# Ch 13. Conditional Independence Testing

Sungwoo Park

Uncertainty Quantification Lab
Seoul National University

January 12th, 2026

# Table of contents

# Contents

# Conditional Independence Testing: Overview

## Problem Setting

Given i.i.d. samples $(X_i, Y_i, W_i)_{i\in[n]}$, test the null hypothesis:

$$H_0 : X \perp\!\!\!\perp Y \mid W$$

where $W$ is a **confounder** variable.

- Is the feature $X$ associated with response $Y$ after accounting for $W$?
- Related to **variable selection** in high-dimensional regression
  - Testing $H_{0,j} : X_j \perp\!\!\!\perp Y \mid X_{-j}$ for feature selection

# Marginal vs. Conditional Independence

## Marginal Independence

$$H_0 : X \perp\!\!\!\perp Y$$

- No conditioning on any variable
- Does not account for confounding effects of $W$
- **Easy** to test via permutation tests : *Theorem 13.1*

# Marginal vs. Conditional Independence

## Conditional Independence

$$H_0 : X \perp\!\!\!\perp Y \mid W$$

- Must account for the confounder $W$
- **Easy** when $W$ is discrete (local permutation test) : *Theorem 13.2*
- **Hard** (impossible) when $W$ is continuous (hardness result) : *Theorem 13.3*

# Q&A: How is the Confounder $W$ Determined?

## Question

How is the confounder $W$ typically determined in practice? It seems like finding an appropriate $W$ would be important, and trivial confounders might also exist.

## Answer

- This chapter **assumes** $W$ **is given** — identifying $W$ is a separate problem
- In practice, $W$ selection is crucial:
  - **Domain knowledge**: Prior understanding of causal relationships
  - **Variable selection**: In high-dimensional settings, $W = X_{-j}$ (all other variables)

# Q&A: Why is Chapter 13 in This Book?

**Question**

Chapter 13 does not seem directly related to conformal prediction. Why is it included in this book?

**Answer**

- This chapter demonstrates the broader applications of **permutation tests** and **exchangeability**
- The book's core techniques extend beyond conformal prediction:
  - Chapter 2's permutation test framework
  - Chapter 4's test-conditional coverage hardness results
  - Chapter 11-12's binning approaches
- Emphasizes the **unifying theme** of "distribution-free inference" and shows how exchangeability-based arguments apply to hypothesis testing

# Contents

# Setup: Marginal Independence Testing

## Problem Setting

- Data: $(X_1, Y_1), \ldots, (X_n, Y_n) \stackrel{\text{i.i.d.}}{\sim} P$
- Null hypothesis: $H_0 : X \perp\!\!\!\perp Y$
- Equivalently: $P \in \mathcal{P}_{X \perp\!\!\!\perp Y}$ (product distributions)

## Test Statistic

Any function $T : (\mathcal{X} \times \mathcal{Y})^n \to \mathbb{R}$ measuring evidence against $H_0$.
**Example (real-valued data):**

$$T\big((X_1, Y_1), \ldots, (X_n, Y_n)\big) = \big|\mathrm{Corr}\big((X_1, \ldots, X_n), (Y_1, \ldots, Y_n)\big)\big|$$

# Permutation Test for Marginal Independence

## Key Insight

Under $H_0 : X \perp\!\!\!\perp Y$, permuting $X_i$'s while fixing $Y_i$'s doesn't change the distribution!

$$(X_\sigma, \mathbf{Y}) \overset{d}{=} (\mathbf{X}, \mathbf{Y}) \quad \text{for all } \sigma \in \mathcal{S}_n$$

- $\mathbf{X} = (X_1, \ldots, X_n)$, $\mathbf{Y} = (Y_1, \ldots, Y_n)$
- $X_\sigma = (X_{\sigma(1)}, \ldots, X_{\sigma(n)})$ for $\sigma \in \mathcal{S}_n$

# Theorem 13.1: Validity of the Permutation Test

## Theorem 13.1

Fix any test statistic $T : (\mathcal{X} \times \mathcal{Y})^n \to \mathbb{R}$. Define the test:

$$\psi(\mathbf{X}, \mathbf{Y}) = \mathbb{1}\{\frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} \mathbb{1}\{T(X_\sigma, \mathbf{Y}) \geq T(\mathbf{X}, \mathbf{Y})\} \leq \alpha\}$$

