前结果。

我们遵循Jin and Cand`es (2023)；Jin andRen (2024) 的设定，并稍作修改以适应在线场景。在每个时刻 t，我们观测测试点 X_t 以及感兴趣的阈值 $c_t \in \mathbb{R}$。目标是根据带标签数据 $Y_t > c_t$ 选择满足 $\{(X_i,\ Y_i,\ c_i)\}_{i=1}^{t-1}$ 的点。

算法 1 使用随机排列进行协变量依赖选择的 PEMI 输入：数据序列 $(Z_1, \ldots, Z_{t-1}, X_t)$，置信水平 $\alpha \in (0, 1)$，排列次数 M，协变量依赖选择规则 $\mathcal{S}$，符合度评分函数 $v(\cdot,\cdot)$。1：计算 $S_t = \mathcal{S}_t(Z_1, \ldots, Z_{t-1}, X_t)$；2：如果 $S_t = 1$ 则 3：从均匀分布(Π_t)中随机抽取 M 个排列 $\Pi_t^{(M)} = (\pi^{(1)}, \ldots, \pi^{(M)})$；ˆ 4：按式 (15) 计算 R_t 和 B_t；5：计算 ˆ$\eta =$ 分位数$\left(\frac{\lceil (1-\alpha)\cdot|\hat{}_{|B_t|}^{R_t|}\rceil}{}; \{v(X_{\pi(t)}, Y_{\pi(t)})\}_{\pi \in B_t} \cup \{+\infty\}\right)$；6：构造预测集 ˆ$\mathcal{C}_{\alpha,t}(\Pi_t^{(M)}) = \{y \in \mathcal{Y}: v(X_t, y) \leq \hat{}\eta\}$。7：结束如果 输出：预测集 ˆ$\mathcal{C}_{\alpha,t}(\Pi_t^{(M)})$。

该选择依赖于任意得分函数 F：$\mathcal{X} \times \mathbb{R} \to \mathbb{R}$，使得 $F(x, y)$ 关于 $y \in \mathbb{R}$ˆ 对任意 $x \in \mathcal{X}$ 都是非递增的。记 $F_i = F(X_i, c_i)$，我们定义一般的 加权共形 *p* 值 如下：

$$p_t^w = \frac{w_t + \sum_{i=1}^{t-1} w_i \cdot \mathbb{1}\{\hat{F}_i \geq \hat{F}_t, Y_i \leq c_i\}}{\sum_{i=1}^{t} w_i},$$

其中 $\{w_i\}_{i=1}^{\infty}$ 是预先固定的非负权重序列。该 p 值是 Jin and Cand`es (2023) 引入的共形 p 值的一种加权形式，并采用了强大的“截断”分数（Jin and Ren，2024）。遵循 Barber等人 (2023) 和 Jin and Cand`es (2023) 中的类似论证，当 $\{(X_i, Y_i, c_i)\}_{i=1}^{t}$ 可交换时，此类 p 值是有效的，即对所有 $\alpha \in [0,1]$，$\mathbb{P}(p_t^w \leq \alpha, Y_t \leq c_t) \leq \alpha$。保形选择框架已应用于药物发现中的潜在候选药物识别（Bai等人，2025）以及可靠语言模型输出的筛选（Gui等人，2024），并扩展至在线设定（Xu and Ramdas，2023）；在这些场景下，选择后的不确定性量化需求自然随之产生。

我们提供了 PEMI 在两大类选择规则下的实现：(1) 对 p 值 p_t^w 设定阈值；(2) 对基于保形 p 值构造的特定 e 值设定阈值。对于我们考虑的保形选择规则，选择决策仅通过二元指示变量 $\mathbb{1}\{Y_i \leq c_i\}$ 依赖过去的结果。因此，在为时间 t 插补候选标签 $y \in \mathcal{Y}$ 后，诱导出的选择事件仅通过 y 是否满足 $y \leq c_t$ 或 $y > c_t$ 来依赖 y。因此，参考集 ˆ$R(y; \Pi_t^{(M)})$ 在每个区域内均为常数，从而得到两个特定于不同机制的参考集 ˆ$R_t^{(0)}$ 和 ˆ$R_t^{(1)}$，分别对应 $y > c_t$ 和 $y \leq c_t$；PEMI 预测集具有统一的形式

$$\hat{\mathcal{C}}_{\alpha,t}(\Pi_t^{(M)}) = \{y > c_t, v(X_t, y) \leq \hat{q}_0\} \cup \{y \leq c_t, v(X_t, y) \leq \hat{q}_1\},$$

其中，ˆq_0，ˆ$q_1 \in \mathbb{R}$ 是 ˆ$R_t^{(0)}$ 和 ˆ$R_t^{(1)}$ 中符合度分数的合适分位数。本小节余下部分将详细介绍如何针对常用的在线检验程序计算这些量。

3.2.1 基于 p 值的选择规则

我们首先考虑通过在特定数值处对保形 p 值进行阈值化来实施的选择程序。随着数据按顺序到达，可以对每个单元分别进行检验，也可以对整个序列中的多个单元进行检验。因此，我们考虑两种主要的阈值方法：

1. 固定阈值：若单元 t 的 p_t^w低于固定阈值 $q \in (0,1)$，则选中该单元。2. 自适应阈值：若单元 t 的 p_t^w 低于由在线多重检验算法确定的自适应阈值，则选中该单元（Javanmard 和 Montanari，2015年；Ramdas 等人，2019）。

第二类涵盖了广泛的在线检验方法，例如 LOND（Javanmard 和 Montanari，2015年）、SAFFRON（Ramdas 等人，2019）以及 ADDIS（Tian 和 Ramdas，2019）。这些方法通常

在每个时间点设置自适应阈值 α_t 以分配错误率，并动态更新该阈值，从而平衡发现能力与错误控制。这些方法中的阈值可以统一为

$$\alpha_t = G\left(t;\, \mathcal{F}_{t-1},\, \boldsymbol{\theta}\right), \quad \text{where} \quad \mathcal{F}_{t-1} = \sigma\big(\{p_i,\, \alpha_i\}_{i=1}^{t-1}\big),$$

对于某个函数 $G(\cdot)$。上文中，$\mathcal{F}_{t-1}$ 是由已实现的 p 值 $\{p_i\}_{i=1}^{t-1}$ 和自适应阈值 $\{\alpha_i\}_{i=1}^{t-1}$ 的历史生成的 σ-代数。最后，$\boldsymbol{\theta}$ 汇集了所有超参数，例如权重 $\{\gamma_t\}_{t\geq 1}$、初始 alpha 财富 W_0，以及 SAFFRON 和 ADDIS 中的阈值参数。

命题 2 给出了此选择规则下 PEMI 集的显式形式。为简洁起见，我们仅展示 $\hat{\mathcal{C}}_{\alpha,\, t}(\Pi_t^{(M)})$。证明见附录 B.6，具体流程总结于算法 2。

命题 2。 令 $\alpha_t \in (0,1)$ 表示时间 t 的选择阈值，该阈值可以如上所述为固定阈值或自适应阈值。考虑一个满足 $p_t^w \leq \alpha_t$ 的已选时间点 $t \in \mathbb{N}^+$。对于任意排列 $\pi \in \Pi_t^{(M)}$，我们定义 $i \in [t]$ 对应的置换得分 $\hat{F}_{\pi(i)} = F(X_{\pi(i)}, c_{\pi(i)})$。对于 $k \in \{0,1\}$ 和 $j \in [t]$，我们定义

$$p_{j,\pi}^{w,k} = \frac{w_j + \sum_{1\leq i\leq j-1, i\neq \pi^{-1}(t)} w_i \mathbb{1}\{\hat{F}_{\pi(i)} \geq \hat{F}_{\pi(j)},\, Y_{\pi(i)} \leq c_{\pi(i)}\} + k\cdot w_{\pi^{-1}(t)} \mathbb{1}\{\hat{F}_t \geq \hat{F}_{\pi(j)}\} \cdot \mathbb{1}\{\pi^{-1}(t) < j\}}{\sum_{i=1}^{j} w_i},$$

并令 $\alpha_{t,\,\pi}^k$ 为在时刻 t 使用排列后的 p 值和给定的在线检验程序计算得到的阈值。那么，有 $\hat{R}_t(y) = \hat{R}_t^{\mathbb{1}\{y\leq c_t\}}$，其中我们定义

$$\hat{R}_t^k = \left\{\pi \in \Pi_t^{(M)} :\, p_{t,\pi}^{w,k} \leq \alpha_{t,\pi}^k \right\} \cup \{\pi_0\}, \quad B_t^k = \left\{\pi \in \hat{R}_t^k : \pi(t) \neq t\right\}, \quad k \in \{0,1\}. \tag{16}$$

那么，*PEMI* 预测集为

$$\hat{\mathcal{C}}_{\alpha,t}(\Pi_t^{(M)}) = \{\, y \in \mathcal{Y} : y > c_t,\, v(X_t, y) \leq \hat{q}_0 \,\} \cup \{\, y \in \mathcal{Y} : y \leq c_t,\, v(X_t, y) \leq \hat{q}_1 \,\}, \tag{17}$$

其中 $\hat{q}_k = \text{分位数}\Big(\frac{\lceil (1-\alpha)\cdot |R_t^k| \rceil}{|B_t^k|};\ \{v(X_{\pi(t)},\, Y_{\pi(t)})\}_{\pi\in B_t^k} \cup \{+\infty\}\Big)$ 对于 $k \in \{0,1\}$。

算法 2 使用随机排列计算保序 p 值和 e 值的 PEMI

输入：数据序列 $(Z_1, \ldots, Z_{t-1}, X_t)$，置信水平 $\alpha \in (0,1)$，排列次数 M，选择规则 $\mathcal{S}$，符合度评分函数 $v(\cdot,\cdot)$，规则 $\in$ {p 值、e 值}。1：计算 $S_t = \mathcal{S}_t(Z_1, \ldots, Z_{t-1}, X_t)$；2：如果 $S_t = 1$，则 3：从均匀分布 (Π_t) 中随机抽取 M 个排列 $\Pi_t^{(M)} = (\pi^{(1)}, \ldots, \pi^{(M)})$；4：计算阈值 α_t；5：如果规则 $=$ 为 p 值，则 6：按照式 (16) 计算 $k = 0, 1$ 的 $\hat{R}_t^k$ 和 $\hat{B}_t^k$；7：如果规则 $=$ 为 e 值，则 8：按照附录 A.2 中的命题 5 计算 $k = 0, 1$ 的 $\hat{R}_t^k$ 和 $\hat{B}_t^k$；9：计算 $k = 0, 1$ 的 $\hat{q}_k = \text{分位数}\Big(\frac{\lceil (1-\alpha)\cdot |R_t^k| \rceil}{|B_t^k|};\ \{v(X_{\pi(t)}, Y_{\pi(t)})\}_{\pi\in B_t^k} \cup \{+\infty\}\Big)$；10：按照式 (17) 计算 $\hat{\mathcal{C}}_{\alpha,t}(\Pi_t^{(M)})$。11：结束如果。输出：预测集 $\hat{\mathcal{C}}_{\alpha,t}(\Pi_t^{(M)})$。

3.2.2 基于 e 值的选择

e 值为基于 p 值的假设检验提供了一种替代方法（Vovk 等人，2005；Ramdas and Wang，2025），近期也应用于保形选择（Xuand Ramdas，2023）。本文针对基于 e 值的程序所作的选择，提出 PEMI 集，其中 e 值基于保形 p 值。

我们研究 e-LOND 算法，以及 XuandRamdas （2023）提出的 e 值形式；该形式可确保在线保形选择的错误发现率控制，因此，当测试样本依次到达且需要实时作出选择决策时，它是一个实用工具。除常见的保形选择设置外，我们还假设可以使用 $n \in \mathbb{N}^+$ 离线数据，并将其用于选择；我们的 PEMI 预测集采用第 2.4 节中的广义方法。在每个时刻 t，对于每个过去的索引$j \in [t-1]$，我们首先构造两个留一保形 p 值：

$$p_j^- := \frac{\sum_{i=-n+1}^{0} \mathbb{1}\{\hat{F}_i \geq \hat{F}_j, Y_i \leq c_i\}}{n+1}, \qquad p_j^+ := \frac{1+\sum_{i=-n+1}^{0} \mathbb{1}\{\hat{F}_i \geq \hat{F}_j, Y_i \leq c_i\}}{n+1}.$$

随后，我们分别对Javanmard 和 Montanari（2015年）的 LOND 程序应用于 $(p_j^-)_{j\in[t-1]}$ 和 $(p_j^+)_{j\in[t-1]}$，从而得到两个选择集 $\hat{\mathfrak{R}}_{t-1}^-$ 和 $\hat{\mathfrak{R}}_{t-1}^+$，并由此得出自适应水平 $\hat{\alpha}^{\text{LOND},-}_t := \alpha\gamma_t(|\hat{\mathfrak{R}}_{t-1}^-|+1)$ 和 $\hat{\alpha}^{\text{LOND},+}_t := \alpha\gamma_t(|\hat{\mathfrak{R}}_{t-1}^+|+1)$。这两个集合仅用于构造 e 值，定义为 $E_t^{\text{e-LOND}} := \mathbb{1}\{p_t \leq \hat{\alpha}^{\text{LOND},+}_t\}/\hat{\alpha}^{\text{LOND},-}_t$，其中 $p_t := \frac{1+\sum_{i=-n+1}^{0} \mathbb{1}\{F_i \geq F_t, Y_i \leq c_i\}}{n+1}$。最后，e-LOND 的选择集按顺序确定：对于每个时间点 i，令 $\hat{\mathfrak{R}}_{i-1}^{\text{e-LOND}}$为截至时间 $t-1$被 e-LOND 选中的时间点集合，当且仅当满足以下条件时，单元 t 才会被选中

$$E_t^{\text{e-LOND}} \geq 1/\alpha_t^{\text{e-LOND}}, \quad \text{where} \quad \alpha_t^{\text{e-LOND}} := \alpha\gamma_t \cdot (|\hat{\mathfrak{R}}_{t-1}^{\text{e-LOND}}|+1).$$

在 e-LOND 选择后计算 PEMI 集，其思路仍基于以下事实：$\hat{R}_t(y; \Pi_t^{(M)})$ 仅通过 y 是否满足 $y \leq c_t$ 来决定。e-LOND 选择事件也可以改写为一个具有复杂数据依赖截断值的 p 值阈值事件；因此，计算也可以遵循类似思路。为简洁起见，具体公式见附录 A.2 中的命题 5。

3.3 基于先前结果的选择

我们考虑的最后一类选择规则基于先前结果的加权分位数。具体而言，当用户指定的分数 $\hat{\mu}(X_t)$ 超过 $\{Y_i\}_{i=1}^{t-1}$ 的加权分位数时，就会选中一个测试点，其中权重 $\{w_i\}_{i=1}^t$ 是一个固定序列，且 $\hat{\mu}\colon \mathcal{X} \to \mathbb{R}$ 是一个预训练评分函数。通过适当的变换，该方法也适用于基于结果的给定函数的加权分位数的选择规则，例如 $Y_i - f(X_i)$、$Y_i \cdot f(X_i)$ 等，其中函数为任意 $f(\cdot)$。

给定权重 $\{w_i\}_{i<t}$，我们将值集合 $\{y_i\}_{i<t}$ 的加权分位数记为 $\text{wQ}(\beta; \{y_i\}_{i=1}^{t-1}) := \inf\{z \in \mathbb{R}: \sum_{i=1}^{t-1} w_i \mathbb{1}\{y_i \leq z\}/\sum_{i=1}^{t-1} w_i \geq \beta\}$，对任意 $\beta \in [0,1]$ 均成立。高效计算的关键如下：通过分数 $\{\hat{\mu}_1, \ldots, \hat{\mu}_{t-1}\}$，将 $y \in \mathbb{R}$ 的取值范围划分为若干区间。在每个区间内，y 与 $\{\hat{\mu}_i\}_{i=1}^{t-1}$ 的排序保持不变，因此参考集 $\hat{R}_t(y)$ 为常数。由此，我们可以在每个区域内选取 y 的一个代表值来计算 $\hat{R}_t(y)$，再将它们合并以得到最终预测集。

命题 3 给出了预测集的显式形式。为简洁起见，我们仅展示 $\hat{\mathcal{C}}_{\alpha,t}(\Pi_t^{(M)})$。详细证明见附录 B.8，算法 3 总结了该过程。

命题3. 假设第 t 个数据被选为 $\hat{\mu}(X_t) \geq \text{wQ}(1-\beta; \{Y_i\}_{i=1}^{t-1})$。我们将 $\{\hat{\mu}(X_i)\}_{i=1}^{t-1}$ 的排序序列记为 $\hat{\mu}_{(1)} \leq \hat{\mu}_{(2)} \leq \cdots \leq \hat{\mu}_{(t-1)}$，它将 y 的范围划分为开区间$I_1, \ldots, I_t$，其中 $I_1 = (-\infty, \hat{\mu}_{(1)})$，$I_j = (\hat{\mu}_{(j-1)}, \hat{\mu}_{(j)})$，对于 $j = 2, \ldots, t-1$，以及 $I_t = (\hat{\mu}_{(t-1)}, +\infty)$。令 $\Pi_t^{(M)}$ 表示 $\{1, \ldots, t\}$ 上的随机置换集。那么，时刻 t 的 *PEMI* 预测集为

$$\hat{\mathcal{C}}_{\alpha,t}(\Pi_t^{(M)}) = \left(\bigcup_{j=1}^{t} \hat{\mathcal{C}}_{\alpha,t}^{(j)}\right) \cup \hat{\mathcal{C}}_{\alpha,t}^{\text{bd}}, \tag{18}$$

这里，我们将边界预测子集定义为

$$\hat{\mathcal{C}}_{\alpha,t}^{\text{bd}} = \left\{ y_k \in \{\hat{\mu}_{(k)}\}_{k=1}^{t-1} : p_t(y_k; \Pi_t^{(M)}) > \alpha \right\}, \tag{19}$$

$where\ p_t(y;\ \Pi_t^\star)$ 定义见(5)。对于每个区间 I_j，区间级预测子集 is

$$\hat{\mathcal{C}}_{\alpha,t}^{(j)} = \Big\{ y \in I_j : v(X_t, y) \le \text{Quantile}\big(\tfrac{\lceil (1-\alpha)\cdot|\hat{R}_t^{(j)}|\rceil}{|B_t^{(j)}|};\ \{v(X_{\pi(t)}, Y_{\pi(t)})\}_{\pi\in B_t^{(j)}} \cup \{+\infty\}\big)\Big\}, \tag{20}$$

其中 $B_t^{(j)} = \{\pi \in \hat{R}_t^{(j)} :\ \pi(t) \neq t\}$，并且，记作我们 $\hat{\mu}_{(0)} = -\infty$ 和 $\hat{\mu}_{(t)} = +\infty$，

$$\hat{R}_t^{(j)} = \big\{ \pi \in \Pi_t^{(M)} : \hat{\mu}(X_{\pi(t)}) \le \hat{\mu}_{(j-1)}, p_\pi^{(1)} \le \alpha \big\} \cup \big\{ \pi \in \Pi_t^{(M)} : \hat{\mu}(X_{\pi(t)}) \ge \hat{\mu}_{(j)}, p_\pi^{(0)} \le \alpha \big\} \cup \{\pi_0\}, \tag{21}$$

其中 $p_\pi^{(k)} = \dfrac{\sum_{1\le i\le t-1,\ i\neq \pi^{-1}(t)} w_i \mathbb{1}\{Y_{\pi(i)} \ge \hat{\mu}\ X_{\pi(t)}\ \} + w_{\pi^{-1}(t)} \cdot k\,()}{\sum_{i=1}^{t-1} w_i}$，$k = 0,\ 1$。

