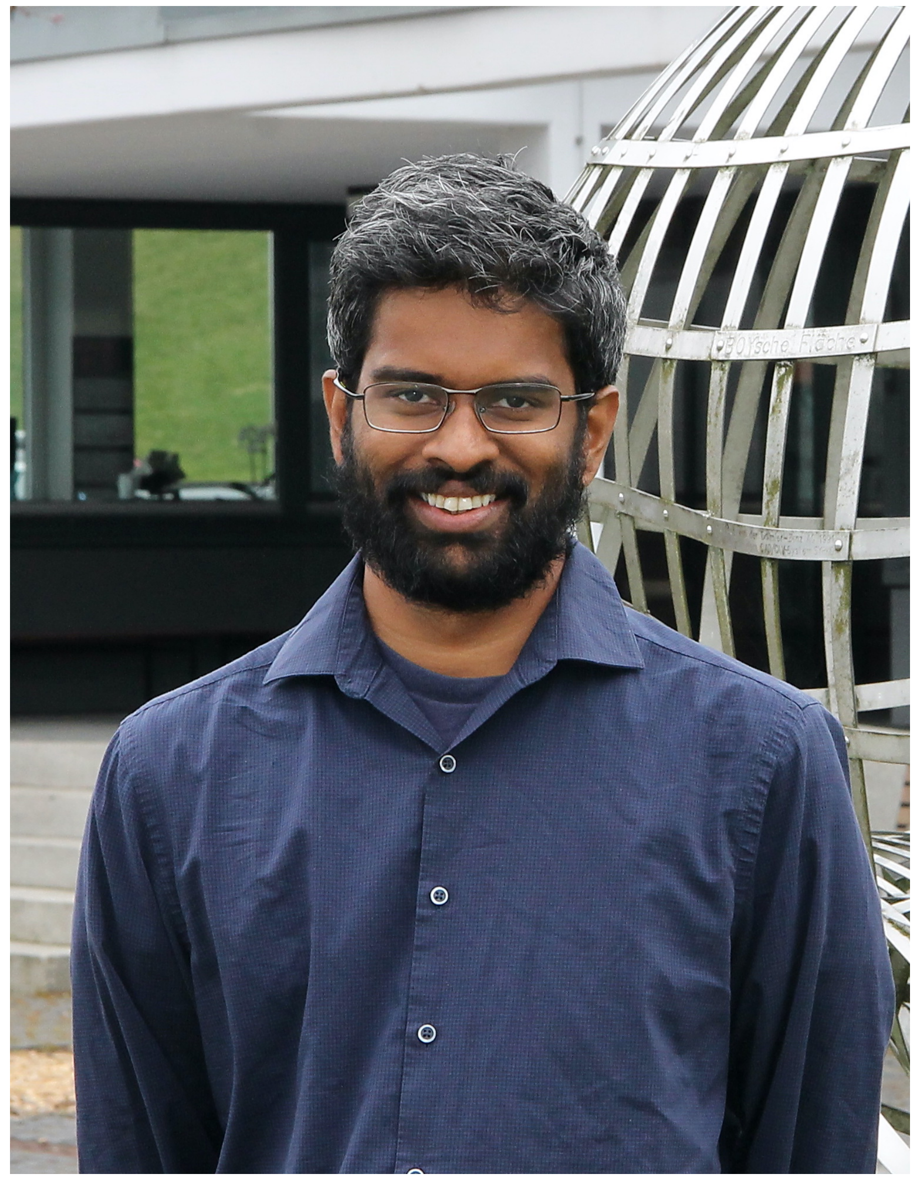
Local Permutation Tests for Conditional Independence

Ilmun Kim

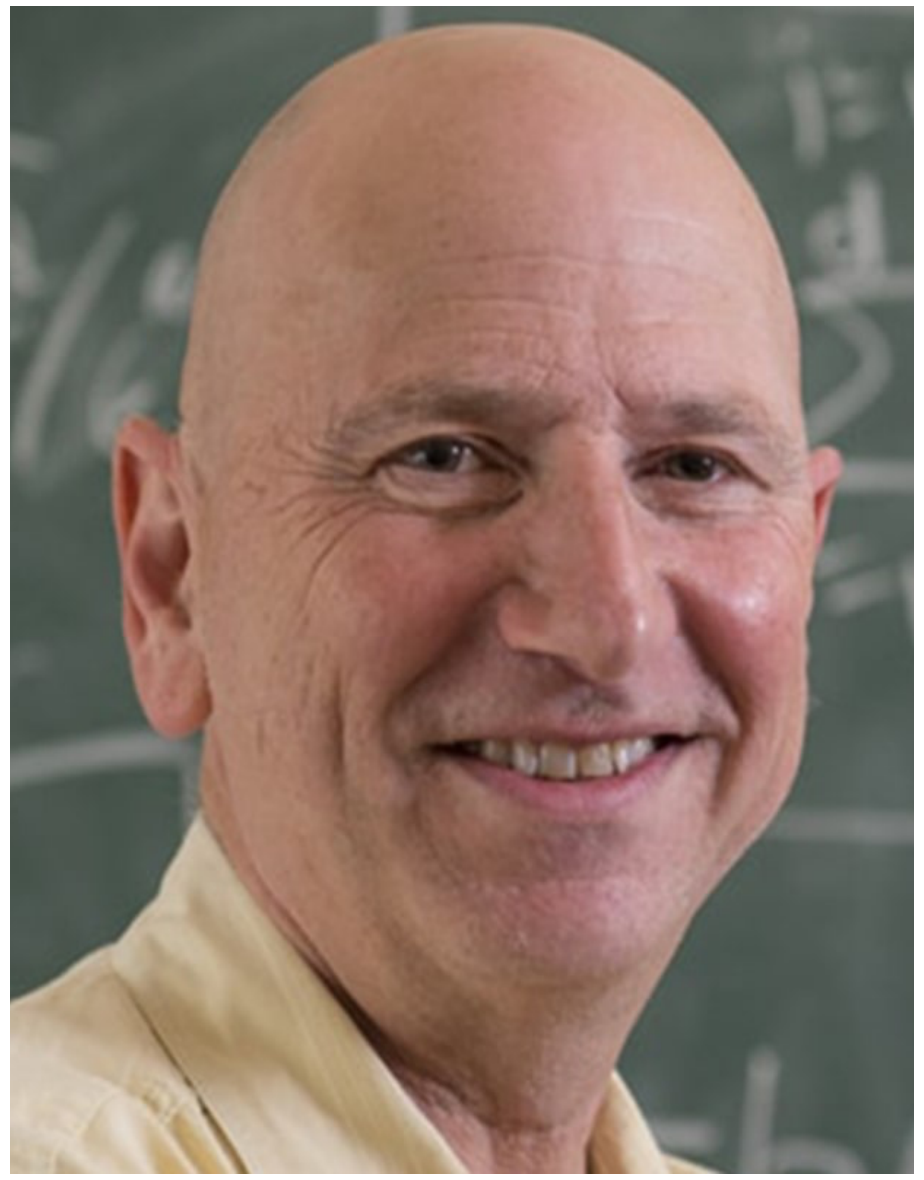
Department of Statistics & Data Science
Yonsei University

