Conformal changepoint localization

Rohan Hore*[1] and Aaditya Ramdas[1]

[1]Department of Statistics and Data Science, Carnegie Mellon University

October 27, 2025

Abstract

We study the problem of *offline changepoint localization*, where the goal is to identify the index at which the data-generating distribution changes. Existing methods often rely on restrictive parametric assumptions or asymptotic approximations, limiting their practical applicability. To address this, we propose a distribution-free framework, CONformal CHangepoint localization (CONCH), which leverages conformal p-values to efficiently construct valid confidence sets for the changepoint. Under mild assumptions of exchangeability within each segment and independence across segments, CONCH guarantees finite-sample coverage. By proving a conformal Neyman–Pearson lemma, we derive principled score functions that yield informative confidence sets. With appropriate score functions, we prove that the normalized length of the confidence set indeed shrinks to zero. We further establish a universality result showing that any distribution-free changepoint localization method can be viewed as an instance of CONCH. Experiments on synthetic and real data confirm that CONCH delivers precise and reliable confidence sets even in challenging settings.

Contents

*Corresponding author: rhore@andrew.cmu.edu

1 Introduction

In this paper, we study the problem of offline changepoint localization, where we are given an ordered sequence of data and are told that the underlying data-generating distribution has changed at some unknown index, called the *changepoint*. Unless explicitly mentioned, in this work, we assume that there is a single changepoint. As a simple illustration, suppose that the data are drawn independently from some distribution P_0 before the changepoint and from a different distribution $P_1 \neq P_0$ thereafter. The objective is to localize the changepoint, i.e., give a confidence set that contains this changepoint with high probability.

Changepoint localization is substantially more challenging than the related task of changepoint detection: merely identifying whether a change has occurred. Yet in domains such as operations engineering, econometrics, and biostatistics, the ability to retrospectively pinpoint the time of distributional change is often critical. Consider, for instance, a manufacturing context: quality measurements of a component may remain stable until a machine begins to malfunction, after which the measurements exhibit a systematic shift. Once the production batch has concluded, it becomes essential to determine when this shift first arose in order to diagnose the source of the malfunction and implement corrective measures.

1.1 Existing approaches

Offline changepoint analysis has been extensively studied due to its wide practical relevance; see Truong et al. [2020], Duggins [2010] for surveys. Classical methods such as CUSUM [Page, 1955] and conformal martingales [Vovk et al., 2003] primarily address the online detection problem rather than retrospective localization.

Likelihood-based procedures assume specific parametric models (e.g., Gaussian mean-shift, linear regression) [Kim and Siegmund, 1989, Quandt, 1958, Gurevich and Vexler, 2006] and mostly focus on detection. More recent post-detection localization techniques [Saha and Ramdas, 2025] still rely on restrictive model assumptions, such as known and non-overlapping pre-change and post-change families.

Several nonparametric methods achieve localization only asymptotically, including SMUCE [Frick et al., 2014], regression-based approaches [Xu et al., 2024], and Gaussian mean-shift intervals [Fotopoulos et al., 2010], among others [Bhattacharyya and Johnson, 1968, Zou et al., 2007]. The construction in Verzelen et al. [2023] attains theoretical optimality but involves non-computable constants, limiting practical use.

Bootstrap-based approaches [Cho and Kirch, 2022] target mean shifts but lack finite-sample validity and are computationally intensive. Rank-based nonparametric tests [Pettitt, 1979, Ross and Adams, 2012] are distribution-free for detection but do not provide confidence sets for localization and often have low power without additional structure. Multi-changepoint algorithms [Anastasiou and Fryzlewicz, 2022, Truong et al., 2020] typically adopt "isolate-detect" strategies and return only point estimates.

Conformal martingale methods [Vovk et al., 2003, Volkhonskiy et al., 2017, Vovk, 2021, Vovk et al., 2021, Nouretdinov et al., 2021, Shin et al., 2023] provide powerful tools for online detection but do not yield confidence sets for localization. Recently, MCP localization [Dandapanthula and Ramdas, 2025] introduced the first truly distribution-free approach to changepoint localization using a matrix of conformal p-values. In practice, however, it often produces wider confidence intervals than appear to be necessary, motivating the need for sharper, yet valid, distribution-free alternatives.

Overall, existing approaches are constrained by model assumptions, focus mainly on detection rather than localization, or trade statistical efficiency for distribution-free validity. In this work, we close this gap by proposing a simple yet principled framework for changepoint localization that is fully distribution-free, finite-sample valid, and yields informative confidence sets. The formal objective of distribution-free confidence sets is introduced in Section 2.

1.2 Our contributions

The main contributions of this work are summarized below:

- We introduce CONCH (CONformal CHangepoint localization), a framework that, given any $\mathbb{R}^{n-1}$-valued changepoint plausibility measure S and a confidence level $1-\alpha$, produces a finite-sample, distribution-free confidence set for the changepoint without making any restrictive assumptions on the pre- and post-change distributions.

- While our framework is valid for any choice of score function, offering great flexibility to the user, its statistical performance can be substantially improved by employing scores tailored to the problem at hand. We derive an expression for the optimal score based on a novel "Conformal Neyman–Pearson" lemma, which may be of independent interest.

- The optimal score requires oracle knowledge, so we propose practically applicable "near-optimal" score functions that yield narrow confidence sets. We also show that, for suitable score functions and under mild conditions, the confidence-set length shrinks to zero, enabling sharp localization.

- We show that CONCH has a universality property: any distribution-free confidence set for the changepoint is an instance of our framework. Moreover, we provide a simple algorithm to calibrate any heuristic confidence set, and an extension of CONCH to the multiple-changepoint setting.

- We demonstrate the practical utility of CONCH on diverse synthetic and real-world datasets. In particular, our method can wrap around any black-box classifier trained to distinguish pre- and post-change samples, producing informative confidence sets even when the change is subtle.

Organization of the paper. The rest of the paper is organized as follows. Section 2 formally defines the problem of distribution-free changepoint localization. Section 3 introduces our general framework, CONCH, and presents algorithms for its practical implementation. Section 4 develops an optimal score and provides guidance on selecting practical score functions that yield narrow confidence sets. Section 5 shows that, with appropriate score choices, the confidence sets shrink, enabling sharp localization. Section 6 establishes a universality result for CONCH. Section 6.1 builds on this foundation to introduce a calibration procedure that turns any localization method into a valid distribution-free one, and Section 6.2 wraps any segmentation algorithm with CONCH to enable multiple-changepoint detection. Section 7 presents empirical evaluations on synthetic and real-world datasets, demonstrating the applicability of our framework.

2 Distribution-free changepoint localization

In this section, we formally describe the problem of distribution-free offline changepoint localization. We begin by introducing some notation. Throughout the paper, $\mathbb{N}$ denotes the set of natural numbers, and for $K \in \mathbb{N}$ we write $[K] := \{1, \ldots, K\}$. For any set S, let $\mathcal{M}(S)$ denote the collection of probability measures on S and let 2^S denote the power set of S. Finally, we use $\stackrel{d}{=}$ to denote equality in distribution.

With this notation in place, consider an ordered sequence of $\mathcal{X}$-valued random variables $\mathbf{X} = (X_1, \ldots, X_n)$ for some $n \in \mathbb{N}$. We assume that there exists an unknown changepoint $\xi \in [n-1]$ such that

$$(X_1, \ldots, X_\xi) \sim \mathcal{P}_{0,\xi}, \qquad (X_{\xi+1}, \ldots, X_n) \sim \mathcal{P}_{1,\xi},$$

where $\mathcal{P}_{0,\xi} \in \mathcal{M}(\mathcal{X}^\xi)$ and $\mathcal{P}_{1,\xi} \in \mathcal{M}(\mathcal{X}^{n-\xi})$ denote the pre-change and post-change distributions, respectively. We write the joint distribution as $\mathcal{P} = \mathcal{P}_{0,\xi} \times \mathcal{P}_{1,\xi}$. In line with the distribution-free perspective, we impose no structural assumptions on $\mathcal{P}_{0,\xi}$ or $\mathcal{P}_{1,\xi}$ beyond the following.

Assumption 1. $\mathcal{P}_{0,\xi}$ and $\mathcal{P}_{1,\xi}$ are exchangeable. Specifically, for any permutations $\pi_L : [\xi] \to [\xi]$ and $\pi_R : [n] \setminus [\xi] \to [n] \setminus [\xi]$, it holds that

$$(X_1, \ldots, X_\xi) \stackrel{d}{=} (X_{\pi_L(1)}, \ldots, X_{\pi_L(\xi)}), \quad (X_{\xi+1}, \ldots, X_n) \stackrel{d}{=} (X_{\pi_R(\xi+1)}, \ldots, X_{\pi_R(n)}).$$

Moreover, the pre-change and post-change segments are independent: $\mathcal{P}_{0,\xi} \perp \mathcal{P}_{1,\xi}$.

In words, Assumption 1 requires that the distribution of $\mathbf{X}$ is invariant under arbitrary permutations of the entries to the left of ξ and, independently, under permutations of those to its right. A canonical example, mentioned in the introduction, is the i.i.d. changepoint model: the *pre-change* observations $(X_1, \ldots, X_\xi)$ are i.i.d. from some P_0, and independently, the *post-change* observations $(X_{\xi+1}, \ldots, X_n)$ are i.i.d. from some P_1.

For any $t \in [n-1]$, let $\mathcal{H}_{0,t}$ denote the hypothesis that t is the true changepoint and that the distributions $\mathcal{P}_{0,t}$ and $\mathcal{P}_{1,t}$ satisfy Assumption 1. We write $\mathbb{P}_t$ and $\mathbb{E}_t$ to denote probability and expectation, respectively, under this model class. We can now formally define what it means to construct a distribution-free confidence set for the changepoint.

Definition 1. Fix $\alpha \in (0, 1)$. A mapping $\mathcal{C}_{1-\alpha} : \mathcal{X}^n \to 2^{[n-1]}$ is called a *distribution-free* *confidence set for changepoint* at level $1-\alpha$ if

$$\mathbb{P}_\xi\big(\, \xi \in \mathcal{C}_{1-\alpha}(\mathbf{X})\,\big) \;\geq\; 1-\alpha. \tag{2.1}$$

Assumption 1 is considerably weaker than the working assumptions underlying most existing changepoint localization methods reviewed in Section 1.1. Prior approaches typically rely on strong parametric models or asymptotic approximations, highlighting the minimal nature of our assumption. While some recent methods [Dandapanthula and Ramdas, 2025] offer distribution-free guarantees under similarly mild conditions, they generally yield diffuse confidence sets. Our approach instead achieves sharper localization while retaining finite-sample validity, making it a significant contribution in this direction. The next section formally introduces our method and its main components.

3 Conformal changepoint localization

This section develops a conformal framework for localizing a changepoint. Conformal p-values, originally introduced by [Vovk et al., 1999, Shafer and Vovk, 2008] in the context of distribution-free predictive inference, have since been extended to a wide range of problems including outlier detection [Bates et al., 2023], post-prediction screening [Jin and Candès, 2023], and conditional two-sample testing [Wu et al., 2024], among others. Building on these developments, we adapt conformal p-values to the changepoint localization problem in an efficient manner, yielding confidence sets for the changepoint with guarantees as in (2.1).

3.1 General framework of CONCH algorithm

Building upon the machinery of conformal p-values, we first present the general framework for distribution-free changepoint localization, namely the Conformal changepoint localization (CONCH) algorithm. Our framework relies on two key components:

- **ChangePoint Plausibility (CPP) score:** We call any mapping $S : \mathcal{X}^n \to \mathbb{R}^{n-1}$ a changepoint plausibility score. Intuitively, for each candidate index $t \in [n-1]$, $S_t = (S(\cdot, \ldots, \cdot))_t$ assigns a score to quantify the chance that t is indeed a changepoint; a larger S_t indicates a stronger plausibility of t being a changepoint.

- **Split-permutation group:** For any $t \in [n-1]$, define the reduced set of permutations

$$\Pi_t := \Big\{\pi \in \mathcal{S}_n : \pi(i) \leq t \;\text{ for all } i \leq t, \;\; \pi(i) > t \;\text{ for all } i > t\Big\}. \tag{3.1}$$

Algorithm 1: CONCH: conformal changepoint localization algorithm

Input: $(X_t)_{t=1}^n$ (dataset), $1-\alpha$ (target coverage) and $S : \mathcal{X}^n \to \mathbb{R}^{n-1}$ (CPP score function)

Output: $\mathcal{C}_{1-\alpha}^{\text{CONCH}}$ (CONCH confidence set at level $1-\alpha$)

for $t \in [n-1]$ **do**
  $\Pi_t \leftarrow \{\pi \in \mathcal{S}_n : \text{for all } i \leq t,\ \pi(i) \leq t \text{ and for all } i > t,\ \pi(i) > t\}$;
  foreach $\pi \in \Pi_t$ **do**
    Evaluate $S_t(\pi(\mathbf{X}))$;
  end
  $p_t \leftarrow \frac{1}{|\Pi_t|} \sum_{\pi \in \Pi_t} \mathbb{1}\{S_t(\pi(\mathbf{X})) \leq S_t(\mathbf{X})\}$;
end
$\mathcal{C}_{1-\alpha}^{\text{CONCH}} \leftarrow \{\, t \in [n-1] : p_t > \alpha \,\}$;
return $\mathcal{C}_{1-\alpha}^{\text{CONCH}}$

Any $\pi \in \Pi_t$ freely permutes indices to the left and right of t independently while never mixing across the split.

Note that, if t is indeed the true changepoint, elements of Π_t preserve the pre-change and post-change exchangeability. Our framework crucially depends on this observation. More precisely, starting from any user-specified CPP score S, we define a conformal p-value p_t for each index $t \in [n-1]$ by looking at the normalized rank of $S_t(\mathbf{X})$ within the set of all permuted scores, $\{S_t(\pi(\mathbf{X})) : \pi \in \Pi_t\}$, i.e.,

$$p_t := \frac{1}{|\Pi_t|} \sum_{\pi \in \Pi_t} \mathbb{1}\{S_t(\pi(\mathbf{X})) \leq S_t(\mathbf{X})\}. \tag{3.2}$$

Intuitively, under $\mathcal{H}_{0,t}$, every permutation $\pi \in \Pi_t$ is equally likely, or equivalently, p_t is super-uniform under the null $\mathcal{H}_{0,t}$, a result we formally establish in Theorem 3.1. For brevity, the proof is deferred to Appendix A.1. Finally, the changepoint confidence set is then given by thresholding these p-values at level α:

$$\mathcal{C}_{1-\alpha}^{\text{CONCH}} := \{t \in [n-1] : p_t > \alpha\},$$

which attains the distribution-free validity in (2.1).

Theorem 3.1. *For each $t \in [n]$, p_t defined in (3.2) is a valid p-value under $\mathcal{H}_{0,t}$. In particular, for any $\alpha \in (0,1)$, $\mathbb{P}_\xi(p_\xi \leq \alpha) \leq \alpha$. Consequently, $\mathcal{C}_{1-\alpha}^{\text{CONCH}}$ is a distribution-free confidence set for changepoint.*

Remark 3.1 (Time-reversal symmetry)**.** The CONCH procedure is invariant under reversing the time index. Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ be the reversal of $\mathbf{X} = (X_1, \ldots, X_n)$, i.e., $Y_i := X_{n-i+1}$. Localizing $\xi \in [n]$ from $\mathbf{X}$ is then equivalent to localizing $n - \xi$ from $\mathbf{Y}$. Indeed, the permutation group Π_t acting on $\mathbf{X}$ corresponds to Π_{n-t} acting on $\mathbf{Y}$, and with the score functions defined accordingly, the CONCH confidence set computed from $\mathbf{Y}$ is exactly the image of the CONCH confidence set from $\mathbf{X}$ under the map $t \mapsto n - t$.

Remark 3.2. We highlight that the CONCH algorithm does not impose any restriction on the choice of CPP score, thereby providing significant flexibility for users to design their own plausibility measure. In particular, the score function may depend non-trivially on the multiset $\{X_1, \ldots, X_n\}$. For readers familiar with the distinction between full and split conformal methods in the setting of predictive inference, this corresponds to an adaptation of the full conformal approach to the changepoint localization setting.

3.2 CONCH-MC: randomized approximation for scalability

To compute the CONCH p-value p_t in (3.2), one must enumerate all permutations in Π_t and compute the corresponding score $S_t(\pi(\mathbf{X}))$ for each π. For large n, this may be computationally expensive. To reduce computational burden, we proceed as follows: to improve efficiency, we sample $\pi^{(1)}, \ldots, \pi^{(M)} \overset{\text{i.i.d.}}{\sim} \text{Unif}(\Pi_t)$, and then use a Monte Carlo approximation to p_t, in particular,

$$\tilde{p}_t := \frac{1 + \sum_{k=1}^{M} \mathbb{1}\left\{S_t(\pi^{(k)}(\mathbf{X})) \le S_t(\mathbf{X})\right\}}{1 + M}. \tag{3.3}$$

This yields a *randomized* confidence set $\{\, t \in [n-1] : \tilde{p}_t > \alpha \,\}$. We refer to this procedure as CONCH-MC, presented formally in Algorithm 2. Similar to the underlying principle of CONCH, any randomly sampled $\pi \in \Pi_t$ preserves pre-change and post-change exchangeability under $\mathcal{H}_{0,t}$, thereby providing us with a valid p-value $\tilde{p}_t$, as we establish formally in Theorem 3.2.

Theorem 3.2. *For any $t \in [n]$, p_t defined in (3.3) is a valid p-value under $\mathcal{H}_{0,t}$. In particular, for any $\alpha \in (0,1)$, $\mathbb{P}_\xi(\tilde{p}_\xi \le \alpha) \le \alpha$. Consequently, $\mathcal{C}^{\text{CONCH-MC}}_{1-\alpha}$ is a distribution-free confidence set for changepoint.*

3.3 Attaining exact validity of CONCH confidence sets

While both p-values p_t and $\tilde{p}_t$ in (3.2) and (3.3) control the Type I error under $\mathcal{H}_{0,t}$ at level α, it is sometimes desirable to attain *exact* level-α validity. Achieving exact validity can yield more powerful or sharper procedures. To this end, we introduce a simple randomized

Algorithm 2: CONCH-MC: CONCH with random permutations

Input: $(X_t)_{t=1}^n$ (dataset), $1-\alpha$ (target coverage), M (number of permutations) and $S : \mathcal{X}^n \to \mathbb{R}^n$ (CPP score function)

Output: $\mathcal{C}_{1-\alpha}^{\text{CONCH-MC}}$ (CONCH-MC confidence set at level $1-\alpha$)

for $t \in [n-1]$ **do**
    $\Pi_t \leftarrow \{\pi \in \mathcal{S}_n : \text{for all } i \le t,\ \pi(i) \le t \text{ and for all } i > t,\ \pi(i) > t\}$;
    for $k \in [M]$ **do**
        Sample $\pi^{(k)} \sim \Pi_t$;
        Evaluate $S_t(\pi^{(k)}(\mathbf{X}))$;
    end
    $\tilde{p}_t \leftarrow \frac{1}{M+1}\Big(1 + \sum_{k=1}^M \mathbb{1}\Big\{S_t(\pi^{(k)}(\mathbf{X})) \le S_t(\mathbf{X})\Big\}\Big)$;
end
$\mathcal{C}_{1-\alpha}^{\text{CONCH-MC}} \leftarrow \{t \in [n-1] : \tilde{p}_t > \alpha\}$
return $\mathcal{C}_{1-\alpha}^{\text{CONCH-MC}}$

refinement of the p-values that guarantees exact validity. Specifically, define

$$\bar{p}_t := \frac{1}{|\Pi_t|}\sum_{\pi \in \Pi_t} \mathbb{1}\{S_t(\pi(\mathbf{X})) < S_t(\mathbf{X})\} + U \cdot \frac{1}{|\Pi_t|}\sum_{\pi \in \Pi_t} \mathbb{1}\{S_t(\pi(\mathbf{X})) = S_t(\mathbf{X})\}, \tag{3.4}$$

where $U \sim \text{Unif}[0,1]$. Plugging these randomized p-values into the general CONCH framework yields the confidence set

$$\bar{\mathcal{C}}_{1-\alpha}^{\text{CONCH}} = \{\, t \in [n-1] : \bar{p}_t > \alpha \,\}, \tag{3.5}$$

which attains confidence exactly $1-\alpha$, as formalized below.