算法3 基于先前结果选择的 PEMI

输入：数据序列 $(Z_1, \ldots, Z_{t-1}, X_t)$，置信水平 $\alpha \in (0, 1)$，排列次数 M，选择规则 $\mathcal{S}$，符合度评分函数 $v(\cdot, \cdot)$。1：计算 $S_t = \mathcal{S}_t(Z_1, \ldots, Z_{t-1}, X_t)$；2：如果 $S_t = 1$，则 3：从 Unif(Π_t) 中随机抽取 M 个排列 $\Pi_t^{(M)} = (\pi^{(1)}, \ldots, \pi^{(M)})$；4：通过排序序列 $\hat{\mu}_{(1)} \le \cdots \le \hat{\mu}_{(t-1)}$ 将 y 的取值范围划分为开区间 $I_1, \ldots, I_t$；5：按照式 (21) 为 $j = 1, \ldots, t$ 构造参考集 $R_t^{(j)}$；6：按照式 (20) 和式 (19) 计算预测集 $\hat{\mathcal{C}}_{\alpha\, t}^{(j)}$（对应 $j = 1, \ldots, t$）和 $\hat{\mathcal{C}}_\alpha^{\text{bd}}{}_{,i}$；7：按区间合并预测集和边界预测集，得到式 (18) 中的最终预测集 $\hat{\mathcal{C}}_{\alpha,t}(\Pi_t^{(M)})$。8：结束 if 输出：预测集 $\hat{\mathcal{C}}_{\alpha,t}(\Pi_t^{(M)})$。

4 药物发现中的真实应用

我们根据第 3 节所述的各种非对称选择规则，通过一个真实的药物发现数据集展示 PEMI 的实用性。在药物发现中，预测模型被广泛用于辅助筛选有前景的候选药物。在此类高风险场景中，量化预测的不确定性对于理解和控制后续开发中的错误至关重要。实际的选择过程通常是在线进行的，候选药物会随时间推移被依次评估，从而产生各种非对称选择规则。

我们聚焦于药物-靶标相互作用（DTI）任务，其目标是为在线过程中选定的药物-靶点对的未知（连续取值）结合亲和力构建预测集。结果表明，在各种选择规则下，PEMI 始终能够提供精确的选择条件覆盖率。相比之下，原始保形预测在大多数情况下无法达到目标覆盖率；而在少数达到覆盖率的情况下，所得预测集的规模又大得异常。

数据集与模型。 我们使用 DAVIS数据集（Davis 等，2011），其中包含 30,060 个药物-靶点对的实值结合亲和力。协变量由药物和疾病靶点的编码结构组成。我们从数据集中随机抽取 20% 的样本，使用 DeepPurpose 库（Huang 等，2020）训练一个三层神经网络，将其作为回归模型 $\hat{\mu}:\ \mathcal{X} \to \mathbb{R}$，用于残差一致性分数 $\mathcal{V}_t$（z_1，……，z_t）$= |y_t - \hat{\mu}$（x_t）$|$。剩余 80% 的数据构成在线数据序列。

选择规则。 我们考虑三种现实的选择规则 $\mathcal{S}$（稍后将详细介绍）：

1. 协变量依赖选择：仅根据协变量信息选择药物 .
   (a) 决策驱动的选择规则。当科学家在固定预算下研究候选药物时，随着更多药物-靶点对被选中，选择策略可能会变得越来越严格，从而使后续选择仅限于预测结合亲和力更高的候选项。

(b) 加权分位数/平均值。当科学家优先考虑来自较近时间点的预测时，选择策略可以基于先前预测亲和力的加权分位数或平均值，其中较近的药物-靶点对获得更高的权重。(c) 基于模型不确定性的选择。当多个模型对候选药物的预测亲和力不一致（方差较高）时，科学家会选择这些候选药物，因为这种不一致可能表明需要进一步的实验研究。

2. 保形选择：基于保形 p 值或 e 值选择药物。

(a) 固定阈值。当科学家逐一评估每个药物-靶点对时，选择规则可以基于设有固定阈值的保序 p 值，只有当候选药物的保序 p 值低于预先设定的水平时，才会进一步研究该候选药物。(b) E-LOND。当科学家旨在识别一系列有前景的药物-靶点对时，选择策略可以遵循 e-LOND 程序，将错误发现率控制在某个 $q \in (0, 1)$ 以下。

3. 基于早期结果的选择：科学家根据过去的结果选择药物，即只有当当前药物-靶点对的预测亲和力 $\hat{}\mu_t$ 超过已实现亲和力 $\{y_s\}_{s<t}$ 的分位数或加权分位数时，才会选择该药物-靶点对；与依据过去的预测相比，基于真实标签进行选择或许能更好地反映经验表现，并减轻模型设定错误的影响。

评估指标。 每个时间点 t 的选择条件覆盖率通过相合估计量 $\hat{}\text{Cov}_t = \hat{}P(Y_t \in \hat{}\mathcal{C}_{\alpha,t}, S_t=1) \,/\, \hat{}P(S_t=1)$ 进行评估，其中 $\hat{}P$是重复实验中的经验比例。为评估效率，我们计算每个时间点上，在发生选择事件的重复实验中预测集长度的中位数（由于较早时间步的带标签样本量较小，可能出现无限长度的集合，因此均值不适用）。此外，我们计算每个时间点所选样本中预测集长度为无限的比例，以评估预测集的有效性。

4.1 协变量依赖选择

4.1.1 决策驱动型选择规则

我们首先考虑一种决策驱动的选择规则，该规则同时纳入过去的选择和当前协变量。具体而言，我们设定 $S_t = \mathbb{1}\{\hat{}\mu_t \geq \tau_1 + \frac{1}{\tau_0}\sum_{i=1}^{t-1} S_i\}$，其中 $\tau_0 = 200$，$\tau_1 = 5.5$。实验设置完全在线，采用 $T = 200$，且不使用离线数据。EXPRESS 论文（Sale and Ramdas, 2025）使用该规则来展示 CAP 方法（Bao et al., 2024a）中的有效性差距；因此，后者不纳入比较。我们采用 Sale and Ramdas (2025) 中使用的置信水平 $\alpha = 0.4$。比较的方法包括 $\hat{}\mathcal{C}_{\alpha,t}(\Pi_t^{(M)})$（我们的方法（上界））和 $\hat{}\mathcal{C}^{\text{rand}}_{\alpha\, t}(\Pi_t^{(M)})$（我们的方法（随机化））、EXPRESS、采用 Humbert et al. (2025) 中三种参数设置的 OnlineSCI（OnlineSCI1/2/3），以及将所有过去时间点用作校准数据的原始共形预测（原始方法）。

图2展示了四种方法在 $N = 10000$ 次运行中的经验选择条件覆盖率、预测集大小中位数，以及预测集长度为无限的比例。在选择条件覆盖率方面，绿色曲线（原始共形预测）始终远低于目标水平。橙色曲线（EXPRESS）虽然具有理论保证，但由于其参考集构造要求严格（基于数据点交换），过于保守，因此也始终远高于目标水平。由于 OnlineSCI 依赖某些超参数，事先可能难以确定如何选择这些参数，且其表现对参数选择十分敏感：OnlineSCI1（粉色）由于初始化不理想，在早期阶段明显欠覆盖；参数选择恰当的 OnlineSCI2（浅蓝色）始终接近目标水平；而 OnlineSCI3（灰色）使用过短的 γ 序列，过于保守且无法收敛。相比之下，PEMI（蓝色和红色）始终高于且接近目标水平。

就预测集大小而言，EXPRESS 过于保守：当 $t > 180$ 时，中位数长度达到

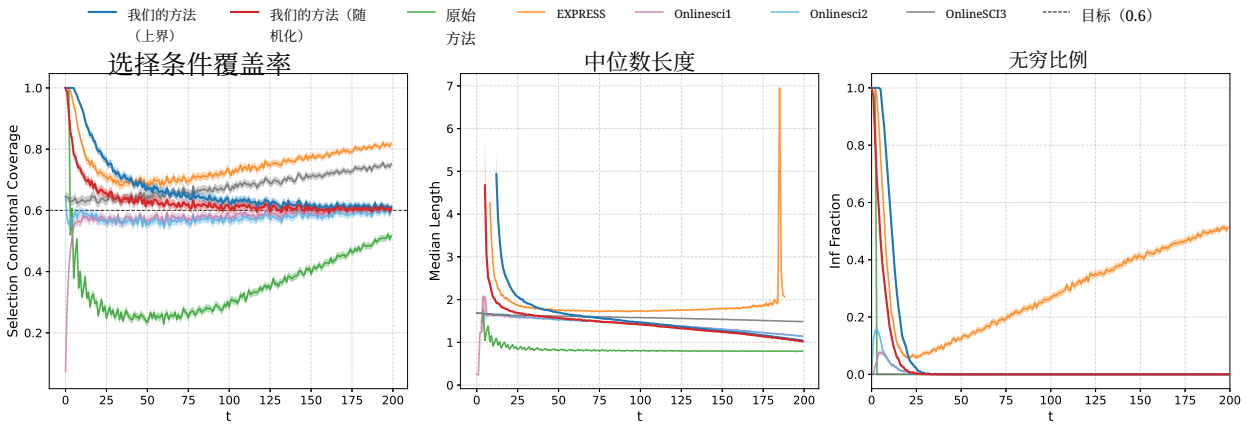

图2： 在药物性质预测中，采用决策驱动的选择规则时，PEMI（我们的方法（上界））、随机化 PEMI（我们的方法（随机化））、原始共形预测（原始方法）、EXPRESS（EXPRESS）和采用三种参数选择的 OnlineSCI（OnlineSCI1/2/3）的选择条件覆盖率（左）、预测集大小中位数（中）以及预测集长度无限的预测集所占比例（右）。紫色虚线表示目标覆盖水平 $1-\alpha=0.6$。

无穷大（即由于该准则过于严格，无法找到参考校准数据），而我们的方法和 OnlineSCI 仍保持合理的效率。最后一个面板中的模式颇具启示性：随着 EXPRESS 对参考集的要求变得更加严格，无限长度集合的比例会随时间增加。相比之下，在带标签点数量适中的情况下（例如 40 个），PEMI 始终能够生成非空洞的预测集。这表明，与数据交换相比，建立在置换检验之上的方法具有统计优势。

4.1.2 预测值的加权分位数或平均值

从现在起，我们将展示 PEMI 在 EXPRESS 框架无法处理的选择规则中的应用。因此，我们将我们的方法与原始保形预测和 OnlineSCI 进行比较（尽管Humbert 等人（2025）的理论结果可能不适用于此处的选择规则，但仍可以执行这些方法）。

该选择规则基于预测值的加权分位数或平均值，其中 $S_t = \mathbb{1}\{\hat{}\mu_t >wQ(1-q_{sel},\ \{\hat{}\mu_i\}_{i=1}^{t-1})\}$ 和 $S_t = \mathbb{1}\{\hat{}\mu_t > \frac{\sum_{i=1}^{t-1} w_{i\hat{}}\ \mu_i}{\sum_{i=1}^{t-1} w_i}\}$，其中我们定义了 $q_{sel}=0.1$ 以及预先指定的权重$w_i \propto 0.5^{(t-i)}$，这些权重会提高与当前测试点更接近的预测值的重要性。为保持一致，我们将置信水平设为 $\alpha=0.4$，并使用绝对残差符合度分数。

图 3 和图 4 分别展示了在这两种选择规则下进行 10,000 次运行的结果。对于选择条件覆盖率，我们的方法在所有时间点均保持在目标水平之上。相比之下，原始共形预测在大多数时间点都严重低于覆盖率要求。OnlineSCI1（使用经过良好调优的参数）在早期阶段低于目标水平，但很快进行适应，并收敛至接近目标的覆盖率；而 OnlineSCI2（使用调优较不充分的参数）在所评估的时间范围内未能适应目标水平。就区间长度而言，与原始共形预测相比，我们的方法和 OnlineSCI 都适度扩大了预测集，以确保有效覆盖率。

4.1.3 基于模型不确定性和在线优化的选择

第三，我们考虑基于多个模型的模型不确定性的选择规则。利用 DeepPurpose 库，我们预训练了三个具有相同神经网络架构和不同$\{f^{(j)}\}_{j=1}^{3}$ 的模型

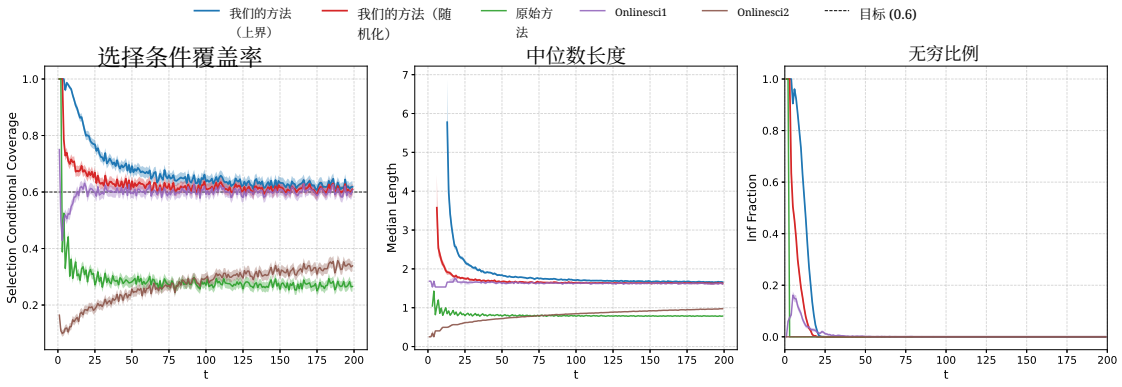

图3： 应用 PEMI（我们的方法（上界））、随机化 PEMI（我们的方法（随机化））、原始保形预测（原始方法）和 OnlineSCIOnlineSCI）时，基于预测值的加权分位数的选择规则下的结果。紫色虚线表示目标覆盖水平 $1-\alpha=0.6$。其他细节同图2。

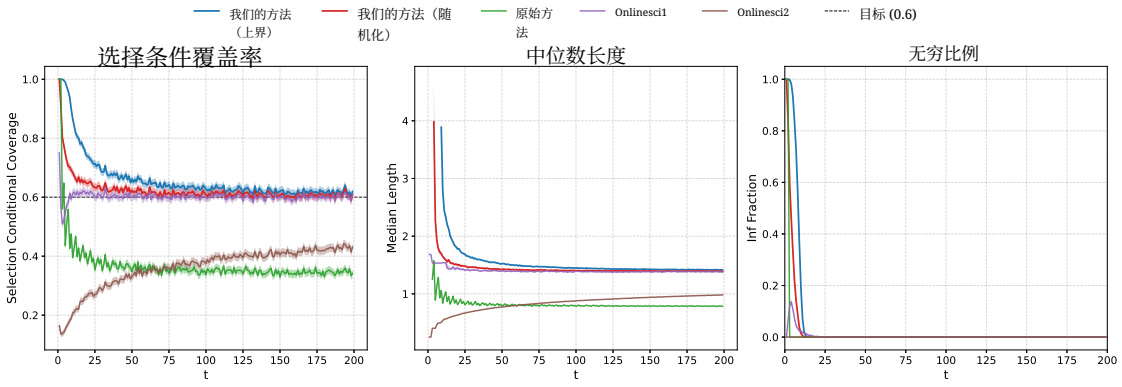

图4： 结果　　基于先前预测值的加权平均数进行选择；其他细节同图3。

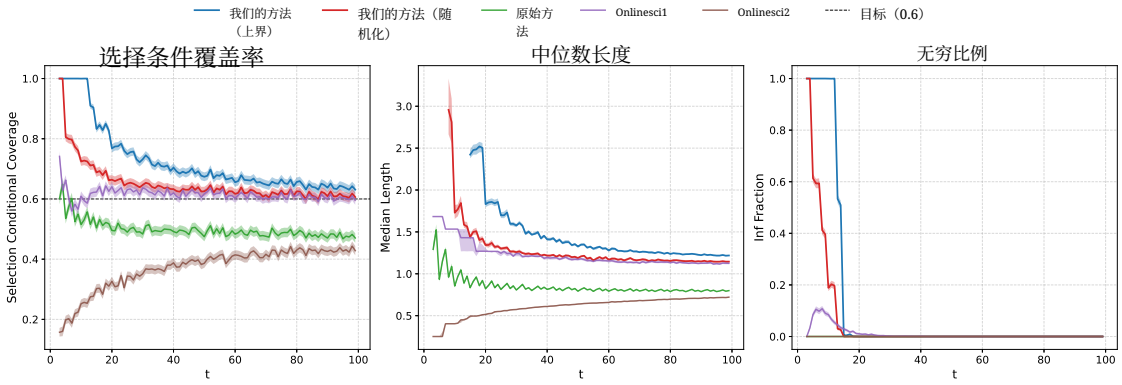

图5： 基于模型不确定性并结合在线优化进行选择时的结果；其他细节同图3。

药物和靶标的特征表示（Morgan 和联合三元组、PubChem 和氨基酸组成，以及 Daylight 和伪氨基酸组成）。对于每个测试点，我们计算方差 $s_t = \mathrm{Var}(\{f^{(j)}(X_t)\}_{j=1}^{3})$，并在满足 $s_t \geq \tau_t$ 时选择该点。阈值 τ_t 通过在线优化程序的最优解确定：$\max_{\tau} \sum_{i=1}^{t-1} s_i \mathbb{1}\{s_i \geq \tau\}$，满足约束 $\frac{1}{t-1}\sum_{i=1}^{t-1} \mathbb{1}\{s_i \geq \tau\} \leq \gamma$。在实际操作中，该在线优化会自适应地设定 τ_t，以最大化纳入点的累积不确定性，同时满足吞吐量约束 γ，从而在每个时间点将有限的检测能力分配给信息量最高的候选项。最后，我们使用第一个模型 $\hat{\mu} = f^{(1)}$，在所有选定的时间点构建预测集。

图5 展示了该选择规则在 $N = 10000$ 次运行中的结果。尽管原始共形预测在几乎所有时间点上都过度自信——因为高不确定性点通常需要更大的预测集——且OnlineSCI的性能再次依赖于超参数的恰当选择，我们的方法所维持的覆盖率始终高于或恰好处于目标水平。事实上，PEMI会略微增大预测集大小以确保有效性。

4.2 共形选择规则

固定阈值。我们首先研究采用固定阈值的保形选择。为证明我们方法的通用性，我们采用加权共形 p 值，其最初旨在缓解由分布偏移造成的覆盖率差距（Jin和Cand`es，2025；Barber等人，2023）。预先固定的权重为 $w_i = 0.99^{t+1-i}$。在每个测试点，若某候选对象的p值低于预先指定的阈值 $q = 0.3$，则将其选中。按照第3.2节的设定，阈值 c_i取为0.7分位数，该分位数来自与样本 i具有相同目标的训练配对的真实结合亲和力。实验重复10000次。

图 6 展示了结果。对于选择条件覆盖率，原始共形预测在大多数时间点都无法达到目标水平，而 OnlineSCI 依赖于超参数选择。相比之下，PEMI 在大多数时间点都非常接近目标水平（但在较早的时间点，由于必须出现无限长度集合，因此不在此列）。为确保覆盖率有效，PEMI 集比原始共形预测更大，但仍小于经过良好调优的 OnlineSCI1 集合。