Then $\psi$ is a **valid test** of $H_{X \perp\!\!\!\perp Y}$:

$$\mathbb{P}_P(\psi(\mathbf{X}, \mathbf{Y}) = 1) \leq \alpha \quad \text{for all } P \in \mathcal{P}_{X \perp\!\!\!\perp Y}$$

## Proof Idea

- Under $H_0$: $P = P_X \times P_Y$
- Conditioning on $\mathbf{Y}$: $X_1, \ldots, X_n$ remain i.i.d. $\sim P_X \Rightarrow$ Checking $X_i$'s exchangeability.

# Recall: Permutation test for exchangeability

**Recall: Theorem 2.4**

For any function $T : \mathcal{Z}^n \to \mathbb{R}$, the p-value $p$ defined as:

$$p := \frac{\sum_{\sigma \in S_n} \mathbb{1}\{T(\mathbf{X}_\sigma) \geq T(\mathbf{X})\}}{n!}$$

satisfies $\mathbb{E}_P[p \leq \tau] \leq \tau$ for all $\tau \in [0,1]$ and $P \in \mathcal{P}_{\text{exch}}$.

Plug-in $T \leftarrow T(\mathbf{X}, \mathbf{Y})$ into Theorem 2.4, conclued the proof.

# Q&A: Alternative Test Statistics

## Question

What are some alternative test statistics $T$ that can be used instead of the absolute Pearson correlation for testing marginal/conditional independence?

## Answer: Alternative Test Statistics

By Theorem 2.4, we can use any test statistic $T$ to construct a valid permutation test. Here are some examples:

1. **Spearman's rank correlation** [7]: $|\mathrm{Corr}(\mathrm{rank}(X), \mathrm{rank}(Y))|$
   - rank is similar to order statistics' index
   - Robust to outliers, captures monotonic relationships

# Q&A: Alternative Test Statistics

## Answer: Alternative Test Statistics

2. **Distance correlation** [8]
   - Using distance covariance $\mathcal{V}^2(X, Y) := ||f_{X,Y}(t, s) - f_X(t)f_Y(s)||^2$
   - Detects nonlinear dependencies, equals zero iff independent
3. **HSIC** (Hilbert-Schmidt Independence Criterion) [5]
   - $HSIC(p_{xy}, \mathcal{F}, \mathcal{G}) := ||C_{xy}||^2_{HS} = \mathbb{E}_{x,x',y,y'}\left[k(x, x')\, l(y, y')\right] + \mathbb{E}_{x,x'}\left[k(x, x')\right]\mathbb{E}_{y,y'}\left[l(y, y')\right] - 2\mathbb{E}_{x,y}\left[\mathbb{E}_{x'}\left[k(x, x')\right]\mathbb{E}_{y'}\left[l(y, y')\right]\right]$
   - RKHS-based measure of dependence

# Contents

# Local Permutation Test: Motivation

## Key Idea

When $W$ is discrete, group data points by their $W$ value:

- Within each group (same $W = w$): $X$ and $Y$ are **independent**
- Permute $X$ values **only within groups**
- This preserves the null distribution!

## Intuition

- Think of $W$ as defining "subpopulations"
- Under $H_0 : X \perp\!\!\!\perp Y \mid W$, within each subpopulation $X$ and $Y$ are independent
- Same logic as marginal independence test, applied locally

# Local Permutation Test: Permutation

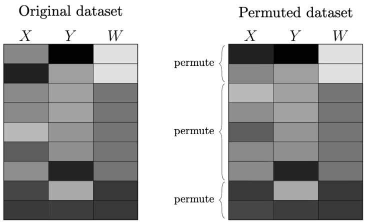


Figure 1: Visualization of a dataset with a discrete confounder along with a local permutation of the covariate [1]

## Allowed Permutations

Define the set of permutations that **preserve** $W$ **values**:

$$\mathcal{S}_n(\mathbf{W}) = \{\sigma \in \mathcal{S}_n : W_{\sigma(i)} = W_i \text{ for all } i \in [n]\}$$