Theorem 3.3. *For each $t \in [n-1]$, $\bar{p}_t$ defined in (3.4) is a valid p-value under $\mathcal{H}_{0,t}$. In particular, for any $\alpha \in (0,1)$,*

$$\mathbb{P}_\xi(\bar{p}_\xi \le \alpha) = \alpha.$$

Consequently, $\mathbb{P}_\xi\Big(\xi \in \bar{\mathcal{C}}_{1-\alpha}^{\text{CONCH}}\Big) = 1-\alpha$ where $\bar{\mathcal{C}}_{1-\alpha}^{\text{CONCH}}$ is as defined in (3.5).

The proof of the result is deferred to Appendix A.1, where we actually prove a stronger version of this theorem. In particular, the validity of the p-value $\bar{p}_\xi$ holds even conditional on the multisets $\{X_1, \ldots, X_\xi\}$ and $\{X_{\xi+1}, \ldots, X_n\}$.

4 Guidelines for choosing the CPP score

The CONCH confidence sets from the earlier section retain validity as in (2.1) for any choice of CPP score, offering substantial flexibility in constructing valid confidence sets. This, however, naturally raises the question: how should one choose a score that yields narrow and informative sets? In what follows, we establish a few general properties regarding the influence of CPP scores on confidence sets and derive an optimal score. However, this optimal score requires oracle knowledge, so we give concrete proposals of practical score functions that closely mimic this ideal. Proofs of all results presented in this section are deferred to Appendix A.2.

We begin with two general properties that explain the influence of CPP score on the resulting CONCH set.

Proposition 4.1. *Fix $n \in \mathbb{N}$ and $\alpha \in (0,1)$.*

(i) **(Symmetry yields trivial p-values).** *Fix $t \in [n-1]$. If the t-th component S_t of the CPP score satisfies*

$$S_t(\cdot) = S_t(\pi(\cdot)) \qquad \text{for every } \pi \in \Pi_t, \tag{4.1}$$

then the p-values p_t and $\tilde{p}_t$ defined in (3.2) and (3.3) are identically 1. Consequently,

$$\mathbb{P}\Big(t \in \mathcal{C}_{1-\alpha}^{\text{CONCH}}\Big) \;=\; \mathbb{P}\Big(t \in \mathcal{C}_{1-\alpha}^{\text{CONCH-MC}}\Big) \;=\; 1.$$

(ii) ***(Conformal data-processing inequality).*** *Let C_1 denote the CONCH confidence set at level $1-\alpha$ based on a CPP score S, and let $\{p_{t,1}\}_{t\in[n-1]}$ be the corresponding conformal p-values. For any non-decreasing function $f : \mathbb{R} \to \mathbb{R}$, let C_2 be the CONCH confidence set at the same level based on the transformed score $f(S)$, with corresponding conformal p-values $\{p_{t,2}\}_{t\in[n-1]}$. Then,*

$$p_{t,1} \le p_{t,2} \quad \text{for all } t \in [n-1],$$

and consequently $C_1 \subseteq C_2$.

Part (i) of the proposition shows that t-wise symmetry, i.e., (4.1), yields trivial conformal p-values regardless of whether $\mathcal{H}_{0,t}$ holds, and therefore leads to overly conservative sets; therefore, such scores should be avoided in practice. Part (ii) of the result establishes a monotonicity property of CONCH: applying any non-decreasing transformation can only enlarge the set. In particular, any *strictly* increasing transformation on the CPP score leaves the confidence set unchanged. These properties help us make practical choices of the CPP score that yield meaningful confidence sets in practice.

For the remainder of this section, we focus on the canonical setting, namely the i.i.d. changepoint model. Specifically, let $\mathcal{P}_{\mathrm{IID}}$ denote the class of distributions for which there exists $\xi \in [n-1]$ such that

$$\mathcal{P}_{0,\xi} = \otimes_{t=1}^{\xi} P_0, \quad \mathcal{P}_{1,\xi} = \otimes_{t=\xi+1}^{n} P_1,$$

where P_0 and P_1 admit densities f_0 and f_1 with respect to a common dominating measure ν on $\mathcal{X}$.

4.1 Optimal CPP score function

In this section, we establish the optimal CPP score function, assuming the knowledge of both densities f_0, f_1, and the true changepoint ξ. By framing the task of identifying an optimal score as a testing problem involving a point null and a point alternative, we can directly apply the classical Neyman–Pearson (NP) lemma. This yields a similar optimality result tailored to the setting of distribution-free changepoint localization, which we call the *second Conformal NP Lemma*. The first instance of such a Conformal NP Lemma appears in Dandapanthula and Ramdas [2025], which establishes an analogous NP optimality result for a conformal p-value-based changepoint test.

For any $t \in [n-1]$, let $\mathcal{X}_{L,t} := \{X_1, \ldots, X_t\}$ and $\mathcal{X}_{R,t} := \{X_{t+1}, \ldots, X_n\}$ denote the (unordered) left and right multisets. Let $\mathcal{P}_{\mathbf{X}}^{(t)} = \mathcal{P}_{0,t} \times \mathcal{P}_{1,t}$ be the law of $\mathbf{X} = (X_1, \ldots, X_n)$ corresponding to a changepoint at t under the i.i.d. model class $\mathcal{P}_{\mathrm{IID}}$, and write $\mathcal{P}^{(t)}_{\mathbf{X}|\mathcal{X}_{L,t},\mathcal{X}_{R,t}}$ for the associated conditional distribution of $\mathbf{X}$ given $(\mathcal{X}_{L,t}, \mathcal{X}_{R,t})$. We define the hypothesis

$$\mathcal{H}_t' : \ \mathbf{X} \mid (\mathcal{X}_{L,t}, \mathcal{X}_{R,t}) \sim \mathcal{P}^{(t)}_{\mathbf{X}|\mathcal{X}_{L,t},\mathcal{X}_{R,t}},$$

for any $t \in [n-1]$, and would like to test $\mathcal{H}_t'$ using conformal p-values. Given a score $s : \mathcal{X}^n \to \mathbb{R}$ and the permutation set Π_t, define the randomized conformal p-value

$$p_t(s) = \frac{1}{|\Pi_t|} \sum_{\pi \in \Pi_t} \mathbb{1}\left\{s(\pi(\mathbf{X})) < s(\mathbf{X})\right\} \ + \ U \cdot \frac{1}{|\Pi_t|} \sum_{\pi \in \Pi_t} \mathbb{1}\left\{s(\pi(\mathbf{X})) = s(\mathbf{X})\right\}, \tag{4.2}$$

where $U \sim \mathrm{Unif}(0,1)$ is independent of $\mathbf{X}$ and the permutations. By Theorem 3.3, $p_t(s)$ is a valid p-value under $\mathcal{H}'_t$. Consequently,

$$\phi_t(\mathbf{X}; s) = \mathbb{1}\{p_t(s) \leq \alpha)\}$$

is a valid test at level α for the null $\mathcal{H}'_t$ with any score function s.

Observe that the test intuitively rejects when the score on the original data is *relatively smaller* than its permuted counterparts. Motivated by this, we seek an optimal score $s^\star$ such that the corresponding test $\phi_t(\mathbf{X}; s^\star)$ achieves maximum power against an alternative $\mathcal{H}'_r$ (with $r \neq t$), which posits that r rather than t is the true changepoint. Equivalently, we aim to identify a score $s^\star$ that tends to take smaller values under the alternative. In line with this intuition, the Conformal Neyman–Pearson lemma, stated below, formally establishes that the likelihood ratio $s^\star(\cdot) = \mathcal{P}_X^{(t)}(\cdot)/\mathcal{P}_X^{(r)}(\cdot)$ defines the optimal test.

Lemma 4.2 (Conformal NP lemma)**.** *Fix $t, r \in [n-1]$ with $t \neq r$. The power, $\mathbb{E}_{\mathcal{H}'_r}[\phi_t(\mathbf{X}; s)]$, is maximized by the score function*

$$s^\star(x_1, \ldots, x_n) := \frac{\prod_{i \leq t} f_0(x_i) \prod_{i>t} f_1(x_i)}{\prod_{i \leq r} f_0(x_i) \prod_{i>r} f_1(x_i)}.$$

Finally, the Conformal NP lemma can be leveraged within the CONCH framework to derive the CPP score that would yield the narrowest confidence set. We observe that the conformal p-value in (3.2) must be valid under $\mathcal{H}_{0,t}$, while be sufficiently small to sharply detect the true changepoint $\xi \neq t$ under $\mathcal{H}_{0,\xi}$. Since only the t-th component of CPP score, S_t, determines p_t, the task of optimizing S_t boils down to finding the optimal test for $\mathcal{H}'_t$ v.s $\mathcal{H}'_\xi$.

We make this connection precise in the theorem below. For notational convenience, we write $C_{1-\alpha}^{\mathrm{CONCH}}(S)$ to denote the randomized CONCH confidence set (3.5) constructed with the CPP score S.

Theorem 4.3. *Any strictly increasing transformation of the CPP score S^{OPT} defined by*

$$S_t^{\mathrm{OPT}}(x_1, \ldots, x_n) = \frac{\prod_{i \leq t} f_0(x_i) \prod_{i>t} f_1(x_i)}{\prod_{i \leq \xi} f_0(x_i) \prod_{i>\xi} f_1(x_i)} \tag{4.3}$$

achieves the minimum expected length of the CONCH confidence set. In particular, for any score function $S : \mathcal{X}^n \to \mathbb{R}^{n-1}$,

$$\mathbb{E}_{\mathcal{H}_{0,\xi} \cap \mathcal{P}_{\mathrm{IID}}}\Big[\,|\bar{C}_{1-\alpha}^{CONCH}(S)|\,\Big] \;\geq\; \mathbb{E}_{\mathcal{H}_{0,\xi} \cap \mathcal{P}_{\mathrm{IID}}}\Big[\,|\bar{C}_{1-\alpha}^{CONCH}(S^{\mathrm{OPT}})|\,\Big].$$

The optimal CPP score function (4.3) depends on the unknown pre-change and post-change densities f_0 and f_1 as well as the true changepoint ξ, and is therefore not directly implementable in practice. In the next subsection, we propose score functions that closely mimic the optimal score, thus providing 'near-optimal' performance in practice.

4.2 Practical choices for CPP score

Motivated by the Conformal NP lemma, we now describe a few principled choices for CPP scores in this setting.

(1) Weighted mean difference. If densities f_0 and f_1 differ merely by a location shift, a natural CPP score is given by

$$S_t(x_1, \cdots, x_n) = \left| \frac{\sum_{i=1}^t w_{t,i} x_i}{\sum_{i=1}^t w_{t,i}} - \frac{\sum_{i>t}^n w_{t,i} x_i}{\sum_{i>t}^n w_{t,i}} \right|. \tag{4.4}$$

The weights $\{w_{t,i}\}$ are introduced to break the t-wise symmetry property, (4.1), and therefore to avoid trivial confidence sets. Intuitively, observations closer to the t-th index should receive more weight when defining the score at t. Reasonable choices for weights include:

$$w_{t,i} = 1 - \frac{|i-t|}{n} \quad \text{or} \quad w_{t,i} = \exp\Big(-|i-t|/n\Big).$$

If $t \in [n-1]$ is believed to be a changepoint, the weighted means on the left and right sides should differ substantially, producing a high CPP score at t as required.

(2) Oracle log likelihood-ratio (LLR). Suppose f_0 and f_1 are known. Then, the optimal CPP score function in (4.3) can be approximated by evaluating the complete likelihood at MLE $\hat{t}$ instead of the true changepoint ξ. Therefore, we may take the CPP score given by

$$S_t(x_1, \cdots, x_n) = \log \left(\frac{\prod_{i \le t} f_0(x_i) \prod_{i>t} f_1(x_i)}{\prod_{i \le \hat{\xi}(\mathbf{x})} f_0(x_i) \prod_{i > \hat{\xi}(\mathbf{x})} f_1(x_i)} \right), \tag{4.5}$$

where we write $\mathbf{x} = (x_1, \ldots, x_n)$, and

$$\hat{\xi}(\mathbf{x}) := \operatorname*{argmax}_{s \in [n-1]} \log\Big(\prod_{i \le s} f_0(x_i) \prod_{i>s} f_1(x_i)\Big) \tag{4.6}$$

is the MLE estimate[1] of the changepoint. If $t \in [n-1]$ is indeed the changepoint, then $\hat{\xi} \approx t$ and S_t will be large, indicating strong plausibility for a change. Since this score closely approximates (4.3), it is expected to sharply localize the true changepoint, as verified in our experiments too.

(3) Learned LLR. When f_0 and f_1 are unknown, for each $t \in [n-1]$, one can plug in estimates (parametric or non-parametric) $\hat{f}_{t,0}$ and $\hat{f}_{t,1}$, and instead consider the CPP score given by

$$S_t(x_1, \cdots, x_n) = \log\left(\frac{\prod_{i \le t} \hat{f}_{t,0}(x_i) \prod_{i>t} \hat{f}_{t,1}(x_i)}{\prod_{i \le \tilde{t}} \hat{f}_{\tilde{t},0}(x_i) \prod_{i>\tilde{\xi}} \hat{f}_{\tilde{\xi},1}(x_i)}\right) \tag{4.7}$$

with $\tilde{\xi} = \operatorname{argmax}_{s \in [n-1]} \log\Big(\prod_{i \le s} \hat{f}_{s,0}(x_i) \prod_{i>s} \hat{f}_{s,1}(x_i)\Big)$ being the corresponding MLE.

(4) Classifier based LLR. Instead of estimating the densities f_0 and f_1 directly, one can train a binary classifier $\hat{g}$ to distinguish post-change from pre-change samples (labeled $Y = 1$ and $Y = 0$, respectively). By Bayes' rule, we have

$$\log \frac{f_1(x)}{f_0(x)} = \log \frac{\mathbb{P}(Y = 1 | X = x)}{\mathbb{P}(Y = 0 \mid X = x)} - \log \frac{\pi_1}{\pi_0},$$

where π_1 and π_0 are class priors. If $\hat{g}$ is trained on balanced data and we write $\hat{g}(x) \in (0, 1)$ to denote the predicted probability of post-change membership, then we obtain the approximation

$$\log \frac{f_1(x)}{f_0(x)} \approx \operatorname{logit} \hat{g}(x) := \log \frac{\hat{g}(x)}{1 - \hat{g}(x)}.$$

The log odds components in (4.5) can then be approximated by the classifier logits to define a practically implementable CPP score. While the choice of classifiers does not affect the validity of our method, a well-trained classifier improves power.

5 Consistency of CONCH confidence sets

In this section, we study the consistency of CONCH confidence sets. In particular, we show that the normalized length $|\mathcal{C}^{\mathrm{CONCH}}_{n,1-\alpha}|/(n-1)$ converges to 0 as $n \to \infty$. To make the setting precise, we generalize the i.i.d. changepoint model: for each n, we observe $(X_{1,n}, \ldots, X_{n,n}) \in$

[1]Note that the MLE estimator $\hat{\xi}(\mathbf{x})$ is a function of the observed data $(x_1, \ldots, x_n)$. Thus, for each permutation π, computing the permuted score $S_t(\pi(\mathbf{X}))$ requires first computing $\hat{\xi}(\pi(\mathbf{X}))$, MLE estimate on the permuted data sequence.

$\mathcal{X}^n$ with a single changepoint at $\xi_n \in [n-1]$, and

$$(X_{1,n}, \ldots, X_{\xi_n,n}) \sim \otimes_{t=1}^{\xi_n} P_0, \qquad (X_{\xi_n+1,n}, \ldots, X_{n,n}) \sim \otimes_{t=\xi_n+1}^{n} P_1,$$

where P_0 and P_1 admit densities f_0 and f_1 with respect to a common dominating measure ν on $\mathcal{X}$. Moreover, we assume that the underlying model satisfies:

(A1) The changepoint lies in the interior: there exists $\tau \in (0,1)$ such that

$$\xi_n/n \longrightarrow \tau \qquad \text{as } n \to \infty.$$

(A2) Let $\mathrm{KL}(P,Q) = \mathbb{E}_{X\sim P}\Big[\log\Big(\frac{\mathsf{d}P}{\mathsf{d}Q}(X)\Big)\Big]$ denote the Kullback–Leibler divergence. We assume

$$0 < \mathrm{KL}(P_0\|P_1),\ \mathrm{KL}(P_1\|P_0) < \infty.$$

(A3) Further, let $\mathrm{V}_{\mathrm{KL}}(P,Q) = \mathrm{Var}_{X\sim P}\Big[\log\Big(\frac{\mathsf{d}P}{\mathsf{d}Q}(X)\Big)\Big]$ denote the Var-entropy measure, and assume that

$$0 < \mathrm{V}_{\mathrm{KL}}(P_0\|P_1),\ \mathrm{V}_{\mathrm{KL}}(P_1\|P_0) < \infty.$$

Given any CPP score function, (3.2) yields conformal p-values $(p_{1,n}, \ldots, p_{n-1,n})$ and the associated CONCH confidence set $\mathcal{C}_{n,1-\alpha}^{\mathrm{CONCH}} := \{\, t \in [n-1] :\ p_{t,n} > \alpha \,\}$. We now state the consistency result for the practical score functions introduced in the previous section.

Theorem 5.1 (Asymptotic sharpness of CONCH)**.** *In the setting above, under (A1)–(A3), for the optimal score* (4.1) *and the oracle LLR score* (4.5),

$$\frac{\left|\mathcal{C}_{n,1-\alpha}^{CONCH}\right|}{n-1} \xrightarrow{P} 0 \qquad \textit{as } n \to \infty.$$

The proof appears in Appendix A.3.2. In Appendix A.3.3, we further establish a stronger statement: under an additional sub-Gaussianity assumption on the log-likelihood ratio, for all indices ℓ with $|\ell - \xi_n| \gg \log n$, the corresponding CONCH p-values $p_{\ell,n}$ converge to 0 uniformly in probability.

6 Universality of the CONCH algorithm

In earlier sections, we have established CONCH as a flexible framework for constructing distribution-free confidence sets for the changepoint. One may naturally ask: is CONCH one of many such distribution-free approaches, or does it truly capture the full class of distribution-free changepoint localization methods? In this section, we give a conclusive answer to this question. In fact, we establish a universality property of CONCH, which states that any procedure satisfying the coverage guarantee in (2.1) can be realized as an instance of our CONCH framework.

Theorem 6.1. *Fix $\alpha \in (0,1)$. Let C be any procedure that maps a dataset $\mathbf{X}$ to a confidence set $C(\mathbf{X})$ such that*

$$\mathbb{P}_\xi\Big(\xi \in C(\mathbf{X})\Big) \geq 1-\alpha.$$

Then, there exists a CPP score function $S : \mathcal{X}^n \to \mathbb{R}^{n-1}$ such that the distribution-free confidence set C coincides exactly the set $\mathcal{C}_{1-\alpha}^{CONCH}$ constructed with the score S.

The proof of this theorem is provided in Appendix A.4. This result establishes CONCH as a canonical framework for distribution-free changepoint inference: a particular choice of CPP score leads to a specific instance within the universal class of valid procedures for changepoint localization.

Moreover, the universality of the CONCH framework gives rise to two natural extensions. First, CONCH can wrap around any heuristic or model-based confidence set, calibrating it to achieve exact, distribution-free coverage guarantees while preserving the original procedure's structural or modeling advantages. Second, in the presence of multiple changepoints, it can be combined with any consistent segmentation algorithm to construct distribution-free confidence sets. We formalize these extensions below.