Joint work with

Matey Neykov
(Northwestern)



Sivaraman Balakrishnan
(CMU)



Larry Wasserman
(CMU)

Conditional Independence (CI) Testing

- Suppose that we observe $\{(X_i, Y_i, Z_i)\}_{i=1}^n \overset{i.i.d.}{\sim} P_{X,Y,Z}$
- We would like to **determine** whether

$$H_0 : X \perp\!\!\!\perp Y \mid Z \quad \text{versus} \quad H_1 : X \not\perp\!\!\!\perp Y \mid Z$$

$$(P_{X,Y,Z} = P_{X|Z}P_{Y|Z}P_Z) \qquad (P_{X,Y,Z} \neq P_{X|Z}P_{Y|Z}P_Z)$$

Applications: Variable Importance

Is the predictor X important to predict the response variable Y?

Applications: Variable Importance

Is the predictor X important to predict the response variable Y?

1. Parametric methods

$$Y = \beta X + \gamma Z + \varepsilon$$

- e.g., F-test

  $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$

- Not meaningful when applied beyond linear models

Applications: Variable Importance

Is the predictor X important to predict the response variable Y?

1. Parametric methods

$$Y = \beta X + \gamma Z + \varepsilon$$

- e.g., F-test

  $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$

- Not meaningful when applied beyond linear models

2. Blackbox methods

- Feature importance from e.g., random forests and XGboost
- Model dependent & does not present uncertainty

Applications: Variable Importance

Is the predictor X important to predict the response variable Y?

1. Parametric methods

$$Y = \beta X + \gamma Z + \varepsilon$$

- e.g., F-test

$$H_0 : \beta = 0 \text{ vs. } H_1 : \beta \neq 0$$

- Not meaningful when applied beyond linear models

2. Blackbox methods

- Feature importance from e.g., random forests and XGboost
- Model dependent & does not present uncertainty

We can formulate it as a **hypothesis testing problem!**

$$H_0 : Y \perp\!\!\!\perp X \mid Z \quad \text{versus} \quad H_1 : Y \not\!\perp\!\!\!\perp X \mid Z$$

Prior methods

Asymptotic method

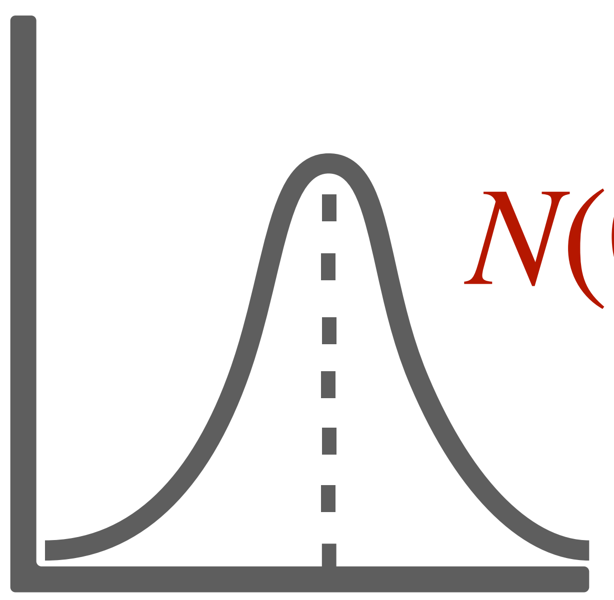
Su and White (2008)
Zhang et al. (2012)
Huang (2010)
Wang et al. (2015)
Strobl et al. (2019)
Zhou et al. (2020)
Shah and Peters (2020)
Huang et al. (2022)
Lundborg et al. (2024)
⋮

Prior methods

Asymptotic method

Su and White (2008)
Zhang et al. (2012)
Huang (2010)
Wang et al. (2015)
Strobl et al. (2019)
Zhou et al. (2020)
Shah and Peters (2020)
Huang et al. (2022)
Lundborg et al. (2024)
⋮

- Pivotal Statistic $\xrightarrow{d}$



$N(0,1)$

- Determine the **critical value** based on the **limiting distribution**

Prior methods

Asymptotic method

Su and White (2008)
Zhang et al. (2012)
Huang (2010)
Wang et al. (2015)
Strobl et al. (2019)
Zhou et al. (2020)
Shah and Peters (2020)
Huang et al. (2022)
Lundborg et al. (2024)
⋮