E-LOND 阈值。其次，我们考虑采用在线多重检验程序进行保形选择。作为示例，我们采用Xu andRamdas (2023）提出的 Ue-LOND 算法（即 e-LOND 的随机化版本）作为在线多重检验程序，该程序能够严格控制错误发现率。此程序中使用的 e 值定义见第3.2.2节。为确保评估所需的选择频率足够高，我们留出 50 个数据点作为离线校准集。与第3.2.2节一致，Ue-LOND 程序会为每个时间点 e_t 计算一个 e 值 α_t 和一个阈值 t；当 $e_t \geq 1/\alpha_t$时，该时间点即被选中。

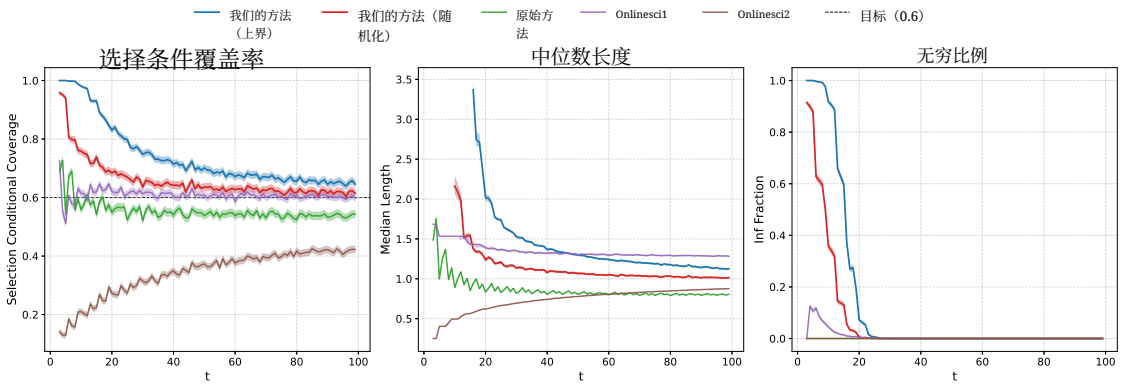

图6： 采用固定阈值进行保形选择的结果；其他细节与图3相同。

对于每个样本 i，我们将 c_i 设为具有相同疾病靶点的训练配对的结合亲和力的 0.3 分位数。值得注意的是，Ue-LOND 程序在此处的选择功效较低。因此，我们降低阈值 c_i，并将实验规模扩大至 100,000 次独立运行，以便进行可靠评估。

图7 展示了该选择规则的结果。Ue-LOND 的检验力有限，因此难以将样本选入参考集，结果也就显得较为保守。尽管如此，PEMI方法仍能保持有效的选择条件覆盖率，而原始共形预测和 OnlineSCI2（其参数调优得较差）的覆盖率则远低于目标水平。调优得更好的 OnlineSCI1 曲线也略低于目标水平，因为它调整阈值的选择时间点很少。对于预测集大小和无限长度集合的比例，我们都观察到这两个指标先上升，随后随时间逐渐下降。这种模式源于选择程序的特性：在序列早期，$\{\alpha_t\}$ 迅速下降，使得难以将样本选入参考集，进而导致集合大小和无限长度集合的比例均上升。随着时间推移，置换参考集不断增大，且 α_t 的衰减速度减慢，因此这两个指标都逐渐下降。

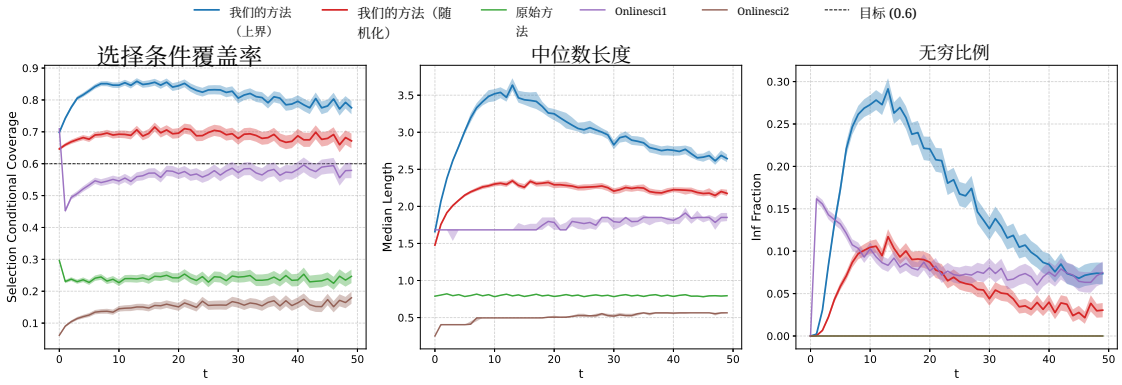

图7：采用 E-LOND 阈值进行保形选择时的结果；其他细节与图3相同。

4.3 基于先前结果的选择

最后，我们评估直接根据先前标签构建的选择规则。当候选项 t 的预测值 $\mu_t = \hat{\mu}(X_t)$ 超过 $\{Y_i\}_{i<t}$ 的分位数和加权分位数时，我们就选择该候选项。权重设为 $w_i \propto 0.5^{(t-i)}$，从而赋予较新的候选项更高的重要性。

图 8 和图 9 分别展示了这两种选择规则在 10,000 次运行中的结果。PEMI 保持了有效覆盖率，而原始共形预测和 OnlineSCI2（调优较差）在大多数时点都表现不佳。参数调优得更好的 OnlineSCI1 在极早期时点仍存在覆盖不足。中间和右侧面板显示，我们的方法具有合理的效率，长度为无穷大的预测集的出现频率会迅速下降。

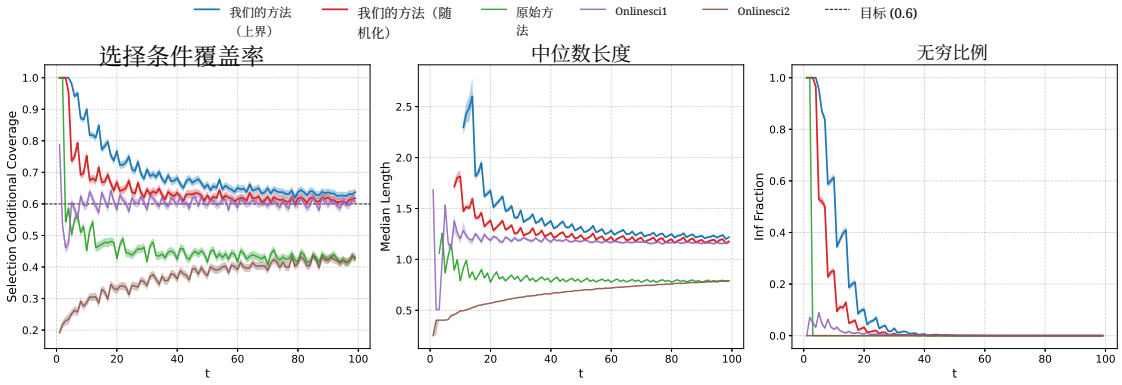

图8： 基于先前结果分位数的选择规则下的结果；其他细节同图 3。

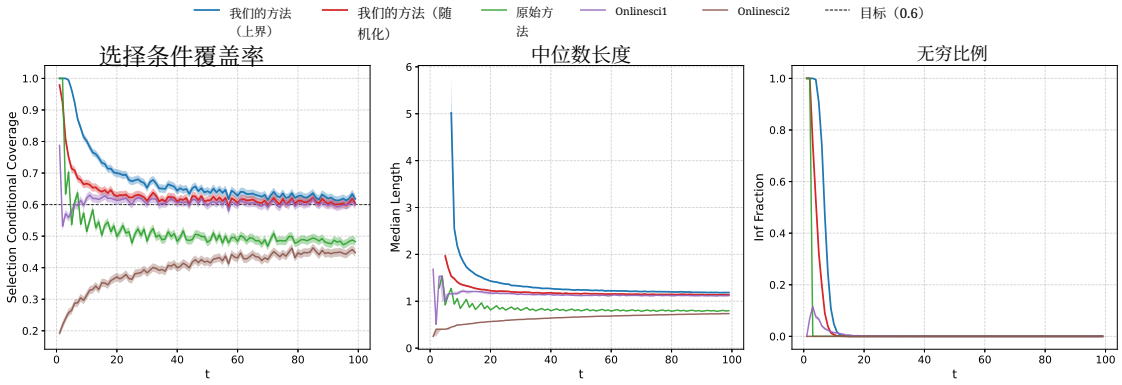

图9： 基于先前结果加权分位数的选择规则下的结果；其他细节同图 3。

5 模拟研究

在本部分中，我们设计了模拟研究，以考察多种因素对我们方法性能的影响，例如预测模型的质量、数据的噪声水平、某些选择规则中权重的衰减率，以及得分函数的质量。

5.1 模型质量与噪声水平

我们首先在可控的预测模型和噪声水平下评估我们的方法，使 $\hat{}\mu(X)$ 与真实标签 Y 之间的差异有所变化。我们生成独立同分布协变量 $X_i \sim$ 均匀分布$[-1,1]$ 和响应变量 $Y_i = \mu(X_i)+\epsilon_i$，其中 $\mu(x) = 3\sin(4\pi x)+4[\max\{0,\, x-0.3\}]^2-4[\max\{0,\, -(x+0.4)\}]^2$，其中 $\epsilon_i \,|\, X_i \sim \mathcal{N}(0,\, [\sigma\cdot(0.5+|X_i|)]^2)$，而 $\sigma \in \{0.1,\, 1.0,\, 10.0\}$ 控制噪声水平。对于模型质量评估，我们使用线性模型和随机森林

来拟合回归模型 $\hat{\mu}(\cdot)$。在非线性真实模型 $\mu(x)$ 下，线性模型会受到明显的模型设定错误影响，而随机森林能够捕捉非线性特征并实现良好拟合。

为简洁起见，我们仅关注权重为 $w_i \propto 0.5^{t-i}$ 、与第4.1.2节类似的加权分位数规则。图10 展示了两种不同质量模型和三个噪声水平下的结果。

模型质量。 比较不同模型可见，在选择条件覆盖率方面，我们的方法在两种模型下均表现良好，而原始共形预测在模型设定错误严重时稳定性较差，这表明 PEMI 对模型质量具有鲁棒性。就预测集大小而言，噪声水平较低时，随机森林自然会生成更短的预测集。然而，噪声水平较高时，两者之间的差异变得微不足道，因为预测集的大部分大小都用于覆盖噪声，而非信号。

噪声水平。 图 10 展示了三种不同噪声水平下的结果。原始共形预测的覆盖率对噪声较为敏感：噪声水平较低时过于保守，噪声水平较高时则无法达到目标。相比之下，我们的方法保持稳健，在所有噪声水平下都能达到预期覆盖率。就预测集大小而言，噪声水平越高，预测集越大。此外，噪声较高时，我们的方法会扩大预测集，以维持有效性。

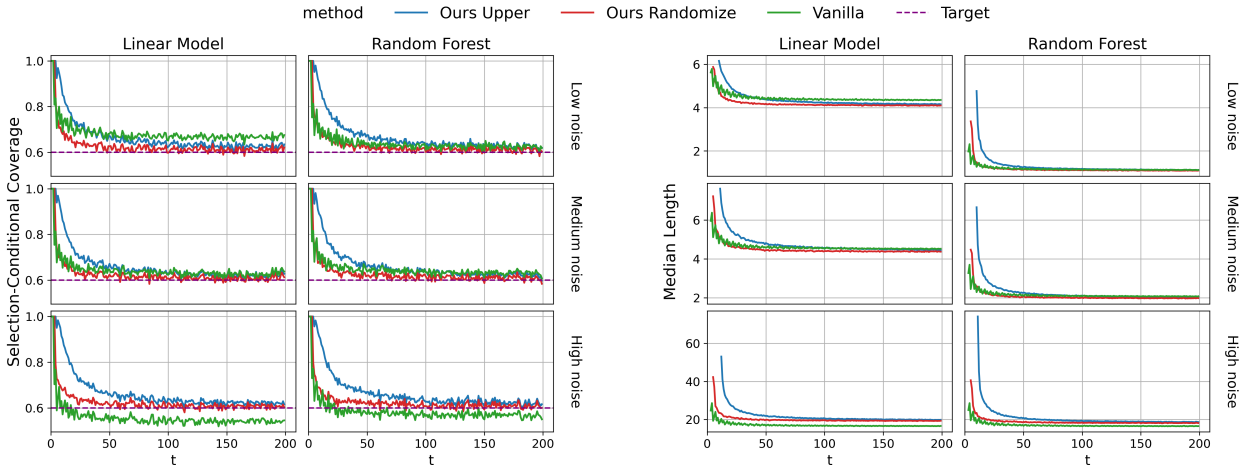

图 10：在不同模型质量和噪声水平下，使用基于预测值加权分位数的选择规则时，应用 PEMI（我们的方法（上界））、随机化 PEMI（我们的方法（随机化））和原始共形预测（原始方法）所得的经验选择条件覆盖率（左）和预测集大小中位数（右）。紫色虚线表示目标覆盖水平 $1-\alpha = 0.6$。无限长度集合的比例很低，因此未显示。

5.2 权重衰减率

接下来，我们在最后一部分使用不同的衰减率评估本方法在加权分位数选择规则下的性能；衰减率会影响选择功效。我们按照 Jin 和 Cand`es (2023) 中设定 3 的数据生成过程，生成独立同分布协变量 $X_i \sim$ 均匀分布 $[-1,1]^{20}$ 和响应变量 $Y_i = \mu(X_i) + \epsilon_i$, $\mu(x) = 5\left(x_1 x_2 + e^{x_4 - 1}\right)$, 其中 $\epsilon_i \sim \mathcal{N}(0, \left[\sigma \cdot (5.5 - |\mu(x_i)|)/2\right]^2)$, 而 σ 控制噪声水平。这里我们使用线性模型，并固定 $\sigma = 1.0$。我们使用与之前相同的选择规则，权重为$w_i \propto \beta^{t-i}$，其中 $\beta \in \{0.1, 0.5, 0.9\}$分别对应快速、中等和缓慢的衰减率。

图11 展示了三种权重衰减率下的结果。在此设置中，权重衰减率决定了选择规则的效能：衰减越快，选择频率越高。这进一步影响我们的选择参考集：过于严格的选择规则会使排列难以进入参考集，从而导致预测集过于保守。事实上，我们观察到

衰减越快，选择条件覆盖率就越快趋近目标水平，同时预测集长度会缩短，无限长度比例也会更快降至零。

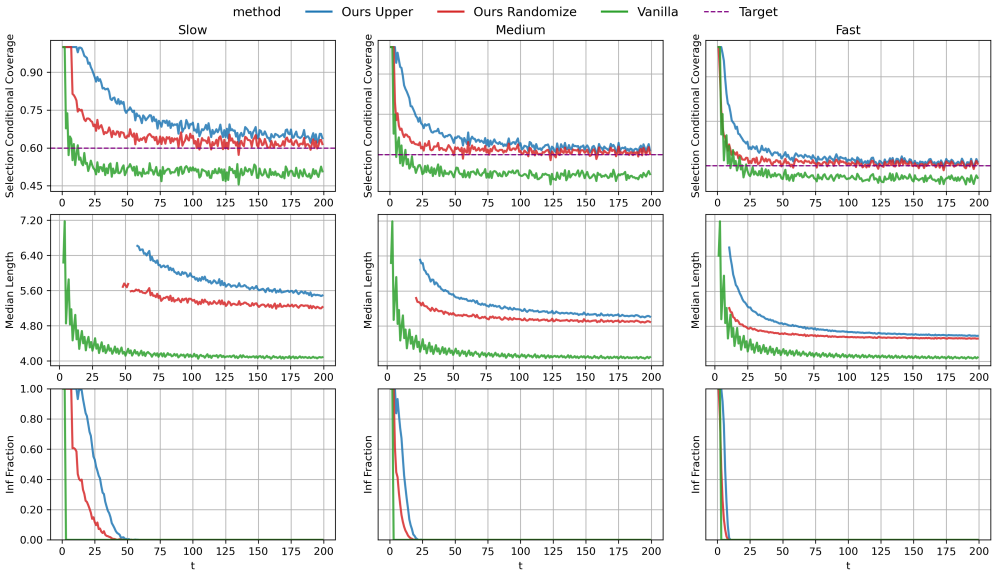

图 11： 在基于预测值的加权分位数的选择规则下，不同权重衰减率设置（列）中的经验选择条件覆盖率（上）、预测集大小中位数（中）以及预测集长度为无限的比例（下）。其他细节与图10相同。

5.3 一致性分数函数

最后，我们在不同的保形分数函数下评估所提出的方法。我们沿用5.1 和 5.2中的数据生成过程，它们分别对应高度非线性和异方差的数据设置，并采用两种噪声水平 $\sigma \in \{1, 5\}$。我们使用与之前相同的加权分位数选择规则。我们将得分函数设为绝对残差评分 $|y - \hat{\mu}(x)|$，其中 $\hat{\mu}(\cdot)$ 分别为线性模型和随机森林；此外，还使用保形分位数回归（CQR）分数（Romano 等人，2019）。

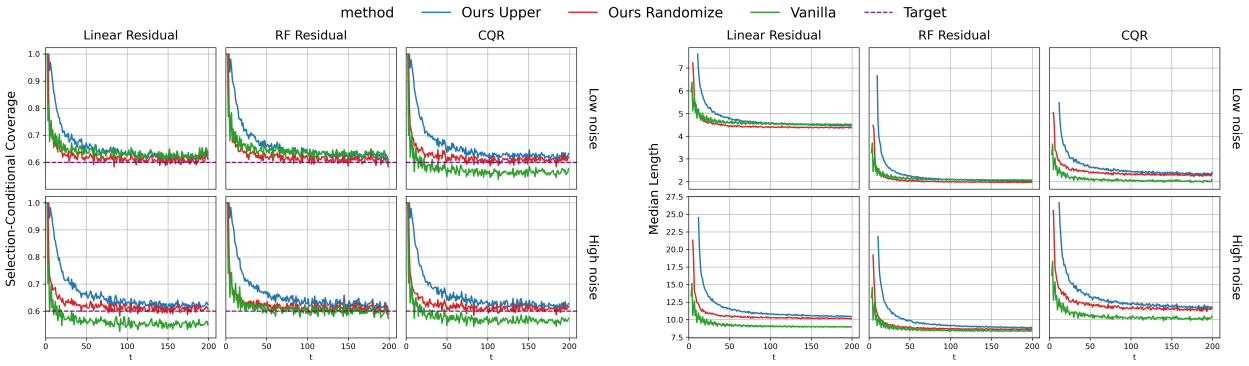

图 12： 在第 5.1 节所述设置下，采用三种不同评分函数（列）和两种噪声水平（行）的结果。其他细节与图 10相同。

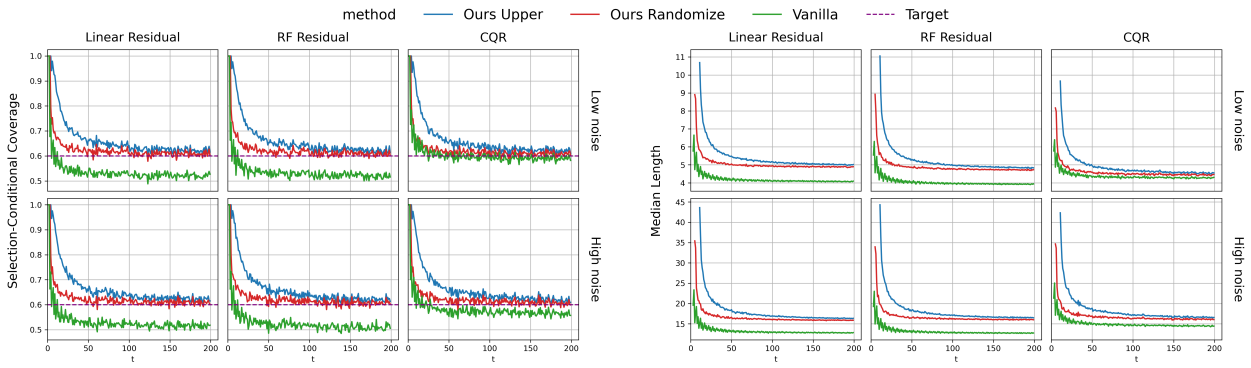

图 13： 第 5.2 节所设定条件下的结果，包含三种不同的评分函数（列）和两种噪声水平（行）。其他细节与图 10相同。

图 12 和图13分别展示了第 5.1 节（非线性）和第 5.2 节（异方差）设置下三种评分函数的结果。在原始共形预测中，与基于残差的集合相比，基于 CQR 的集合在不同运行中的性能指标更加稳定。然而，在大多数情况下，采用 CQR 的原始共形预测仍然存在覆盖不足的问题。尽管 CQR 旨在提供近似的 X条件覆盖率，但对于动态选择而言，它不足以实现选择条件覆盖率。相比之下，我们的方法在所有评分函数下都能始终实现有效覆盖率。我们还观察到，在低噪声设置下，CQR 通常会产生略小的集合；而在噪声较大的设置下，集合则更宽。