6.1 Calibration of heuristic confidence sets

Suppose we are given a confidence set $C : \mathcal{X}^n \to 2^{[n-1]}$ that may or may not be valid, even asymptotically. For instance, it could be one obtained from a Bayesian or bootstrap-based method. Guided by the general CONCH framework, we can construct a CPP score function from such a set and thereby obtain a distribution-free, finite-sample valid confidence set. Two natural constructions of CPP score are as follows:

- **Set membership score.** Define $S_t(x_1, \ldots, x_n) = \mathbb{1}\{t \in C(x_1, \ldots, x_n)\}$, which records only whether t is included in the given confidence set.
- **Set distance score.** Define $S_t(x_1, \ldots, x_n) = \min_{\ell \in C(x_1,\ldots,x_n)} |t - \ell|$, which refines the membership score by measuring the distance of t to the nearest index in the set.

Running the CONCH algorithm with either score yields a valid confidence set by Theo-

Algorithm 3: CONCH-CAL: CONCH calibration algorithm

Input: $(X_t)_{t=1}^n$ (dataset), $t_0 \in [n-1]$ (point estimate), and $\mathrm{pval} : \mathcal{X}^n \to [0,1]^{n-1}$ (p-value function)

Output: $\mathcal{C}_{1-\alpha}^{\text{CONCH-CAL}}$ (CONCH-CAL confidence set at level $1-\alpha$)

Define $\hat{S} : \mathcal{X}^n \to \mathbb{R}^{n-1}$ as in (6.1) ;
for $t \in [n-1]$ **do**
    Compute CONCH p-value p_t as in (3.2) with S_t replaced by $\hat{S}_t$;
end
$\mathcal{C}_{1-\alpha}^{\text{CONCH-CAL}} \leftarrow \{\, t \in [n-1] : p_t > \alpha \,\}$;
return $\mathcal{C}_{1-\alpha}^{\text{CONCH-CAL}}$

rem 3.1. Moreover, by Proposition 4.1 (*ii*), the set distance score always produces a narrower confidence set than the set membership score.

However, both score functions are relatively coarse and often lead to wide confidence sets. In particular, for indices close to n or 0, these scores frequently induce t-wise symmetry (cf. (4.1)), resulting in artificially inflated p-values in that region (by Proposition 4.1 (*i*)). Since this behavior is undesirable in practice, we next introduce a more informative CPP score that yields sharper confidence sets.

Most existing model-based or resampling-based approaches produce a point estimate t_0. Moreover, in many cases, they first construct a p-value function $\mathrm{pval} : \mathcal{X}^n \to [0,1]^{n-1}$, which is then thresholded to form the confidence set C. Both components (t_0, pval) can be combined to define a more informative CPP score,

$$\hat{S}_t(x_1, \ldots, x_n) = \frac{\mathrm{pval}(x_1, \ldots, x_n; t)}{\mathrm{pval}(x_1, \ldots, x_n; t_0)}. \tag{6.1}$$

Applying CONCH with this score yields what we refer to as the CONCH-CAL algorithm, formally presented in Algorithm 3. By construction, this produces a valid distribution-free confidence set while retaining the original method's assessment of the changepoint. In practice, this allows analysts to exploit the strengths of bootstrap or Bayesian methods, such as their interpretability, while simultaneously ensuring exact finite-sample coverage.

We note that the point estimate t_0 depends on the ordered sequence $(x_1, \ldots, x_n)$, and thus the denominator $\mathrm{pval}(\cdot, \ldots, \cdot; t_0)$ is not invariant under permutations. Although one could in principle use $\mathrm{pval}(\cdot, \ldots, \cdot; t)$ directly as the CPP score in CONCH, this approach typically inherits the same conservativeness observed with set-membership and set-distance scores.

Algorithm 4: CONCH-SEG: Segmentwise CONCH for Multiple Changepoints

Input: $(X_t)_{t=1}^n$ (data); $\hat{K}$ and $0 = \hat{\xi}_0 < \hat{\xi}_1 < \cdots < \hat{\xi}_{\hat{K}} < n = \hat{\xi}_{\hat{K}+1}$ (estimated changepoints); $S : \cup_{m \in \mathbb{N}} \mathcal{X}^m \to \mathbb{R}^m$ (CPP score)

Output: $\mathcal{C}_{1-\alpha}^{\text{CONCH-SEG}}$ (overall confidence set at level $1 - \alpha$)

Compute $(\tilde{X}_0, \dots, \tilde{X}_{\hat{K}})$ as in (6.3);

Initialize $\mathcal{C} \leftarrow \varnothing$;

for $\ell \in [\hat{K}]$ **do**

  $(L_\ell, R_\ell) \leftarrow (\tilde{X}_{\ell-1},\, \tilde{X}_\ell)$;

  Let $X^{(\ell)} \leftarrow (X_{L_\ell}, \dots, X_{R_\ell})$;

  Define score $S^{(\ell)} : \mathcal{X}^{R_\ell - L_\ell + 1} \to \mathbb{R}^{R_\ell - L_\ell}$;

  Compute CONCH p-values $\{p_t : t \in [L_\ell,\, R_\ell - 1]\}$ as in (3.2), using $S^{(\ell)}$ on $X^{(\ell)}$;

  Set $\mathcal{C}_\ell \leftarrow \{\, t \in [L_\ell,\, R_\ell - 1] : p_t > \alpha \,\}$;

  Update $\mathcal{C} \leftarrow \mathcal{C} \cup \mathcal{C}_\ell$;

end

return $\mathcal{C}_{1-\alpha}^{\textit{CONCH-SEG}} \leftarrow \mathcal{C}$

6.2 Localization of Multiple Changepoints

We extend CONCH to settings with multiple changepoints. The key observation is that, given any sufficiently consistent segmentation algorithm, the sequence can be partitioned into disjoint segments so that each segment contains, with high probability, at most one changepoint. We can then run CONCH independently within each segment and aggregate the resulting sets to obtain an overall confidence set.

Formally, suppose there exist $K \in [n]$ and changepoints

$$0 = \xi_0 < \xi_1 < \cdots < \xi_K < n = \xi_{K+1},$$

such that for each $\ell \in \{0, 1, \dots, K\}$,

$$(X_{\xi_\ell + 1}, \dots, X_{\xi_{\ell+1}}) \sim \mathcal{P}^{(\ell)}, \qquad \mathcal{P}^{(\ell)} \in \mathcal{M}\Big(\mathcal{X}^{\,\xi_{\ell+1} - \xi_\ell}\Big). \tag{6.2}$$

Consistent with Assumption 1, we assume each $\mathcal{P}^{(\ell)}$ is exchangeable and that the collection $\{\mathcal{P}^{(0)}, \dots, \mathcal{P}^{(K)}\}$ is pairwise independent. Further, suppose a segmentation algorithm returns (a) an estimate $\hat{K}$ of the number of changepoints, and (b) an ordered sequence of estimated changepoints

$$0 = \hat{\xi}_0 < \hat{\xi}_1 < \cdots < \hat{\xi}_{\hat{K}} < n = \hat{\xi}_{\hat{K}+1},$$

such that $\hat{K} \approx K$ and $\hat{\xi}_\ell \approx \xi_\ell$ for all $\ell \in [K]$.

Based on these estimates, we discretize the timeline $\{1,\dots,n\}$ into $\hat{K}$ data-dependent segments centered at the $\hat{\xi}_\ell$'s. Specifically, for $\ell \in \{0,\dots,\hat{K}\}$, define

$$\tilde{X}_\ell := \begin{cases} 1, & \text{if } \ell = 0, \\ \left\lfloor \frac{1}{2}(\hat{\xi}_\ell + \hat{\xi}_{\ell+1}) \right\rfloor, & \text{if } \ell \in [\hat{K}-1], \\ n, & \text{if } \ell = \hat{K}. \end{cases} \tag{6.3}$$

We then take the ℓ-th segment around $\hat{\xi}_\ell$ to be $[\tilde{X}_{\ell-1}, \tilde{X}_\ell]$ for $\ell \in [\hat{K}]$. Running CONCH independently on each segment and aggregating the segmentwise sets yields our overall confidence set. The resulting procedure, denoted CONCH-SEG, is summarized in Algorithm 4.

Kernel-based changepoint detection (KCPD) methods [Harchaoui and Cappé, 2007, Arlot et al., 2019] provide consistent estimators of changepoints under mild conditions [e.g., Garreau and Arlot, 2018, Diaz-Rodriguez and Jia, 2025]. Consequently, the CONCH framework can be seamlessly wrapped around a KCPD routine to construct confidence sets in the multiple-changepoint setting.

In Appendix B.2, we consider a Gaussian mean-shift model with multiple changepoints and show that CONCH-SEG, when wrapped around a KCPD algorithm, sharply localizes the changepoints, demonstrating that this extension is both practical and statistically powerful.

Appendix A.5 establishes asymptotic validity of a cross-fitted variant of CONCH-SEG. A direct analysis is hard because segmentation and CONCH are applied to the same data, potentially violating Assumption 1. Cross-fitting uses disjoint folds to estimate changepoints and then, run CONCH within the estimated segments. This restores the required independence and exchangeability structure and yields asymptotically valid coverage guarantee.

7 Experiments

In this section, we evaluate the performance of CONCH through synthetic simulations and real-data applications to images (CIFAR-100, MNIST, DomainNet) and text (SST-2). Throughout, when we refer to CONCH confidence sets, we mean those produced by the practical CONCH-MC procedure (Algorithm 2). Although the CONCH-MC p-values are typically conservative under the null, the method nevertheless attains high power and yields informative, narrow confidence sets with sharp changepoint localization across all settings considered.

7.1 Numerical simulations

7.1.1 Detecting Gaussian mean-shift

We begin with the most well-studied setting for changepoint analysis, namely the Gaussian mean-shift model, to illustrate the behavior of our proposed CONCH framework. Specifically, we generate a sequence of $n = 1000$ i.i.d. observations with a changepoint at $\xi = 400$: the pre-change distribution is $\mathcal{P}_{0,\xi} = \bigotimes_{t=1}^{\xi} \mathcal{N}(-1, 1)$, while the post-change distribution is $\mathcal{P}_{1,\xi} = \bigotimes_{t=\xi+1}^{n} \mathcal{N}(1, 1)$. In this setup, changepoint localization reduces to detecting a mean shift in the Gaussian family with scale parameter 1.

We evaluate CONCH using four choices of CPP scores, introduced earlier in Section 4.2:

(i) weighted mean difference, with a specified weight function,

(ii) oracle log-likelihood ratio (LLR),

(iii) parametrically learned LLR, assuming knowledge of the Gaussian family,

(iv) nonparametrically learned LLR, using kernel density estimates.

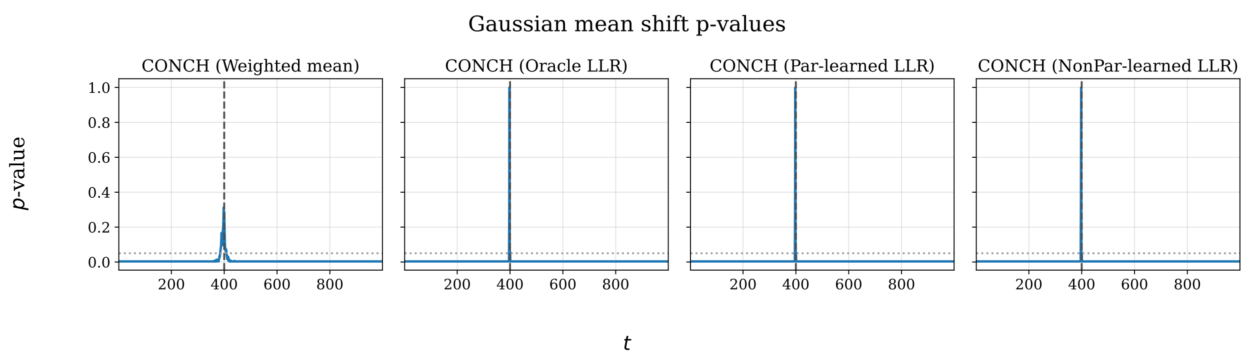

Figure 1: Distribution of conformal p-values under the Gaussian mean-shift model for different methods.

Figure 1 displays the distribution of the resulting p-values produced by each method. CONCH produces sharply localized confidence sets across all score choices. The weighted-mean score results in the widest interval, $[385, 408]$, whereas all three LLR-based scores (oracle, parametrically learned and non-parametrically learned) yield a much narrower set $\{397, 398, 400\}$.

Overall, these results highlight two key features: (i) the validity of CONCH is preserved regardless of the choice of score, and (ii) more informative scores lead to substantially sharper localization.

Appendix B.1 presents an additional comparison between the CONCH intervals and those from Dandapanthula and Ramdas [2025].

7.1.2 Refinement of resampling-based confidence sets using CONCH-CAL

We demonstrate that the CONCH-CAL procedure (Algorithm 3) can refine confidence sets that were not originally designed with distribution-free validity guarantees. The Gaussian mean-shift model has been extensively studied, and several bootstrap-based methods provide asymptotically valid intervals that perform well in practice. However, under mild model misspecification, these intervals can become overly wide or may miss the true changepoint ξ.

If a confidence set is valid, then the set-membership score from Section 6.1 should reproduce the same set. In contrast, CONCH-CAL leverages p-values, and a finer notion of CPP score building on them. This provides a principled mechanism to recalibrate and refine existing confidence sets, thereby yielding sharper, distribution-free intervals.

In both experiments, we use the same residual bootstrap scheme to construct the initial confidence sets: for each replicate, the changepoint is re-estimated on a resampled sequence formed from centered residuals, producing an empirical distribution of $\hat{\tau}$ from which percentile-based intervals and p-values are obtained.

We consider two settings: (i) the Gaussian mean-shift model with $n = 500$ and $\xi = 200$, where $\mathcal{P}_{0,\xi} = \bigotimes_{t=1}^{\xi} \mathcal{N}(-1, 3)$ and $\mathcal{P}_{1,\xi} = \bigotimes_{t=\xi+1}^{n} \mathcal{N}(1, 3)$, and (ii) a Laplace mean-shift model with $n = 500$ i.i.d. observations and the same changepoint, where $\mathcal{P}_{0,\xi} = \bigotimes_{t=1}^{\xi} \text{Laplace}(-1, 3)$ and $\mathcal{P}_{1,\xi} = \bigotimes_{t=\xi+1}^{n} \text{Laplace}(1, 3)$.

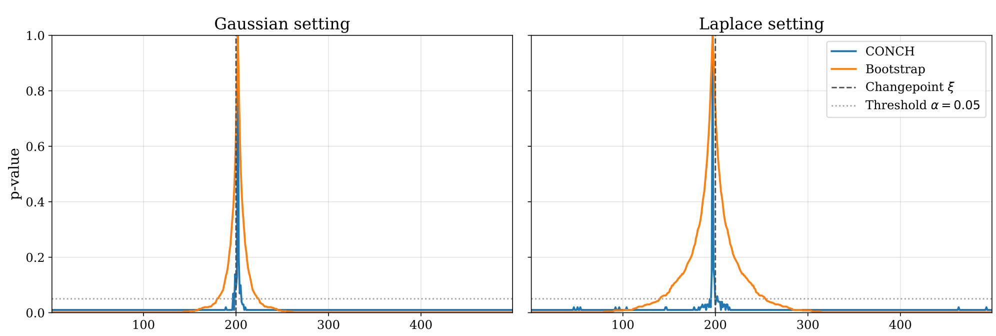

Figure 2: Refinement of bootstrap-based confidence sets using CONCH-CAL under Gaussian and Laplace mean-shift models.

In the Gaussian case, the bootstrap interval $[180, 224]$ is refined by CONCH-CAL to a tight and accurate interval $[197, 205]$. In the Laplace case, the bootstrap interval $[140, 258]$, inflated by heavy-tailed noise, is reduced to $[196, 202]$ after calibration. The bootstrap p-values in the Laplace setting are notably more spread out, while those from CONCH-CAL remain sharply concentrated near the true changepoint, highlighting its robustness and stability across distributional regimes.

7.2 Real data experiments

7.2.1 DomainNet: detecting domain shift

In this experiment, we tackle the problem of detecting a domain shift using the publicly available DomainNet dataset [Peng et al., 2019], which consists of six diverse domains (real, sketch, painting, clipart, infograph, and quickdraw). Among these, we use the *real* and *sketch* domains to construct a changepoint detection setting. Moreover, we convert all images to grayscale to remove color cues and further increase the similarity between classes. Specifically, before the changepoint ($\xi = 350$), we observe samples from the real domain, and after ξ, we observe samples from the sketch domain, totaling 800 samples (Figure 3).

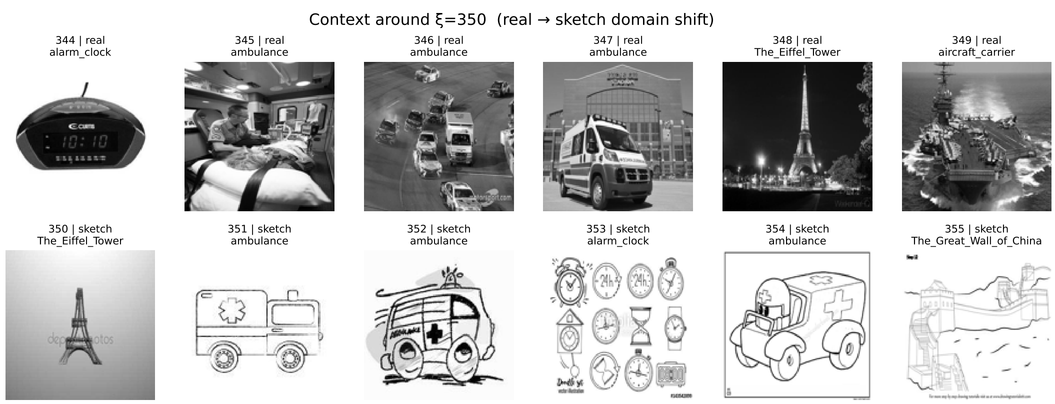

Figure 3: Illustration of the DomainNet changepoint setup: samples switch from the real to the sketch domain at $\xi = 350$ ($n = 800$). Images are drawn from the DomainNet dataset, which was collected via online search; class labels may not perfectly align with visual semantics, making the domain-shift detection problem more challenging.

We first train a CNN-based classifier to distinguish real images from hand-drawn sketches. Although the classifier provides substantial discriminative information, it does not directly translate into distribution-free guarantees for changepoint localization. The CONCH framework bridges this gap by converting classifier outputs into a principled, distribution-free procedure, yielding a narrow confidence set $[350, 351]$ that consistently contains the true changepoint (Figure 4).

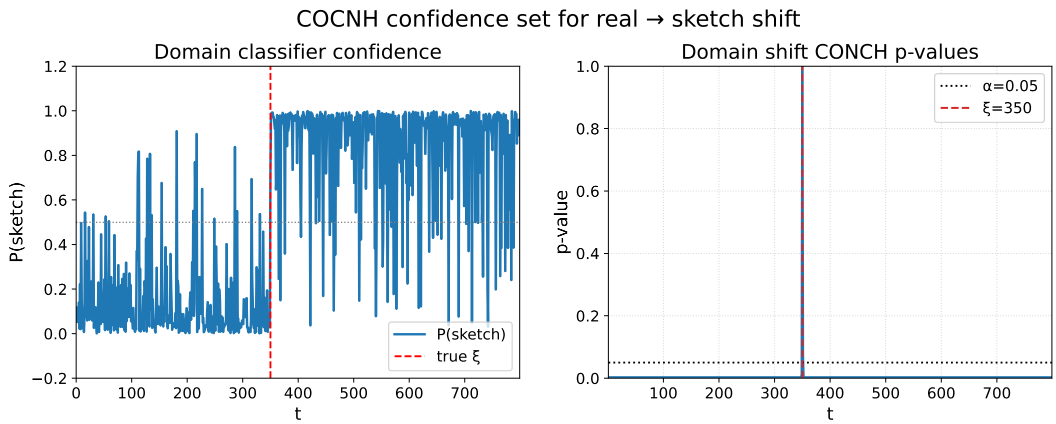

Figure 4: p-values for domain shift detection between real and sketch domains: classifier scores (left) and CONCH p-values (right)

7.2.2 SST-2: detecting sentiment change using large language models

We next demonstrate our method on text data, showing that it can localize changepoints in language settings. Using the Stanford Sentiment Treebank (SST-2) dataset of movie reviews labeled with binary sentiment [Socher et al., 2013], we simulate a shift from predominantly positive to negative sentiment. Such a setup mirrors real-world scenarios, e.g., detecting changes in customer feedback or public opinion.