Model-X framework

Candés et al. (2018)
Berrett et al. (2020)
Bellot and van der Schaar (2019)
Katsevish and Ramdas (2022)
Watson and Wright (2021)
Liu et al. (2022)
⋮

Prior methods

Asymptotic method

Su and White (2008)
Zhang et al. (2012)
Huang (2010)
Wang et al. (2015)
Strobl et al. (2019)
Zhou et al. (2020)
Shah and Peters (2020)
Huang et al. (2022)
Lundborg et al. (2024)
⋮

Model-X framework

Candés et al. (2018)
Berrett et al. (2020)
Bellot and van der Schaar (2019)
Katsevish and Ramdas (2022)
Watson and Wright (2021)
Liu et al. (2022)
⋮

- Assume $F_{X|Z}$ is known
- Generate $\tilde{X}$ from $F_{X|Z}$
- Construct an exchangeable sequence under the null

$$T, T_1, \ldots, T_B$$

Prior methods

Asymptotic method

Su and White (2008)
Zhang et al. (2012)
Huang (2010)
Wang et al. (2015)
Strobl et al. (2019)
Zhou et al. (2020)
Shah and Peters (2020)
Huang et al. (2022)
Lundborg et al. (2024)
⋮

Model-X framework

Candés et al. (2018)
Berrett et al. (2020)
Bellot and van der Schaar (2019)
Katsevish and Ramdas (2022)
Watson and Wright (2021)
Liu et al. (2022)
⋮

Local permutation

Margaritis (2005)
Fukumizu et al. (2008)
Doran et al. (2014)
Sen et al. (2017)
Neykov et al. (2021)
⋮

Prior methods

Asymptotic method

Su and White (2008)
Zhang et al. (2012)
Huang (2010)
Wang et al. (2015)
Strobl et al. (2019)
Zhou et al. (2020)
Shah and Peters (2020)
Huang et al. (2022)
Lundborg et al. (2024)
⋮

Model-X framework

Candés et al. (2018)
Berrett et al. (2020)
Bellot and van der Schaar (2019)
Katsevish and Ramdas (2022)
Watson and Wright (2021)
Liu et al. (2022)
⋮

Local permutation

Margaritis (2005)
Fukumizu et al. (2008)
Doran et al. (2014)
Sen et al. (2017)
Neykov et al. (2021)
⋮

We aim to enhance our understanding of

local permutation tests:

"when does it work and when does it fail?"

Local Permutation Tests

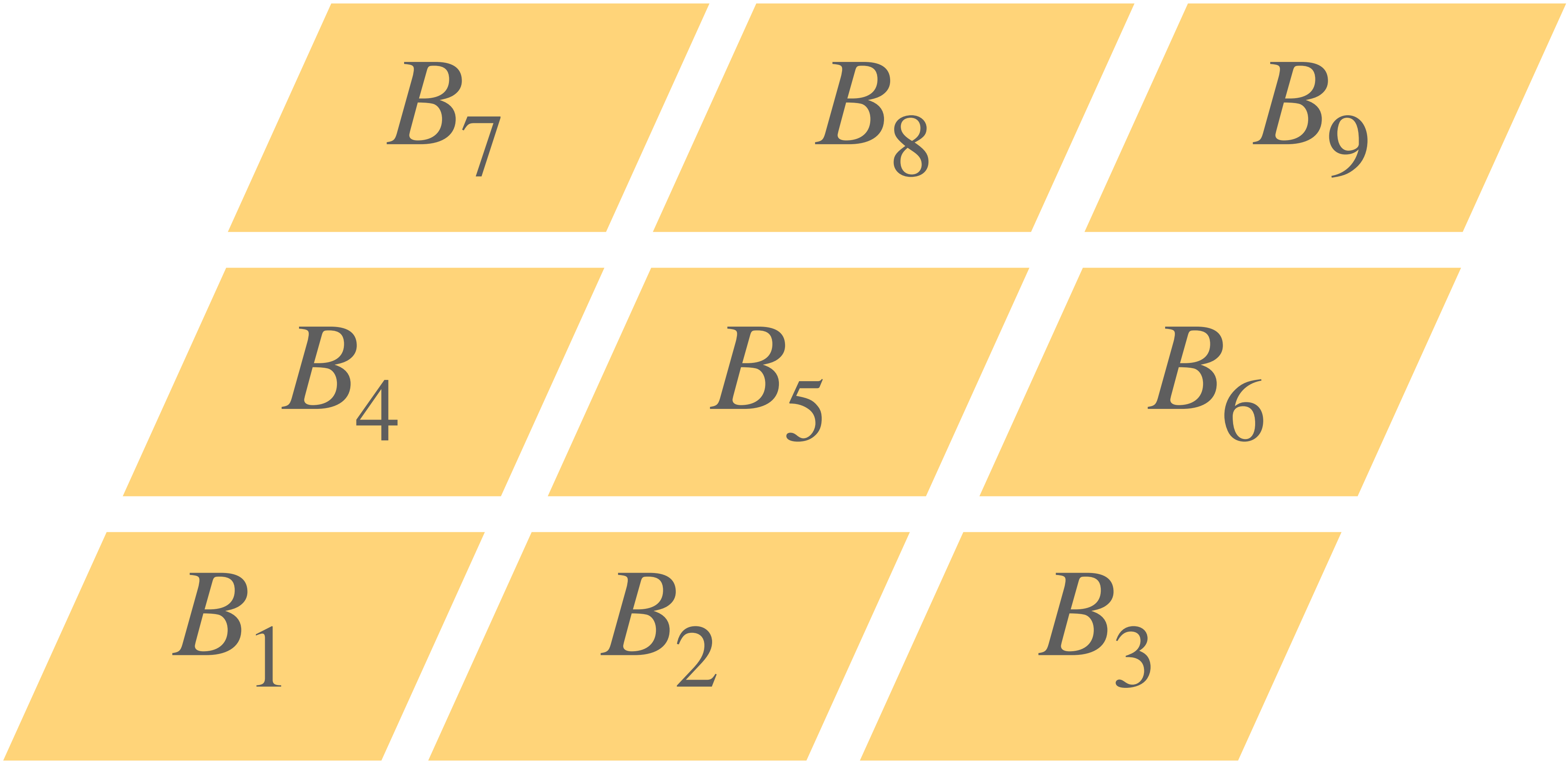
- Partition the domain of Z into $B_1, \ldots, B_M$ bins