6 讨论

本文提出了 PEMI，这是一个适用于任何对带标签数据不具对称性的选择规则的选择性共形预测通用框架。其核心思想是将选择后推断应用于置换检验：我们使用能够导向同一选择事件的排列子集来校准预测集，从而消除了早期方法所需的唯一对称性条件。基于蒙特卡洛置换检验，我们针对多种常用的在线选择规则，推导了 PEMI 的高效计算实现。我们通过一个真实的药物发现数据集和大量模拟研究，展示了 PEMI 的有效性。

仍有若干研究问题有待解决。首先，许多针对在线数据的共形预测方法都能够应对分布漂移（Gibbs 和 Candes，2021；Barber等人，2023；Bao et al., 2024a），近期也有研究将类似思想用于选择性覆盖保证（Humbert 等人，2025），这引出了 PEMI 如何适应分布偏移的问题。其次，可以通过更一般的置换检验实现选择条件覆盖率。除了使用全部排列或均匀抽取的子集之外，其他选择——例如一组随机交换——也能构成有效的置换检验（Ramdas 等人，2023；Barber等人，2023）。具体而言，JOMI 方法（Jin and Ren，2024）可以视为将 Π_t 设为这样一组排列：它将数据点 t 与所有其他数据点交换；尽管其对称性条件不成立，Barber等人（2023）证明，随机抽取一次交换确实能在非对称一致性分数下得到有效的共形预测集，这与本文的设定类似。因此，研究类似思想能否实现有效的选择条件覆盖率颇有意义。最后，对于选择事件罕见的选择程序（例如第 4.2 节中的程序），难以获得足够大的参考集和长度有限的预测集。这或许反映了在带标签数据有限且选择事件罕见时，实现选择条件覆盖率所面临的根本困难，值得开展理论研究。

致谢

Y. J. 将 感谢 Rina Barber 就置换检验的一般性进行了有益讨论 s.

参考文献

Angelopoulos, A. N., Barber, R. F. 和 Bates, S. (2025). 保形预测的理论基础。Bai, T., Tang, P., Xu, Y., Svetnik, V., Yang, B., Khalili, A., Yu, X. 和 Yang, A. (2025). 面向药物发现中高效且准确的化合物筛选的保形选择。Bao, Y., Huo, Y., Ren, H. 和 Zou, C. (2024a). Cap：一种具有 FCR 控制的在线选择性保形预测通用算法。*arXiv* 预印本 *arXiv:2403.07728*。Bao, Y., Huo, Y., Ren, H. 和 Zou, C. (2024b). 具有错误覆盖声明率控制的选择性保形推断。《*Biometrika*》，111(3):727–742。Barber, R. F., Candes, E. J., Ramdas, A. 和 Tibshirani, R. J. (2023). 超越可交换性的保形预测。Benjamini, Y. 和 Yekutieli, D. (2005). 针对选定参数的假发现率校正多个置信区间。《美国统计学会杂志》，100(469):71–81。Chakrabarty, D., Zhou, Y. 和 Lukose, R. (2008). 在线背包问题。载于互联网与网络经济学研讨会（*WINE*），第1–9页。Davis, M. I., Hunt, J. P., Herrgard, S., Ciceri, P., Wodicka, L. M., Pallares, G., Hocker, M., Treiber, D. K. 和 Zarrinkar, P. P. (2011). 激酶抑制剂选择性的全面分析。《自然·生物技术》，29(11):1046—1051。Gazin, U., Heller, R., Marandon, A. 和 Roquain, E. (2025). 在错误覆盖率控制下选择信息性保形预测集。《皇家统计学会杂志*B*辑：统计方法学》，第qkae120页。Gibbs, I. 和 Candes, E. (2021年). 分布偏移下的自适应保形推断。《神经信息处理系统进展》，34:1660–1672。Gui, Y., Jin, Y. 和 Ren, Z. (2024). 保形对齐：知道何时可以有保证地信任基础模型。《神经信息处理系统进展》，37:73884–73919。Huang, K., Fu, T., Glass, L. M., Zitnik, M., Xiao, C. 和 Sun, J. (2020). DeepPurpose：用于药物-靶标相互作用预测的深度学习库。《生物信息学》，36(22–23):5545–5547。Humbert, P., Gazin, U., Heller, R. 和 Roquain, E. (2025). 在线选择性保形推断：自适应评分、收敛速率与最优性。*arXiv* 预印本 *arXiv:2508.10336*。Javanmard, A. 和 Montanari, A. (2015年). 论假发现率的在线控制。Jin, Y. 和 Cand`es, E. J. (2025). 通过加权保形p值在协变量偏移下进行无模型选择性推断。《*Biometrika*》，第asaf066页。Jin, Y. 和 Cand`es, E. J. (2023). 利用保形 p 值通过预测进行选择。Jin, Y. 和 Ren, Z. (2024). 焦点对象置信度：具有选择条件覆盖率的共形预测。*arXiv* 预印本 *arXiv:2403.03868*。Lee, J. D., Sun, D. L., Sun, Y. 和 Taylor, J. E. (2016). 精确选择后推断及其在套索回归中的应用。《统计年鉴》，44(3)。

Loftus, T. J., Shickel, B., Ruppert, M. M., Balch, J. A., Ozrazgat-Baslanti, T., Tighe, P. J., Efron, P. A., Hogan, W. R., Rashidi, P., Upchurch Jr, G. R., 等（2022）。医疗保健中的不确定性感知深度学习：一篇范围综述。*PLOS* 数字健康，1(8):e0000085。Markovic, J., Xia, L., 和 Taylor, J.（2017）。选择性推断的统一方法及其在交叉验证中的应用。*arXiv* 预印本 *arXiv:1703.06559*。Ramdas, A., Barber, R. F., Cand` es, E. J., 和 Tibshirani, R. J.（2023）。使用任意置换分布的置换检验。*Sankhya A*，85(2):1156–1177。Ramdas, A. 和 Wang, R.（2025）。使用 e 值进行假设检验。*Foundations and Trends® in Statistics*，1(1-2):1–390。Ramdas, A., Zrnic, T., Wainwright, M., 和 Jordan, M.（2019）。SAFFRON：一种用于在线控制假发现率的自适应算法。Reid, S., Taylor, J., 和 Tibshirani, R.（2017）。高斯样本中信号大小的选择后点估计与区间估计。*Canadian Journal of Statistics*，45(2):128–148。Romano, Y., Patterson, E., 和 Cand`es, E. J.（2019）。保形分位数回归。Sale, Y. 和 Ramdas, A.（2025）。在线选择性保形预测：错误与解决方案。*arXiv* 预印本 *arXiv:2503.16809*。Svensson, F., Aniceto, N., Norinder, U., Cortes-Ciriano, I., Spjuth, O., Carlsson, L., 和 Bender, A.（2018）。用于定量构效关系建模的保形回归——量化预测不确定性。*Journal of Chemical Information and Modeling*，58(5):1132–1140。Taylor, J. 和 Tibshirani, R.（2016）。l1 惩罚似然模型的选择后推断。Tian, J. 和 Ramdas, A.（2019）。ADDIS：一种用于在保守零假设下在线控制错误发现率的自适应丢弃算法。Tibshirani, R. J., Rinaldo, A., Tibshirani, R., 和 Wasserman, L.（2018）。模型选择后的均匀渐近推断与自助法。Vovk, V., Gammerman, A., 和 Shafer, G.（2005）。随机世界中的算法学习。Springer。Weinstein, A. 和 Ramdas, A.（2019）。错误覆盖率和错误符号率的在线控制。Xu, Z. 和 Ramdas, A.（2023）。使用 e 值进行在线多重检验。Xu, Z., Wang, R., 和 Ramdas, A.（2024）。基于 e 值的置信区间的选择后推断。*Electronic Journal of Statistics*，18(1):2292–2338。

A 延迟细节

A.1 命题1的完整版本

命题4。*As* 假设 $S_t = 1$。令 $\Pi_t^{(M)}$ 为来自 Π_t 上均匀分布的独立同分布样本。定义

$$\hat{R}_t = \{\pi \in \Pi_t^{(M)} : S_t(\pi) = 1\} \cup \{\pi_0\}, \qquad B_t = \{\pi \in \Pi_t^{(M)} : S_t(\pi) = 1,\ \pi(t) \neq t\}.$$

PEMI 预测集如下：

(a) $\hat{\mathcal{C}}_{\alpha,t}(\Pi_t^{(M)}) = \left\{y \in \mathcal{Y} : v(X_t, y) \leq \text{Quantile}\left(\beta; \{v(X_{\pi(t)}, Y_{\pi(t)})\}_{\pi \in B_t} \cup \{+\infty\}\right)\right\}$ *for* $\beta = \frac{\lceil (1-\alpha)\cdot|\hat{R}_t|\rceil}{|B_t|}$.

(b) $\hat{\mathcal{C}}_{\alpha,t}^{\text{rand}}(\Pi_t^{(M)}) = \{y \in \mathcal{Y} :\ v(X_t, y) \leq q\}$，其中随机阈值 $q \in \mathbb{R}$ 是从

$$q \sim \begin{cases} \text{Quantile}\left(\frac{r^*-1}{|B_t|}; \{v(X_{\pi(t)}, Y_{\pi(t)})\}_{\pi \in B_t} \cup \{+\infty\}\right), & \textit{with probability } p; \\ \text{Quantile}\left(\frac{r^*}{|B_t|}; \{v(X_{\pi(t)}, Y_{\pi(t)})\}_{\pi \in B_t} \cup \{+\infty\}\right), & \textit{with probability } 1-p. \end{cases}$$

这里，我们定义概率 $p = \frac{\alpha\cdot|B_t| - R_t| - (|B_t| - r^* + 1 - e_B^*)\, e_R^*}{e_R^*}$ 对于 $e_R^* = \#\{\pi \in \hat{R}_t :\ v(X_{\pi(t)}, Y_{\pi(t)}) = v_{(r^*)}\}$，$e_B^* = \#\{\pi \in B_t :\ v(X_{\pi(t)}, Y_{\pi(t)}) = v_{(r^*)}\}$，$r^* = \min\{i :\ v_{(i)} = v_{(\lceil(1-\alpha)\cdot|R_t|\rceil)}\}$，$v_{(1)} \leq \cdots \leq v_{(|B_t|)}$ 是 $\{v(X_{\pi(t)}, Y_{\pi(t)}) : \pi \in B_t\}$ 的顺序统计量。

A.2 e-LOND 选择下的 PEMI 集

命题 5 给出了 PEMI 集 $\hat{\mathcal{C}}_{\alpha,t}(\Pi_t^{(M)})$ 的显式形式。详细证明见附录 B.7。

命题 5。考虑一个满足 $E_t \geq 1/\alpha_t^{e\text{-}LOND}$ 的选定单元 $t \in \mathbb{N}^+$。对于 $\{-n+1, \ldots, t\}$ 上的任意排列 π，我们将排列 $F_{\pi(i)} = F(X_{\pi(i)}, c_{\pi(i)})$ 后的分数记为 $i \in \{-n+1, \ldots, t\}$。对于 $k \in \{0, 1\}$，我们定义排列 π 下每个 $j \in [t]$ 的两个留一保形 *p* 值为

$$p_{j,\pi}^{k,-} = \frac{\sum_{-n+1 \leq i \leq 0, i \neq \pi^{-1}(t)} \mathbb{1}\left\{\hat{F}_{\pi(i)} \geq \hat{F}_{\pi(j)},\ Y_{\pi(i)} \leq c_{\pi(i)}\right\} + k \cdot \mathbb{1}\left\{\hat{F}_t \geq \hat{F}_{\pi(j)}\right\} \cdot \mathbb{1}\left\{\pi^{-1}(t) \leq 0\right\}}{n+1},$$

$$p_{j,\pi}^{k,+} = \frac{1 + \sum_{-n+1 \leq i \leq 0, i \neq \pi^{-1}(t)} \mathbb{1}\left\{\hat{F}_{\pi(i)} \geq \hat{F}_{\pi(j)},\ Y_{\pi(i)} \leq c_{\pi(i)}\right\} + k \cdot \mathbb{1}\left\{\hat{F}_t \geq \hat{F}_{\pi(j)}\right\} \cdot \mathbb{1}\left\{\pi^{-1}(t) \leq 0\right\}}{n+1}.$$

对于每个 $i \in [t]$，令 $\hat{\alpha}_{i,\pi}^{k,-}$ 以及 $\hat{\alpha}_{i,\pi}^{k,+}$ 分别表示使用 $\{p_{j,\pi}^{k,-}\}_{1 \leq j \leq i-1}$ 和 $\{p_{j,\pi}^{k,+}\}_{1 \leq j \leq i-1}$ 通过 *LOND*算法计算得到的相应阈值。然后，我们定义排列 π 下的 *e* 值为 $E_{i,\pi}^k := \mathbb{1}\{p_{i,\pi}^{k,+} \leq \hat{\alpha}_{i,\pi}^{k,+}\}/\hat{\alpha}_{i,\pi}^{k,-}$，对所有 $i \in [t]$ 均成立。类似地，基于这些 *e* 值和 *e-LOND* 算法，我们可以计算时刻 t 的阈值 $\hat{\alpha}_{t,\pi}^k$。于是，对于 $y > c_t$，有 $\hat{R}_t(y) = \hat{R}_t^0$；对于 $y \leq c_t$，有 $\hat{R}_t(y) = \hat{R}_t^1$，其中

$$\hat{R}_t^k = \left\{\pi \in \Pi_t^{(M)} : E_{t,\pi}^k \geq 1/\hat{\alpha}_{t,\pi}^k\right\} \cup \{\pi_0\}, \quad B_t^k = \{\pi \in \hat{R}_t^k : \pi(t) \neq t\}.$$

最后，*PEMI* 预测集可通过以下方式计算：

$$\hat{\mathcal{C}}_{\alpha,t}(\Pi_t^{(M)}) = \{\, y \in \mathcal{Y} : y > c_t,\ v(X_t, y) \leq \hat{q}_0 \,\} \cup \{\, y \in \mathcal{Y} : y \leq c_t,\ v(X_t, y) \leq \hat{q}_1 \,\}, \tag{22}$$

其中 $\hat{q}_k = \text{分位数}\left(\frac{\lceil(1-\alpha)\cdot|R_t^k|\rceil}{|B_t^k|};\ \{v(X_{\pi(t)},\ Y_{\pi(t)})\}_{\pi \in B_t^k} \cup \{+\infty\}\right)$ 对于 $k = 0,\ 1$。

B 技术证明

B.1 定理1的证明

回顾一下，$\mathcal{D}_t = (Z_1,\ Z_2,\ \ldots,\ Z_t)$ 是截至时间 t 的有序数据集，其真实标签 Y_t 未知。令 $[\mathcal{D}_t] = [Z_1, \ldots, Z_{t-1}, Z_t]$ 表示 $\mathcal{D}_t$的无序集。我们将无序集的任一实现值记为

$[d_t]=[z_1,\ldots,z_{t-1},z_t]$。这里，给定 $[\mathcal{D}_t]=[d_t]$ 后，唯一的随机性在于 $\mathcal{D}_t$ 中的元素与 $[d_t]$中的值如何对应。我们将 $\{1,\ldots,t\}$上的所有排列构成的集合记为 Π_t。为简化记号，我们也将 $S_t(\pi;[d_t])=\mathcal{S}_t(z_{\pi(1)},z_{\pi(2)},\ldots,x_{\pi(t)})$ 和 $V_t(\pi;[d_t])=\mathcal{V}_t(z_{\pi(1)},z_{\pi(2)},\ldots,z_{\pi(t)})$ 分别记作。给定任意实现值 $[d_t]$，我们令 $\hat{}\pi$ 表示与观测数据对应的随机排列，因此 $(Z_1,\ldots,Z_t)=(z_{\hat{\pi}(1)},\ldots,z_{\hat{\pi}(t)})$。因此，观测数据的选择决策满足

$$S_t=\mathcal{S}_t(Z_1,\ldots,Z_{t-1},X_t)=S_t(\hat{\pi};[d_t]).$$

我们分别证明定理1 中的三个陈述。

情形 $\Pi_t^\star=\Pi_t$的证明。 我们首先关注$\mathcal{C}_{\alpha,\ t}(\Pi_t)$。只需证明，对任意实现值$[d_t]$,

$$\mathbb{P}\big(p_t(Y_t;\Pi_t)\le\alpha\mid[\mathcal{D}_t]=[d_t],S_t=1\big)\le\alpha. \tag{23}$$

首先，我们将满足选择规则的置换参考集定义为

$$R_t=\{\pi\in\Pi_t:S_t(\pi;[d_t])=1\}.$$

此处，根据定义，R_t 完全由 $[d_t]$决定，而非数据排序。然后我们断言，在给定 $[\mathcal{D}_t]=[d_t]$的条件下，

$$\big\{V_t^{Y_t}(\pi):\pi\in\hat{R}_t(Y_t;\Pi_t)\big\}=\big\{V_t(\pi;[d_t]):\pi\in R_t\big\}, \tag{24}$$

其中两边均允许元素重复，且两个集合均为无序集合。换言之，（24）意味着无论排列$\hat{}\pi$ 对应于哪一个，左端始终只依赖于 $[d_t]$。在给定 $[\mathcal{D}_t]=[d_t]$的条件下，我们有

$$V_t^{Y_t}(\pi)=\mathcal{V}_t(Z_{\pi(1)},Z_{\pi(2)},\ldots,Z_{\pi(t)})=\mathcal{V}_t(z_{\pi\circ\hat{\pi}(1)},z_{\pi\circ\hat{\pi}(2)},\ldots,z_{\pi\circ\hat{\pi}(t)})=V_t(\pi\circ\hat{\pi};[d_t]), \tag{25}$$

$$S_t^{Y_t}(\pi)=\mathcal{S}_t(Z_{\pi(1)},Z_{\pi(2)},\ldots,X_{\pi(t)})=\mathcal{S}_t(z_{\pi\circ\hat{\pi}(1)},z_{\pi\circ\hat{\pi}(2)},\ldots,x_{\pi\circ\hat{\pi}(t)})=S_t(\pi\circ\hat{\pi};[d_t]). \tag{26}$$