We observe $n = 1000$ reviews with a changepoint at $\xi = 400$: before ξ, reviews are i.i.d. positive (P_0); after ξ, reviews are i.i.d. negative (P_1). For example:

- $t = 399$ (positive): "juicy writer"
- $t = 400$ (positive): "intricately structured and well-realized drama"
- $t = 401$ (negative): "painfully "
- $t = 402$ (negative): "than most of jaglom's self-conscious and gratingly irritating films"

First, we find a DistilBERT model fine-tuned for sentiment classification [Sanh et al., 2019], and then the corresponding model logits are used to build a CPP score for our CONCH method, which yields a 95% confidence set $[400, 401]$ (Figure 5, left panel), effectively pinpointing the changepoint. Even under a subtler scenario, where sentiment shifts only from 60% positive to 40% positive, we obtain a nontrivial 95% confidence set $[326, 463]$ (Figure 5, right panel), demonstrating sharp localization of the changepoint in complex settings.

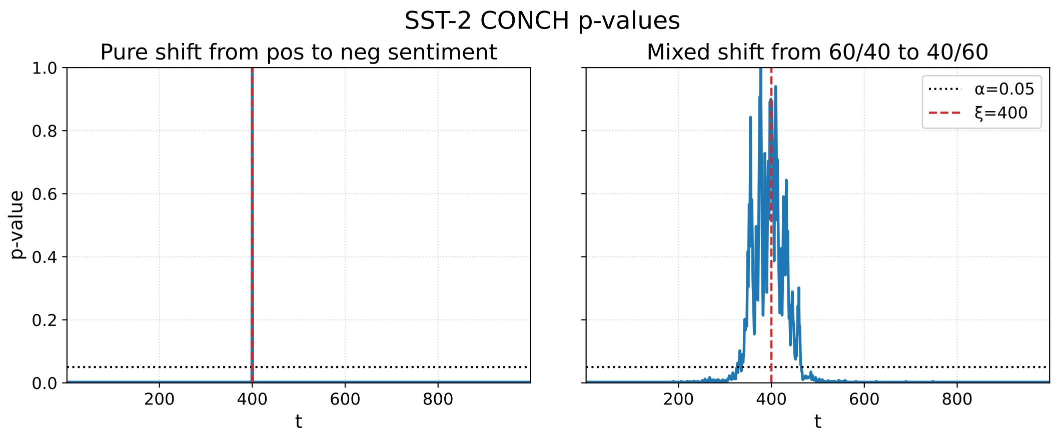

Figure 5: CONCH p-values for sentiment shift in SST-2: from positive to negative reviews at $\xi = 400$ (left), and from 60% positive to 40% positive (right).

Additional experiments Appendix B presents supplementary experiments spanning several changepoint detection settings. We first illustrate CONCH-SEG for multiple changepoint localization in Gaussian mean-shift models, then consider a two-urn example (urn-shift detection), a digit-class shift on MNIST, and a class-shift experiment on CIFAR-100, demonstrating the robustness and flexibility of our approach.

8 Conclusion

In this work, we introduced CONCH, a framework for distribution-free offline changepoint localization. Our approach leverages conformal p-values to construct confidence sets with finite-sample, distribution-free guarantees. We provided design guidelines, including principled choices of score functions and a Monte Carlo approximation to the full-permutation p-value, that enhance both the power and practicality of the framework. With an appropriate score function, the CONCH confidence sets shrink with sample size, as one would expect from any useful localization procedure.

We established a universality result positioning CONCH as a canonical method for distribution-free offline changepoint localization. This, in turn, paves the way for (i) a simple calibration procedure that can wrap around any localization algorithm to yield valid confidence sets, and (ii) an extension of the CONCH framework to multiple-changepoint settings.

While our extension of the conformal localization framework to the multiple-changepoint setting is natural, it relies crucially on first obtaining a consistent segmentation, which is a nontrivial task in its own right. A promising direction for future work is to adapt techniques such as wild binary segmentation [Fryzlewicz, 2014], and extend CONCH directly to the multiple-changepoint case, thereby broadening its scope and applicability.

References

Krizhevsky Alex. Learning multiple layers of features from tiny images. *https://www. cs. toronto. edu/kriz/learning-features-2009-TR. pdf*, 2009.

Andreas Anastasiou and Piotr Fryzlewicz. Detecting multiple generalized change-points by isolating single ones. *Metrika*, 85(2):141–174, 2022.

Anastasios N Angelopoulos, Rina Foygel Barber, and Stephen Bates. Theoretical foundations of conformal prediction. *arXiv preprint arXiv:2411.11824*, 2024.

Sylvain Arlot, Alain Celisse, and Zaid Harchaoui. A kernel multiple change-point algorithm via model selection. *Journal of machine learning research*, 20(162):1–56, 2019.

Stephen Bates, Emmanuel Candès, Lihua Lei, Yaniv Romano, and Matteo Sesia. Testing for outliers with conformal p-values. *The Annals of Statistics*, 51(1):149–178, 2023.

Gouri K Bhattacharyya and Richard A Johnson. Nonparametric tests for shift at an unknown time point. *The Annals of Mathematical Statistics*, pages 1731–1743, 1968.

A. E. Brockwell. Universal residuals: A multivariate transformation. *Statistics & Probability Letters*, 77(14):1473–1478, 2007.

Haeran Cho and Claudia Kirch. Bootstrap confidence intervals for multiple change points based on moving sum procedures. *Computational Statistics & Data Analysis*, 175:107552, 2022.

Sanjit Dandapanthula and Aaditya Ramdas. Offline changepoint localization using a matrix of conformal p-values. *arXiv preprint arXiv:2505.00292*, 2025.

Li Deng. The MNIST database of handwritten digit images for machine learning research. *Signal Processing Magazine, IEEE*, 29:141–142, 11 2012.

Jairo Diaz-Rodriguez and Mumin Jia. Consistent kernel change-point detection under m-dependence for text segmentation. *arXiv preprint arXiv:2510.03437*, 2025.

Jonathan W Duggins. *Parametric Resampling Methods for Retrospective Changepoint Analysis*. PhD thesis, Virginia Polytechnic Institute and State University, 2010.

Stergios B Fotopoulos, Venkata K Jandhyala, and Elena Khapalova. Exact asymptotic distribution of change-point mle for change in the mean of Gaussian sequences. *The Annals of Applied Statistics*, pages 1081–1104, 2010.

Klaus Frick, Axel Munk, and Hannes Sieling. Multiscale change point inference. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(3):495–580, 2014.

Piotr Fryzlewicz. Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42(6):2243–2281, 2014. ISSN 00905364.

Damien Garreau and Sylvain Arlot. Consistent change-point detection with kernels. *arXiv preprint arXiv:1612.04740*, 2018.

Gregory Gurevich and Albert Vexler. Guaranteed maximum likelihood splitting tests of a linear regression model. *Statistics*, 40(6):465–484, 2006.

Zaid Harchaoui and Olivier Cappé. Retrospective multiple change-point estimation with kernels. In *2007 IEEE/SP 14th Workshop on Statistical Signal Processing*, pages 768–772. IEEE, 2007.

Matthew T Harrison. Conservative hypothesis tests and confidence intervals using importance sampling. *Biometrika*, 99(1):57–69, 2012.

Ying Jin and Emmanuel J Candès. Selection by prediction with conformal p-values. *Journal of Machine Learning Research*, 24(244):1–41, 2023.

Hyune-Ju Kim and David Siegmund. The likelihood ratio test for a change-point in simple linear regression. *Biometrika*, 76(3):409–423, 1989.

Erich Leo Lehmann and Joseph P Romano. *Testing statistical hypotheses*. Springer, 2005.

Ilia Nouretdinov, Vladimir Vovk, and Alex Gammerman. Conformal changepoint detection in continuous model situations. In *Conformal and Probabilistic Prediction and Applications*, pages 300–302. Proceedings of Machine Learning Research, 2021.

Ewan Stafford Page. A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4):523–527, 1955.

Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1406–1415, 2019.

Anthony N Pettitt. A non-parametric approach to the change-point problem. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 28(2):126–135, 1979.

Richard E Quandt. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 53(284):873–880, 1958.

Gordon J Ross and Niall M Adams. Two nonparametric control charts for detecting arbitrary distribution changes. *Journal of Quality Technology*, 44(2):102–116, 2012.

Aytijhya Saha and Aaditya Ramdas. Post-detection inference for sequential changepoint localization. *arXiv preprint arXiv:2502.06096*, 2025.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, 2019.

Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008.

Jaehyeok Shin, Aaditya Ramdas, and Alessandro Rinaldo. E-detectors: A nonparametric framework for sequential change detection. *The New England J of Stat. in Data Sci.*, 2 (2):229–260, 2023. ISSN 2693-7166.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642, 2013.

Charles Truong, Laurent Oudre, and Nicolas Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020.

Nicolas Verzelen, Magalie Fromont, Matthieu Lerasle, and Patricia Reynaud-Bouret. Optimal change-point detection and localization. *The Annals of Statistics*, 51(4):1586–1610, 2023.

Denis Volkhonskiy, Evgeny Burnaev, Ilia Nouretdinov, Alexander Gammerman, and Vladimir Vovk. Inductive conformal martingales for change-point detection. In *Conformal and Probabilistic Prediction and Applications*, pages 132–153. PMLR, 2017.

Vladimir Vovk. Testing randomness online. *Statistical Science*, 36(4):595–611, 2021.

Vladimir Vovk, Ilia Nouretdinov, and Alexander Gammerman. Testing exchangeability on-line. In *Proceedings of the 20th International Conference on Machine Learning (ICML)*, pages 768–775, 2003.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*. Springer, 2005.

Vladimir Vovk, Ivan Petej, Ilia Nouretdinov, Ernst Ahlberg, Lars Carlsson, and Alex Gammerman. Retrain or not retrain: Conformal test martingales for change-point detection. In *Conformal and Probabilistic Prediction and Applications*, pages 191–210. PMLR, 2021.

Volodya Vovk, Alexander Gammerman, and Craig Saunders. Machine-learning applications of algorithmic randomness. In *Proceedings of the Sixteenth International Conference on Machine Learning*, pages 444–453, 1999.

Xiaoyang Wu, Lin Lu, Zhaojun Wang, and Changliang Zou. Conditional testing based on localized conformal p-values. *arXiv preprint arXiv:2409.16829*, 2024.

Haotian Xu, Daren Wang, Zifeng Zhao, and Yi Yu. Change-point inference in high-dimensional regression models under temporal dependence. *The Annals of Statistics*, 52 (3):999–1026, 2024.

Changliang Zou, Yukun Liu, Peng Qin, and Zhaojun Wang. Empirical likelihood ratio test for the change-point problem. *Statistics & Probability Letters*, 77(4):374–382, 2007.

A Proofs

In this section, we present proofs of the main results and a few additional results. For notational convenience, throughout we write $\mathbf{x} \in \mathcal{X}^n$ to denote the tuple $(x_1, \ldots, x_n)$.

A.1 Proving coverage guarantees of CONCH confidence sets

A.1.1 Proof of Theorem 3.1

First, observe that under the null $\mathcal{H}_{0t}$, $\pi(\mathbf{X}) \overset{d}{=} \mathbf{X}$ for any $\pi \in \Pi_t$. We define a function $p_t : \mathcal{X}^n \to [0, 1]$ by

$$p_t(\mathbf{x}) := \frac{1}{|\Pi_t|} \sum_{\pi \in \Pi_t} \mathbb{1} \left\{ S_t(\pi(\mathbf{x})) \leq S_t(\mathbf{x}) \right\}.$$

Further, note that $p_t \equiv p_t(\mathbf{X})$. Therefore,

$$\begin{aligned}
\mathbb{P}_t\left(p_t(\mathbf{X}) \leq \alpha\right) &= \frac{1}{|\Pi_t|} \sum_{\pi \in \Pi_t} \mathbb{P}_t\left(p_t(\pi(\mathbf{X})) \leq \alpha\right) \\
&= \mathbb{E}_t\left[\frac{1}{|\Pi_t|} \sum_{\pi \in \Pi_t} \mathbb{1}\left\{p_t(\pi(\mathbf{X})) \leq \alpha\right\}\right] \\
&= \mathbb{E}_t\left[\frac{1}{|\Pi_t|} \sum_{\pi \in \Pi_t} \mathbb{1}\left\{\frac{1}{|\Pi_t|} \sum_{\pi' \in \Pi_t} \mathbb{1}\left\{S_t(\pi'(\mathbf{x})) \leq S_t(\pi(\mathbf{x}))\right\} \leq \alpha\right\}\right] \leq \alpha,
\end{aligned}$$

where the penultimate step follows by noting that $\pi \circ \Pi_t = \Pi_t$, and the last inequality follows by Harrison [2012, Lemma 3]. This completes the proof. □

A.1.2 Proof of Theorem 3.2

Given permutations $\pi_{1,t}, \ldots, \pi_{M,t} \in \Pi_t$, we define the fucntion

$$\tilde{p}_t(\mathbf{x}; \pi_{1,t}, \ldots, \pi_{M,t}) := \frac{1 + \sum_{k=1}^{M} \mathbb{1}\left\{s_t(\pi_{k,t}(\mathbf{x})) \leq s_t(\mathbf{x})\right\}}{1 + M},$$

Consider an additional uniform draw $\pi_{0,t}$ from Π_t.

Hence, note that with $\pi_{1,t}, \ldots, \pi_{M,t} \overset{iid}{\sim} \mathrm{Unif}(\Pi_t)$, we have that

$$(\pi_{1,t}, \ldots, \pi_{M,t}) \overset{d}{=} (\pi_{0,t} \circ \pi_{1,t}, \ldots, \pi_{0,t} \circ \pi_{M,t}).$$

Moreover, conditional on $\pi_{0,t}, \pi_{1,t}, \ldots, \pi_{M,t}$, $\mathbf{X} \overset{d}{=} \pi_{0,t}(\mathbf{X})$ under the null $\mathcal{H}_{0,t}$. Consequently,

$$\tilde{p}_t(\mathbf{X}; \pi_{1,t}, \ldots, \pi_{M,t}) \overset{d}{=} \tilde{p}_t(\mathbf{X}; \pi_{0,t} \circ \pi_{1,t}, \ldots, \pi_{0,t} \circ \pi_{M,t}) \overset{d}{=} \tilde{p}_t(\pi_{0,t}(\mathbf{X}); \pi_{0,t} \circ \pi_{1,t}, \ldots, \pi_{0,t} \circ \pi_{M,t}).$$

Finally, note that for $\tilde{p}_t$, defined in (3.3), $\tilde{p}_t \equiv \tilde{p}_t(\mathbf{X}; \pi_{1,t}, \ldots, \pi_{M,t})$, and therefore,

$$\begin{aligned}
\tilde{p}_t(\mathbf{X}; \pi_{1,t}, \ldots, \pi_{M,t}) &\overset{d}{=} \tilde{p}_t(\pi_{0,t}(\mathbf{X}); \pi_{0,t} \circ \pi_{1,t}, \ldots, \pi_{0,t} \circ \pi_{M,t}) \\
&= \frac{1 + \sum_{k=1}^{M} \mathbb{1}\left\{s_t(\pi_{k,t}(\mathbf{X})) \leq s_t(\pi_{0,t}(\mathbf{X}))\right\}}{M+1} \\
&= \frac{\sum_{k=0}^{M} \mathbb{1}\left\{s_t(\pi_{k,t}(\mathbf{X})) \leq s_t(\pi_{0,t}(\mathbf{X}))\right\}}{M+1},
\end{aligned}$$

i.e., the rank of $s_t(\pi_{0,t}(\mathbf{X}))$ in the exchangeable collection $\{s_t(\pi_{0,t}(\mathbf{X})), s_t(\pi_{1,t}(\mathbf{X})), \ldots, s_t(\pi_{M,t}(\mathbf{X}))\}$. Consequently,

$$\mathbb{P}_t\left(\tilde{p}_t = \tilde{p}_t(\mathbf{X}; \pi_{1,t}, \ldots, \pi_{M,t}) \leq \alpha\right) \leq \alpha.$$

This completes the proof. $\square$

A.1.3 Proof of Theorem 3.3

We begin by letting F denote the distribution of $S_t(\pi(\mathbf{X}))$ conditional on the multisets $M_{\text{left}} := \{X_1, \ldots, X_t\}$ and $M_{\text{right}} := \{X_{t+1}, \ldots, X_n\}$, where $\pi \sim \text{Unif}(\Pi_t)$. Then

$$\bar{p}_t = \lim_{y \uparrow S_t(\mathbf{X})} F(y) + U\Big(F(S_t(\mathbf{X})) - \lim_{y \uparrow S_t(\mathbf{X})} F(y)\Big).$$

Under $\mathcal{H}_{0,t}$, we have $S_t(\mathbf{X}) \overset{d}{=} S_t(\pi(\mathbf{X}))$ conditional on M_{left} and M_{right}. Hence, by Dandapanthula and Ramdas [2025, Lemma E.1], the p-value $\bar{p}_t$, conditional on M_{left} and M_{right}, follows Unif$[0,1]$ (see also Brockwell, 2007). Therefore,

$$\mathbb{P}_t(\bar{p}_t \leq \alpha) = \mathbb{E}_t[\mathbb{P}_t(\bar{p}_t \leq \alpha \mid \{X_1, \ldots, X_n\})] = \mathbb{E}_t[\alpha] = \alpha.$$

This completes the proof.

A.2 Proving properties of the CPP score and optimality results

A.2.1 Proof of Proposition 4.1

The first part of the result follows immediately by noting that when S_t satisfies the t-wise symmetry assumption in (4.1), then by the definitions of conformal p-values in (3.2) and (3.3), p_t and $\tilde{p}_t$ are identically equal to 1, as required.

For the second part, fix $t \in [n-1]$. We prove the result for CONCH, noting that it holds identically for CONCH-MC. By definition (see (3.2)),

$$p_{t,1} = \frac{1}{|\Pi_t|} \sum_{\pi \in \Pi_t} \mathbb{1}\left\{S_t(\pi(\mathbf{X})) \leq S_t(\mathbf{X})\right\}, \quad p_{t,2} = \frac{1}{|\Pi_t|} \sum_{\pi \in \Pi_t} \mathbb{1}\left\{f(S_t(\pi(\mathbf{X}))) \leq f(S_t(\mathbf{X}))\right\}.$$

Since f is non-decreasing,

$$S_t(\pi(\mathbf{X})) \leq S_t(\mathbf{X}) \implies f(S_t(\pi(\mathbf{X}))) \leq f(S_t(\mathbf{X})),$$

and therefore $p_{t,1} \leq p_{t,2}$. As this holds for all $t \in [n-1]$, it further follows that $C_1 \subseteq C_2$.