Local Permutation Tests

- **Partition** the domain of Z into $B_1, \ldots, B_M$ bins
- **Transform** $\{(X_i, Y_i, Z_i)\}_{i=1}^n \mapsto \{(X_i, Y_i, \widetilde{Z}_i)\}_{i=1}^n$ where $\widetilde{Z}_i = k$ if $Z_i \in B_k$

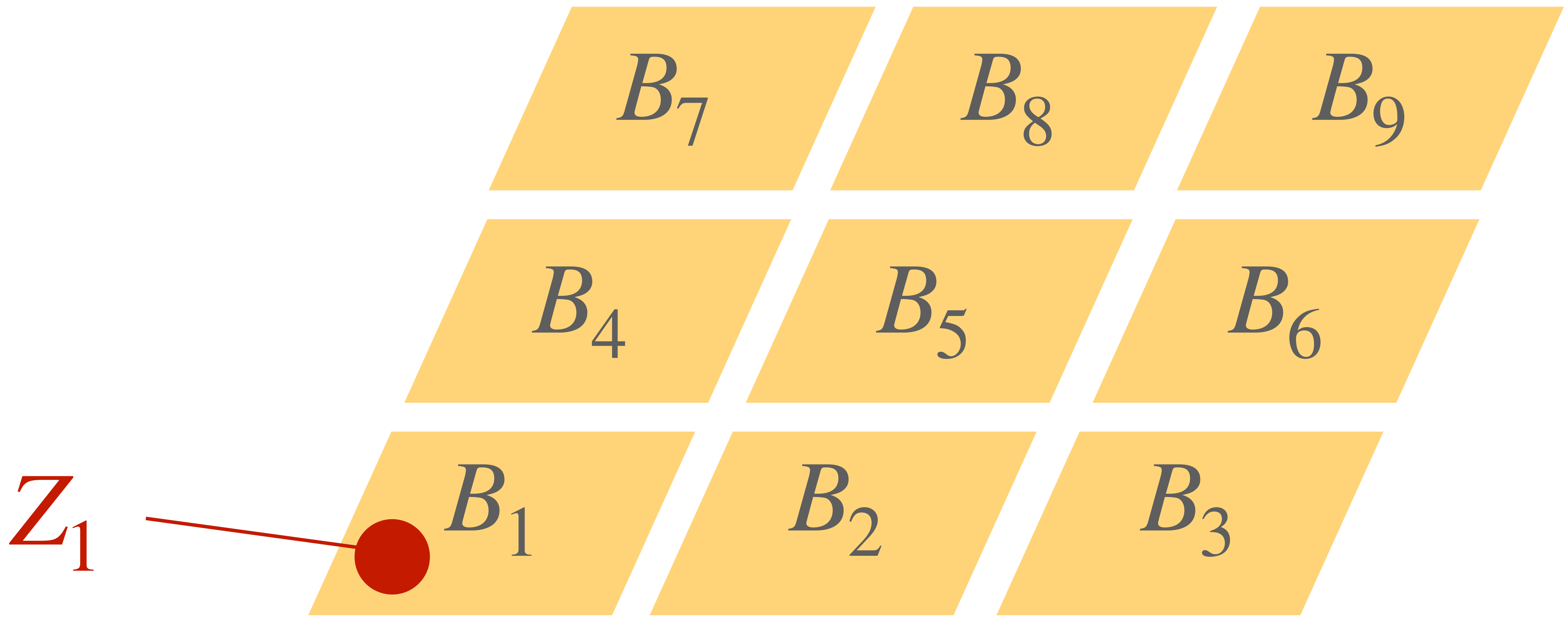
Local Permutation Tests

- Partition the domain of Z into $B_1, \ldots, B_M$ bins
- Transform $\{(X_i, Y_i, Z_i)\}_{i=1}^n \mapsto \{(X_i, Y_i, \widetilde{Z}_i)\}_{i=1}^n$ where $\widetilde{Z}_i = k$ if $Z_i \in B_k$