然后我们可以证明上述断言：

$$\begin{aligned}
\{V_t^{Y_t}(\pi):\pi\in\hat{R}_t(Y_t;\Pi_t)\}&\overset{(a)}{=}\{V_t(\pi\circ\hat{\pi};[d_t]):\pi\in\hat{R}_t(Y_t;\Pi_t)\}\\
&\overset{(b)}{=}\{V_t(\pi\circ\hat{\pi};[d_t]):\pi\in\Pi_t,\ S_t^{Y_t}(\pi)=1\}\\
&\overset{(c)}{=}\{V_t(\pi\circ\hat{\pi};[d_t]):\pi\in\Pi_t,\ S_t(\pi\circ\hat{\pi};[d_t])=1\}\\
&\overset{(d)}{=}\{V_t(\sigma;[d_t]):\sigma\in\Pi_t,\ S_t(\sigma;[d_t])=1\}\\
&\overset{(e)}{=}\{V_t(\pi;[d_t]):\pi\in R_t\}.
\end{aligned}$$

上文中，步骤 (a) 由（25）得出，步骤 (b) 根据 $\hat{}\ R_t(Y_t;\Pi_t)$ 的定义得出，步骤 (c) 由（26）得出，步骤 (d) 根据 Π_t 的群结构得出，步骤 (e) 则根据 R_t 的定义得出。至此，我们完成了公式（24）的证明。

为方便起见，我们将任意排列 π 在 $[d_t]$ 上的 p 值记为

$$p_t(\pi;[d_t])=\frac{\sum_{\pi'\in R_t}\mathbb{1}\{V_t(\pi;[d_t])\le V_t(\pi';[d_t])\}}{|R_t|}$$

并去除对 Π_t的依赖。回想一下，π_0 表示恒等排列，因此 $\pi_0\circ\hat{}\pi=\hat{}\pi$。根据公式（24），我们还可以得出 $|\hat{}R_t(Y_t;\Pi_t)|=|R_t|$。因此，给定 $[\mathcal{D}_t]=[d_t]$，我们有

$$p_t(Y_t;\Pi_t)=\frac{\sum_{\pi\in\hat{R}_t(Y_t;\Pi_t)}\mathbb{1}\{V_t^{Y_t}(\pi_0)\le V_t^{Y_t}(\pi)\}}{|\hat{R}_t(Y_t;\Pi_t)|}=\frac{\sum_{\pi\in R_t}\mathbb{1}\{V_t(\hat{\pi};[d_t])\le V_t(\pi;[d_t])\}}{|R_t|}=p_t(\hat{\pi};[d_t]). \tag{27}$$

回到公式（23）的证明，我们有

$$
\begin{aligned}
&\mathbb{P}\big(p_t(Y_t;\Pi_t)\le\alpha \,\big|\, [\mathcal{D}_t]=[d_t], S_t=1\big)\\
&=\mathbb{P}\big(p_t(Y_t;\Pi_t)\le\alpha \,\big|\, [\mathcal{D}_t]=[d_t], S_t(\hat{\pi};[d_t])=1\big)\\
&\overset{(a)}{=}\frac{\mathbb{P}\left(p_t(Y_t;\Pi_t)\le\alpha,\ S_t(\hat{\pi};[d_t])=1 \,|\, [\mathcal{D}_t]=[d_t]\right)}{\mathbb{P}\left(S_t(\hat{\pi};[d_t])=1 \,|\, [\mathcal{D}_t]=[d_t]\right)}\\
&\overset{(b)}{=}\frac{\mathbb{P}\left(p_t(\hat{\pi};[d_t])\le\alpha,\ S_t(\hat{\pi};[d_t])=1 \,|\, [\mathcal{D}_t]=[d_t]\right)}{\mathbb{P}\left(S_t(\hat{\pi};[d_t])=1 \,|\, [\mathcal{D}_t]=[d_t]\right)}\\
&\overset{(c)}{=}\frac{\frac{1}{|\Pi_t|}\sum_{\pi\in\Pi_t}\mathbb{1}\left\{p_t(\pi;[d_t])\le\alpha\right\}S_t(\pi;[d_t])}{\frac{1}{|\Pi_t|}\sum_{\pi\in\Pi_t}S_t(\pi;[d_t])}\\
&\overset{(d)}{=}\frac{1}{|R_t|}\sum_{\pi\in R_t}\mathbb{1}\left\{\frac{\sum_{\pi'\in R_t}\mathbb{1}\left\{V_t(\pi;[d_t])\le V_t(\pi';[d_t])\right\}}{|R_t|}\le\alpha\right\}\\
&\overset{(e)}{\le}\alpha.
\end{aligned}
$$

上文中，步骤 (a) 根据贝叶斯法则，步骤 (b) 由 (27) 得出，步骤 (c) 根据 $\hat{}\pi$ 从全排列集合 Π_t中均匀抽取这一事实得出，步骤 (d) 代入 p 值和参考集的显式定义 R_t，步骤 (e) 则由其定义得出。

然后，根据塔式性质，我们得出 (23)，从而得到覆盖保证。随机化版本的有效性证明与下文针对 $\Pi_t^\star=\Pi_t^{(M)}$ 的随机化版本的证明思路相同，此处省略。

情形 $\Pi_t^\star=\Pi_t^{(M)}$ 的证明。 我们的证明依赖于以下引理。

引理 1。 设 $\hat{}\pi,\pi^{(1)},\ldots,\pi^{(M)}$ 是从均匀分布(Π_t)中独立同分布抽取的随机排列。那么随机变量$(\hat{}\pi,\ \pi^{(1)}\circ\hat{}\pi,\ \ldots,\ \pi^{(M)}\circ\hat{}\pi)$ 是可交换的。

引理的证明 *1*。要证明 $(\hat{}\pi,\ \pi^{(1)}\circ\hat{}\pi,\ldots,\ \pi^{(M)}\circ\hat{}\pi)$ 的可交换性，只需证明它们的联合分布在对指标进行任意重新排序后保持不变。形式上，令 $[\hat{}\pi^M]=[\hat{}\pi,\ \pi^{(1)}\circ\hat{}\pi,\ldots,\ \pi^{(M)}\circ\hat{}\pi]$ 表示这些随机变量构成的无序集，并令 $[\sigma^M]=[\sigma^{(0)},\sigma^{(1)},\ldots,\sigma^{(M)}]$表示该无序集的任意一个取值。那么，只需证明对于 τ 的任意排列$\{0,1,\ldots,M\}$，都有

$$
\mathbb{P}\left(\hat{\pi}=\sigma^{(0)},\ \pi^{(1)}\circ\hat{\pi}=\sigma^{(1)},\ \ldots,\ \pi^{(M)}\circ\hat{\pi}=\sigma^{(M)}\right)=\mathbb{P}\left(\hat{\pi}=\sigma^{\tau(0)},\ \pi^{(1)}\circ\hat{\pi}=\sigma^{\tau(1)},\ \ldots,\ \pi^{(M)}\circ\hat{\pi}=\sigma^{\tau(M)}\right). \tag{28}
$$

由于 $\hat{}\pi,\pi^{(1)},\ldots,\pi^{(M)}$ 是从 Π_t 中独立均匀抽取的，我们可以直接计算等式左侧的值:

$$
\begin{aligned}
&\mathbb{P}\left(\hat{\pi}=\sigma^{(0)},\ \pi^{(1)}\circ\hat{\pi}=\sigma^{(1)},\ \ldots,\ \pi^{(M)}\circ\hat{\pi}=\sigma^{(M)}\right)\\
&=\mathbb{P}\left(\hat{\pi}=\sigma^{(0)},\ \pi^{(1)}=\sigma^{(1)}\circ(\sigma^{(0)})^{-1},\ \ldots,\ \pi^{(M)}=\sigma^{(M)}\circ(\sigma^{(0)})^{-1}\right)\\
&\overset{(a)}{=}\mathbb{P}(\hat{\pi}=\sigma^{(0)})\cdot\prod_{m=1}^{M}\mathbb{P}\left(\pi^{(m)}=\sigma^{(m)}\circ(\sigma^{(0)})^{-1}\right)\\
&\overset{(b)}{=}\left(\frac{1}{|\Pi_t|}\right)^{M+1}.
\end{aligned}
$$

步骤 (a) 源于 $(\hat{}\pi,\ \pi^{(1)},\ \ldots,\ \pi^{(M)})$ 的独立性。步骤 (b) 成立，因为 $\hat{}\pi,\ \pi^{(1)},\ \ldots,\ \pi^{(M)}$ 中的每一个都是在 Π_t 上服从均匀分布的随机变量，因此，取 π 中任一特定值的概率为。此外，根据 的群结构，排列 $\pi^{(1)}()$，...，$()\,\pi^{(1)}$ 都是 中的固定元素。因此，我们可以直接计算步骤 (b) 中的概率。

取 Π_t 中任一特定值的概率为 $1/|\Pi_t|$。此外，根据 Π_t 的群结构，排列 $\{\sigma^{(0)},\ \sigma^{(1)}\circ(\sigma^{(0)})^{-1},\ \ldots,\ \sigma^{(M)}\circ(\sigma^{(0)})^{-1}\}$ 都是 Π_t 中的固定元素。因此，我们可以直接计算步骤 (b) 中的概率。

类似地，对于任意排列 τ，我们有

$$\begin{aligned}
&\mathbb{P}\left(\hat{\pi}=\sigma^{\tau(0)},\ \pi^{(1)}\circ\hat{\pi}=\sigma^{\tau(1)},\ \ldots,\ \pi^{(M)}\circ\hat{\pi}=\sigma^{\tau(M)}\right)\\
&=\mathbb{P}\left(\hat{\pi}=\sigma^{\tau(0)},\ \pi^{(1)}=\sigma^{\tau(1)}\circ(\sigma^{\tau(0)})^{-1},\ \ldots,\ \pi^{(M)}=\sigma^{\tau(M)}\circ(\sigma^{\tau(0)})^{-1}\right)\\
&=\mathbb{P}(\hat{\pi}=\sigma^{\tau(0)})\cdot\prod_{m=1}^{M}\mathbb{P}\left(\pi^{(m)}=\sigma^{\tau(m)}\circ(\sigma^{\tau(0)})^{-1}\right)\\
&=\left(\frac{1}{|\Pi_t|}\right)^{M+1}.
\end{aligned}$$

因此，我们可以得出结论（28）。于是，$(\hat{\pi},\ \pi^{(1)}\circ\hat{\pi},\ldots,\ \pi^{(M)}\circ\hat{\pi})$ 的联合分布在排列下保持不变，因此这些随机排列是可交换的。由此完成引理 1 的证明。

□

关于 $\mathcal{C}_{\alpha,t}(\Pi_t^{(M)})$ 的详细证明。 回顾 $\Pi_t^{(M)}=(\pi^{(1)},\ldots,\pi^{(M)})$。令 $[\hat{\pi}^M]=[\hat{\pi},\ \pi^{(1)}\circ\hat{\pi},\ldots,\ \pi^{(M)}\circ\hat{\pi}]$表示随机排列构成的无序集，而 $[\sigma^M]=[\sigma^{(0)},\sigma^{(1)},\ldots,\sigma^{(M)}]$ 表示该无序集的任意一个实现。为证明（7），只需证明

$$\mathbb{P}\big(p_t(Y_t;\Pi_t^{(M)})\le\alpha\mid[\mathcal{D}_t]=[d_t],[\hat{\pi}^M]=[\sigma^M],S_t=1\big)\le\alpha. \tag{29}$$

首先，我们定义无序集 $[\sigma^M]$ 中的置换参考集如下

$$R_t^M=\left\{\sigma\in[\sigma^M]:S_t(\sigma;[d_t])=1\right\}.$$

在此，R_t^M完全由 由 $[d_t]$和 $[\sigma^M]$确定。然后，在给定 $[\mathcal{D}_t]=[d_t]$和 $[\hat{\pi}^M]=[\sigma^M]$的条件下，我们有 ：

$$\begin{aligned}
\{V_t^{Y_t}(\pi):\pi\in\hat{R}_t(Y_t;\Pi_t^{(M)})\}&=\{V_t(\pi\circ\hat{\pi};[d_t]):\pi\in\hat{R}_t(Y_t;\Pi_t^{(M)})\}\\
&=\{V_t(\pi\circ\hat{\pi};[d_t]):\pi\in\Pi_t^{(M)}\cup\{\pi_0\},\ S_t^{Y_t}(\pi)=1\}\\
&=\{V_t(\pi\circ\hat{\pi};[d_t]):\pi\in\Pi_t^{(M)}\cup\{\pi_0\},\ S_t(\pi\circ\hat{\pi};[d_t])=1\}\\
&=\{V_t(\sigma;[d_t]):\sigma\in[\sigma^M],\ S_t(\sigma;[d_t])=1\}\\
&=\{V_t(\pi;[d_t]):\pi\in R_t^M\}.
\end{aligned}$$

为方便起见，我们将 $[d_t]$上的 p 值定义为

$$p_t(\pi;[d_t])=\frac{\sum_{\pi'\in R_t^M}\mathbb{1}\{V_t(\pi;[d_t])\le V_t(\pi';[d_t])\}}{|R_t^M|}.$$

与 $\Pi_t^\star=\Pi_t$ 的证明类似，在给定 $[\mathcal{D}_t]=[d_t]$ 和 $[\hat{\pi}^M]=[\sigma^M]$的条件下，我们有:

$$\begin{aligned}
p_t(Y_t;\Pi_t^{(M)})&=\frac{\sum_{\pi\in\hat{R}_t(Y_t;\Pi_t^{(M)})}\mathbb{1}\{V_t^{Y_t}(\pi_0)\le V_t^{Y_t}(\pi)\}}{|\hat{R}_t(Y_t;\Pi_t^{(M)})|}\\
&=\frac{\sum_{\pi\in R_t^M}\mathbb{1}\{V_t(\hat{\pi};[d_t])\le V_t(\pi;[d_t])\}}{|R_t^M|}=p_t(\hat{\pi};[d_t]).
\end{aligned}\tag{30}$$

回到对（29）的证明，我们有

$$
\begin{aligned}
&\mathbb{P}\big(p_t(Y_t;\Pi_t^{(M)}) \le \alpha \mid [\mathcal{D}_t]=[d_t],[\hat{\pi}^M]=[\sigma^M],S_t=1\big)\\
&=\mathbb{P}\big(p_t(Y_t;\Pi_t^{(M)}) \le \alpha \mid [\mathcal{D}_t]=[d_t],[\hat{\pi}^M]=[\sigma^M],S_t(\hat{\pi};[d_t])=1\big)\\
&\overset{(a)}{=}\frac{\mathbb{P}\left(p_t(Y_t;\Pi_t^{(M)}) \le \alpha,\ S_t(\hat{\pi};[d_t])=1 \mid [\mathcal{D}_t]=[d_t],[\hat{\pi}^M]=[\sigma^M]\right)}{\mathbb{P}\left(S_t(\hat{\pi};[d_t])=1 \mid [\mathcal{D}_t]=[d_t],[\hat{\pi}^M]=[\sigma^M]\right)}\\
&\overset{(b)}{=}\frac{\mathbb{P}\left(p_t(\hat{\pi};[d_t]) \le \alpha,\ S_t(\hat{\pi};[d_t])=1 \mid [\mathcal{D}_t]=[d_t],[\hat{\pi}^M]=[\sigma^M]\right)}{\mathbb{P}\left(S_t(\hat{\pi};[d_t])=1 \mid [\mathcal{D}_t]=[d_t],[\hat{\pi}^M]=[\sigma^M]\right)}\\
&\overset{(c)}{=}\frac{\frac{1}{1+M}\sum_{m=0}^{M}\mathbb{1}\left\{p_t(\sigma^{(m)};[d_t]) \le \alpha\right\}S_t(\sigma^{(m)};[d_t])}{\frac{1}{1+M}\sum_{m=0}^{M}S_t(\sigma^{(m)};[d_t])}\\
&\overset{(d)}{=}\frac{1}{|R_t^M|}\sum_{\sigma^{(m)}\in R_t^M}\mathbb{1}\left\{\frac{\sum_{\sigma^{(n)}\in R_t^M}\mathbb{1}\left\{V_t(\sigma^{(n)};[d_t]) \le V_t(\sigma^{(m)};[d_t])\right\}}{|R_t^M|} \le \alpha\right\}\\
&\overset{(e)}{\le}\alpha.
\end{aligned}
$$

上文中，步骤 (a) 根据贝叶斯法则得出，步骤 (b) 由 (30) 得出，步骤 (c) 根据 ($\hat{}\pi$，$\pi^{(1)}\circ\hat{}\pi,\ldots,\pi^{(M)}\circ\hat{}\pi$) 可交换这一事实得出，步骤 (d) 代入 p 值和参考集 R_t^M的明确定义得出，步骤 (e) 则根据其定义得出。

由塔式性质可得（29），进而得到（7）。

证明 $\mathcal{C}^{\text{rand}}_{\alpha,t}(\Pi_t^{(M)})$。这里的论证与 $\mathcal{C}_{\alpha,t}(\Pi_t^{(M)})$ 的证明类似。我们的目标是证明

$$
\mathbb{P}\big(p_t^{\text{rand}}(Y_t;\Pi_t^{(M)}) \le \alpha \mid [\mathcal{D}_t]=[d_t],[\hat{\pi}^M]=[\sigma^M],S_t=1\big) \le \alpha. \tag{31}
$$

这里，我们采用与 R_t^M 相同的形式进行证明（b）。因此，在给定 $[\mathcal{D}_t]=[d_t]$ 和 $[\hat{}\pi^M]=[\sigma^M]$ 的条件下，我们有：

$$
\{V_t^{Y_t}(\pi):\pi\in\hat{R}_t(Y_t;\Pi_t^{(M)})\}=\{V_t(\pi;[d_t]):\pi\in R_t^M\}.
$$

然后，我们定义 $[d_t]$ 上的随机化 p 值：

$$
p_t^{\text{rand}}(\pi;[d_t])=\frac{\sum_{\pi'\in R_t^M}\mathbb{1}\{V_t(\pi;[d_t])<V_t(\pi';[d_t])\}+U_t\cdot\sum_{\pi'\in R_t^M}\mathbb{1}\{V_t(\pi;[d_t])=V_t(\pi';[d_t])\}}{|R_t^M|}.
$$