A.2.2 Proof of Lemma 4.2 (conformal NP lemma)

In the setup of Section 4.1, we consider the following hypothesis testing problem:

$$\mathcal{H}_0' : \mathbf{X} \mid \mathcal{X}_{L,t}, \mathcal{X}_{R,t} \sim \mathcal{P}^{[t]}_{\mathbf{X}|\mathcal{X}_{L,t},\mathcal{X}_{R,t}} \quad \text{v.s.} \quad \mathcal{H}_1' : \mathbf{X} \mid \mathcal{X}_{L,t}, \mathcal{X}_{R,t} \sim \mathcal{P}^{[r]}_{\mathbf{X}|\mathcal{X}_{L,t},\mathcal{X}_{R,t}}.$$

Given samples $\mathbf{X} \in \mathbb{R}^n$, observe that

$$\frac{\mathsf{d}\Big(\mathcal{P}^{[r]}_{\mathbf{X}|\mathcal{X}_{L,t},\mathcal{X}_{R,t}}\Big)}{\mathsf{d}\Big(\mathcal{P}^{[t]}_{\mathbf{X}|\mathcal{X}_{L,t},\mathcal{X}_{R,t}}\Big)}(\mathbf{X}) \propto \frac{\mathsf{d}(\mathcal{P}^{[r]}_{\mathbf{X}})}{\mathsf{d}(\mathcal{P}^{[r]}_{\mathbf{X}})} = \frac{\prod_{i\le r} f_0(X_i)\prod_{i>r} f_1(X_i)}{\prod_{i\le t} f_0(X_i)\prod_{i>t} f_1(X_i)} = s^\star(\mathbf{X})^{-1}.$$

By the Neyman–Pearson lemma [Lehmann and Romano, 2005, Theorem 3.2.1 (ii)], any test $\phi(\mathbf{X})$ that attains exact validity at level α under $\mathcal{H}_0'$ and satisfies

$$\phi(\mathbf{X}) = \begin{cases} 1 & \text{if } s^\star(\mathbf{X})^{-1} > \tau_\alpha, \\ 0 & \text{if } s^\star(\mathbf{X})^{-1} < \tau_\alpha, \end{cases} \tag{A.1}$$

for an appropriate threshold $\tau_\alpha \in \mathbb{R}$, is most powerful for testing $\mathcal{H}_0'$ against $\mathcal{H}_1'$.

As discussed in Section 4.1, the test $\phi_t(\cdot; s) = \mathbb{1}\{p_t(s) \le \alpha\}$ controls the Type I error exactly at level α under $\mathcal{H}_0'$ for any score function s. Therefore, to establish the optimality of $s^\star$, it suffices to show that $\phi_t(\cdot; s^\star)$ admits the form given in (A.1).

Define $\mathbf{X}_\pi = \pi(\mathbf{X})$ for $\pi \sim \mathrm{Unif}(\Pi_t)$, and let $F_{s^\star(\mathbf{X}_\pi)}$ denote the conditional cumulative distribution function of $s^\star(\mathbf{X}_\pi)$ given $\mathbf{X}$. Set

$$\tau_\alpha := \inf\{y \in \mathbb{R} : F_{s^\star(\mathbf{X}_\pi)}(y) \ge \alpha\}.$$

By the definition of p_t in (4.2), we have

$$\begin{aligned} s^\star(\mathbf{X})^{-1} > \tau_\alpha &\implies p_t(s^\star) \le \alpha, \\ s^\star(\mathbf{X})^{-1} < \tau_\alpha &\implies p_t(s^\star) > \alpha, \end{aligned}$$

as required. This completes the proof. □

A.2.3 Proof of Theorem 4.3

By Proposition (*ii*), the CONCH confidence sets are invariant under any strictly increasing transformation of the score function. Therefore, it suffices to prove this result for S^{OPT}.

We observe that only the t-th coordinate of CPP score S_t determines the CONCH p-value $\bar{p}_t$ defined in (3.4). Therefore, with the notation laid out in Section 4.1, we can write

$$n - \mathbb{E}_{\mathcal{H}_{0,\xi} \cap \mathcal{P}_{\mathrm{IID}}}[\bar{C}^{\mathrm{CONCH}}_{1-\alpha}(S)] = \sum_{t=1}^{n} \mathbb{E}_{\mathcal{H}_{0,\xi} \cap \mathcal{P}_{\mathrm{IID}}}[\mathbb{1}\{p_t(S_t) \leq \alpha\}].$$

Finally, note that for any $j \in [n-1]$, $\mathcal{H}_{0,j} \cap \mathcal{P}_{\mathrm{IID}} = \mathcal{P}^{[\xi]}$. Hence, by tower law, we have that

$$\mathbb{E}_{\mathcal{P}^{[\xi]}}[\mathbb{1}\{p_t(S_t) \leq \alpha\}] = \mathbb{E}\Big[\mathbb{E}\Big[\mathbb{1}\{p_t(S_t) \leq \alpha\} \;\Big|\; \{X_1, \ldots, X_\xi\}, \{X_{\xi+1}, \ldots, X_n\}\Big]\Big]$$

Moreover, the p-value $\bar{p}_t = p_t(S_t)$ must be valid under $\mathcal{H}'_t$ applying Lemma 4.2, the optimal form of S_t^{OPT} follows readily.

A.3 Asymptotic sharpness of CONCH confidence sets

In this section, we establish the consistency of CONCH confidence sets. In the setting of Section 5 and under assumptions (A1)–(A3), we first show that the average of the CONCH p-values converges to 0 in L^1. Under the additional assumption that the log-likelihood ratio is sub-Gaussian under both pre-and post-change distributions, we further prove a uniform decay: p-values for indices at least $\log n$ away from the true changepoint ξ_n converge to 0 uniformly. Before going to the details, we define a few notation.

Define the log-likelihood ratio

$$\delta(x) \;=\; \log\!\left(\frac{f_0(x)}{f_1(x)}\right),$$

so that

$$\mathrm{KL}(P_0 \| P_1) = \mathbb{E}_{X \sim P_0}\Big[\delta(X)\Big], \qquad \mathrm{KL}(P_1 \| P_0) = \mathbb{E}_{X \sim P_1}\Big[-\delta(X)\Big],$$

and the Var-Entropy measures are given by

$$\sigma_0^2 := \mathrm{V}_{\mathrm{KL}}(P_0 \| P_1) = \mathrm{Var}_{X \sim P_0}(\delta(X)), \qquad \sigma_1^2 := \mathrm{V}_{\mathrm{KL}}(P_1 \| P_0) = \mathrm{Var}_{X \sim P_1}(\delta(X)).$$

We also let $\sigma_\star = \max\{\sigma_0, \sigma_1\}$, and define the Jeffreys divergence

$$\mathrm{J}(P_0, P_1) = \mathrm{KL}(P_0 \| P_1) + \mathrm{KL}(P_1 \| P_0).$$

Let $\hat{\xi}_n : \mathcal{X}^n \to [n-1]$ be a given changepoint estimator and write $\hat{\xi}_n(\mathbf{X}) \equiv \hat{\xi}_n$. Given $\hat{\xi}_n$, we define the score

$$S_t^{(n)}(x_1, \ldots, x_n) = \log\left(\frac{\prod_{i\le t} f_0(x_i)\prod_{i>t} f_1(x_i)}{\prod_{i\le \hat{\xi}_n} f_0(x_i)\prod_{i>\hat{\xi}_n} f_1(x_i)}\right). \tag{A.2}$$

Based on this score, define permutation p-values $(\tilde{p}_{1,n}, \ldots, \tilde{p}_{n-1,n})$ by

$$\tilde{p}_{t,n} := \frac{1}{|\Pi_t|}\sum_{\pi\in\Pi_t} \mathbb{1}_{S_t^{(n)}(\pi(\mathbf{X}))\le S_t^{(n)}(\mathbf{X})}. \tag{A.3}$$

The only distinction from the p-values $\{p_{t,n}\}$ used in Section 4.2 is that here, when evaluating $S_t^{(n)}(\pi(\mathbf{x}))$, we *do not* recompute $\hat{\xi}_n$ under the permutation. This 'frozen-estimator' version will serve as an intermediate p-value in our proofs, making the study of consistency of CONCH confidence sets much simpler.

A.3.1 Auxiliary lemmas

We begin with a few lemmas for the intermediate p-values $\{\tilde{p}_{t,n}\}$, establish consistency of the changepoint estimators from Section 4.2, and also relate the p-values $\{\tilde{p}_{t,n}\}$ with the CONCH p-values $\{p_{t,n}\}$. To state the lemmas, we first define the empirical quantities

$$\hat{\mu}_{k,n,L} := \frac{1}{k}\sum_{i=1}^{k}\delta(X_i), \qquad \hat{\mu}_{k,n,R} := \frac{1}{n-k}\sum_{i=k+1}^{n} -\delta(X_i), \qquad \hat{v}_n := \frac{1}{n}\sum_{i=1}^{n}\delta^2(X_i).$$

Lemma A.1. *For $k\in[n-1]$, we have deterministically*

$$\tilde{p}_{k,n} \le \left(\frac{n}{\max\{k, n-k\}}\cdot\frac{\hat{v}_n}{|k-\hat{\xi}_n|\,\Delta^2_{k,\hat{\xi}_n}}\cdot\mathbb{1}_{\Delta_{k,\hat{\xi}_n}<0}\right) + \mathbb{1}_{\Delta_{k,\hat{\xi}_n}\ge 0}, \tag{A.4}$$

where

$$\Delta_{k,\hat{\xi}_n} := \frac{S_k^{(n)}(\mathbf{X})}{|k-\hat{\xi}_n|} - \hat{\mu}_{k,n,L}\mathbb{1}_{k\ge\hat{\xi}_n} - \hat{\mu}_{k,n,R}\mathbb{1}_{k<\hat{\xi}_n}.$$

Proof. Fix $t_n\in[n-1]$ and set $\hat{m}_n = |t_n-\hat{\xi}_n|$. Without loss of generality assume $t_n > \hat{\xi}_n$. Then

$$S_{t_n}^{(n)}(\mathbf{x}) = \sum_{i=\hat{\xi}_n(\mathbf{x})+1}^{t_n}\delta(x_i),$$

a sum of δ over a block of $\hat{m}_n$ observations. Let $\mathcal{X}_L := \{X_1, \ldots, X_{t_n}\}$ be the multiset of the first t_n observations. The permutation p-value (A.3) is

$$\tilde{p}_{t_n,n} := \mathbb{P}_{\pi\sim\mathrm{Unif}(\Pi_{t_n})}\left(S_{t_n}^{(n)}(\pi(\mathbf{X}))\le S_{t_n}^{(n)}(\mathbf{X}) \;\middle|\; X_1, \ldots, X_n\right).$$

Since $t_n \geq \hat{\xi}_n$, sampling π uniformly from Π_{t_n} is equivalent to sampling without replacement (WOR) from $\mathcal{X}_L$. Thus

$$\tilde{p}_{t_n,n} = \mathbb{P}\left(\frac{1}{\hat{m}_n} \sum_{j=1}^{\hat{m}_n} \delta(\tilde{X}_j) \leq \frac{S_{t_n}^{(n)}(\mathbf{X})}{\hat{m}_n} \;\middle|\; X_1, \ldots, X_n \right),$$

where $\tilde{X}_1, \ldots, \tilde{X}_{\hat{m}_n}$ are drawn WOR from $\mathcal{X}_L$. Then

$$\mathbb{E}\left[\frac{1}{\hat{m}_n} \sum_{j=1}^{\hat{m}_n} \delta(\tilde{X}_j) \;\middle|\; \mathbf{X} \right] = \hat{\mu}_{t_n,n,L}, \quad \operatorname{Var}\left(\frac{1}{\hat{m}_n} \sum_{j=1}^{\hat{m}_n} \delta(\tilde{X}_j) \;\middle|\; \mathbf{X} \right) = \frac{v_{t_n,n}}{\hat{m}_n} \cdot \frac{n - \hat{m}_n}{n-1},$$

where $v_{t_n,n} = \frac{1}{t_n} \sum_{i=1}^{t_n} \delta^2(X_i) - (\hat{\mu}_{t_n,n,L})^2$. Note that

$$v_{t_n,n} \leq \frac{1}{t_n} \sum_{i=1}^{t_n} \delta^2(X_i) \leq \frac{n}{t_n} \hat{v}_n.$$

Now

$$\begin{aligned} &\mathbb{P}\left(\frac{1}{\hat{m}_n} \sum_{j=1}^{\hat{m}_n} \delta(\tilde{X}_j) \leq \frac{S_{t_n}^{(n)}(\mathbf{X})}{\hat{m}_n} \;\middle|\; X_1, \ldots, X_n \right) \\ &\qquad \leq \mathbb{P}\left(\frac{1}{\hat{m}_n} \sum_{j=1}^{\hat{m}_n} \delta(\tilde{X}_j) - \hat{\mu}_{t_n,n,L} \leq \frac{S_{t_n}^{(n)}(\mathbf{X})}{\hat{m}_n} - \hat{\mu}_{t_n,n,L} \;\middle|\; X_1, \ldots, X_n \right). \end{aligned}$$

If $\Delta_{t_n,\hat{\xi}_n} = \frac{S_{t_n}^{(n)}(\mathbf{X})}{\hat{m}_n} - \hat{\mu}_{t_n,n,L} < 0$, then by Chebyshev's inequality,

$$\tilde{p}_{t_n,n} \leq \frac{v_{t_n,n}}{\hat{m}_n \Delta^2_{t_n,\hat{\xi}_n}} \cdot \frac{n - \hat{m}_n}{n-1} \;\leq\; \frac{n}{t_n} \frac{\hat{v}_n}{\hat{m}_n \Delta^2_{t_n,\hat{\xi}_n}}.$$

If $\Delta_{t_n,\hat{\xi}_n} \geq 0$, then $\tilde{p}_{t_n,n} \leq 1$. This proves the claim. The case $t_n < \hat{\xi}_n$ is analogous after replacing $\hat{\mu}_{t_n,n,L}$ by $\hat{\mu}_{t_n,n,R}$ in the definition of $\Delta_{t_n,\hat{\xi}_n}$. □

Lemma A.2. *In the setting of Section 5, suppose* $0 < \sigma_\star < \infty$. *Then, for any* $\eta \in (0, 1)$,

$$\mathbb{P}\left(\hat{v}_n \leq \frac{2\left(\sigma_\star^2 + \mathrm{J}(P_0, P_1)^2\right)}{\eta} \right) \geq 1 - \eta.$$

Proof. Fix $\eta \in (0,1)$. We can write

$$\hat{v}_n = \frac{\xi_n}{n} \cdot \frac{1}{\xi_n} \sum_{i=1}^{\xi_n} \delta^2(X_i) \;+\; \frac{n-\xi_n}{n} \cdot \frac{1}{n-\xi_n} \sum_{i=\xi_n+1}^{n} \delta^2(X_i).$$

Recall $\delta^2(X_1), \ldots, \delta^2(X_{\xi_n})$ are i.i.d. with

$$\mathbb{E}_{X\sim P_0}\Big[\delta^2(X)\Big] = \sigma_0^2 + \mathrm{KL}(P_0\|P_1)^2.$$

By Markov's inequality,

$$\mathbb{P}\left(\frac{1}{\xi_n}\sum_{i=1}^{\xi_n} \delta^2(X_i) \geq \frac{2\Big(\sigma_0^2 + \mathrm{KL}(P_0\|P_1)^2\Big)}{\eta}\right) \leq \frac{\eta}{2}.$$

Similarly,

$$\mathbb{P}\left(\frac{1}{n-\xi_n}\sum_{i=\xi_n+1}^{n} \delta^2(X_i) \geq \frac{2\Big(\sigma_1^2 + \mathrm{KL}(P_1\|P_0)^2\Big)}{\eta}\right) \leq \frac{\eta}{2}.$$

Using σ_0, $\sigma_1 \leq \sigma_\star$ and $\mathrm{KL}(P_0\|P_1)$, $\mathrm{KL}(P_1\|P_0) \leq \mathrm{J}(P_0, P_1)$, and a union bound, it follows that the event,

$$\hat{v}_n \leq \frac{2\Big(\sigma_\star^2 + \mathrm{J}(P_0,P_1)^2\Big)}{\eta},$$

holds with probability at least $1-\eta$. $\square$

Lemma A.3. *Let $\hat{\xi}_{0,n}$ be defined as*

$$\hat{\xi}_{0,n} = \operatorname*{argmax}_{s\in[n-1]} \log\Bigg(\prod_{i\leq s} f_0(X_{i,n}) \prod_{i>s} f_1(X_{i,n})\Bigg).$$

Then $|\hat{\xi}_{0,n} - \xi_n| = \mathrm{O}_P(\log n)$.

Proof. We write $(X_1, \ldots, X_n)$ to denote $(X_{1,n}, \ldots, X_{n,n})$. First note that $\hat{\xi}_{0,n}$ is equivalently the maximizer of $L(s)$ over $s \in [n-1]$, where

$$L(s) = \sum_{i=1}^{s} \log f_0(X_{i,n}) + \sum_{i=s+1}^{n} \log f_1(X_i).$$

Therefore, for any $t > \xi_n$,

$$L(t) = L(\xi_n) + \sum_{s=\xi_n+1}^{t} \delta(X_s).$$

We note that $\{\delta(X_s)\}_{s=\xi_n+1,\ldots,t}$ are i.i.d. sub-Gaussian with variance factor K_1 and mean $-\mathrm{KL}(P_1\|P_0)$. Hence the empirical average of $\{\delta(X_s)\}_{s=\xi_n+1,\ldots,t}$ is sub-Gaussian with variance factor $K_1/\sqrt{t-\xi_n}$. Consequently,

$$\mathbb{P}\left(\frac{1}{t-\xi_n}\sum_{s=\xi_n+1}^{t}\delta(X_s)>0\right)\leq \exp\biggl(-\frac{(t-\xi_n)\{\mathrm{KL}(P_1\|P_0)\}^2}{2K_1^2}\biggr). \tag{A.5}$$

Hence, by a union bound, for any $c_1>0$,

$$\begin{aligned}\mathbb{P}(\hat{\xi}_{0,n}\geq \xi_n+c_1\log n) &= \mathbb{P}\Bigl(\exists\ t>\xi_n+c_1\log n \text{ such that } L(t)>L(\xi_n)\Bigr)\\ &\leq \sum_{t\geq \xi_n+c_1\log n}\mathbb{P}\left(L(t)>L(\xi_n)\right)\\ &= \sum_{t\geq \xi_n+c_1\log n}\mathbb{P}\left(\frac{1}{t-\xi_n}\sum_{s=\xi_n+1}^{t}\delta(X_s)>0\right).\end{aligned}$$

Applying (A.5), we obtain

$$\begin{aligned}\mathbb{P}(\hat{\xi}_{0,n}\geq \xi_n+c_1\log n) &\leq \sum_{t\geq \xi_n+c_1\log n}\exp\biggl(-\frac{(t-\xi_n)\{\mathrm{KL}(P_1\|P_0)\}^2}{2K_1^2}\biggr)\\ &\leq n\exp\biggl(-\frac{c_1\log n\cdot\{\mathrm{KL}(P_1\|P_0)\}^2}{2K_1^2}\biggr).\end{aligned}$$

Choosing $c_1>\frac{2K_1^2}{\{\mathrm{KL}(P_1\|P_0)\}^2}$, we obtain

$$\mathbb{P}(\hat{\xi}_{0,n}\geq \xi_n+c_1\log n)\to 0 \qquad \text{as } n\to\infty.$$

By a similar argument, for an appropriate constant c_2,

$$\mathbb{P}(\hat{\xi}_{0,n}\leq \xi_n-c_2\log n)\to 0 \qquad \text{as } n\to\infty.$$

Together, these imply $|\hat{\xi}_{0,n}-\xi_n|=\mathrm{O}_P(\log n)$, as required. □

Lemma A.4. *Let $\{p_{t_n,n}\}$ be the CONCH p-values defined via (3.2) based on CPP score (4.3), optimal score, or (4.5) i.e., oracle LLR score. Then for all $t\in[n-1]$,*

$$p_{t,n}\ \leq\ \tilde{p}_{t,n}.$$

Proof. First, consider the log transform of the optimal score defined in (4.3):

$$\log\left(\frac{\prod_{i\le t} f_0(x_i)\prod_{i>t} f_1(x_i)}{\prod_{i\le \xi} f_0(x_i)\prod_{i>\xi} f_1(x_i)}\right),$$

which coincides with (A.2) if we take $\hat{\xi}_n \equiv \xi$ for all $n \in \mathbb{N}$. By Proposition 4.1, the CONCH p-values constructed using optimal score, (4.3) don't change under a log transformation. Therefore, in this case, $p_{t,n} = \tilde{p}_{t,n}$ for all t.