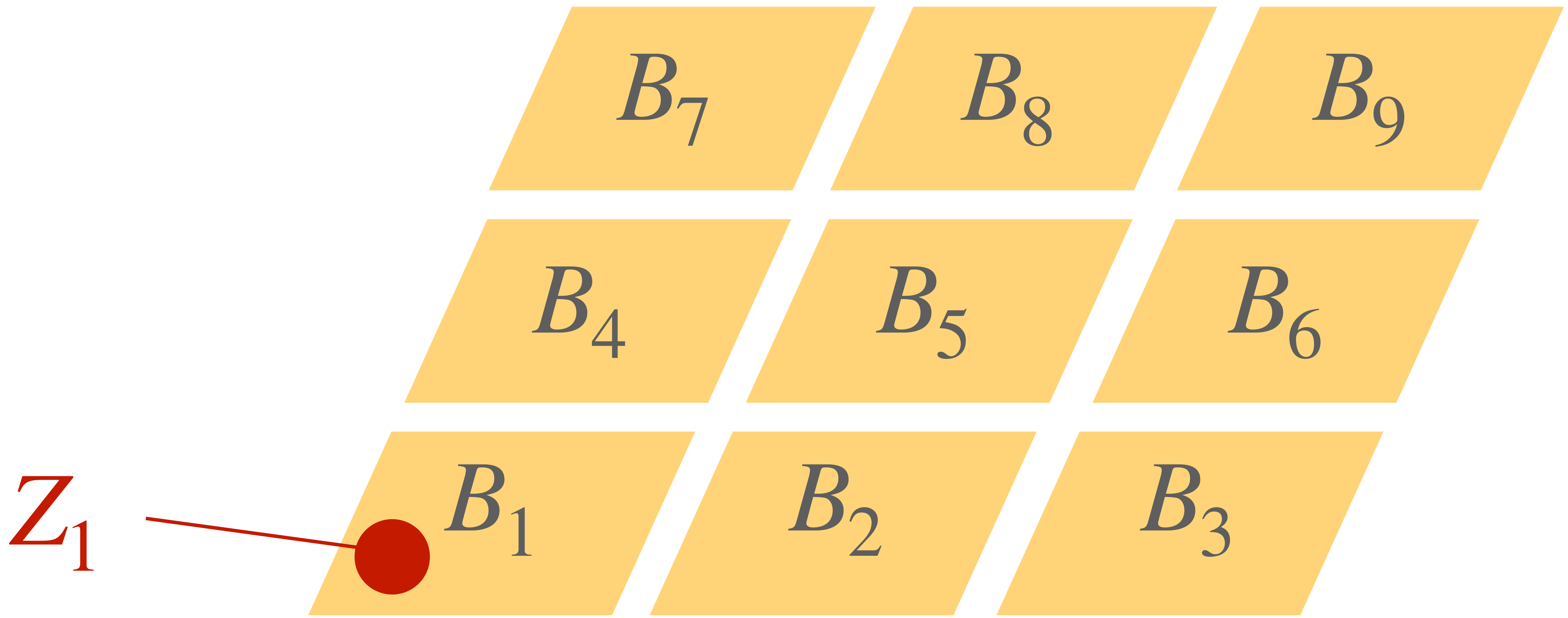
Local Permutation Tests

- **Partition** the domain of Z into $B_1, \ldots, B_M$ bins
- **Transform** $\{(X_i, Y_i, Z_i)\}_{i=1}^n \mapsto \{(X_i, Y_i, \widetilde{Z}_i)\}_{i=1}^n$ where $\widetilde{Z}_i = k$ if $Z_i \in B_k$



Local Permutation Tests

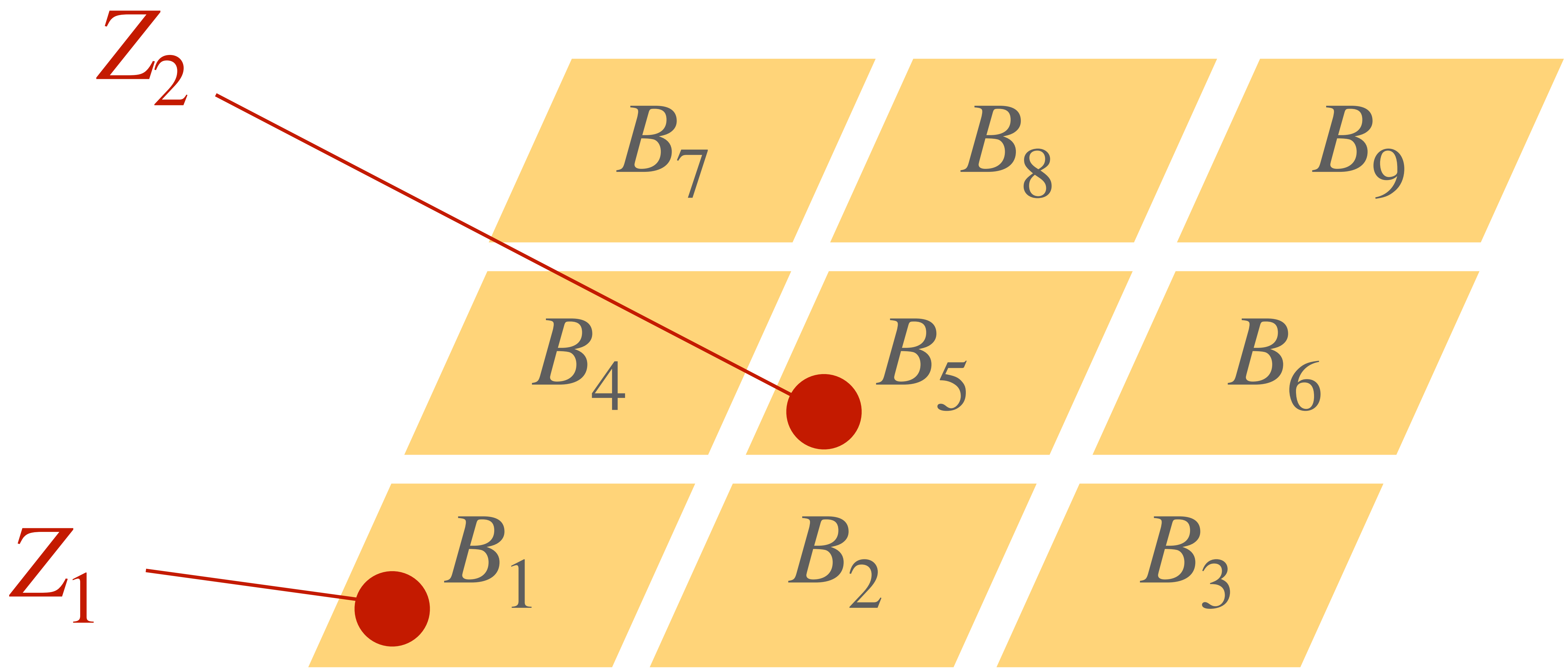
- **Partition** the domain of Z into $B_1, \ldots, B_M$ bins
- **Transform** $\{(X_i, Y_i, Z_i)\}_{i=1}^n \mapsto \{(X_i, Y_i, \widetilde{Z}_i)\}_{i=1}^n$ where $\widetilde{Z}_i = k$ if $Z_i \in B_k$