与 $\mathcal{C}_{\alpha,t}(\Pi_t^{(M)})$ 的证明类似，我们有

$$
p_t^{\text{rand}}(Y_t;\Pi_t^{(M)})=p_t^{\text{rand}}(\hat{\pi};[d_t]).
$$

回到对（31）的证明，我们有

$$
\begin{aligned}
&\mathbb{P}\big(p_t^{\text{rand}}(Y_t;\Pi_t^{(M)}) \le \alpha \mid [\mathcal{D}_t]=[d_t],[\hat{\pi}^M]=[\sigma^M],S_t=1\big)\\
&=\mathbb{P}\big(p_t^{\text{rand}}(Y_t;\Pi_t^{(M)}) \le \alpha \mid [\mathcal{D}_t]=[d_t],[\hat{\pi}^M]=[\sigma^M],S_t(\hat{\pi};[d_t])=1\big)\\
&\overset{(a)}{=}\frac{\mathbb{P}\left(p_t^{\text{rand}}(Y_t;\Pi_t^{(M)}) \le \alpha,\ S_t(\hat{\pi};[d_t])=1 \mid [\mathcal{D}_t]=[d_t],[\hat{\pi}^M]=[\sigma^M]\right)}{\mathbb{P}\left(S_t(\hat{\pi};[d_t])=1 \mid [\mathcal{D}_t]=[d_t],[\hat{\pi}^M]=[\sigma^M]\right)}\\
&\overset{(b)}{=}\frac{\mathbb{P}\left(p_t^{\text{rand}}(\hat{\pi};[d_t]) \le \alpha,\ S_t(\hat{\pi};[d_t])=1 \mid [\mathcal{D}_t]=[d_t],[\hat{\pi}^M]=[\sigma^M]\right)}{\mathbb{P}\left(S_t(\hat{\pi};[d_t])=1 \mid [\mathcal{D}_t]=[d_t],[\hat{\pi}^M]=[\sigma^M]\right)}
\end{aligned}
$$

$$\overset{(c)}{=} \frac{\sum_{\sigma^{(m)}\in R_t^M} \mathbb{P}\left\{ \frac{\sum_{\sigma^{(n)}\in R_t^M} \mathbb{1}\left\{V_t(\sigma^{(n)};[d_t])<V_t(\sigma^{(m)};[d_t])\right\}+U_t\cdot\sum_{\sigma^{(n)}\in R_t^M} \mathbb{1}\left\{V_t(\sigma^{(n)};[d_t])=V_t(\sigma^{(m)};[d_t])\right\}}{|R_t^M|} \le \alpha \mid [d_t],[\sigma^M] \right\}}{|R_t^M|}$$

$$\overset{(d)}{=} \alpha.$$

上文的步骤 (a)、(b)、(c) 遵循与 $\mathcal{C}_{\alpha,t}(\Pi_t^{(M)})$ 的证明相同的论证，而步骤 (d) 则由随机化 p 值的一项标准结果得出（Vovk 等人，2005，命题 2.4）。

然后，同样地，根据塔式性质，我们得出（31），进而得到（8）。

B.2 定理2的证明

为简化符号，在本证明中我们记为 $n = n_{\text{off}}$。我们将完整数据 $[\mathcal{D}_t] = [Z_{-n+1},\dots,Z_t]$ 的无序“数据袋”及其实现值 $[d_t] = [z_{-n+1},\dots,z_t]$ 定义为。回顾一下，$\hat{\pi}$ 表示与观测数据对应的随机排列，因此（

$$Z_1 \dots Z_t = z_{\hat{\pi}(1)} \dots z_{\hat{\pi}(t)} \qquad [\pi^M] = [\pi \quad \pi^{(1)}\circ\pi \quad \dots \quad \pi^{(M)}\circ\pi] \qquad \{-n+1,\dots,t\}$$

）。令 $\hat{\pi}$ 表示索引上的随机排列所构成的无序集，并令 $[\sigma^{(0)},\sigma^{(1)},\dots,\sigma^{(M)}]$ 表示其实现值。我们还记作 $S_t(\pi;[d_t]) = \mathcal{S}_t(z_{\pi(-n+1)},z_{\pi(2)},\dots,x_{\pi(t)})$，以及 $V_t(\pi;[d_t]) = \mathcal{V}(z_{\pi(-n+1)},z_{\pi(2)},\dots,z_{\pi(t)})$。因此，观测数据的选择决策满足

$$S_t = \mathcal{S}_t(Z_{-n+1},\dots,Z_{t-1},X_t) = S_t(\hat{\pi};[d_t]).$$

要证明（10），只需证明

$$\mathbb{P}\big(p_t^{\text{off}}(Y_t) \le \alpha \mid [\mathcal{D}_t] = [d_t], [\hat{\pi}^M] = [\sigma^M], S_t = 1\big) \le \alpha. \tag{32}$$

首先，我们定义无序集 $[\tau^M]$ 中满足选择规则的置换参考集

$$R_t^{\text{off}} = \left\{\sigma \in [\sigma^M] : S_t(\sigma;[d_t]) = 1\right\}.$$

这里，R_t^{off} 完全由 $[d_t]$ 和 $[\sigma^M]$ 决定。与定理 1 中 (b) 的证明类似，我们有，在给定 $[\mathcal{D}_t] = [d_t]$ 和 $[\hat{\pi}^M] = [\sigma^M]$ 的条件下，

$$\{V_t^{Y_t}(\pi) : \pi \in \hat{R}_t^{\text{off}}(Y_t)\} = \{V_t(\pi;[d_t]) : \pi \in R_t^{\text{off}}\}.$$

根据这一论断，我们有

$$p_t^{\text{off}}(Y_t) = p_t^{\text{off}}(\hat{\pi};[d_t]), \quad \text{where} \quad p_t^{\text{off}}(\pi;[d_t]) = \frac{\sum_{\pi'\in R_t^{\text{off}}} \mathbb{1}\{V_t(\pi;[d_t]) \le V_t(\pi';[d_t])\}}{|R_t^{\text{off}}|}.$$

回到证明（32），我们有

$$\begin{aligned}
&\mathbb{P}\big(p_t^{\text{off}}(Y_t) \le \alpha \mid [\mathcal{D}_t] = [d_t], [\hat{\pi}^M] = [\sigma^M], S_t = 1\big) \\
&= \mathbb{P}\big(p_t^{\text{off}}(Y_t) \le \alpha \mid [\mathcal{D}_t] = [d_t], [\hat{\pi}^M] = [\sigma^M], S_t(\hat{\pi};[d_t]) = 1\big) \\
&= \frac{\mathbb{P}\left(p_t^{\text{off}}(Y_t) \le \alpha,\ S_t(\hat{\pi};[d_t]) = 1 \mid [\mathcal{D}_t] = [d_t], [\hat{\pi}^M] = [\sigma^M]\right)}{\mathbb{P}\left(S_t(\hat{\pi};[d_t]) = 1 \mid [\mathcal{D}_t] = [d_t], [\hat{\pi}^M] = [\sigma^M]\right)} \\
&= \frac{\mathbb{P}\left(p_t^{\text{off}}(\hat{\pi};[d_t]) \le \alpha,\ S_t(\hat{\pi};[d_t]) = 1 \mid [\mathcal{D}_t] = [d_t], [\hat{\pi}^M] = [\sigma^M]\right)}{\mathbb{P}\left(S_t(\hat{\pi};[d_t]) = 1 \mid [\mathcal{D}_t] = [d_t], [\hat{\pi}^M] = [\sigma^M]\right)} \\
&= \frac{\frac{1}{1+M}\sum_{m=0}^{M} \mathbb{1}\left\{p_t^{\text{off}}(\sigma^{(m)};[d_t]) \le \alpha\right\} S_t(\sigma^{(m)};[d_t])}{\frac{1}{1+M}\sum_{m=0}^{M} S_t(\sigma^{(m)};[d_t])}
\end{aligned}$$

$$= \frac{1}{|R_t^{\text{off}}|} \sum_{\sigma^{(m)} \in R_t^{\text{off}}} \mathbb{1}\left\{ \frac{\sum_{\sigma^{(n)} \in R_t^{\text{off}}} \mathbb{1}\left\{V_t(\sigma^{(n)}; [d_t]) \le V_t(\sigma^{(m)}; [d_t])\right\}}{|R_t^{\text{off}}|} \le \alpha \right\}$$

$$\le \alpha.$$

上文中，每一步均遵循与定理 1 中证明 (b) 相同的思路。

然后，利用塔式性质可得（32），因此（10）。

B.3 定理3的证明

首先，回顾用于 FCR 控制的预测集构造。分类体系定义为 $\mathfrak{S} = \{ s_1 \ \dots \ s_t \}$ $s_1 = \mathcal{S}\ X_1 \quad s_2 = \mathcal{S}\ Z_1\ X_2 \quad \dots \quad s_t = \mathcal{S}\ Z_1\ \dots\ Z_{t-1}\ X_t$ $\Pi_t^{(M)} = (,,)$，其中 ( )、(,)、(,,)。回顾一下，$(\pi^{(1)}, \dots, \pi^{(M)})$ 是从索引$\{1, \dots, t\}$上的全排列集合中均匀抽取的随机排列。因此，参考集定义为

$$\hat{R}_t^{\text{str}}(y) = \{\pi \in \Pi_t^{(M)} : S_1^y(\pi) = s_1,\ \dots,\ S_t^y(\pi) = s_t\} \cup \{\pi_0\}.$$

将该参考集代入我们的框架，得到的预测集为

$$\hat{\mathcal{C}}_{\alpha,t}^{\text{str}} = \left\{y \in \mathcal{Y} : p_t^{\text{str}}(y) > \alpha\right\}, \quad \text{where} \quad p_t^{\text{str}}(y) = \frac{\sum_{\pi \in \hat{R}_t^{\text{str}}(y)} \mathbb{1}\{V_t^y(\pi_0) \le V_t^y(\pi)\}}{|\hat{R}_t^{\text{str}}(y)|}. \tag{33}$$

此外，我们沿用附录 B.1 中的记号。我们将数据的无序集定义为 $[\mathcal{D}_t] = [Z_1, \dots, Z_t]$，其实现值为 $[d_t] = [z_1, \dots, z_t]$。给定任意实现值 $[d_t]$，令 $\hat{}\pi$ 表示与观测数据对应的随机排列，使得 $(Z_1, \dots, Z_t) = (z_{\hat{\pi}(1)}, \dots, z_{\hat{\pi}(t)})$。因此，观测数据的选择决策满足 $S_t = S_t(\hat{}\pi; [d_t])$。令 $[\hat{}\pi^M] = [\hat{}\pi,\ \pi^{(1)} \circ \hat{}\pi, \dots,\ \pi^{(M)} \circ \hat{}\pi]$ 表示索引 $\{1, \dots, t\}$ 上的随机排列的无序集，并令 $[\sigma^{(0)}, \sigma^{(1)}, \dots, \sigma^{(M)}]$ 表示其实现值。我们也记 $S_t(\pi; [d_t]) = \mathcal{S}_t(z_{\pi(1)}, z_{\pi(2)}, \dots, x_{\pi(t)})$ 以及 $V_t(\pi; [d_t]) = \mathcal{V}_t(z_{\pi(1)}, z_{\pi(2)}, \dots, z_{\pi(t)})$。

引理 2. 由 (33) 定义的预测集满足强选择条件覆盖：对于任意 $t \ge 1$ 以及任意 $(s_1, \dots, s_{t-1}) \in \{0,1\}^{t-1}$，都有

$$\mathbb{P}(Y_t \in \hat{\mathcal{C}}_{\alpha,t}^{str} \mid S_1 = s_1, \dots, S_{t-1} = s_{t-1}, S_t = 1) \ge 1 - \alpha. \tag{34}$$

证明。 沿用前面的证明，要证明 (34)，只需证明

$$\mathbb{P}\big(p_t^{\text{str}}(Y_t) \le \alpha \mid [\mathcal{D}_t] = [d_t], [\hat{\pi}^M] = [\sigma^M], S_1 = s_1, \dots, S_{t-1} = s_{t-1}, S_t = 1\big) \le \alpha. \tag{35}$$

首先，我们定义无序集 $[\sigma^M]$ 中满足选择规则的置换参考集如下

$$R_t^{\text{str}} = \left\{\sigma \in [\sigma^M] : S_1(\sigma; [d_t]) = s_1, \dots, S_{t-1}(\sigma; [d_t]) = s_{t-1}, S_t(\sigma; [d_t]) = 1\right\}.$$

这里， R_t^{str}完全由 $[d_t]$和 $[\sigma^M]$确定。然后，在给定 $[\mathcal{D}_t] = [d_t]$和 $[\hat{}\pi^M] = [\sigma^M]$的条件下，有:

$$\begin{aligned}
&\{V_t^{Y_t}(\pi) : \pi \in \hat{R}_t^{\text{str}}(Y_t)\} \\
&= \{V_t(\pi \circ \hat{\pi}; [d_t]) : \pi \in \hat{R}_t^{\text{str}}(Y_t)\} \\
&= \{V_t(\pi \circ \hat{\pi}; [d_t]) : \pi \in \Pi_t^{(M)} \cup \{\pi_0\}, S_1^{Y_t}(\pi) = s_1,\ \dots,\ S_{t-1}^{Y_t}(\pi) = s_{t-1}, S_t^{Y_t}(\pi) = 1\} \\
&= \{V_t(\pi \circ \hat{\pi}; [d_t]) : \pi \in \Pi_t^{(M)} \cup \{\pi_0\}, S_1(\pi \circ \hat{\pi}; [d_t]) = s_1, \dots, S_{t-1}(\pi \circ \hat{\pi}; [d_t]) = s_{t-1}, S_t(\pi \circ \hat{\pi}; [d_t]) = 1\} \\
&= \{V_t(\sigma; [d_t]) : \sigma \in [\sigma^M], S_1(\sigma; [d_t]) = s_1, \dots, S_{t-1}(\sigma; [d_t]) = s_{t-1}, S_t(\sigma; [d_t]) = 1\}
\end{aligned}$$

$= \{V_t(\pi;\ [d_t])\ :\ \pi \in R_t^{\mathrm{str}}\}.$

根据这一结论，沿用定理 1 中 (b) 的证明中的相同论证，我们得到，给定$[\mathcal{D}_t] = [d_t]$ 和 $[\hat{}\pi^M] = [\sigma^M]$,

$$p_t^{\mathrm{str}}(Y_t) = p_t^{\mathrm{str}}(\hat{\pi}; [d_t]), \quad \text{where} \quad p_t^{\mathrm{str}}(\pi; [d_t]) = \frac{\sum_{\pi' \in R_t^{\mathrm{str}}} \mathbb{1}\{V_t(\pi; [d_t]) \le V_t(\pi'; [d_t])\}}{|R_t^{\mathrm{str}}|}.$$

回到 (35) 的证明，我们有

$$\begin{aligned}
&\mathbb{P}\big(p_t^{\mathrm{str}}(Y_t) \le \alpha \mid [\mathcal{D}_t] = [d_t], [\hat{\pi}^M] = [\sigma^M], S_1 = s_1, \ldots, S_{t-1} = s_{t-1}, S_t = 1\big) \\
&= \mathbb{P}\big(p_t^{\mathrm{str}}(Y_t) \le \alpha \mid [\mathcal{D}_t] = [d_t], [\hat{\pi}^M] = [\sigma^M], S_1(\hat{\pi}; [d_t]) = s_1, \ldots, S_{t-1}(\hat{\pi}; [d_t]) = s_{t-1}, S_t(\hat{\pi}; [d_t]) = 1\big) \\
&= \frac{\mathbb{P}\left(p_t^{\mathrm{str}}(Y_t) \le \alpha,\ S_1(\hat{\pi}; [d_t]) = s_1, \ldots, S_t(\hat{\pi}; [d_t]) = 1 \mid [\mathcal{D}_t] = [d_t], [\hat{\pi}^M] = [\sigma^M]\right)}{\mathbb{P}\left(S_1(\hat{\pi}; [d_t]) = s_1, \ldots, S_t(\hat{\pi}; [d_t]) = 1 \mid [\mathcal{D}_t] = [d_t], [\hat{\pi}^M] = [\sigma^M]\right)} \\
&= \frac{\mathbb{P}\left(p_t^{\mathrm{str}}(\hat{\pi}; [d_t]) \le \alpha,\ S_1(\hat{\pi}; [d_t]) = s_1, \ldots, S_t(\hat{\pi}; [d_t]) = 1 \mid [\mathcal{D}_t] = [d_t], [\hat{\pi}^M] = [\sigma^M]\right)}{\mathbb{P}\left(S_1(\hat{\pi}; [d_t]) = s_1, \ldots, S_t(\hat{\pi}; [d_t]) = 1 \mid [\mathcal{D}_t] = [d_t], [\hat{\pi}^M] = [\sigma^M]\right)} \\
&= \frac{\frac{1}{1+M}\sum_{m=0}^{M} \mathbb{1}\left\{p_t^{\mathrm{str}}(\sigma^{(m)}; [d_t]) \le \alpha\right\} \mathbb{1}\left\{S_1(\sigma^{(m)}; [d_t]) = s_1, \ldots, S_t(\sigma^{(m)}; [d_t]) = 1\right\}}{\frac{1}{1+M}\sum_{m=0}^{M} \mathbb{1}\left\{S_1(\sigma^{(m)}; [d_t]) = s_1, \ldots, S_t(\sigma^{(m)}; [d_t]) = 1\right\}} \\
&= \frac{1}{|R_t^{\mathrm{str}}|} \sum_{\sigma^{(m)} \in R_t^{\mathrm{str}}} \mathbb{1}\left\{\frac{\sum_{\sigma^{(n)} \in R_t^{\mathrm{str}}} \mathbb{1}\left\{V_t(\sigma^{(n)}; [d_t]) \le V_t(\sigma^{(m)}; [d_t])\right\}}{|R_t^{\mathrm{str}}|} \le \alpha\right\} \\
&\le \alpha.
\end{aligned}$$

由塔式性质，我们得出 (35)，因此 (34)。 □

定理 3 的证明。 与 Sale andRamdas (2025，命题 5.2) 的证明类似，FCR 控制直接由上文确立的强选择条件覆盖推出。假设在时间 T，有任意序列 $(s_1, \ldots, s_T)$。那么，

$$\begin{aligned}
&\mathbb{E}\left[\frac{\sum_{t=1}^{T} S_t \mathbb{1}\{Y_t \notin \hat{\mathcal{C}}_{\alpha,t}^{\mathrm{str}}\}}{\sum_{j=1}^{T} S_j} \,\middle|\, S_1 = s_1, \ldots, S_T = s_T\right] \\
&= \frac{1}{\sum_{j=1}^{T} s_j} \mathbb{E}\left[\sum_{t=1}^{T} S_t \mathbb{1}\{Y_t \notin \hat{\mathcal{C}}_{\alpha,t}^{\mathrm{str}}\} \,\middle|\, S_1 = s_1, \ldots, S_T = s_T\right] \\
&= \frac{1}{\sum_{j=1}^{T} s_j} \mathbb{E}\left[\sum_{t \le T: s_t = 1} \mathbb{1}\{Y_t \notin \hat{\mathcal{C}}_{\alpha,t}^{\mathrm{str}}\} \,\middle|\, S_1 = s_1, \ldots, S_T = s_T\right] \\
&= \frac{1}{\sum_{j=1}^{T} s_j} \sum_{t \le T: s_t = 1} \mathbb{P}\left(Y_t \notin \hat{\mathcal{C}}_{\alpha,t}^{\mathrm{str}} \,\middle|\, S_1 = s_1, \ldots, S_T = s_T\right) \\
&\overset{(a)}{=} \frac{1}{\sum_{j=1}^{T} s_j} \sum_{t \le T: s_t = 1} \mathbb{P}\left(Y_t \notin \hat{\mathcal{C}}_{\alpha,t}^{\mathrm{str}} \,\middle|\, S_1 = s_1, \ldots, S_{t-1} = s_{t-1}, S_t = 1\right) \\
&\overset{(b)}{\le} \frac{1}{\sum_{j=1}^{T} s_j} \sum_{t \le T: s_t = 1} \alpha
\end{aligned}$$

$= \alpha.$

上文中，步骤 (a) 源于我们的假设：给定 $(S_1, \ldots, S_t)$，时间 t 之后的选择决策与 $\hat{} \ Y_t$ 是否被 $C_{\alpha,t}$ 覆盖相互独立；步骤 (b) 则源于强选择条件覆盖的定义。