# Local Permutation Test: Definition

**Test Definition (Equation 13.4)**

$$\psi(\mathbf{X}, \mathbf{Y}, \mathbf{W}) = \mathbb{1}\{\frac{1}{|\mathcal{S}_n(\mathbf{W})|} \sum_{\sigma \in \mathcal{S}_n(\mathbf{W})} \mathbb{1}\{T(X_\sigma, \mathbf{Y}, \mathbf{W}) \geq T(\mathbf{X}, \mathbf{Y}, \mathbf{W})\} \leq \alpha\}$$

- Only permute $X$ values among data points with *same* $W$ value
- This is called a “**local**” permutation test

# Theorem 13.2: Validity of Local Permutation Test

**Theorem 13.2**

Fix any test statistic $T : (\mathcal{X} \times \mathcal{Y} \times \mathcal{W})^n \to \mathbb{R}$. Let $(X_i, Y_i, W_i)_{i\in[n]} \overset{\text{i.i.d.}}{\sim} P$. Then $\psi$ is a **valid test** of $H_{X \perp\!\!\!\perp Y|W}$:

$$\mathbb{P}_P(\psi(\mathbf{X}, \mathbf{Y}, \mathbf{W}) = 1) \leq \alpha \quad \text{for all } P \in \mathcal{P}_{X \perp\!\!\!\perp Y|W}$$

- **Distribution-free**: No assumptions on $P$ (beyond conditional independence)
- **Test statistic-free**: Works for any choice of $T$
- Proof uses same techniques as Chapter 2 (permutation test validity)

# When is the Local Permutation Test Powerful?

## Factor 1: Number of "Clashes"

- **Clashes**: pairs $i \neq j$ with $W_i = W_j$
- Many clashes $\Rightarrow$ rich set of permutations $\Rightarrow$ potential for power
- No clashes (all $W_i$ unique) $\Rightarrow |\mathcal{S}_n(\mathbf{W})| = 1 \Rightarrow$ **no power**

## Factor 2: Choice of Test Statistic $T$

- $T$ must discriminate between null and alternative
- Standard considerations for hypothesis testing

## Critical Observation

If $W$ is **continuous**, all $W_i$'s are distinct a.s. $\Rightarrow$ **No power!**

# Proof Sketch: Theorem 13.2

1. Conditional on **W** and **Y**
   Under $H_0 : X \perp\!\!\!\perp Y \mid W$, conditioning on $(\mathbf{W}, \mathbf{Y})$:

   $$X_1, \ldots, X_n \text{ are independent (but not identically distributed)}$$

   with $X_i \sim P_{X|W}(\cdot|W_i)$.
2. Exchangeability Within Groups
   For any $\sigma \in \mathcal{S}_n(\mathbf{W})$: since $W_{\sigma(i)} = W_i$,

   $$(X_\sigma, \mathbf{Y}) \mid \mathbf{W} \stackrel{d}{=} (\mathbf{X}, \mathbf{Y}) \mid \mathbf{W} \cdots (13.6)$$

# Proof Sketch: Theorem 13.2

3. Using quantile technique
   We can check that $\mathcal{S}_n(\mathbf{w}) = \{\sigma \circ \sigma' : \sigma' \in \mathcal{S}_n(\mathbf{w})\}$ for any $\sigma \in \mathcal{S}_n(\mathbf{w})$. So,

   $$\hat{q}(\mathbf{x}_\sigma, \mathbf{y}, \mathbf{w}) = \text{Quantile}(T((\mathbf{x}_\sigma)_{\sigma'}, \mathbf{y}, \mathbf{w})_{\sigma' \in \mathcal{S}_n(\mathbf{w})}; 1-\alpha) = \hat{q}(\mathbf{x}, \mathbf{y}, \mathbf{w})$$