$\boxed{\widetilde{Z}_1 = 1}$

Local Permutation Tests

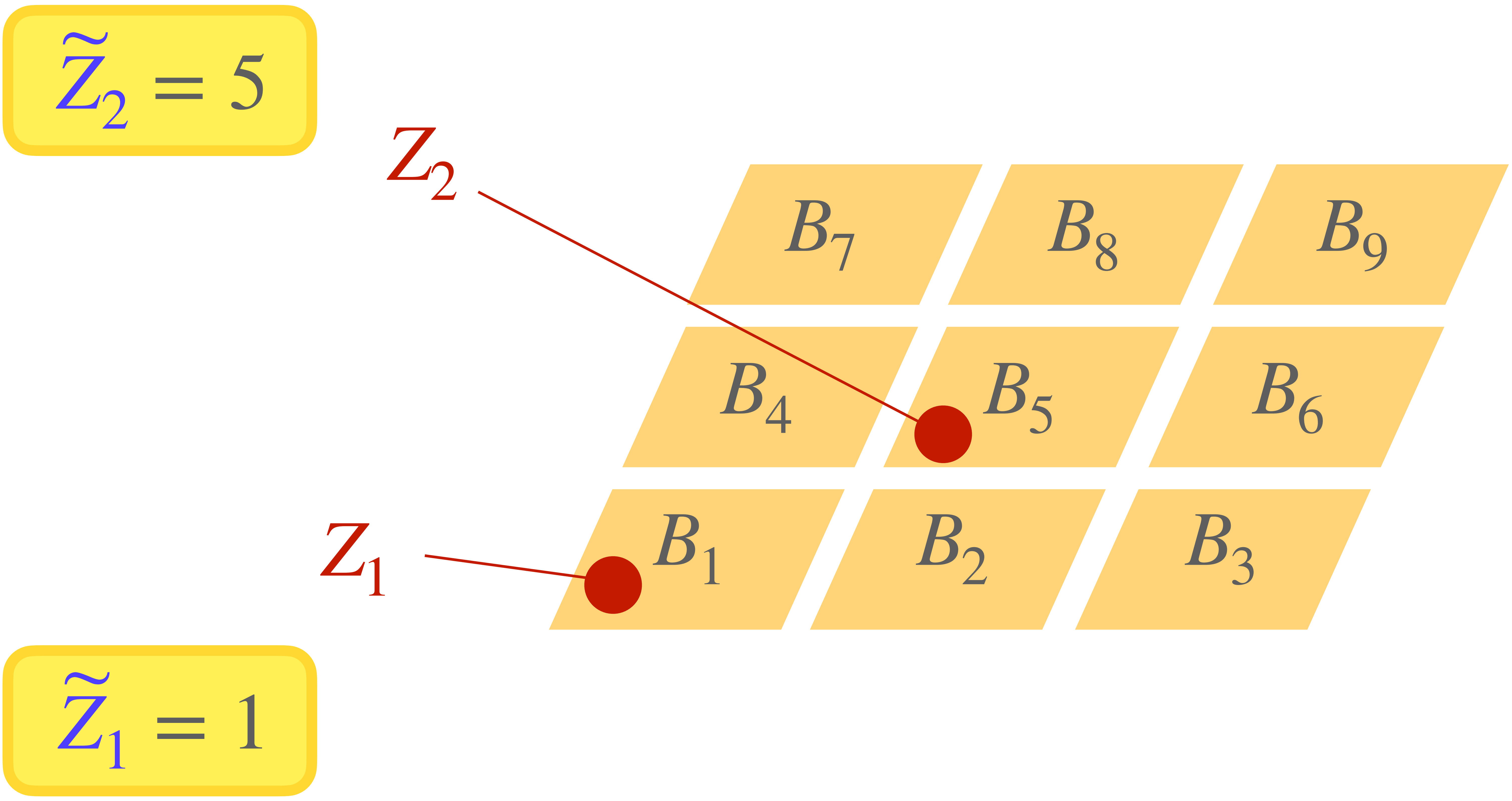
- **Partition** the domain of Z into $B_1, \ldots, B_M$ bins
- **Transform** $\{(X_i, Y_i, Z_i)\}_{i=1}^n \mapsto \{(X_i, Y_i, \widetilde{Z}_i)\}_{i=1}^n$ where $\widetilde{Z}_i = k$ if $Z_i \in B_k$



