Ch 13. Conditional Independence Testing

Sungwoo Park

Uncertainty Quantification Lab
Seoul National University

January 13th, 2026

Table of contents

Contents

Sungwoo Park

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

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)
- **Hard** (impossible) when W is continuous (hardness result)

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

Sungwoo Park

Setup: Marginal Independence Testing

Problem Setting

- Data: $(X_1, Y_1), \ldots, (X_n, Y_n) \overset{\text{i.i.d.}}{\sim} P$
- 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):

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

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!

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

Notation

- $\mathbf{X} = (X_1, \ldots, X_n)$, $\mathbf{Y} = (Y_1, \ldots, Y_n)$
- $X_\sigma = (X_{\sigma(1)}, \ldots, X_{\sigma(n)})$ 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 ψ 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. $\sim P_X$
- Apply Theorem 2.4 (permutation test validity from Chapter 2)

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

1 **Spearman's rank correlation**: $|\text{Corr}(\text{rank}(X), \text{rank}(Y))|$
   - Robust to outliers, captures monotonic relationships
2 **Distance correlation** (Székely et al., 2007)
   - Detects nonlinear dependencies, equals zero iff independent
3 **HSIC** (Hilbert-Schmidt Independence Criterion)
   - Kernel-based measure of dependence
4 **Mutual information estimators**
   - Information-theoretic 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: Definition

Allowed Permutations

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

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

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 ψ is a **valid test** of $H_{X \perp\!\!\!\perp Y | W}$:

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

Key Points

- **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)

Proof Sketch: Theorem 13.2

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

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

   with $X_i \sim P_{X\mid W}(\cdot \mid W_i)$.

2. Exchangeability Within Groups
   For any $\sigma \in \mathcal{S}_n(\mathbf{W})$: since $W_{\sigma(i)} = W_i$,

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

3. Apply Permutation Test Logic
   Average over permutations $\sigma \in \mathcal{S}_n(\mathbf{W})$ to get valid p-value.

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!**

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)
  - **Causal discovery**: Learn DAG structure to identify confounders
- Wrong W choice $\Rightarrow$ spurious association or missed association

Contents

Sungwoo Park

The Problem with Continuous W

Continuous Confounder Challenge

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

Question

Can we design *any* valid test with nontrivial power for continuous W?

Answer: NO!

Theorem 13.3 shows this is **impossible** in a distribution-free setting.

Theorem 13.3: Hardness Result

Theorem 13.3

If the test ψ has Type I error bounded by α 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:

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

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

4. $\Rightarrow$ Trivially $X \perp\!\!\!\perp Y \mid W$ under $\hat{P}_M$!

5. Valid test must have $\mathbb{P}_{\hat{P}_M}(\psi = 1) \leq \alpha$

6. By Lemma 4.15: this extends to any P with nonatomic P_W

Connection to Previous Hardness Results

Recurring Theme in the Book

Discrete case: Distribution-free inference is possible (Chapter 4.3, 11.3, 13.2)

Continuous case: Distribution-free inference faces fundamental limits (Chapter 4.4, 11.4, 13.3)

Same Proof Technique

- Chapter 4: Test-conditional coverage
- Chapter 11: Distribution-free regression
- Chapter 12: ECE calibration
- Chapter 13: Conditional independence testing

All use the **sample-resample lemma** (Lemma 4.15)!

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** (Azadkia & Chatterjee, 2021)
   - Asymptotically valid, no smoothness assumptions needed
2. **Model-X framework** (Candès et al., 2018)
   - 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** (Barber & Candès, 2015)
   - Uses "pairwise exchangeability" for FDR control
4. **Generalized Covariance Measure** (Shah & Peters, 2020)
   - 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

Lipschitz Condition (Equation 13.7)

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

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

for all $w, w' \in \mathcal{W}$.

Smoothness Assumption

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

$$\mathcal{P}^L_{X \perp\!\!\!\perp Y | W} = \{ P \in \mathcal{P}_{X \perp\!\!\!\perp Y | W} : (13.7) \text{ holds with constant } L \}$$

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

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

Test Definition (Equation 13.8)

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

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 α)
- 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:

$$d_{\mathrm{H}}^2(P, \tilde{P}) \leq (Lh)^2$$

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:

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

Sungwoo Park

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 $d_{\mathrm{H}}^2(\cdot, \cdot)$ is convex, but $d_{\mathrm{H}}(\cdot, \cdot)$ itself is not. Is my understanding correct?

Answer

- $d_{\mathrm{H}}^2(P, Q)$ is **convex** in each argument (jointly convex)
- $d_{\mathrm{H}}(P, Q)$ is generally **not convex**
- What the proof actually uses (Equation 13.9):

$$d_{\mathrm{H}}(P_{X|W}(\cdot|w), P_X^{(k)}) \leq Lh$$

- This follows from $P_X^{(k)}$ being a convex combination, using d_{H}^2 convexity and Jensen's inequality
- The proof logic is correct; d_{H}^2 convexity is the key property

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?

Answer

- Chapter 11's blurring: Smooths the target Y to avoid hardness
- Chapter 13's binning: Discretizes W to enable testing

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?

Answer

- **Kernel-based approaches are possible**:
  - Weighted permutations among data points with similar W values
  - Kernel $K(w, w')$ determines permutation probabilities
  - See: Sen et al. (2017), Kim et al. (2022)
- Key trade-off: More permutations $\Leftrightarrow$ larger approximation error
- Both binning and kernel methods aim to approximate discrete case

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.