Next, we turn to the oracle LLR score. Fix $n \in \mathbb{N}$, and recall that the oracle LLR score is

$$V_t^{(n)}(\mathbf{x}) := \log\left(\frac{\prod_{i\le t} f_0(x_i)\prod_{i>t} f_1(x_i)}{\prod_{i\le \hat{\xi}_n(\mathbf{x})} f_0(x_i)\prod_{i>\hat{\xi}_n(\mathbf{x})} f_1(x_i)}\right),$$

where

$$\hat{\xi}_n(\mathbf{x}) := \operatorname*{argmax}_{s\in[n-1]} \log\biggl(\prod_{i\le s} f_0(x_i)\prod_{i>s} f_1(x_i)\biggr). \tag{A.6}$$

Now, it follows that for any permutation π, we have that

$$\prod_{i\le \hat{\xi}_n(\pi(\mathbf{x}))} f_0((\pi(\mathbf{x}))_i)\prod_{i>\hat{\xi}_n(\pi(\mathbf{x}))} f_1((\pi(\mathbf{x}))_i) \ge \prod_{i\le \hat{\xi}_n(\mathbf{x})} f_0((\pi(\mathbf{x}))_i)\prod_{i>\hat{\xi}_n(\mathbf{x})} f_1((\pi(\mathbf{x}))_i),$$

and consequently, for any $t \in [n-1]$ and any permutation π of $[n]$,

$$V_t^{(n)}(\pi(\mathbf{X})) \le S_t^{(n)}(\pi(\mathbf{X})), \qquad V_t^{(n)}(\mathbf{X}) = S_t^{(n)}(\mathbf{X}).$$

Hence, for all $t \in [n-1]$, deterministically,

$$p_{t,n} \le \tilde{p}_{t,n},$$

as required. $\square$

A.3.2 Proof of consistency results

Theorem A.5. *Suppose, in the setting of Section 5, that (A1)–(A3) hold and* $|\hat{\xi}_n - \xi_n| = o_P(n^{1/2})$. *Then the p-values* $\{\tilde{p}_{t,n}\}$ *in* (A.3) *satisfy*

$$\mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n-1}\tilde{p}_{i,n}\right] \longrightarrow 0 \qquad \text{as } n\to\infty.$$

Consequently,

$$\frac{1}{n-1}\sum_{i=1}^{n-1}\tilde{p}_{i,n} \xrightarrow{P} 0 \qquad \textit{as } n\to\infty.$$

Proof. Fix $\eta \in (0,1)$. It suffices to prove that

$$\lim_{n\to\infty}\mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n-1}\tilde{p}_{i,n}\right] \le 3\eta.$$

Since $\xi_n/n \to \tau$ and $|\hat{\xi}_n - \xi_n| = \mathrm{o}_P(n^{1/2})$, there exists $N_0 \in \mathbb{N}$ such that for all $n \ge N_0$,

$$\tau^2 \le \xi_n/n \le \sqrt{\tau}, \qquad \mathbb{P}\Big(|\hat{\xi}_n - \xi_n| > n^{1/2}\Big) \le \eta.$$

Fix $\kappa > 1$ (to be chosen later), and partition indices into $\mathcal{I} := \{i : \ |i-\xi_n| \ge \kappa\, n^{1/2}\}$ and its complement $[n-1]\setminus\mathcal{I}$. Since $\tilde{p}_{i,n} \le 1$,

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n-1}\tilde{p}_{i,n}\right] &= \mathbb{E}\left[\frac{1}{n-1}\sum_{i\in\mathcal{I}}\tilde{p}_{i,n}\right] + \mathbb{E}\left[\frac{1}{n-1}\sum_{i\in[n-1]\setminus\mathcal{I}}\tilde{p}_{i,n}\right] \\
&\le \mathbb{E}\left[\frac{1}{n-1}\sum_{i\in\mathcal{I}}\tilde{p}_{i,n}\right] + \frac{2\kappa\, n^{1/2}+1}{n-1} \\
&\le \mathbb{E}\Bigg[\frac{1}{n-1}\sum_{i\in\mathcal{I}}\tilde{p}_{i,n}\ \Bigg|\ |\hat{\xi}_n - \xi_n| \le n^{1/2}\Bigg] + \eta + \frac{2\kappa\, n^{1/2}+1}{n-1}. \qquad \text{(A.7)}
\end{aligned}$$

Fix $t_n \in \mathcal{I}$ with $t_n > \xi_n$ (the case $t_n < \xi_n$ is symmetric). By the tower property,

$$\begin{aligned}
\mathbb{E}\Big[\tilde{p}_{t_n,n}\ \Big|\ |\hat{\xi}_n-\xi_n| \le n^{1/2}\Big] &= \mathbb{E}\Big[\mathbb{E}\Big[\tilde{p}_{t_n,n}\ \Big|\ \hat{\xi}_n,\ |\hat{\xi}_n-\xi_n| \le n^{1/2}\Big]\ \Big|\ |\hat{\xi}_n-\xi_n| \le n^{1/2}\Big] \\
&\le \max_{|k_n-\xi_n|\le n^{1/2}} \mathbb{E}\Big[\tilde{p}_{t_n,n}\ \Big|\ \hat{\xi}_n = k_n\Big].
\end{aligned}$$

Write $\hat{m}_n = |t_n - \hat{\xi}_n| = t_n - \hat{\xi}_n$. By Lemma A.1,

$$\tilde{p}_{t_n,n} \le \frac{n}{t_n}\,\frac{\hat{v}_n}{\hat{m}_n\,\Delta^2_{t_n,\hat{\xi}_n}}\mathbb{1}_{\Delta_{t_n,\hat{\xi}_n}<0} + \mathbb{1}_{\Delta_{t_n,\hat{\xi}_n}\ge 0},$$

where $\Delta_{t_n,\hat{\xi}_n} := \frac{S^{(n)}_{t_n}(\mathbf{X})}{\hat{m}_n} - \hat{\mu}_{t_n,n,L}$. We bound the pieces to control $\mathbb{E}[\tilde{p}_{t_n,n}]$.

First, $\hat{m}_n \ge (\kappa-1)\, n^{1/2}$. Next, for all $n \ge N_0$,

$$\frac{n}{t_n} \le \frac{n}{\xi_n} \le \frac{1}{\tau^2}.$$

By Lemma A.2, with probability at least $1-\eta$,

$$\hat{v}_n \le \frac{4\Big(\sigma_\star^2 + \mathrm{J}(P_0,P_1)^2\Big)}{\eta}.$$

We now obtain a high-probability lower bound on $\Delta^2_{t_n,\hat{\xi}_n}$. Note that $\delta(X_1),\dots,\delta(X_{\xi_n})$ are i.i.d. with mean $\mathrm{KL}(P_0\|P_1)$ and variance σ_0^2, while $\delta(X_{\xi_n+1}),\dots,\delta(X_n)$ are i.i.d. with mean $\mathrm{KL}(P_1\|P_0)$ and variance σ_1^2. On $\{\hat{\xi}_n = k_n\}$ with $\xi_n - n^{1/2} \le k_n \le \xi_n + n^{1/2}$,

$$\frac{S^{(n)}_{t_n}(\mathbf{X})}{\hat{m}_n} - \hat{\mu}_{t_n,n,L} = \sum_{i=1}^{t_n} a_i\,\delta(X_i), \qquad a_i = -\frac{1}{t_n}\mathbb{1}_{i\le k_n} + \frac{k_n}{\hat{m}_n t_n}\mathbb{1}_{i>k_n}.$$

The variance of $\sum_{i=1}^{t_n} a_i\,\delta(X_i)$ is bounded by $\sigma_\star^2\sum_{i=1}^{t_n} a_i^2$. Since $k_n/t_n \le 1$ and $\hat{m}_n < t_n$,

$$\sum_{i=1}^{t_n} a_i^2 = \frac{k_n}{t_n^2} + \frac{k_n^2}{\hat{m}_n t_n^2} \le \frac{1}{t_n} + \frac{1}{\hat{m}_n} \le \frac{2}{\hat{m}_n} \le \frac{2}{(\kappa-1)\,n^{1/2}}.$$

Using $\mathbb{E}_{P_1}\delta(X) = -\mathrm{KL}(P_1\|P_0)$ and $\mathbb{E}_{P_0}\delta(X) = \mathrm{KL}(P_0\|P_1)$, we obtain

$$\mathbb{E}\Bigg[\sum_{i=1}^{t_n} a_i\,\delta(X_i)\Bigg] = \begin{cases} -\dfrac{\xi_n}{t_n}\,\mathrm{J}(P_0,P_1), & k_n > \xi_n, \\ -\dfrac{\xi_n}{t_n}\,\mathrm{J}(P_0,P_1) + \dfrac{(\hat{m}_n - m_n)_+}{\hat{m}_n}\,\mathrm{J}(P_0,P_1), & k_n \le \xi_n, \end{cases}$$

where $a_+ = \max\{a,0\}$. Moreover,

$$\frac{(\hat{m}_n - m_n)_+}{\hat{m}_n} \le \frac{1}{\kappa-1}, \qquad \mathbb{E}\Bigg[\sum_{i=1}^{t_n} a_i\,\delta(X_i)\Bigg] \le -\Big(\frac{\xi_n}{t_n} - \frac{1}{\kappa-1}\Big)\mathrm{J}(P_0,P_1).$$

Choose κ large enough so that $\frac{1}{\kappa-1} \le \tau^2/2$. Then for all $n \ge N_0$,

$$\mathbb{E}\Bigg[\sum_{i=1}^{t_n} a_i\,\delta(X_i)\Bigg] \le -\frac{\tau^2}{2}\,\mathrm{J}(P_0,P_1) < 0.$$

By Chebyshev's inequality, with probability at least $1-\eta$,

$$\sum_{i=1}^{t_n} a_i\,\delta(X_i) \;\le\; -\frac{\tau^2}{2}\,\mathrm{J}(P_0,P_1) + \tau\sigma_\star\sqrt{\frac{1}{\eta\,n^{1/2}}}.$$

Hence there exists $N_1 > N_0$ such that for all $n \geq N_1$, with probability at least $1 - \eta$,

$$\sum_{i=1}^{t_n} a_i \, \delta(X_i) \;\leq\; -\frac{\tau^2}{4} \, \mathrm{J}(P_0, P_1).$$

Equivalently, with probability at least $1 - \eta$,

$$\Delta_{t_n, \hat{\xi}_n} = \frac{S_{t_n}^{(n)}(\mathbf{X})}{\hat{m}_n} - \hat{\mu}_{t_n, n, L} \;\leq\; -\frac{\tau^2}{4} \, \mathrm{J}(P_0, P_1), \qquad \hat{\xi}_n = k_n.$$

Let $\mathcal{A}$ be the event

$$\hat{v}_n \leq \frac{4\Big(\sigma_\star^2 + \mathrm{J}^2(P_0, P_1)\Big)}{\eta}, \qquad \Delta_{t_n, \hat{\xi}_n}^2 \geq \frac{\tau^4}{16} \, \mathrm{J}^2(P_0, P_1), \qquad \Delta_{t_n, \hat{\xi}_n} < 0.$$

By the discussion above, $\mathbb{P}(\mathcal{A} \cap \{\hat{\xi}_n = k_n\}) \geq 1 - 2\eta$, and on $\mathcal{A} \cap \{\hat{\xi}_n = k_n\}$,

$$\tilde{p}_{t_n, n} \leq \frac{1}{\tau^2} \cdot \frac{4\Big(\sigma_\star^2 + \mathrm{J}^2(P_0, P_1)\Big)}{\eta} \cdot \frac{\tau^2}{2n^{1/2}} \cdot \frac{16}{\tau^4 \mathrm{J}^2(P_0, P_1)}.$$

Consequently, for any $n \geq N_1$ and for any k_n such that $|k_n - \xi_n| \leq n^{1/2}$,

$$\mathbb{E}\Big[\tilde{p}_{t_n, n} \;\Big|\; \hat{\xi}_n = k_n\Big] \leq \frac{32 \left(\sigma_\star^2 + \mathrm{J}^2(P_0, P_1)\right)}{\tau^4 \, \eta \, \mathrm{J}^2(P_0, P_1) \, n^{1/2}} + 2\eta.$$

Averaging over $t_n \in \mathcal{I}$ and by (A.7),

$$\mathbb{E}\left[\frac{1}{n-1} \sum_{i=1}^{n-1} \tilde{p}_{i,n}\right] \leq \frac{32 \left(\sigma_\star^2 + \mathrm{J}^2(P_0, P_1)\right)}{\tau^4 \, \eta \, \mathrm{J}^2(P_0, P_1) \, n^{1/2}} + 3\eta + \frac{2\kappa \, n^{1/2} + 1}{n-1}.$$

Therefore,

$$\lim_{n \to \infty} \mathbb{E}\left[\frac{1}{n-1} \sum_{i=1}^{n-1} \tilde{p}_{i,n}\right] \leq 3\eta.$$

Since η is arbitrary, the result follows. □

Proof of Theorem 5.1 Fix $\eta > 0$. By Markov's inequality, we have that

$$\mathbb{P}\left(\frac{|\,\mathcal{C}_{n,1-\alpha}^{\mathrm{CONCH}}\,|}{n-1} \geq \eta\right) \leq \frac{\mathbb{E}\left[|\,\mathcal{C}_{n,1-\alpha}^{\mathrm{CONCH}}\,|\right]}{(n-1)\eta}.$$

Therefore, it suffices to show that

$$\frac{1}{n-1}\,\mathbb{E}\left[|\,\mathcal{C}^{\mathrm{CONCH}}_{n,1-\alpha}\,|\right] \longrightarrow 0, \qquad \text{as } n\to\infty.$$

Next, note that by applying Markov's inequality once more,

$$\frac{1}{n-1}\,\mathbb{E}\left[|\,\mathcal{C}^{\mathrm{CONCH}}_{n,1-\alpha}\,|\right] \;=\; \frac{1}{n-1}\sum_{i=1}^{n-1}\mathbb{P}\left(p_{i,n}>\alpha\right) \;\le\; \frac{1}{(n-1)\alpha}\sum_{i=1}^{n-1}\mathbb{E}[p_{i,n}]$$

The result now follows by Theorem A.5 and Lemma A.4. □

A.3.3 Uniform decay of CONCH p-values away from the changepoint

Theorem A.6. *In the setting of Section 5, suppose (A1)–(A3) hold. Further, assume that the log-likelihood ratio $\delta(X)$ is sub-Gaussian under P_0 with variance proxy K_0 and under P_1 with variance proxy K_1. Let $\{p_{i,n}\}$ denote the CONCH p-values computed using either the optimal score (4.3) or the oracle LLR score (4.5). Then there exists a constant $\kappa>0$ (independent of n) such that, as $n\to\infty$,*

$$\max_{|t_n-\xi_n|\,\ge\,\kappa\log n}\; p_{t_n,n} \xrightarrow{P} 0.$$

Proof. Fix $\eta,\alpha_0>0$, and let $\kappa>0$ be chosen later. By Lemma A.4, for each $i\in[n-1]$, $p_{i,n}\le\tilde{p}_{i,n}$, so it suffices to show

$$\lim_{n\to\infty}\mathbb{P}\Big(\max_{t_n\,\ge\,\xi_n+\kappa\log n}\;\tilde{p}_{t_n,n}>\alpha_0\Big)\le 2\eta,$$

where $\hat{\xi}_n$ is either ξ_n or the MLE estimate (4.6). By symmetry, the same control holds when the maximum is taken over the indices $\{t_n\le\xi_n-\kappa\log n\}$, and a union bound then yields the claim. Throughout, we write $(X_1,\dots,X_n)$ for $(X_{1,n},\dots,X_{n,n})$ and set $K_\star:=\max\{K_0,K_1\}$.

Since $\xi_n/n\to\tau$ and, for both choices of $\hat{\xi}_n$, $|\hat{\xi}_n-\xi_n|=\mathrm{O}_P(\log n)$, there exists N_0 such that for all $n>N_0$,

$$\mathbb{P}\Big(|\hat{\xi}_n-\xi_n|>C_0\,\log n\Big)\le\eta, \qquad \frac{\xi_n}{n}\ge\tau^2. \tag{A.8}$$

Therefore,

$$\mathbb{P}\Big(\max_{t_n\,\ge\,\xi_n+\kappa\log n}\;\tilde{p}_{t_n,n}>\alpha_0\Big)\le\mathbb{P}\Big(\max_{t_n\,\ge\,\xi_n+\kappa\log n}\;\tilde{p}_{t_n,n}>\alpha_0,\;|\hat{\xi}_n-\xi_n|\le C_0\,\log n\Big)+\eta.$$

Fix $t_n \geq \xi_n + \kappa \log n$ with $\kappa > C_0$, and write $\hat{m}_n = |t_n - \hat{\xi}_n|$. By Lemma A.1,

$$\tilde{p}_{t_n,n} \leq \frac{n}{t_n} \frac{\hat{v}_n}{\hat{m}_n \, \Delta^2_{t_n,\hat{\xi}_n}} \mathbb{1}_{\Delta_{t_n,\hat{\xi}_n}<0} + \mathbb{1}_{\Delta_{t_n,\hat{\xi}_n}\geq 0},$$

where $\Delta_{t_n,\hat{\xi}_n} := \frac{S^{(n)}_{t_n}(\mathbf{X})}{\hat{m}_n} - \hat{\mu}_{t_n,n,L}$. We bound each component in this bound above.

First, $\hat{m}_n \geq (\kappa - C_0)\log n$. Next, for all $n > N_0$ and $t_n \geq \xi_n + \kappa \log n$,

$$\frac{n}{t_n} \leq \frac{n}{\xi_n} \leq \frac{1}{\tau^2}.$$

By Lemma A.2, with probability at least $1-\eta$,

$$\hat{v}_n \leq \frac{4\Big(\sigma^2_\star + \mathrm{J}(P_0,P_1)^2\Big)}{\eta}. \tag{A.9}$$

On the complement of this event, if

$$\Delta_{t_n,\hat{\xi}_n} \leq -\left(\frac{4\Big(\sigma^2_\star + \mathrm{J}(P_0,P_1)^2\Big)}{\eta\alpha_0(\kappa - C_0)\log n}\right)^{1/2},$$

then we have $\tilde{p}_{t_n,n} \leq \alpha_0$. Therefore, now it suffices to bound

$$\mathbb{P}\left(\Delta_{t_n,\hat{\xi}_n} \leq -\left(\frac{4\Big(\sigma^2_\star + \mathrm{J}(P_0,P_1)^2\Big)}{\eta\alpha_0(\kappa - C_0)\log n}\right)^{1/2}, \ |\hat{\xi}_n - \xi_n| \leq C_0 \log n\right).$$

Unlike in the proof of A.5, sub-Gaussianity here yields an exponential tail, which then yields the uniform control.