$\widetilde{Z}_1 = 1$

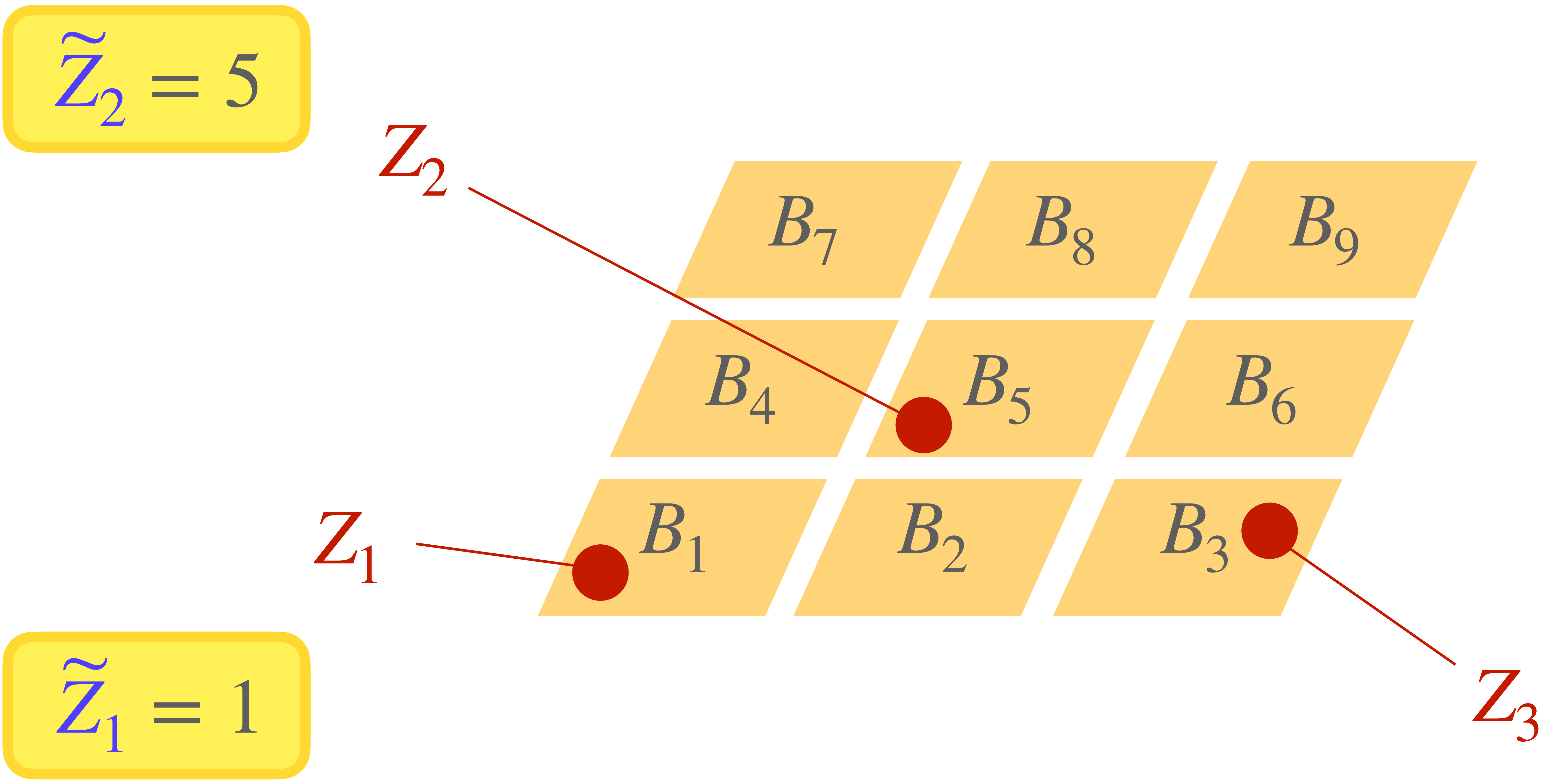
Local Permutation Tests

- **Partition** the domain of Z into $B_1, \ldots, B_M$ bins
- **Transform** $\{(X_i, Y_i, Z_i)\}_{i=1}^n \mapsto \{(X_i, Y_i, \widetilde{Z}_i)\}_{i=1}^n$ where $\widetilde{Z}_i = k$ if $Z_i \in B_k$