   Therefore,

   $$\begin{aligned} &\frac{1}{|S_n(\mathbf{w})|} \sum_{\sigma \in S_n(\mathbf{w})} \mathbf{1}\{\psi(x_\sigma, y, \mathbf{w}) = 1\} \\ &= \frac{1}{|S_n(\mathbf{w})|} \sum_{\sigma \in S_n(\mathbf{w})} \mathbf{1}\{T(x_\sigma, y, \mathbf{w}) > \hat{q}(x, y, \mathbf{w})\} \\ &= \frac{1}{|S_n(\mathbf{w})|} \sum_{\sigma \in S_n(\mathbf{w})} \mathbf{1}\Big\{T(x_\sigma, y, \mathbf{w}) > \text{Quantile}\big((T(x_{\sigma'}, y, \mathbf{w}))_{\sigma' \in S_n(\mathbf{w})}; 1-\alpha\big) \\ &\leq \alpha \end{aligned}$$

# Proof Sketch: Theorem 13.2

4. (13.6) implies that

$$\mathbb{P}_P\big(\psi(\mathbf{X},\mathbf{Y},\mathbf{W}) = 1 \mid \mathbf{W}\big) = \mathbb{P}_P\big(\psi(\mathbf{X}_\sigma,\mathbf{Y},\mathbf{W}) = 1 \mid \mathbf{W}\big)$$

holds for each $\sigma \in \mathcal{S}_n(\mathbf{W})$. Taking an average over $\sigma \in \mathcal{S}_n(\mathbf{W})$,

$$\begin{aligned}
&\mathbb{P}_P\big(\psi(\mathbf{X},\mathbf{Y},\mathbf{W}) = 1 \mid \mathbf{W}\big) \\
&= \frac{1}{|\mathcal{S}_n(\mathbf{W})|} \sum_{\sigma \in \mathcal{S}_n(\mathbf{W})} \mathbb{P}_P\big(\psi(X_\sigma,\mathbf{Y},\mathbf{W}) = 1 \mid \mathbf{W}\big) \\
&= \mathbb{E}_{\mathbb{P}}\left[\frac{1}{|\mathcal{S}_n(\mathbf{W})|} \sum_{\sigma \in \mathcal{S}_n(\mathbf{W})} \mathbf{1}\{\psi(X_\sigma,\mathbf{Y},\mathbf{W}) = 1\} \,\middle|\, \mathbf{W}\right] \le \alpha
\end{aligned}$$

# Contents

# The Problem with Continuous $W$

When $W$ has a **nonatomic** (continuous) distribution:
$\mathbb{P}(W_i = W_j) = 0$ for $i \neq j$

- No two data points share the same $W$ value
- Local permutation test has no power

# Theorem 13.3: Hardness Result

## Theorem 13.3

If the test $\psi$ has Type I error bounded by $\alpha$ for any distribution:

$$\mathbb{P}_P(\psi(\mathbf{X}, \mathbf{Y}, \mathbf{W}) = 1) \leq \alpha \quad \text{for all } P \in \mathcal{P}_{X \perp\!\!\!\perp Y | W}$$

Then the **power** of the test is not better than random for any $P$ with $P_W$ nonatomic:

$\mathbb{P}_P(\psi(\mathbf{X}, \mathbf{Y}, \mathbf{W}) = 1) \leq \alpha$ for **all** $P \in \mathcal{P}_{(X,Y,W)}$ with $P_W$ nonatomic

## Interpretation

For continuous $W$: **any** valid test has power $\leq \alpha$ (no better than random guessing!)

# Proof Sketch: Theorem 13.3

**Key Technique: Sample-Resample (Lemma 4.15)**

Same technique as hardness results in Chapters 4, 11, 12!

1. Fix arbitrary sequence $(X^{(i)}, Y^{(i)}, W^{(i)})_{i\in[M]}$ with distinct $W^{(i)}$
2. Empirical distribution $\hat{P}_M = \frac{1}{M}\sum_{i=1}^{M} \delta_{(X^{(i)},Y^{(i)},W^{(i)})}$
3. Under $\hat{P}_M$: $X, Y$ are deterministic functions of $W \Rightarrow$ Trivially $X \perp\!\!\!\perp Y \mid W$ under $\hat{P}_M$!
4. Valid test must have $\mathbb{P}_{\hat{P}_M}(\psi = 1) \leq \alpha$
5. By Sample-Resample Method: this extends to any $P$ with nonatomic $P_W$