Note that $\delta(X_1), \ldots, \delta(X_{\xi_n})$ are i.i.d. sub-Gaussian with variance factor K_0, and $\delta(X_{\xi_n+1}), \ldots, \delta(X_{t_n})$ are i.i.d. sub-Gaussian with variance factor K_1. On $\{\hat{\xi}_n = k_n\}$ with $\xi_n - C_0 \log n \leq k_n \leq \xi_n + C_0 \log n$,

$$\frac{S^{(n)}_{t_n}(\mathbf{X})}{\hat{m}_n} - \hat{\mu}_{t_n,n,L} = \sum_{i=1}^{t_n} a_i \, \delta(X_i), \qquad a_i = -\frac{1}{t_n}\mathbb{1}_{i\leq k_n} + \frac{k_n}{\hat{m}_n t_n}\mathbb{1}_{i>k_n},$$

and $\sum_{i=1}^{t_n} a_i \, \delta(X_i)$ is sub-Gaussian with variance factor $K_\star\sqrt{\sum_{i=1}^{t_n} a_i^2}$. Since $k_n/t_n \leq 1$ and $\hat{m}_n < t_n$,

$$\sum_{i=1}^{t_n} a_i^2 = \frac{k_n}{t_n^2} + \frac{k_n^2}{\hat{m}_n t_n^2} \leq \frac{1}{t_n} + \frac{1}{\hat{m}_n} \leq \frac{2}{\hat{m}_n} \leq \frac{2}{(\kappa - C_0)\log n}.$$

Moreover, recalling $\mathbb{E}_{P_1}\delta(X) = -\mathrm{KL}(P_1\|P_0)$ and $\mathbb{E}_{P_0}\delta(X) = \mathrm{KL}(P_0\|P_1)$, we get

$$\mathbb{E}\Bigg[\sum_{i=1}^{t_n} a_i\,\delta(X_i)\Bigg] = \begin{cases} -\dfrac{\xi_n}{t_n}\,\mathrm{J}(P_0,P_1), & k_n > \xi_n, \\ -\dfrac{\xi_n}{t_n}\,\mathrm{J}(P_0,P_1) + \dfrac{(\hat{m}_n - m_n)_+}{\hat{m}_n}\,\mathrm{J}(P_0,P_1), & k_n \le \xi_n, \end{cases}$$

where $a_+ = \max\{a,0\}$. Hence, on $\{\hat{\xi}_n = k_n\}$,

$$\frac{(\hat{m}_n - m_n)_+}{\hat{m}_n} \le \frac{C_0}{\kappa - C_0}, \qquad \mathbb{E}\Bigg[\sum_{i=1}^{t_n} a_i\,\delta(X_i)\Bigg] \le -\Big(\frac{\xi_n}{t_n} - \frac{C_0}{\kappa - C_0}\Big)\mathrm{J}(P_0,P_1).$$

Note that $\xi_n/t_n \ge \xi_n/n \ge \tau^2$ for all $n > N_0$. Next, there exists $N_1 \in \mathbb{N}$ with $N_1 \ge N_0$ such that for all $n > N_1$,

$$\log(2C_0\log n + 1) \le \log n, \qquad \frac{4\Big(\sigma_\star^2 + \mathrm{J}(P_0,P_1)^2\Big)}{\eta\alpha_0(\kappa - C_0)\log n} \le \frac{16K_\star^2}{\kappa - C_0}.$$

Thus, for $n > N_1$,

$$\left(\frac{4\Big(\sigma_\star^2 + \mathrm{J}(P_0,P_1)^2\Big)}{\eta\alpha_0(\kappa - C_0)\log n}\right)^{1/2} \le \frac{4K_\star}{\sqrt{\kappa - C_0}}.$$

Define

$$u := -\frac{4K_\star}{\sqrt{\kappa - C_0}} + \tau^2\,\mathrm{J}(P_0,P_1) - \frac{C_0}{\kappa - C_0}\,\mathrm{J}(P_0,P_1).$$

Choose $\kappa > C_0$ so that

$$u \ge \tfrac{\tau}{4}\,\mathrm{J}(P_0,P_1), \qquad \frac{\tau^2\,\mathrm{J}^2(P_0,P_1)}{64K_\star^2}\,(\kappa - C_0) > 3.$$

By sub-Gaussianity of $\sum_{i=1}^{t_n} a_i\,\delta(X_i)$, for $n > \max\{N_0, N_1\}$,

$$\begin{aligned}
&\mathbb{P}\left(\Delta_{t_n,\hat{\xi}_n} \ge -\left(\frac{4\Big(\sigma_\star^2 + \mathrm{J}(P_0,P_1)^2\Big)}{\eta\alpha_0(\kappa - C_0)\log n}\right)^{1/2},\ \hat{\xi}_n = k_n\right) \\
&\qquad \le \mathbb{P}\Bigg(\sum_{i=1}^{t_n} a_i\,\delta(X_i) \ge -\frac{4K_\star}{\sqrt{\kappa - C_0}}\Bigg) \le \mathbb{P}\Bigg(\sum_{i=1}^{t_n} a_i\,\delta(X_i) - \mathbb{E}\sum_{i=1}^{t_n} a_i\,\delta(X_i) \ge \frac{\tau}{4}\,\mathrm{J}(P_0,P_1)\Bigg) \\
&\qquad \le \exp\Bigg(-\frac{\tau^2\,\mathrm{J}^2(P_0,P_1)}{64K_\star^2}\,(\kappa - C_0)\,\log n\Bigg) \ \le\ e^{-3\log n} = n^{-3}.
\end{aligned}$$

Hence, by a union bound over at most $2C_0 \log n + 1$ values of k_n, for $n > \max\{N_0, N_1\}$,

$$\mathbb{P}\left(\Delta^2_{t_n,\hat{\xi}_n} \leq \left(\frac{4\left(\sigma_\star^2 + \mathrm{J}(P_0,P_1)^2\right)}{\eta\alpha_0(\kappa - C_0)\log n}\right), \quad |\hat{\xi}_n - \xi_n| \leq C_0 \log n\right) \leq n^{-2}.$$

Another union bound, summing this probability over at most n indices t_n, yields

$$\mathbb{P}\left(\max_{t_n \geq \xi_n + \kappa \log n} \Delta^2_{t_n,\hat{\xi}_n} \leq \left(\frac{4\left(\sigma_\star^2 + \mathrm{J}(P_0,P_1)^2\right)}{\eta\alpha_0(\kappa - C_0)\log n}\right), \quad |\hat{\xi}_n - \xi_n| \leq C_0 \log n\right) \leq \frac{1}{n}. \tag{A.10}$$

Combining the steps above, by (A.8) and (A.9), we obtain

$$\mathbb{P}\left(\max_{t_n \geq \xi_n + \kappa \log n} \tilde{p}_{t_n,n} > \alpha_0\right) \leq 2\eta + \frac{1}{n},$$

and consequently,

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{t_n \geq \xi_n + \kappa \log n} \tilde{p}_{t_n,n} > \alpha_0\right) \leq 2\eta,$$

as required. This completes the proof. $\square$

A.4 Proof of Theorem 6.1 (Universality Theorem)

The proof of this universality theorem is inspired by the classical universality result for full-conformal procedures in the predictive inference framework (see Vovk et al., 2005, Chapter 2.4; Angelopoulos et al., 2024, Theorem 9.6).

Firstly, based on the given confidence set C. Fix $n \in \mathbb{N}$

$$S_t(\mathbf{x}) = \mathbb{1}\left\{t \in C(\mathbf{x})\right\} \in \{0, 1\},$$

for any $t \in [n-1]$. We will show that the CONCH confidence set constructed from this score, denoted $\mathcal{C}^{\mathrm{CONCH}}_{1-\alpha}$, coincides exactly with the given confidence set C.

We start with showing that $\mathcal{C}^{\mathrm{CONCH}}_{1-\alpha}(\mathbf{X}) \supseteq C(\mathbf{X})$; that is, if $t \in C(\mathbf{X})$, then it holds that $p_t > \alpha$, where p_t is as defined in (3.2). This is immediate by observing that if $t \in C(\mathbf{X})$, then $S_t(\mathbf{X}) = 1$, and consequently,

$$p_t = \frac{1}{|\Pi_t|} \sum_{\pi \in \Pi_t} \mathbb{1}\left\{S_t(\pi(\mathbf{X})) \leq S_t(\mathbf{X})\right\} = \frac{1}{|\Pi_t|} \sum_{\pi \in \Pi_t} \mathbb{1}\left\{S_t(\pi(\mathbf{X})) \leq 1\right\} = 1.$$

Next, we show that $\mathcal{C}_{1-\alpha}^{\mathrm{CONCH}}(\mathbf{X}) \subseteq C(\mathbf{X})$, i.e., if $t \notin C(\mathbf{X})$, then $p_t \leq \alpha$. To that end, we first claim that for any $t \in [n-1]$ and any vector $\mathbf{x} \in \mathcal{X}^n$,

$$\frac{1}{|\Pi_t|} \sum_{\pi \in \Pi_t} \mathbb{1}\left\{t \in C(\pi(\mathbf{x}))\right\} \geq 1 - \alpha. \tag{A.11}$$

We now prove this claim. Fix $t \in [n-1]$, and sample π uniformly from the set of permutations Π_t. Define $\tilde{\mathbf{X}} := (\tilde{X}_1, \ldots, \tilde{X}_n) := \pi(\mathbf{x})$. Conditional on the multisets $\{x_1, \ldots, x_t\}$ and $\{x_{t+1}, \ldots, x_n\}$, we have

$$\tilde{X}_1, \ldots, \tilde{X}_t \text{ are exchangeable, and } \tilde{X}_{t+1}, \ldots, \tilde{X}_n \text{ are exchangeable.}$$

Moreover, conditional on the multisets, $(\tilde{X}_1, \ldots, \tilde{X}_t)$ and $(\tilde{X}_{t+1}, \ldots, \tilde{X}_n)$ are independent, implying that the sampling process of $\tilde{\mathbf{X}}$ satisfies Assumption 1. Consequently,

$$\mathbb{P}_{\pi \sim \mathrm{Unif}(\Pi_t)}\Big(t \in C(\tilde{\mathbf{X}}) \,\Big|\, \{x_1, \ldots, x_t\}, \{x_{t+1}, \ldots, x_n\}\Big) \geq 1 - \alpha,$$

or equivalently, (A.11) holds.

Returning to the main proof, observe that if $t \notin C(\mathbf{X})$, then $S_t(\mathbf{X}) = 0$. Consequently,

$$\begin{aligned} p_t &= \frac{1}{|\Pi_t|} \sum_{\pi \in \Pi_t} \mathbb{1}\left\{S_t(\pi(\mathbf{X})) \leq S_t(\mathbf{X})\right\} \\ &= \frac{1}{|\Pi_t|} \sum_{\pi \in \Pi_t} \mathbb{1}\left\{S_t(\pi(\mathbf{X})) \leq 0\right\} = \frac{1}{|\Pi_t|} \sum_{\pi \in \Pi_t} \mathbb{1}\left\{t \notin C(\pi(\mathbf{X}))\right\} \leq \alpha, \end{aligned}$$

where the last step follows from (A.11). This completes the proof. □

A.5 Asymptotic validity of CONCH-SEG algorithm

In this section, we establish asymptotic coverage for a cross-fitted variant of CONCH-SEG. A direct analysis of CONCH-SEG from Section 6.2 is hard because segmentation and CONCH are applied to the same observations, potentially violating Assumption 1. To decouple these steps, we partition the index set into two disjoint folds

$$\mathcal{I}_1 := \{t \in [n] : t \text{ odd}\}, \quad \text{and} \quad \mathcal{I}_2 := \{t \in [n] : t \text{ even}\}.$$

For $r \in \{1, 2\}$, we run the segmentation algorithm on $\mathcal{I}_r$ to obtain $\hat{K}^{(r)}$ and

$$0 = \hat{\xi}_0^{(r)} < \hat{\xi}_1^{(r)} < \cdots < \hat{\xi}_{\hat{K}^{(r)}}^{(r)} < n = \hat{\xi}_{\hat{K}^{(r)}+1}^{(r)}.$$

Algorithm 5: CONCH-SEG-`crossfit`

Input: $(X_t)_{t=1}^n$ (data); $S : \bigcup_{m \in \mathbb{N}} \mathcal{X}^m \to \mathbb{R}^m$ (CPP score function); segmentation algorithm `SEG`

Output: $\mathcal{C}_{1-\alpha}^{\text{CONCH-SEG-}\texttt{crossfit}}$

$\mathcal{I}_1 \leftarrow \{t \le n : \ t \text{ odd}\}, \ \mathcal{I}_2 \leftarrow \{t \le n : \ t \text{ even}\}$;
$\mathcal{C} \leftarrow \varnothing$;
for $r \in \{1, 2\}$ **do**
  $(\hat{K}^{(r)}, \hat{\xi}_1^{(r)}, \ldots, \hat{\xi}_{\hat{K}^{(r)}}^{(r)}) \leftarrow \texttt{SEG}\Big((X_t)_{t \in \mathcal{I}_r}\Big)$;
  Compute $(\tilde{X}_0^{(r)}, \ldots, \tilde{X}_{\hat{K}^{(r)}}^{(r)})$ by (6.3) based on $(\hat{\xi}_1^{(r)}, \ldots, \hat{\xi}_{\hat{K}^{(r)}}^{(r)})$;
  $J_\ell^{(r)} \leftarrow [\, \tilde{X}_{\ell-1}^{(r)}, \, \tilde{X}_\ell^{(r)} \,] \cap \mathcal{I}_{3-r}$ for $\ell \in [\hat{K}^{(r)}]$;
  for $\ell \in [\hat{K}^{(r)}]$ **do**
    Let $X^{(r,\ell)}$ be the subsequence $(X_t)_{t \in J_\ell^{(r)}}$ ordered by increasing index t;
    Define $S^{(r,\ell)} : \mathcal{X}^{|J_\ell^{(r)}|} \to \mathbb{R}^{|J_\ell^{(r)}|-1}$;
    Compute CONCH p-values $\{p_t : t \in J_\ell^{(r)} \setminus \{\max J_\ell^{(r)}\}\}$ as in (3.2) using score $S^{(r,\ell)}$ on $X^{(r,\ell)}$;
    $\mathcal{C}_\ell^{(r)} \leftarrow \{\, t \in J_\ell^{(r)} : p_t > \alpha \,\}$;
    $\mathcal{C} \leftarrow \mathcal{C} \cup \mathcal{C}_\ell^{(r)}$;
  end
end
return $\mathcal{C}_{1-\alpha}^{\textit{CONCH-SEG-crossfit}} \leftarrow \mathcal{C}$

Next, we form the segment boundaries as in (6.3), based on the sequence $(\hat{\xi}_0^{(r)}, \ldots, \hat{\xi}_{\hat{K}^{(r)}}^{(r)})$, to obtain

$$1 = \tilde{X}_0^{(r)} < \tilde{X}_1^{(r)} < \cdots < \tilde{X}_{\hat{K}^{(r)}-1}^{(r)} < \tilde{X}_{\hat{K}^{(r)}}^{(r)} = n,$$

and define disjoint segments $J_\ell^{(r)} := [\, \tilde{X}_{\ell-1}^{(r)}, \, \tilde{X}_\ell^{(r)} \,] \cap \mathcal{I}_{3-r}$ for $\ell \in [\hat{K}^{(r)}]$. We then run CONCH on the restricted segment $J_\ell^{(r)}$ to produce a segmentwise confidence set, and aggregate across segments and folds. In particular, when segmentation is performed on $\mathcal{I}_1$ (respectively, $\mathcal{I}_2$), CONCH is run on $\mathcal{I}_2$ (respectively, $\mathcal{I}_1$). This construction, formalized in Algorithm 5, restores the exchangeability structure required for validity, as established in the following theorem.

Theorem A.7 (Asymptotic coverage of CONCH-SEG-`crossfit`)**.** *Fix $\alpha \in (0, 1)$ and consider the multiple-changepoint model in (6.2). Suppose the segmentation algorithm, for each $r \in \{1, 2\}$, satisfies:*

(a) ***Consistent changepoint count.*** $\mathbb{P}\Big(\hat{K}^{(r)} = K\Big) \to 1$ *as* $n \to \infty$.

(b) ***Vanishing normalized localization error.***

$$\max_{k\in[\hat{K}^{(r)}]}\min_{j\in[K]}\frac{|\hat{\xi}_k^{(r)}-\xi_j|}{\min_{j\in[K-1]}(\xi_{j+1}-\xi_j)}\xrightarrow{P}0\quad \textit{as } n\to\infty.$$

(c) ***No clustering of estimated changepoints.*** *There exists $\eta>0$, such that*

$$\mathbb{P}\left(\frac{\min_{k\in[\hat{K}^{(r)}-1]}\left(\hat{\xi}_{k+1}^{(r)}-\hat{\xi}_k^{(r)}\right)}{\min_{j\in[K-1]}(\xi_{j+1}-\xi_j)}>\eta\right)\to 1\quad \textit{as } n\to\infty.$$

Then, for any $k\in[K]$,

$$\mathbb{P}\Big(\xi_k\in\mathcal{C}_{1-\alpha}^{\textit{CONCH-SEG-}\texttt{crossfit}}\Big)\geq 1-\alpha-\mathrm{o}(1)\qquad \textit{as } n\to\infty,$$

where the probability is taken under the model class (6.2).

Proof. Fix $k\in[K]$ and, without loss of generality, assume ξ_k is odd (the even case is symmetric). Recall that in the cross-fitted construction, fold $r=2$ applies the segmentation algorithm to the even indices and runs CONCH on the odd indices. Define the event

$$\mathcal{G}:=\Big\{\hat{K}^{(2)}=K,\ \ \xi_{k-1}<\tilde{X}_{k-1}^{(2)}<\xi_k<\tilde{X}_k^{(2)}<\xi_{k+1}\Big\},$$

namely, the (even-fold) midpoint boundaries $\tilde{X}_{k-1}^{(2)}$ and $\tilde{X}_k^{(2)}$ lie on opposite sides of the true changepoint ξ_k. On $\mathcal{G}$, ξ_k falls inside $\Big[\tilde{X}_{k-1}^{(2)},\tilde{X}_k^{(2)}\Big]$, so by construction,

$$\Big\{\xi_k\in\mathcal{C}_{1-\alpha}^{\text{CONCH-SEG-}\texttt{crossfit}}\Big\}=\Big\{\xi_k\in\mathcal{C}_k^{(2)}\Big\},$$

where $\mathcal{C}_k^{(2)}$ is the CONCH confidence set computed on $J_k^{(2)}$, the odd-index subsequence of $[\tilde{X}_{k-1}^{(2)},\tilde{X}_k^{(2)}]$.

Hence, by the law of total probability and tower law, we have

$$\begin{aligned}\mathbb{P}\Big(\xi_k\in\mathcal{C}_{1-\alpha}^{\text{CONCH-SEG-}\texttt{crossfit}}\Big)&\geq\mathbb{P}\Big(\xi_k\in\mathcal{C}_k^{(2)}\ \Big|\ \mathcal{G}\Big)\,\mathbb{P}(\mathcal{G})\\&=\mathbb{E}\Big[\mathbb{P}\Big(\xi_k\in\mathcal{C}_k^{(2)}\ \Big|\ \mathcal{G},(X_t)_{t\in\mathcal{I}_2}\Big)\ \Big|\ \mathcal{G}\Big]\,\mathbb{P}(\mathcal{G})\end{aligned}\tag{A.12}$$

Conditional on $\mathcal{G}$ and $(X_t)_{t\in\mathcal{I}_2}$, the CONCH algorithm is run on an independent set of observations with indices in $J_k^{(2)}$. Therefore, $(X_t)_{t\in J_k^{(2)}}$ satisfies Assumption 1 with a single changepoint at ξ_k. Consequently, by Theorem 3.1, we have

$$\mathbb{P}\Big(\xi_k \in \mathcal{C}_k^{(2)} \;\Big|\; \mathcal{G}, (X_t)_{t \in \mathcal{I}_2}\Big) \geq 1 - \alpha \quad \text{almost surely.}$$

Plugging this into (A.12) yields

$$\mathbb{P}\Big(\xi_k \in \mathcal{C}_{1-\alpha}^{\text{CONCH-SEG-}\texttt{crossfit}}\Big) \geq (1-\alpha)\mathbb{P}(\mathcal{G}) \geq (1-\alpha) - \mathbb{P}(\mathcal{G}^c).$$

Now, it remains to show $\mathbb{P}(\mathcal{G}) \to 1$. To that end, let $\Delta := \min_{j \in [K-1]}(\xi_{j+1} - \xi_j)$ and define the event

$$\mathcal{A} := \Big\{\hat{K}^{(2)} = K, \min_{k \in [\hat{K}^{(2)}-1]} \Big(\hat{\xi}_{k+1}^{(2)} - \hat{\xi}_k^{(r)}\Big) > \eta\Delta \\ \text{and } \max_{j \in [K]} \min_{i \in [K]} |\hat{\xi}_j^{(2)} - \xi_i| \leq \min\{\Delta/8, \eta\Delta/2\}\Big\}.$$

By the theorem hypothesis, $\mathbb{P}(\mathcal{A}) \to 1$. Now on $\mathcal{A}$, for each j choose $i_j \in [K]$ attaining the inner minimum, i.e.,

$$\min_{i \in [K]} |\hat{\xi}_j^{(2)} - \xi_i| \;=\; |\hat{\xi}_j^{(2)} - \xi_{i_j}|.$$

Firstly observe that $i_{j+1} \neq i_j$ for all $j \in [K]$. If that's not the case, then

$$\eta\Delta < \hat{\xi}_{j+1}^{(2)} - \hat{\xi}_j^{(2)} \leq |\hat{\xi}_{j+1}^{(2)} - \xi_{i_{j+1}}| + |\hat{\xi}_j^{(2)} - \xi_{i_j}| \leq \eta\Delta,$$

leading to a contradiction.