由塔式性质可知，FCR 控制（12）。

B.4 定理 4的证明

令 $\mathcal{D}_j = \mathcal{D}_{\text{calib}} \cup \{Z_{n+j}\}$ 和 $\mathcal{D}_j^c = \{X_{n+l}\}_{l \in [m] \setminus \{j\}}$。与附录 B.1类似，我们将无序集记为 $[\mathcal{D}_j] = [Z_1, \ldots, Z_n, Z_{n+j}]$，其实现值记为 $[d_j] = [z_1, \ldots, z_n, z_{n+j}]$。这里 $z_i = (x_i, y_i)$。给定任意实现值 $[d_j]$，令 $\hat{}\pi$ 表示与观测数据对应的随机排列，使得 $(Z_1, \ldots, Z_n, Z_{n+j}) = (z_{\hat{\pi}(1)}, \ldots, z_{\hat{\pi}(n)}, z_{\hat{\pi}(n+j)})$。因此，观测数据的选择子集满足

$$\hat{S} = \mathcal{S}(\mathcal{D}_{\text{calib}}, \mathcal{D}_{\text{test}}) = S^{\hat{\pi}}([d_j]).$$

回顾一下，$\Pi_{n+j}^{(M)} = (\pi^{(1)}, \ldots, \pi^{(M)})$ 是从索引 $\{1, \ldots, t\}$ 上的全排列集合中均匀抽取的随机排列。令 $[\hat{}\pi^M] = [\hat{}\pi, \ \pi^{(1)} \circ \hat{}\pi, \ldots, \ \pi^{(M)} \circ \hat{}\pi]$ 表示这些随机排列构成的无序集，其实现的排列为 $[\sigma^M] = [\sigma^{(0)}, \sigma^{(1)}, \ldots, \sigma^{(M)}]$。接着，我们将 $d_{\text{calib}}^{\pi} = (z_{\pi(1)}, \ldots, z_{\pi(n)})$ 和 $d_{\text{test}}^{\pi} = (X_{n+1}, \ldots, x_{\pi(n+j)}, \ldots, X_{n+m})$ 记为。此外，我们将排列后的数据的选择子集记为 $S^{\pi} = \mathcal{S}(d_{\text{calib}}^{\pi}, d_{\text{test}}^{\pi})$ 和 $V_{n+j}(\pi; [d_j]) = \mathcal{V}(z_{\pi(1)}, \ldots, z_{\pi(n)}, z_{\pi(n+j)})$。

要证明（14），只需证明

$$\mathbb{P}\big(p_{n+j}(Y_{n+j}) \le \alpha \mid j \in \hat{S}, [\mathcal{D}_j] = [d_j], \mathcal{D}_j^c, [\hat{\pi}^M] = [\sigma^M]\big) \le \alpha. \tag{36}$$

我们定义 在无序集 $[\sigma^M]$ 中满足选择规则的置换参考集为

$$R_{n+j}^M = \left\{\sigma \in [\sigma^M] : j \in S^{\sigma}\right\}.$$

这里，R_{n+j}^M 完全由 $[d_j]$、$\mathcal{D}_j^c$ 和 $[\sigma^M]$ 确定。然后，在给定 $[\mathcal{D}_t] = [d_t]$、$\mathcal{D}_j^c$ 和 $[\hat{}\pi^M] = [\sigma^M]$ 的条件下，我们有：

$$\begin{aligned}
\{V_{n+j}^{Y_{n+j}}(\pi) : \pi \in \hat{R}_{n+j}^M(Y_{n+j})\} &= \{V_{n+j}(\pi \circ \hat{\pi}; [d_j]) : \pi \in \hat{R}_{n+j}(Y_{n+j})\} \\
&= \{V_{n+j}(\pi \circ \hat{\pi}; [d_j]) : \pi \in \Pi_{n+j}^{(M)} \cup \{\pi_0\}, \ j \in \hat{S}^{\pi}(Y_{n+j})\} \\
&= \{V_{n+j}(\pi \circ \hat{\pi}; [d_j]) : \pi \in \Pi_{n+j}^{(M)} \cup \{\pi_0\}, \ j \in S^{\pi \circ \hat{\pi}}\} \\
&= \{V_{n+j}(\sigma; [d_j]) : \sigma \in [\sigma^M], \ j \in S^{\sigma}\} \\
&= \{V_{n+j}(\pi; [d_j]) : \pi \in R_{n+j}^M\}.
\end{aligned}$$

根据定理 1 中（b）的证明所用的相同论证，我们有，在给定 $[\mathcal{D}_t] = [d_t]$、$\mathcal{D}_j^c$ 和 $[\hat{}\pi^M] = [\sigma^M]$ 的条件下

$$p_{n+j}(Y_{n+j}) = p_{n+j}(\hat{\pi}; [d_j]), \quad \text{where} \quad p_{n+j}(\pi; [d_j]) = \frac{\sum_{\pi' \in R_{n+j}^M} \mathbb{1}\{V_{n+j}(\pi; [d_j]) \le V_{n+j}(\pi'; [d_j])\}}{|R_{n+j}^M|}.$$

回到式（36的证明，我们得到

$$\begin{aligned}
&\mathbb{P}\big(p_{n+j}(Y_{n+j}) \le \alpha \mid j \in \hat{S}, [\mathcal{D}_j] = [d_j], \mathcal{D}_j^c, [\hat{\pi}^M] = [\sigma^M]\big) \\
&= \mathbb{P}\big(p_{n+j}(Y_{n+j}) \le \alpha \mid j \in S^{\hat{\pi}}, [\mathcal{D}_j] = [d_j], \mathcal{D}_j^c, [\hat{\pi}^M] = [\sigma^M]\big)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\mathbb{P}\big(p_{n+j}(Y_{n+j}) \le \alpha, j \in S^{\hat{\pi}} \mid [\mathcal{D}_j] = [d_j], \mathcal{D}_j^c, [\hat{\pi}^M] = [\sigma^M]\big)}{\mathbb{P}\left(j \in S^{\hat{\pi}} \mid [\mathcal{D}_j] = [d_j], \mathcal{D}_j^c, [\hat{\pi}^M] = [\sigma^M]\right)} \\
&= \frac{\mathbb{P}\big(p_{n+j}(\hat{\pi}; [d_j]) \le \alpha, j \in S^{\hat{\pi}} \mid [\mathcal{D}_j] = [d_j], \mathcal{D}_j^c, [\hat{\pi}^M] = [\sigma^M]\big)}{\mathbb{P}\left(j \in S^{\hat{\pi}} \mid [\mathcal{D}_j] = [d_j], \mathcal{D}_j^c, [\hat{\pi}^M] = [\sigma^M]\right)} \\
&= \frac{\frac{1}{1+M}\sum_{m=0}^{M} \mathbb{1}\left\{p_{n+j}(\sigma^{(m)}; [d_j]) \le \alpha\right\} \cdot \mathbb{1}\{j \in S^{\sigma^{(m)}}\}}{\frac{1}{1+M}\sum_{m=0}^{M} \mathbb{1}\{j \in S^{\sigma^{(m)}}\}} \\
&= \frac{1}{|R_{n+j}^M|} \sum_{\sigma^{(m)} \in R_{n+j}^M} \mathbb{1}\left\{\frac{\sum_{\sigma^{(n)} \in R_{n+j}^M} \mathbb{1}\left\{V_{n+j}(\sigma^{(n)}; [d_j]) \le V_{n+j}(\sigma^{(m)}; [d_j])\right\}}{|R_{n+j}^M|} \le \alpha\right\} \\
&\le \alpha.
\end{aligned}
$$

然后根据塔式性质，我们得出式（36），因此得出式（14）。

B.5 命题1的证明

首先，回顾一般框架中的预测集：$\hat{\mathcal{C}}_{\alpha,t}(\Pi_t^{(M)}) = \{y \in \mathcal{Y} : p_t(y; \Pi_t^{(M)}) > \alpha\}$. 在协变量依赖设定中，注意到选择规则不依赖于候选值 y，因此 p 值的定义可以进一步简化。令 $\hat{R}_t = \{\pi \in \Pi_t^{(M)} : S_t(\pi) = 1\} \cup \{\pi_0\}$, $B_t = \{\pi \in \Pi_t^{(M)} : S_t(\pi) = 1, \pi(t) \ne t\}$, 并记 $V_t^y(\pi) = v(X_{\pi(t)}, Y_{\pi(t)})$ 和 $S_t^y(\pi) = S_t(\pi)$。于是，p 值可以展开为以下更明确的形式：

$$
\begin{aligned}
p_t(y; \Pi_t^{(M)}) &= \frac{\sum_{\pi \in \hat{R}_t} \mathbb{1}\{V_t^y(\pi_0) \le V_t^y(\pi)\}}{|\hat{R}_t|} \\
&= \frac{1 + \sum_{\pi \in \Pi_t^{(M)}} \mathbb{1}\{V_t^y(\pi_0) \le V_t^y(\pi)\} S_t(\pi)}{1 + \sum_{\pi \in \Pi_t^{(M)}} S_t(\pi)} \\
&= \frac{1 + \sum_{\pi \in \Pi_t^{(M)}} \mathbb{1}\{v(X_t, y) \le v(X_{\pi(t)}, Y_{\pi(t)})\} S_t(\pi)}{1 + \sum_{\pi \in \Pi_t^{(M)}} S_t(\pi)} \\
&= \frac{1 + \sum_{\pi \in \Pi_t^{(M)}} \mathbb{1}\{\pi(t) = t\} S_t(\pi)}{1 + \sum_{\pi \in \Pi_t^{(M)}} S_t(\pi)} + \frac{\sum_{\pi \in \Pi_t^{(M)}, \pi(t) \ne t} \mathbb{1}\{v(X_t, y) \le v(X_{\pi(t)}, Y_{\pi(t)})\} S_t(\pi)}{1 + \sum_{\pi \in \Pi_t^{(M)}} S_t(\pi)} \\
&= \frac{|\hat{R}_t| - |B_t|}{|\hat{R}_t|} + \frac{\sum_{\pi \in B_t} \mathbb{1}\{v(X_t, y) \le v(X_{\pi(t)}, Y_{\pi(t)})\}}{|\hat{R}_t|}.
\end{aligned}
$$

预测集 $\hat{\mathcal{C}}_{\alpha,t}(\Pi_t^{(M)})$ 由所有满足 $p_t(y; \Pi_t^{(M)}) \ge \alpha$ 的 y 组成。展开不等式可得

$$
\begin{aligned}
p_t(y; \Pi_t^{(M)}) > \alpha &\Leftrightarrow \frac{|\hat{R}_t| - |B_t|}{|\hat{R}_t|} + \frac{\sum_{\pi \in B_t} \mathbb{1}\{v(X_t, y) \le v(X_{\pi(t)}, Y_{\pi(t)})\}}{|\hat{R}_t|} > \alpha \\
&\Leftrightarrow \sum_{\pi \in B_t} \mathbb{1}\{v(X_t, y) \le v(X_{\pi(t)}, Y_{\pi(t)})\} > (\alpha - 1) \cdot |\hat{R}_t| + |B_t| \\
&\Leftrightarrow \sum_{\pi \in B_t} \mathbb{1}\{v(X_t, y) \le v(X_{\pi(t)}, Y_{\pi(t)})\} \ge \lfloor (\alpha - 1) \cdot |\hat{R}_t| + |B_t| \rfloor + 1.
\end{aligned}
$$

令 $K := \lfloor (\alpha - 1) \cdot |\hat{R}_t| + |B_t| \rfloor + 1$。那么，上述条件要求 $v(X_t, y)$ 不大于 $\{v(X_{\pi(t)}, Y_{\pi(t)}) : \pi \in B_t\}$ 中第 $(|B_t| - K + 1)$ 小的值。该排序还可以进一步表示为

q 的项　　分位数。形式上，与顺序统计量对应的分位数水平定义为

$$\begin{aligned}\beta &= \frac{|B_t| - K + 1}{|B_t|} \\ &= \frac{|B_t| - (\lfloor(\alpha-1)\cdot|\hat{R}_t| + |B_t|\rfloor + 1) + 1}{|B_t|} \\ &= \frac{\lceil(1-\alpha)\cdot|\hat{R}_t|\rceil}{|B_t|}.\end{aligned}$$

因此，预测集可以写成

$$\hat{\mathcal{C}}_{\alpha,t}(\Pi_t^{(M)}) = \left\{y \in \mathcal{Y} : v(X_t, y) \le \text{Quantile}\left(\beta;\ \{v(X_{\pi(t)}, Y_{\pi(t)})\}_{\pi\in B_t} \cup \{+\infty\}\right)\right\}.$$

证明完毕。

（b）的证明 回顾一下，在此设置下，随机化 p 值定义为

$$\begin{aligned}&p_t^{\text{rand}}(y; \Pi_t^{(M)}) \\ &= \frac{U_t \cdot \sum_{\pi\in\hat{R}_t(y;\Pi_t^{(M)})} \mathbb{1}\{V_t^y(\pi_0) = V_t^y(\pi)\} + \sum_{\pi\in\hat{R}_t(y;\Pi_t^{(M)})} \mathbb{1}\{V_t^y(\pi_0) < V_t^y(\pi)\}}{|\hat{R}_t(y;\Pi_t^{(M)})|} \\ &= \frac{U_t \cdot \left(1 + \sum_{\pi\in\Pi_t^{(M)}} \mathbb{1}\{V_t^y(\pi_0) = V_t^y(\pi)\}S_t(\pi)\right) + \sum_{\pi\in\Pi_t^{(M)}} \mathbb{1}\{V_t^y(\pi_0) < V_t^y(\pi)\}S_t(\pi)}{1 + \sum_{\pi\in\Pi_t^{(M)}} S_t(\pi)} \\ &= \frac{U_t \cdot \left(1 + \sum_{\pi\in\Pi_t^{(M)}} \mathbb{1}\{v(X_t, y) = v(X_{\pi(t)}, Y_{\pi(t)})\}S_t(\pi)\right) + \sum_{\pi\in\Pi_t^{(M)}} \mathbb{1}\{v(X_t, y) < v(X_{\pi(t)}, Y_{\pi(t)})\}S_t(\pi)}{1 + \sum_{\pi\in\Pi_t^{(M)}} S_t(\pi)}.\end{aligned}$$

这里，$U_t \sim$均匀分布$[0,1]$。由于即使在 $\pi(t) \neq t$ 的情况下，也可能发生 $v(X_t, y) = v(X_{\pi(t)}, Y_{\pi(t)})$，为方便记号，我们进一步定义

$$E_R(v(X_t, y)) = \left\{\pi \in \hat{R}_t : v(X_t, y) = v(X_{\pi(t)}, Y_{\pi(t)})\right\}.$$

其中 $\hat{R}_t = \{\pi \in \Pi_t^{(M)} :\ S_t(\pi) = 1\} \cup \{\pi_0\}$，类似地，$B_t = \{\pi \in \Pi_t^{(M)} :\ S_t(\pi) = 1,\ \pi(t) \neq t\}$ 如前所述。因此，p 值可以重写为

$$p_t^{\text{rand}}(y; \Pi_t^{(M)}) = \frac{U_t \cdot (|E_R(v(X_t, y))|)}{|\hat{R}_t|} + \frac{\sum_{\pi\in B_t} \mathbb{1}\{v(X_t, y) < v(X_{\pi(t)}, Y_{\pi(t)})\}}{|\hat{R}_t|}.$$

预测集 $\hat{\mathcal{C}}_{\alpha,t}^{\text{rand}}(\Pi_t^{(M)})$ 由所有满足 $p_t^{\text{rand}}(y; \Pi_t^{(M)}) > \alpha$ 的 y 组成。展开该不等式可得

$$U_t \cdot (|E_R(v(X_t, y))|) + \sum_{\pi\in B_t} \mathbb{1}\{v(X_t, y) < v(X_{\pi(t)}, Y_{\pi(t)})\} > \alpha \cdot |\hat{R}_t|.$$

令 $v_{(1)} \le v_{(2)} \le \cdots \le v_{(|B_t|)}$ 表示 $\{v(X_{\pi(t)}, Y_{\pi(t)})\}_{\pi\in B_t}$ 的排序值。根据 (a) 的证明，我们可以发现，在此情形下，用于确定性 p 值预测集的分位数阈值对应于这些有序值中的一个上分位数。这里，我们仍定义 $K := \lfloor(\alpha-1)\cdot|\hat{R}_t| + |B_t|\rfloor + 1$，并用 $v_{(|B_t|-K+1)}$ 表示该上分位数。为方便起见，我们记 $r^* = \min\left\{i : v_{(i)} = v_{(|B_t|-K+1)}\right\} = \min\{i : v_{(i)} = v_{(\lceil(1-\alpha)\cdot|\hat{R}_t|\rceil)}\}$，它是所有等于该上分位数的分数中的最小索引。我们还设

$$e_R^* = \#\{\pi \in \hat{R}_t : v(X_{\pi(t)}, Y_{\pi(t)}) = v_{(r^*)}\}, \quad e_B^* = \#\{\pi \in B_t : v(X_{\pi(t)}, Y_{\pi(t)}) = v_{(r^*)}\}.$$

根据 $v_{(r^*)}$ 的定义，从上面的不等式可以发现，若 $v(X_t, y) > v_{(r^*)}$，则对所有 U_t，$p_t(y) < \alpha$；若 $v(X_t, y) < v_{(r^*)}$，则对所有 U_t，$p_t(y) > \alpha$；而若 $v(X_t, y) = v_{(r^*)}$，则 $p_t(y) > \alpha$ 是否成立取决于 U_t 的取值。换言之，构造预测集的阈值以一定概率取上分位数，否则取下分位数。接下来，我们可以直接求解关于 U_t 的不等式，计算出选择上分位数的精确概率 p。

将 $v(X_t, y) = v_{(r^*)}$ 代入阈值可得

$$p_t^{\text{rand}}(y; \Pi_t^{(M)}) > \alpha \iff U_t > \frac{\alpha \cdot |\hat{R}_t| - (|B_t| - r^* + 1 - e_B^*)}{e_R^*} \equiv p,$$

其中，$p \in [0, 1]$ 是取上分位数阈值的概率。因此，随机化机制决定阈值取下分位数还是上分位数。具体而言，

$$q = \begin{cases} \text{Quantile}\left(\frac{r^*-1}{|B_t|};\ \{v(X_{\pi(t)}, Y_{\pi(t)})\}_{\pi \in B_t} \cup \{+\infty\}\right), & \text{if } U_t \le p \\ \text{Quantile}\left(\frac{r^*}{|B_t|};\ \{v(X_{\pi(t)}, Y_{\pi(t)})\}_{\pi \in B_t} \cup \{+\infty\}\right), & \text{if } U_t > p \end{cases}$$

或者，等价地，

$$q \sim \begin{cases} \text{Quantile}\left(\frac{r^*-1}{|B_t|};\ \{v(X_{\pi(t)}, Y_{\pi(t)})\}_{\pi \in B_t} \cup \{+\infty\}\right), & \text{with probability } p \\ \text{Quantile}\left(\frac{r^*}{|B_t|};\ \{v(X_{\pi(t)}, Y_{\pi(t)})\}_{\pi \in B_t} \cup \{+\infty\}\right), & \text{with probability } 1-p \end{cases}$$

因此，最终预测集为

$$\hat{\mathcal{C}}_{\alpha,t}^{\text{rand}}(\Pi_t^{(M)}) = \{y \in \mathcal{Y} : v(X_t, y) \le q\}.$$

证明至此完成。

B.6 命题2的证明

我们分别考虑 $y \le c_t$ 和 $y > c_t$。关键步骤是证明，对于任意 $y \le c_t$，对应的参考集 $\hat{} R_t(y)$ 等于 $\hat{} R_t^1$；对于任意 $y > c_t$，$\hat{}R_t(y)$ 等于 $\hat{} R_t^0$。这样，我们就可以分别将两种情况反演为预测区间的形式，然后合并两个结果，得到最终预测集。