# Sample-Resample Method

**Recall: Lemma 4.15**

Let $P$ be a distribution on $\mathcal{Z}$, and let $m, M \geq 1$. Let $Q$ denote the distribution on $\mathcal{Z}^m$ obtained by the following process to generate $(Z_1, \cdots, Z_m)$:

1. Sample $(Z^{(1)}, \cdots, Z^{(M)})$ i.i.d. from $P$, and define the empirical distribution $\hat{P}_M = \frac{1}{M}\sum_{i=1}^{M} \delta_{Z^{(i)}}$
2. Resample $(Z_1, \cdots, Z_m)$ i.i.d. from $\hat{P}_M$

Then

$$d_{TV}(P^m, Q) \leq \frac{m(m-1)}{2M}$$

# Proof (5) Sketch: Theorem 13.3

Construct $\mathbb{P}_{\hat{P}_M}$ by sampling the M data points i.i.d. from $P$ as Lemma 4.15, Sample $(X_1, Y_1, W_1), \cdots, (X_M, Y_M, W_M)$ i.i.d. from $\hat{P}_M$. So,

$$\mathbb{P}_Q(\psi = 1) \leq \alpha \text{ a.s.}$$

We also konw that $W^{(1)}, \cdots, W^{(M)}$ are distinct a.s.,

$$\mathbb{P}_{\hat{P}_M}(\psi = 1 | \hat{P}_M) \leq \alpha \text{ a.s.}$$

Using both inequality and Lemma 4.15,

$$\mathbb{P}_P(\psi = 1) \leq \alpha + \frac{n(n-1)}{2M}$$

and taking $M$ large to get desired result.

# Q&A: Alternative Approaches Without Permutation Tests

**Question**

Are there methods to test conditional independence without using permutation tests?

**Answer: Alternative Approaches (from Bibliographic Notes)**

1. **Conditional Chatterjee's correlation** [2]
   - Asymptotically valid, no smoothness assumptions needed
2. **Model-X framework** [4]
   - Assumes knowledge of $P_{X|W}$
   - Resample $X$ under the null using known conditional distribution

# Q&A: Alternative Approaches Without Permutation Tests

**Question**

Are there methods to test conditional independence without using permutation tests?

**Answer: Alternative Approaches (from Bibliographic Notes)**

3. **Knockoff filter** [3]
   - Uses "pairwise exchangeability" for FDR control
4. **Generalized Covariance Measure** [6]
   - Uses residuals from regression of $X$ on $W$ and $Y$ on $W$

# Contents

# Escaping the Hardness Result

## The Dilemma

- Distribution-free: Valid for *all* $P \in \mathcal{P}_{X \perp\!\!\!\perp Y | W}$
- But: No power for continuous $W$!

## Solution: Restrict the Null Hypothesis

- Add a **smoothness assumption** on the conditional distribution
- Control Type I error against a *smaller* class of nulls
- Allows nontrivial power while maintaining (approximate) validity

# Smoothness Assumption

## Definition: Hellinger Distance

For distributions $P_1, P_2$ with densities $f_1, f_2$:

$$d_{\mathrm{H}}(P_1, P_2) := \left( \frac{1}{2} \int_{\mathcal{X}} \left( \sqrt{f_1(x)} - \sqrt{f_2(x)} \right)^2 d\mu(x) \right)^{1/2}$$

## Restricted Null

$$\mathcal{P}^L_{X \perp\!\!\!\perp Y|W} = \{P \in \mathcal{P}_{X \perp\!\!\!\perp Y|W} : L\text{-Lipschitz w.r.t. Hellinger distance}\}$$

## Lipschitz Condition (Equation 13.7)

Assume the map $w \mapsto P_{X|W}(\cdot|w)$ is $L$-Lipschitz w.r.t. Hellinger distance:

$$d_{\mathrm{H}}(P_{X|W}(\cdot|w), P_{X|W}(\cdot|w')) \leq L\|w - w'\|$$

# Binned Local Permutation Test

## Key Idea

- Partition $\mathcal{W}$ into bins: $\mathcal{W} = \cup_k \mathcal{W}_k$
- Bin diameter bounded: $\max_k \sup_{w,w' \in \mathcal{W}_k} \|w - w'\| \leq h$