We further claim that $i_1 < i_2 < \ldots < i_K$. If instead, $i_{j+1} < i_j$ for some $j \in [K-1]$, then

$$\begin{aligned} \hat{\xi}_{j+1}^{(2)} - \hat{\xi}_j^{(2)} &= \Big(\hat{\xi}_{j+1}^{(2)} - \xi_{i_{j+1}}\Big) + \Big(\xi_{i_{j+1}} - \xi_{i_j}\Big) + \Big(\xi_{i_j} - \hat{\xi}_j^{(2)}\Big) \\ &\leq \frac{\Delta}{8} - \Delta + \frac{\Delta}{8} = -\frac{3\Delta}{4}, \end{aligned}$$

contradicting $\hat{\xi}_{j+1}^{(2)} > \hat{\xi}_j^{(2)}$. Hence, we must have $i_j = j$ for all j, and

$$|\hat{\xi}_j^{(2)} - \xi_j| \leq \Delta/8, \qquad \text{for } j \in [K]. \tag{A.13}$$

Consequently, it follows that for $j \in [K]$,

$$\begin{aligned}\hat{\xi}^{(2)}_{j+1} - \hat{\xi}^{(2)}_j &= \left(\hat{\xi}^{(2)}_{j+1} - \xi_{j+1}\right) + \left(\xi_{j+1} - \xi_j\right) + \left(\xi_j - \hat{\xi}^{(2)}_j\right) \\ &\geq -\frac{\Delta}{8} + \Delta - \frac{\Delta}{8} = \frac{3\Delta}{4}.\end{aligned}$$

Let $\tilde{X}^{(2)}_j$ denote the midpoint between $\hat{\xi}^{(2)}_j$ and $\hat{\xi}^{(2)}_{j+1}$ (and set $\tilde{X}^{(2)}_0 = 1$, $\tilde{X}^{(2)}_K = n$ as in (6.3)). Then

$$\tilde{X}^{(2)}_j - \hat{\xi}^{(2)}_j = \frac{1}{2}\left(\hat{\xi}^{(2)}_{j+1} - \hat{\xi}^{(2)}_j\right) \geq \frac{3\Delta}{8}, \qquad \hat{\xi}^{(2)}_j - \tilde{X}^{(2)}_{j-1} = \frac{1}{2}\left(\hat{\xi}^{(2)}_j - \hat{\xi}^{(2)}_{j-1}\right) \geq \frac{3\Delta}{8}.$$

Together with (A.13), this implies for each $j \in [K]$ that

$$\tilde{X}^{(2)}_{j-1} < \xi_j < \tilde{X}^{(2)}_j.$$

Hence $\mathcal{A} \subseteq \mathcal{G}$, and thus $\mathbb{P}(\mathcal{G}) \to 1$. Therefore,

$$\mathbb{P}\left(\xi_k \in \mathcal{C}^{\text{CONCH-SEG-}\texttt{crossfit}}_{1-\alpha}\right) \geq (1-\alpha) - \mathbb{P}(\mathcal{G}^c) = 1 - \alpha - \mathrm{o}(1),$$

as claimed. □

In particular, kernel-based changepoint detection (KCPD) yields estimators that satisfy the consistency conditions (a) and (b), stated in the above theorem, under mild regularity assumptions [see, e.g., Garreau and Arlot, 2018, Diaz-Rodriguez and Jia, 2025]. This supports the large-sample validity of the CONCH-SEG-`crossfit` procedure when wrapped around a KCPD algorithm. However, in our experiments, whether or not we employ cross-fitting has little effect on power; see Appendix B.2 for empirical evidence.

B Additional Experiments

B.1 Gaussian mean-shift: comparison with Dandapanthula and Ramdas [2025]

In the Gaussian mean-shift setting described in Section 7.1.1, we compare our framework against the changepoint localization method of Dandapanthula and Ramdas [2025], which also constructs distribution-free confidence sets for changepoints using a matrix of conformal p-values. Figure 6 displays the p-value distributions from both methods. Their approach yields a confidence set over $[362, 432]$, which is broader than the widest interval obtained by

CONCH (using the weighted-mean score).

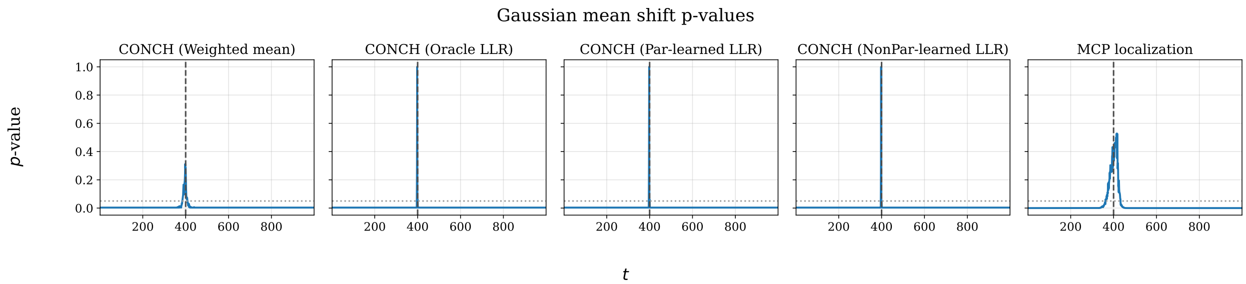

Figure 6: p-value distributions from Dandapanthula and Ramdas [2025] and CONCH under the Gaussian mean-shift model.

B.2 Gaussian mean-shift: localization of multiple changepoints

We consider a multiple-changepoint Gaussian mean-shift model to illustrate the performance of CONCH-SEG (Algorithm 4). In particular, we generate $n = 1500$ observations with true changepoints at $\xi_1 = 150$, $\xi_2 = 500$, $\xi_3 = 820$, and $\xi_4 = 1100$, segment means $\mu_1 = -1$, $\mu_2 = 0.5$, $\mu_3 = 1.5$, $\mu_4 = -2$, and $\mu_5 = -1$, and common variance $\sigma^2 = 1$, i.e.,

$$X_t \sim \mathcal{N}(\mu_j, \sigma^2) \quad \text{for } t \in (\xi_{j-1}, \xi_j], \qquad \text{and} \;\; \xi_0 = 0, \;\xi_5 = n.$$

Initial changepoint estimates are obtained via KCPD [Garreau and Arlot, 2018] with a Gaussian kernel, yielding the sequence $(150, 497, 820, 1091)$. As expected, when the pre- and post-change distributions around a changepoint are more distinct, KCPD detects the changepoint more accurately, whereas the estimate is offset by a small margin when the adjacent distributions are more similar.

We then apply CONCH-SEG, wrapping the CONCH framework around KCPD and using the parametric CPP score specialized to the Gaussian family with known variance (see (4.7)). The resulting confidence set is

$$[145, 152] \;\cup\; [483, 515] \;\cup\; [820, 821] \;\cup\; [1084, 1103].$$

Figure 7 displays the observed sequence (left) and the corresponding conformal p-values (right). The procedure sharply localizes all the changepoints. The sets around ξ_2 and ξ_4 are wider than the sets around ξ_1 and ξ_3 because the distributions on either side are more similar.

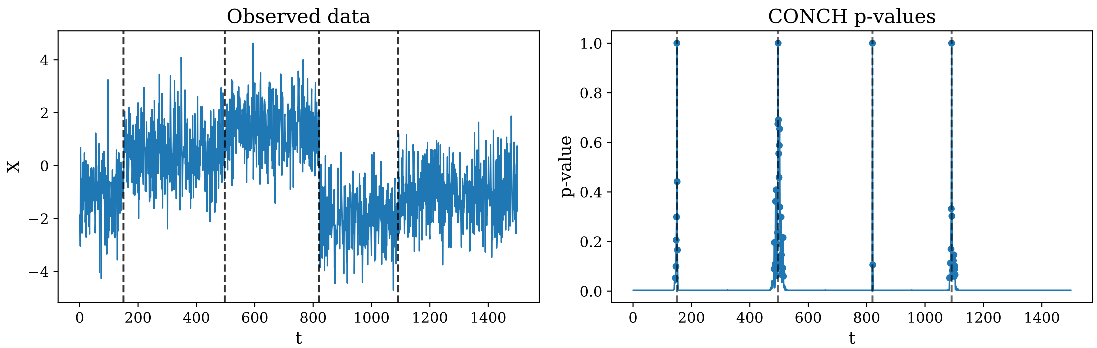

Figure 7: Left: observed sequence with multiple changepoints. Right: conformal p-values from CONCH-SEG using the parametric LLR score.

B.3 Two urns model: effect of dissimilarity between $\mathcal{P}_{0,\xi}$ and $\mathcal{P}_{1,\xi}$ on confidence set length

While we have demonstrated the performance of CONCH on a variety of changepoint detection tasks, our experiments so far have focused on i.i.d. settings, that is, changepoint models within $\mathcal{P}_{\mathrm{IID}}$. In what follows, we go beyond the i.i.d. assumption and show that the CONCH framework requires only exchangeability to produce valid confidence sets.

To illustrate this, we evaluate the performance of the CONCH confidence sets on a two-urn model with finite populations. Specifically, we consider two urns, each containing 2500 balls colored either red or blue. The proportions of red balls in the first and second urns are $0.5 - \delta$ and $0.5 + \delta$, respectively, for some $\delta \in (0, 0.5)$. We draw balls without replacement: the first $\xi = 350$ draws come from urn 1, and the remaining from urn 2, yielding a total of $n = 800$ observations. Our goal is to detect the changepoint ξ. We use the weighted mean difference as the CPP score, and for each $\delta \in \{0.05, 0.10, \ldots, 0.50\}$, we run the CONCH-MC algorithm (Algorithm 2) with $M = 300$ permutations to obtain confidence sets.

When δ is small, the pre-change and post-change distributions are nearly indistinguishable. Consequently, no method can sharply localize the changepoint, including CONCH confidence sets. As δ increases, the two distributions become more distinct. In the extreme case $\delta = 1$, the first urn contains only blue balls and the second only red balls, allowing perfect localization with absolute confidence. Accordingly, the average length of the CONCH confidence sets decreases with δ, as shown in the right panel of Figure 8, where the shaded region denotes one standard error around the mean. Across the whole collection of δ values, the true change-point $\xi = 350$ lies within the reported confidence set, demonstrating the validity of our procedure (left panel of Figure 8).

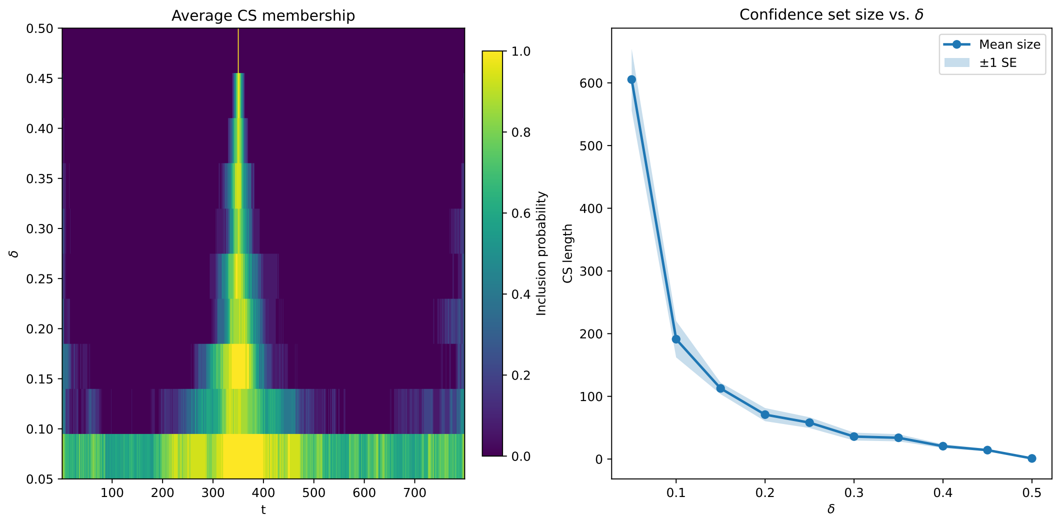

Figure 8: Two-urn changepoint experiment: CONCH confidence sets across δ values. Left: confidence sets always contain $\xi = 350$; right: average confidence set length decreases as dissimilarity δ increases.

B.4 MNIST: detect change in digits

We conduct a simulation based on the MNIST handwritten digits dataset [Deng, 2012] to evaluate the performance of CONCH for a digit shift localization. In particular, suppose we observe a sequence of $1,000$ images: the first $\xi = 400$ observations consist of i.i.d. samples of the digit "1", and the latter observations are i.i.d. samples of the digit "7" (Figure 9).

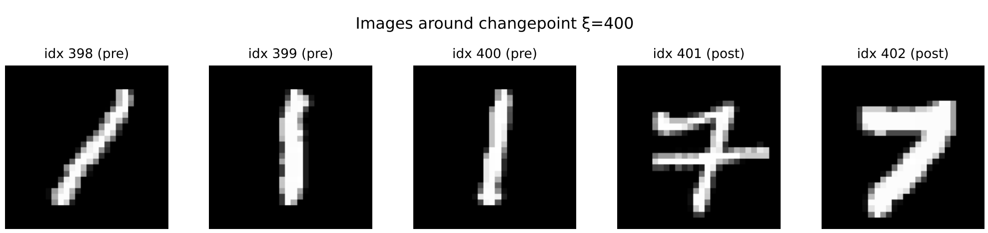

Figure 9: Illustration of MNIST changepoint setup: digit class shifts from '1' to '7' at $\xi = 400$ ($n = 1000$).

As in our main experiments, we use a classifier based log-likelihood ratio as CPP score in our CONCH algorithm. Specifically, we employ a pretrained convolutional neural network classifier to distinguish between the two digits; its logits define the CPP score, which is then passed to CONCH to produce a confidence interval for the changepoint.

Although the handwritten digits "1" and "7" often exhibit substantial visual similarity, our approach accurately detects the changepoint, yielding an exceptionally narrow, in fact singleton confidence set $\{400\}$ (Figure 10). We remark that the sharp localization here is partially a consequence of the strong classifier, which can confidently distinguish between the two digits. In the next section, we investigate how classifier strength influences the width of CONCH confidence sets on the CIFAR-100 dataset.

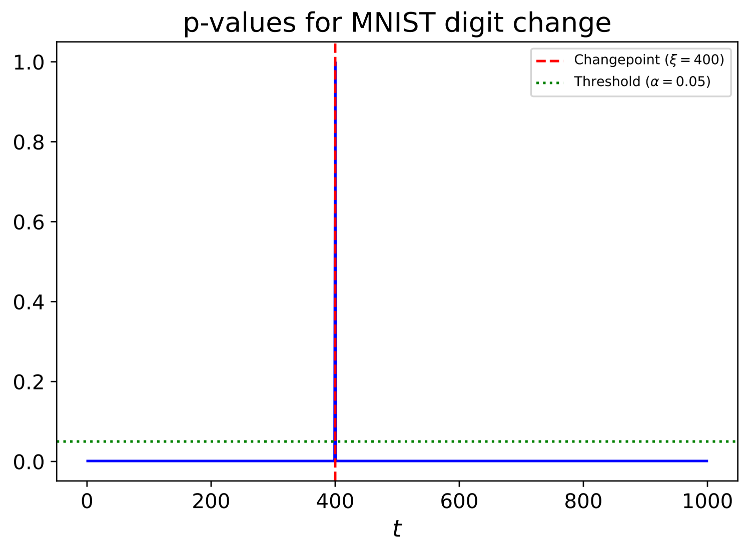

Figure 10: p-values for digit shift detection in MNIST: from digit '1' to digit '7' at $\xi = 400$

B.5 CIFAR100: classifier strength affects power of CONCH

We simulate a class-shift scenario using the CIFAR-100 image dataset [Alex, 2009] to evaluate CONCH under a challenging setting. Specifically, we construct a sequence of $n = 1,000$ observations with a changepoint at $\xi = 400$: the pre-change distribution $\mathbb{P}_{0,\xi}$ consists of i.i.d. images of bears, while the post-change distribution $\mathbb{P}_{1,\xi}$ consists of i.i.d. images of beavers (Figure 11). Because bears and beavers share many visual attributes, accurately localizing the changepoint is a non-trivial task.

Figure 11: Illustration of CIFAR-100 changepoint setup: sequence shifts from bear images to beaver images at $\xi = 400$ ($n = 1000$).

We pre-train a small three-block convolutional network with a lightweight classification head. We first train this network for 5 epochs to obtain a weak classifier and then train it further for an additional 20 epochs to obtain a stronger classifier. The resulting logits from each model define a CPP score, which we pass to CONCH to produce a changepoint confidence interval.

Figure 12 reports the p-value distributions and confidence sets produced by CONCH. As anticipated, the stronger classifier yields sharper separation between the two classes, leading to a much narrower confidence set $[398, 408]$ compared to the weaker model's wider interval $[393, 427]$. This experiment highlights both the sensitivity of CONCH to classifier quality and its ability to localize changepoints even under subtle visual differences between classes.

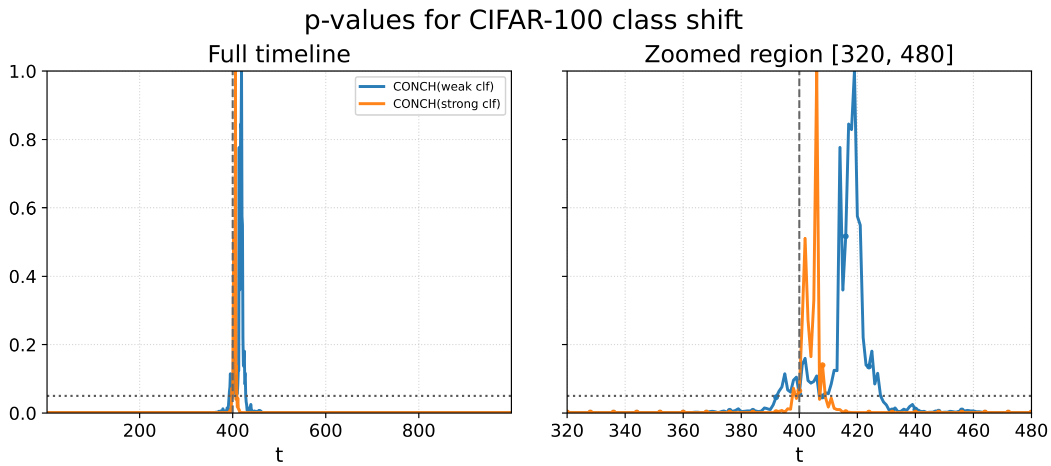

Figure 12: CONCH p-values for CIFAR-100 class shift (bear $\rightarrow$ beaver): weak vs. strong classifiers over the full timeline (left) and a zoomed window around $\xi = 400$ (right).