情况 1： $y \le c_t$。在这种情况下，对于任意 $\pi \in \Pi_t^{(M)}$，有

$$\begin{aligned} p_{t,\pi}^w(y) &= \frac{w_t + \sum_{i=1}^{t-1} w_i \cdot \mathbb{1}\{\hat{F}_{\pi(i)} \ge \hat{F}_{\pi(t)}, Y_{\pi(i)} \le c_{\pi(i)}\}}{\sum_{i=1}^t w_i} \\ &= \frac{w_t + \sum_{1 \le i \le t-1, i \ne \pi^{-1}(t)} w_i \cdot \mathbb{1}\{\hat{F}_{\pi(i)} \ge \hat{F}_{\pi(t)},\ Y_{\pi(i)} \le c_{\pi(i)}\} + w_{\pi^{-1}(t)} \cdot \mathbb{1}\{\hat{F}_t \ge \hat{F}_{\pi(t)}, Y_t \le c_t\}}{\sum_{i=1}^t w_i} \\ &= \frac{w_t + \sum_{1 \le i \le t-1, i \ne \pi^{-1}(t)} w_i \cdot \mathbb{1}\{\hat{F}_{\pi(i)} \ge \hat{F}_{\pi(t)},\ Y_{\pi(i)} \le c_{\pi(i)}\} + w_{\pi^{-1}(t)} \cdot \mathbb{1}\{\hat{F}_t \ge \hat{F}_{\pi(t)}, y \le c_t\}}{\sum_{i=1}^t w_i} \\ &= \frac{w_t + \sum_{1 \le i \le t-1, i \ne \pi^{-1}(t)} w_i \cdot \mathbb{1}\{\hat{F}_{\pi(i)} \ge \hat{F}_{\pi(t)},\ Y_{\pi(i)} \le c_{\pi(i)}\} + w_{\pi^{-1}(t)} \cdot \mathbb{1}\{\hat{F}_t \ge \hat{F}_{\pi(t)}\}}{\sum_{i=1}^t w_i} \\ &= p_{t,\pi}^{w,1}. \end{aligned}$$

由于现有方法，例如 LOND (Javanmard 和 Montanari，2015年)、SAFFRON (Ramdas 等人，2019)，以及 ADDIS (Tian 和 Ramdas，2019)，在计算自适应阈值时仅使用历史信息中的先前 p 值和阈值，因此根据上述 p 值形式可知

当 $y \le c_t$ 时，信息 $\alpha_{t,\ \pi}(y) = \alpha^1_{t,\ \pi}$。

因此，对于任意 $y \le c_t$，我们都有

$$\begin{aligned}\hat{R}_t(y) &= \{\pi \in \Pi_t^{(M)} : S_t^y(\pi) = 1\} \cup \{\pi_0\} = \{\pi \in \Pi_t^{(M)} : p^w_{t,\pi}(y) \le \alpha_{t,\pi}(y)\} \cup \{\pi_0\} \\ &= \{\pi \in \Pi_t^{(M)} : p^{w,1}_{t,\pi} \le \alpha^1_{t,\pi}\} \cup \{\pi_0\} = \hat{R}^1_t.\end{aligned}$$

情况 2： $y > c_t$。在此情况下，对于任意 $\pi \in \Pi_t^{(M)}$，都有

$$\begin{aligned}p^w_{t,\pi}(y) &= \frac{w_t + \sum_{i=1}^{t-1} w_i \cdot \mathbb{1}\{\hat{F}_{\pi(i)} \ge \hat{F}_{\pi(t)}, Y_{\pi(i)} \le c_{\pi(i)}\}}{\sum_{i=1}^t w_i} \\ &= \frac{w_t + \sum_{1\le i\le t-1, i\ne\pi^{-1}(t)} w_i \cdot \mathbb{1}\{\hat{F}_{\pi(i)} \ge \hat{F}_{\pi(t)},\ Y_{\pi(i)} \le c_{\pi(i)}\} + w_{\pi^{-1}(t)} \cdot \mathbb{1}\{\hat{F}_t \ge \hat{F}_{\pi(t)}, Y_t \le c_t\}}{\sum_{i=1}^t w_i} \\ &= \frac{w_t + \sum_{1\le i\le t-1, i\ne\pi^{-1}(t)} w_i \cdot \mathbb{1}\{\hat{F}_{\pi(i)} \ge \hat{F}_{\pi(t)},\ Y_{\pi(i)} \le c_{\pi(i)}\} + w_{\pi^{-1}(t)} \cdot \mathbb{1}\{\hat{F}_t \ge \hat{F}_{\pi(t)}, y \le c_t\}}{\sum_{i=1}^t w_i} \\ &= \frac{w_t + \sum_{1\le i\le t-1, i\ne\pi^{-1}(t)} w_i \cdot \mathbb{1}\{\hat{F}_{\pi(i)} \ge \hat{F}_{\pi(t)},\ Y_{\pi(i)} \le c_{\pi(i)}\}}{\sum_{i=1}^t w_i} \\ &= p^{w,0}_{t,\pi}.\end{aligned}$$

类似地，当 $y > c_t$ 时，我们有 $\alpha_{t,\ \pi}(y) = \alpha^0_{t,\ \pi}$。因此，对于任意 $y > c_t$，我们都有

$$\begin{aligned}\hat{R}_t(y) &= \{\pi \in \Pi_t^{(M)} : S_t^y(\pi) = 1\} \cup \{\pi_0\} = \{\pi \in \Pi_t^{(M)} : p^w_{t,\pi}(y) \le \alpha_{t,\pi}(y)\} \cup \{\pi_0\} \\ &= \{\pi \in \Pi_t^{(M)} : p^{w,0}_{t,\pi} \le \alpha^0_{t,\pi}\} \cup \{\pi_0\} = \hat{R}^0_t.\end{aligned}$$

因此，对于所有 $y \le c_t$，参考集 $\hat{}\ R_t(y)$ 都相同，且等于 $\hat{}\ R^1_t$；对于所有 $y > c_t$，参考集也都相同，且等于 $\hat{}\ R^0_t$。按照命题 1 中的论证，我们可以对该过程求逆，从而分别得到对应于 $y \le c_t$ 和 $y > c_t$ 的两个预测子集，并将它们合并以形成最终预测集（17）。证明完毕。

B.7 命题 5 的证明

与附录 B.6 类似，我们分别考虑 $y \le c_t$ 和 $y > c_t$ 两种情况。我们将证明，对于任意$y \le c_t$，参考集满足 $\hat{}\ R_t(y) = \hat{}R^1_t$；对于任意 $y > c_t$，参考集满足 $\hat{}\ R_t(y) = \hat{}R^0_t$。

情形 1： $y \le c_t$。在此情形下，对于任意 $\pi \in \Pi_t^{(M)}$，均有

$$\begin{aligned}&p_{t,\pi}(y) \\ &= \frac{1 + \sum_{i=-n+1}^{0} \mathbb{1}\{\hat{F}_{\pi(i)} \ge \hat{F}_{\pi(t)}, Y_{\pi(i)} \le c_{\pi(i)}\}}{n+1} \\ &= \frac{1 + \sum_{-n+1\le i\le 0, i\ne\pi^{-1}(t)} \mathbb{1}\{\hat{F}_{\pi(i)} \ge \hat{F}_{\pi(t)},\ Y_{\pi(i)} \le c_{\pi(i)}\} + \mathbb{1}\{\hat{F}_t \ge \hat{F}_{\pi(t)},\ Y_t \le c_t\} \cdot \mathbb{1}\{\pi^{-1}(t) \le 0\}}{n+1} \\ &= \frac{1 + \sum_{-n+1\le i\le 0, i\ne\pi^{-1}(t)} \mathbb{1}\{\hat{F}_{\pi(i)} \ge \hat{F}_{\pi(t)},\ Y_{\pi(i)} \le c_{\pi(i)}\} + \mathbb{1}\{\hat{F}_t \ge \hat{F}_{\pi(t)},\ y \le c_t\} \cdot \mathbb{1}\{\pi^{-1}(t) \le 0\}}{n+1} \\ &= \frac{1 + \sum_{-n+1\le i\le 0, i\ne\pi^{-1}(t)} \mathbb{1}\{\hat{F}_{\pi(i)} \ge \hat{F}_{\pi(t)},\ Y_{\pi(i)} \le c_{\pi(i)}\} + \mathbb{1}\{\hat{F}_t \ge \hat{F}_{\pi(t)}\} \cdot \mathbb{1}\{\pi^{-1}(t) \le 0\}}{n+1} \\ &= p^1_{t,\pi}.\end{aligned}$$

按照相同的论证，我们可以将 $p_{t,\pi}^{+}(y)$ 和 $p_{t,\pi}^{-}(y)$ 表示为 $p_{t,\pi}^{1,+}$ 以及 $p_{t,\pi}^{1,-}$，其中 $y \le c_t$。此外，通过在 LOND算法中应用这些 p 值，我们可以计算出相应的自适应阈值 $\hat{\alpha}_{t,\pi}^{1}$、$\hat{\alpha}_{t,\pi}^{1,-}$ 以及 $\hat{\alpha}_{t,\pi}^{1,+}$。

因此，对于任意 $y \le c_t$，e 值和参考集可定义为

$$E_{t,\pi}(y) = \frac{\mathbb{1}\left\{p_{t,\pi}(y) \le \hat{\alpha}_t^{\mathrm{LOND},+}(y)\right\}}{\hat{\alpha}_t^{\mathrm{LOND},-}(y)} = \frac{\mathbb{1}\left\{p_{t,\pi}^1 \le \hat{\alpha}_t^{1,+}\right\}}{\hat{\alpha}_t^{1,-}} = E_{t,\pi}^1,$$

$$\hat{R}_t(y) = \left\{\pi \in \Pi_t^{(M)} : E_{t,\pi}(y) \ge \frac{1}{\hat{\alpha}_{t,\pi}(y)}\right\} \cup \{\pi_0\} = \left\{\pi \in \Pi_t^{(M)} : E_{t,\pi}^1 \ge \frac{1}{\hat{\alpha}_{t,\pi}^1}\right\} \cup \{\pi_0\} = \hat{R}_t^1.$$

情况 2： $y > c_t$。在此情况下，对于任意 $\pi \in \Pi_t^{(M)}$，均有

$$\begin{aligned}
&p_{t,\pi}(y) \\
&= \frac{1 + \sum_{i=-n+1}^{0} \mathbb{1}\{\hat{F}_{\pi(i)} \ge \hat{F}_{\pi(t)}, Y_{\pi(i)} \le c_{\pi(i)}\}}{n+1} \\
&= \frac{1 + \sum_{-n+1 \le i \le 0, i \ne \pi^{-1}(t)} \mathbb{1}\left\{\hat{F}_{\pi(i)} \ge \hat{F}_{\pi(t)},\ Y_{\pi(i)} \le c_{\pi(i)}\right\} + \mathbb{1}\left\{\hat{F}_t \ge \hat{F}_{\pi(t)},\ Y_t \le c_t\right\} \cdot \mathbb{1}\left\{\pi^{-1}(t) \le 0\right\}}{n+1} \\
&= \frac{1 + \sum_{-n+1 \le i \le 0, i \ne \pi^{-1}(t)} \mathbb{1}\left\{\hat{F}_{\pi(i)} \ge \hat{F}_{\pi(t)},\ Y_{\pi(i)} \le c_{\pi(i)}\right\} + \mathbb{1}\left\{\hat{F}_t \ge \hat{F}_{\pi(t)},\ y \le c_t\right\} \cdot \mathbb{1}\left\{\pi^{-1}(t) \le 0\right\}}{n+1} \\
&= \frac{1 + \sum_{-n+1 \le i \le 0, i \ne \pi^{-1}(t)} \mathbb{1}\left\{\hat{F}_{\pi(i)} \ge \hat{F}_{\pi(t)},\ Y_{\pi(i)} \le c_{\pi(i)}\right\}}{n+1} \\
&= p_{t,\pi}^0.
\end{aligned}$$

按照相同的论证，当 $y > c_t$ 时，我们可以将 $p_{t,\pi}^{+}(y)$ 和 $p_{t,\pi}^{-}(y)$ 表示为 $p_{t,\pi}^{0,+}$ 以及 $p_{t,\pi}^{0,-}$。此外，将这些 p 值应用于 LOND 算法后，我们可以计算出相应的自适应阈值 $\hat{\alpha}_{t,\pi}^{0}$、$\hat{\alpha}_{t,\pi}^{0,-}$ 和 $\hat{\alpha}_{t,\pi}^{0,+}$。

因此，对于任意 $y > c_t$，e 值和参考集可定义为

$$E_{t,\pi}(y) = \frac{\mathbb{1}\left\{p_{t,\pi}(y) \le \hat{\alpha}_t^{\mathrm{LOND},+}(y)\right\}}{\hat{\alpha}_t^{\mathrm{LOND},-}(y)} = \frac{\mathbb{1}\left\{p_{t,\pi}^0 \le \hat{\alpha}_t^{0,+}\right\}}{\hat{\alpha}_t^{0,-}} = E_{t,\pi}^0,$$

$$\hat{R}_t(y) = \left\{\pi \in \Pi_t^{(M)} : E_{t,\pi}(y) \ge \frac{1}{\hat{\alpha}_{t,\pi}(y)}\right\} \cup \{\pi_0\} = \left\{\pi \in \Pi_t^{(M)} : E_{t,\pi}^0 \ge \frac{1}{\hat{\alpha}_{t,\pi}^0}\right\} \cup \{\pi_0\} = \hat{R}_t^0.$$

因此，对于所有 $y \le c_t$，参考集 $\hat{R}_t(y)$ 均相同，且等于 $\hat{R}_t^1$；对于所有 $y > c_t$，参考集也均相同，且等于 $\hat{R}_t^0$。按照命题 1中的论证，我们可以反转该过程，得到分别对应于 $y \le c_t$ 和 $y > c_t$ 的两个预测子集，并将它们合并以形成最终预测集（22）。证明至此完成。

B.8 命题的证明 3

回顾一下，我们基于先前结果定义的选择规则为 $\hat{S}_t = \mathbb{1}\{\hat{\mu}(X_t) \ge \text{加权分位数}(1-\alpha,\ \{Y_i\}_{i=1}^{t-1})\}$。为方便起见，我们将此选择规则改写为 p 值形式：

$$p_t = \frac{\sum_{i=1}^{t-1} w_i \mathbb{1}\{Y_i \ge \hat{\mu}(X_t)\}}{\sum_{i=1}^{t-1} w_i}, \quad \hat{S}_t = \mathbb{1}\{p_t \le \alpha\}.$$

假设 $y \in I_j = (\mu_{(j-1)},\ \mu_{(j)})$。令 $l = \pi^{-1}(t)$ 为满足 $\pi(l) = t$ 的唯一整数。那么对于每个排列 $\pi \in \Pi_t^{(M)}$，均有

$$\begin{aligned} p_\pi(y) &= \frac{\sum_{i=1}^{t-1} w_i \mathbb{1}\{Y_{\pi(i)} \geq \hat{\mu}(X_{\pi(t)})\}}{\sum_{i=1}^{t-1} w_i} \\ &= \frac{\sum_{\substack{i=1 \\ i\neq l}}^{t-1} w_i \mathbb{1}\{Y_{\pi(i)} \geq \hat{\mu}(X_{\pi(t)})\} + w_l \cdot \mathbb{1}\{y \geq \hat{\mu}(X_{\pi(t)})\}}{\sum_{i=1}^{t-1} w_i}. \end{aligned}$$

由于 $\{\hat{\mu}_{(i)}\}_{i=1}^{t-1}$ 是预测值的有序序列，$\hat{\mu}(X_{\pi(t)})$ 不可能属于开区间 $(\mu_{(j-1)},\ \mu_{(j)})$。因此，我们将分别考虑两种情况：$\hat{\mu}(X_{\pi(t)}) \leq \hat{\mu}_{(j-1)}$ 和 $\hat{\mu}(X_{\pi(t)}) \geq \hat{\mu}_{(j)}$。

情况 1： $\hat{\mu}(X_{\pi(t)}) \leq \hat{\mu}_{(j-1)}$。由于我们有 $y \in (\mu_{(j-1)},\ \mu_{(j)})$，在此情况下，$y$ 总是大于 $\hat{\mu}(X_{\pi(t)})$。因此，对于任意排列 $\pi \in \Pi_t^{(M)}$，都有

$$p_\pi(y) = \frac{\sum_{\substack{i=1 \\ i\neq l}}^{t-1} w_i \mathbb{1}\{Y_{\pi(i)} \geq \hat{\mu}(X_{\pi(t)})\} + w_l \cdot \mathbb{1}\{y \geq \hat{\mu}(X_{\pi(t)})\}}{\sum_{i=1}^{t-1} w_i} = \frac{\sum_{\substack{i=1 \\ i\neq l}}^{t-1} w_i \mathbb{1}\{Y_{\pi(i)} \geq \hat{\mu}(X_{\pi(t)})\} + w_l}{\sum_{i=1}^{t-1} w_i} = p_\pi^1$$

情况 2： $\hat{\mu}(X_{\pi(t)}) \geq \hat{\mu}_{(j)}$。由于我们有 $y \in (\mu_{(j-1)},\ \mu_{(j)})$，在此情况下，$y$ 总是小于 $\hat{\mu}(X_{\pi(t)})$。因此，对于任意排列 $\pi \in \Pi_t^{(M)}$，都有

$$p_\pi(y) = \frac{\sum_{\substack{i=1 \\ i\neq l}}^{t-1} w_i \mathbb{1}\{Y_{\pi(i)} \geq \hat{\mu}(X_{\pi(t)})\} + w_l \cdot \mathbb{1}\{y \geq \hat{\mu}(X_{\pi(t)})\}}{\sum_{i=1}^{t-1} w_i} = \frac{\sum_{\substack{i=1 \\ i\neq l}}^{t-1} w_i \mathbb{1}\{Y_{\pi(i)} \geq \hat{\mu}(X_{\pi(t)})\}}{\sum_{i=1}^{t-1} w_i} = p_\pi^0$$

因此，对于任意 $y \in (\mu_{(j-1)},\ \mu_{(j)})$，我们有

$$\begin{aligned} \hat{R}_t(y) &= \{\pi \in \Pi_t^{(M)} : S_t^y(\pi) = 1\} \cup \{\pi_0\} \\ &= \{\pi \in \Pi_t^{(M)} : p_\pi(y) \leq \alpha\} \cup \{\pi_0\} \\ &= \left\{\pi \in \Pi_t^{(M)} : \hat{\mu}(X_{\pi(t)}) \leq \hat{\mu}_{(j-1)}, p_\pi^1 \leq \alpha\right\} \cup \left\{\pi \in \Pi_t^{(M)} : \hat{\mu}(X_{\pi(t)}) \geq \hat{\mu}_{(j)}, p_\pi^0 \leq \alpha\right\} \cup \{\pi_0\} \\ &= \hat{R}_t^{(j)} \end{aligned}$$

类似地，对于所有开区间 $I_j = (\mu_{(j-1)}, \mu_{(j)})$ $j = 1, \ldots, t$，我们可以分别构造参考集 $\hat{R}_t^{(j)}, , j = 1, \ldots, t$。按照命题 1 中的论证，我们可以反转该过程，并得到对应于 $y \in I_j, j = 1, \ldots, t$ 的预测子集。然后，对于每个边界点 $y_k = \hat{\mu}_{(k)} (k = 1, \ldots, t-1)$，我们直接填补 y_k，并在 $p_t^{\text{rand}}(y_k; \Pi_t^{(M)}) > \alpha$ 时将其纳入预测集。最后，在考虑 $\mathcal{Y}$ 的所有划分后，我们可以将它们合并为如式（18）所示的最终预测集。证明完毕。