## Allowed Permutations

$\mathcal{S}_n^{\text{bin}}(\mathbf{W}) = \{\sigma \in \mathcal{S}_n : W_{\sigma(i)} \text{ and } W_i \text{ are in the same bin for all } i\}$

## Test Definition (Equation 13.8)

$$\psi(\mathbf{X}, \mathbf{Y}, \mathbf{W})$$
$$:= \mathbb{1}\{\frac{1}{|\mathcal{S}_n^{\text{bin}}(\mathbf{W})|} \sum_{\sigma \in \mathcal{S}_n^{\text{bin}}(\mathbf{W})} \mathbb{1}\{T(X_\sigma, \mathbf{Y}, \mathbf{W}) \geq T(\mathbf{X}, \mathbf{Y}, \mathbf{W})\} \leq \alpha\}$$

# Theorem 13.4: Validity of Binned Test

**Theorem 13.4**

Fix any $T : (\mathcal{X} \times \mathcal{Y} \times \mathcal{W})^n \to \mathbb{R}$. Let $(X_i, Y_i, W_i)_{i \in [n]} \overset{\text{i.i.d.}}{\sim} P$.
For any partition with $\max_k \sup_{w,w' \in \mathcal{W}_k} \|w - w'\| \leq h$:

$$\mathbb{P}_P(\psi(\mathbf{X}, \mathbf{Y}, \mathbf{W}) = 1) \leq \alpha + Lh\sqrt{2n} \quad \text{for all } P \in \mathcal{P}^L_{X \perp\!\!\!\perp Y | W}$$

- Type I error: $\alpha + Lh\sqrt{2n}$ (not exactly $\alpha$)
- Choose bin size $h = o(n^{-1/2}) \Rightarrow$ nearly valid test
- Trade-off: smaller $h$ = more valid, but fewer permutations = less power

# Proof Sketch: Theorem 13.4

1. Construct Auxiliary Distribution $\tilde{P}$
   - Express $P$ as mixture: $P = \sum_k \pi_k \cdot P^{(k)}$ where $\pi_k = \mathbb{P}(W \in \mathcal{W}_k)$
   - Define $\tilde{P} = \sum_k \pi_k \cdot (P_X^{(k)} \times P_{(Y,W)}^{(k)})$
   - Under $\tilde{P}$: $X \perp\!\!\!\perp (Y, W) \mid K$ where $K = k(W)$ (bin index)
2. $\tilde{P}$ Satisfies Conditional Independence
   - By construction: $X \perp\!\!\!\perp (Y, W) \mid K$ under $\tilde{P}$
   - Local permutation test with discrete confounder $K$ is **exactly valid**
   - $\Rightarrow \mathbb{P}_{\tilde{P}}(\psi = 1) \leq \alpha$

# Proof Sketch: Theorem 13.4

3. Bound Distance Between $P$ and $\tilde{P}$
   Using smoothness assumption and Hellinger distance properties:

$$d_{\mathrm{H}}(P_{X|W}(\cdot|w), P_X^{(k)}) \leq Lh \quad \text{for all } w \in \mathcal{W}_k$$

   This gives:

$$d_{\mathrm{H}}^2(P, \tilde{P}) = \mathbb{E}_P\left[d_H^2(P_{X|W}(\cdot|W), P_X^{(k(W))})\right] \leq (Lh)^2$$

4. Extend to $n$ Samples
   - Subadditivity of squared Hellinger distance: $d_{\mathrm{H}}^2(P^n, \tilde{P}^n) \leq n(Lh)^2$
   - Total variation bound: $d_{\mathrm{TV}}(P^n, \tilde{P}^n) \leq \sqrt{2} d_{\mathrm{H}}(P^n, \tilde{P}^n) \leq Lh\sqrt{2n}$

   Therefore:

$$\mathbb{P}_P(\psi = 1) \leq \mathbb{P}_{\tilde{P}}(\psi = 1) + Lh\sqrt{2n} \leq \alpha + Lh\sqrt{2n}$$

# Q&A: Hellinger Distance Convexity

## Question

In the proof of Theorem 13.4, it mentions that $d_{\mathrm{H}}(\cdot,\cdot)$ is convex in each argument. I believe $d^2_{\mathrm{H}}(\cdot,\cdot)$ is convex, but $d_{\mathrm{H}}(\cdot,\cdot)$ itself is not. Is my understanding correct?

## Answer

- $d^2_{\mathrm{H}}(P,Q)$ is **convex** in each argument (jointly convex)
- $d_{\mathrm{H}}(P,Q)$ is generally **not convex** but **concave** in each argument
- What the proof actually uses: $P_X^{(k)}$ is convex conbination of $P_{X|W}(\cdot|w)$

# Q&A: Hellinger Distance Convexity

## Answer

$$
\begin{aligned}
d_{\mathrm{H}}(P_{X|W}(\cdot|w), P_X^{(k)}) &= d_{\mathrm{H}}(P_{X|W}(\cdot|w), \sum_{w' \in \mathcal{W}_k} \lambda_{w'} P_{X|W}(\cdot|w')) \\
&\leq \sum_{w' \in \mathcal{W}_k} \lambda_{w'} \cdot d_{\mathrm{H}}(P_{X|W}(\cdot|w), P_{X|W}(\cdot|w')) \\
&\leq Lh
\end{aligned}
$$

# Q&A: Kernel Blurring for Continuous $W$?

**Question**

The local permutation test faces issues when $W$ is continuous, similar to Chapter 11. Can kernel-based blurring techniques from Chapter 11 be applied here?

Under smoothness assumption, using Hellinger distance is similar to bulrring.

- Blurring : Making target $\mu_P(x)$ as smoothed/blurred version (weighted expectation):

$$\hat{\mu}_P(x) = \frac{\mathbb{E}\left[\mu_P(X) \cdot H(x, X)\right]}{\mathbb{E}\left[H(x, X)\right]}$$

- Hellinger distance : Making conditional distribution $P_{X|W}(\cdot|w)$ as Binned weighted Expectation distribution:

$$P_X^{(k)} = \int_{\mathcal{W}_k} P_{X|W}(\cdot|w')d\mu_k(w')$$

# Contents

# Connections to Other Chapters

| Topic | Discrete | Continuous |
|---|---|---|
| Test-conditional coverage (Ch 4) | Possible | Hard |
| Distribution-free regression (Ch 11) | Possible | Hard |
| ECE calibration (Ch 12) | Estimable | Not estimable |
| Conditional indep. testing (Ch 13) | Valid test | No power |

## Unifying Theme

- **Discrete setting**: Repeated observations enable inference
- **Continuous setting**: Fundamental impossibility results
- **Smoothness assumptions**: Bridge between the two
- **Binning**: Convert continuous to approximately discrete

# Thank you!

# References I


[1] Anastasios N Angelopoulos, Rina Foygel Barber, and Stephen Bates.
Theoretical foundations of conformal prediction.
2024.

[2] Mona Azadkia and Sourav Chatterjee.
A simple measure of conditional dependence.
*The Annals of Statistics*, 49(6):3070–3102, 2021.

[3] Rina Foygel Barber and Emmanuel J Candes.
Controlling the false discovery rate via knockoffs.
*The Annals of Statistics*, 43(5):2055–2085, 2015.

# References II

[4] Emmanuel Candes, Yingying Fan, Lucas Janson, and Jinchi Lv.
Panning for gold: Model-x knockoffs for high-dimensional controlled variable selection.
*Journal of the Royal Statistical Society: Series B*, 80(3):551–577, 2018.

[5] Arthur Gretton, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf.
Measuring statistical dependence with Hilbert-Schmidt norms.
In *Algorithmic Learning Theory (ALT 2005)*, volume 3734 of *Lecture Notes in Computer Science*, pages 63–77, Berlin, Heidelberg, 2005. Springer.

# References III

[6] Rajen D Shah and Jonas Peters.
The hardness of conditional independence testing and the generalised covariance measure.
*The Annals of Statistics*, 48(3):1514–1538, 2020.

[7] Charles Spearman.
The proof and measurement of association between two things.
*The American Journal of Psychology*, 15(1):72–101, 1904.

[8] Gábor J. Székely, Maria L. Rizzo, and Nail K. Bakirov.
Measuring and testing dependence by correlation of distances.
*The Annals of Statistics*, 35(6):2769–2794, 2007.