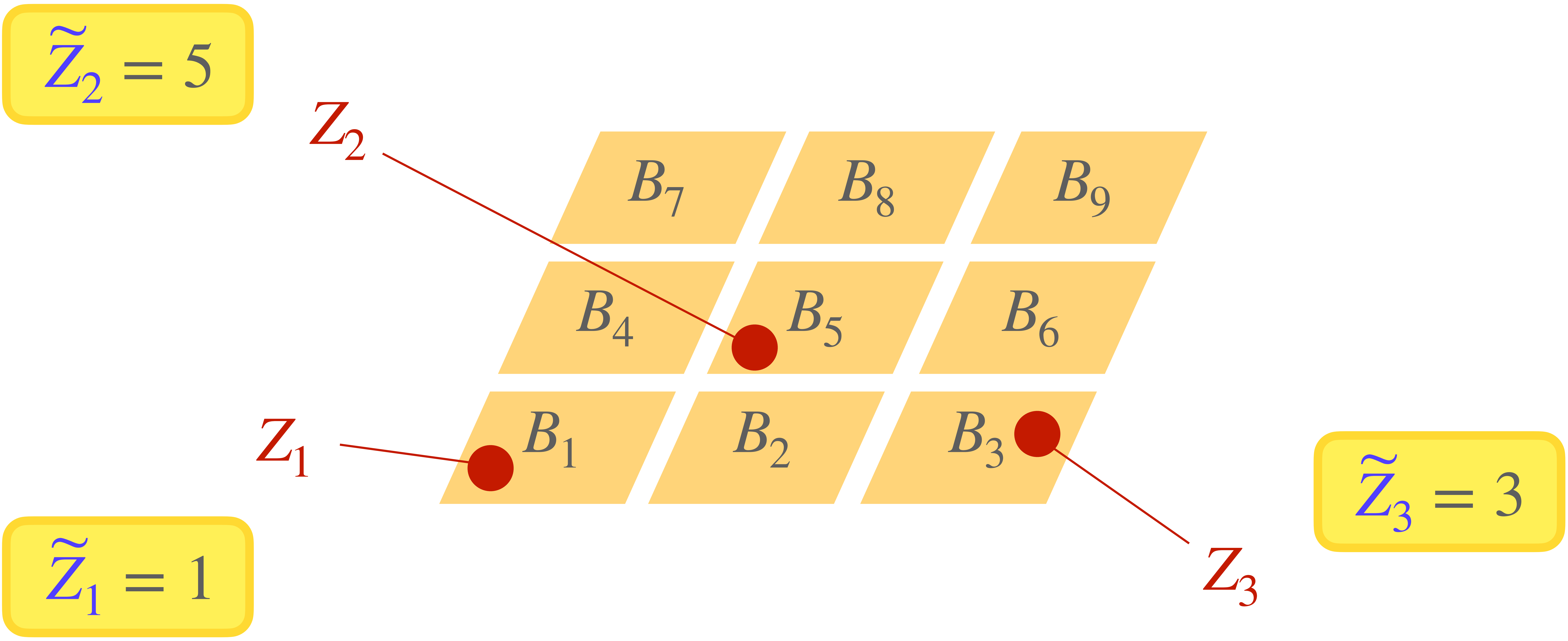
Local Permutation Tests

- **Partition** the domain of Z into $B_1, \ldots, B_M$ bins
- **Transform** $\{(X_i, Y_i, Z_i)\}_{i=1}^n \mapsto \{(X_i, Y_i, \widetilde{Z}_i)\}_{i=1}^n$ where $\widetilde{Z}_i = k$ if $Z_i \in B_k$



Local Permutation Tests

- **Partition** the domain of Z into $B_1, \ldots, B_M$ bins
- **Transform** $\{(X_i, Y_i, Z_i)\}_{i=1}^n \mapsto \{(X_i, Y_i, \widetilde{Z}_i)\}_{i=1}^n$ where $\widetilde{Z}_i = k$ if $Z_i \in B_k$



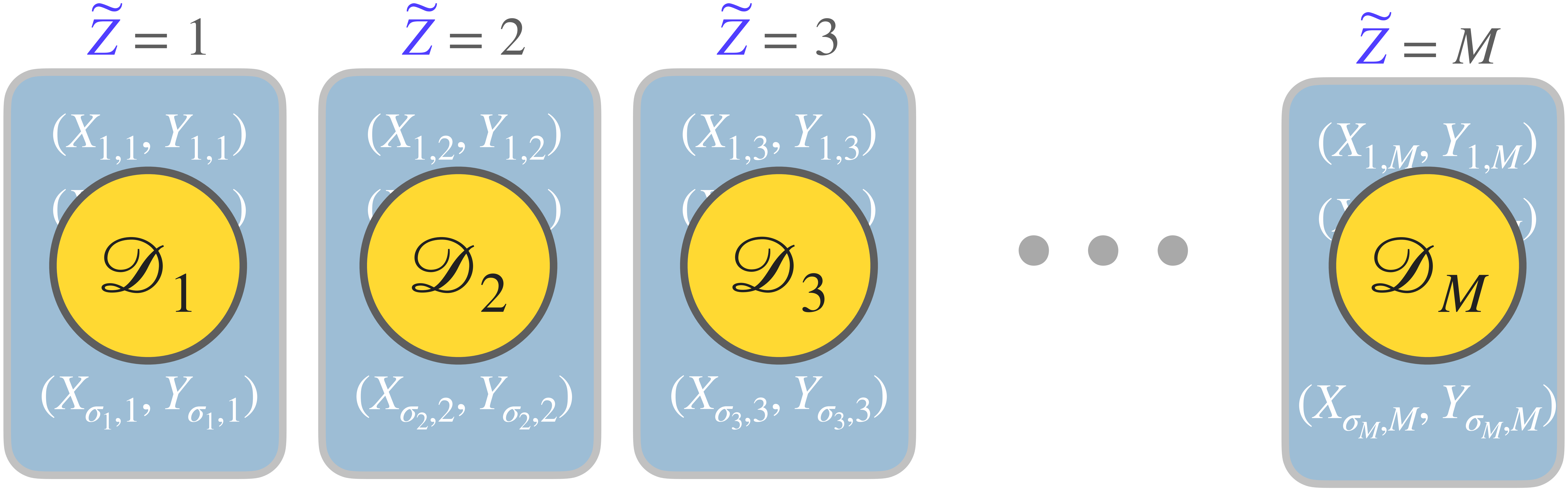
Local Permutation Tests

- **Partition** the domain of Z into $B_1, \ldots, B_M$ bins
- **Transform** $\{(X_i, Y_i, Z_i)\}_{i=1}^n \mapsto \{(X_i, Y_i, \widetilde{Z}_i)\}_{i=1}^n$ where $\widetilde{Z}_i = k$ if $Z_i \in B_k$

$\widetilde{Z} = 1$	$\widetilde{Z} = 2$	$\widetilde{Z} = 3$		$\widetilde{Z} = M$
$(X_{1,1}, Y_{1,1})$	$(X_{1,2}, Y_{1,2})$	$(X_{1,3}, Y_{1,3})$		$(X_{1,M}, Y_{1,M})$
$(X_{2,1}, Y_{2,1})$	$(X_{2,2}, Y_{2,2})$	$(X_{2,3}, Y_{2,3})$	$\cdots$	$(X_{2,M}, Y_{2,M})$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$(X_{\sigma_1,1}, Y_{\sigma_1,1})$	$(X_{\sigma_2,2}, Y_{\sigma_2,2})$	$(X_{\sigma_3,3}, Y_{\sigma_3,3})$		$(X_{\sigma_M,M}, Y_{\sigma_M,M})$

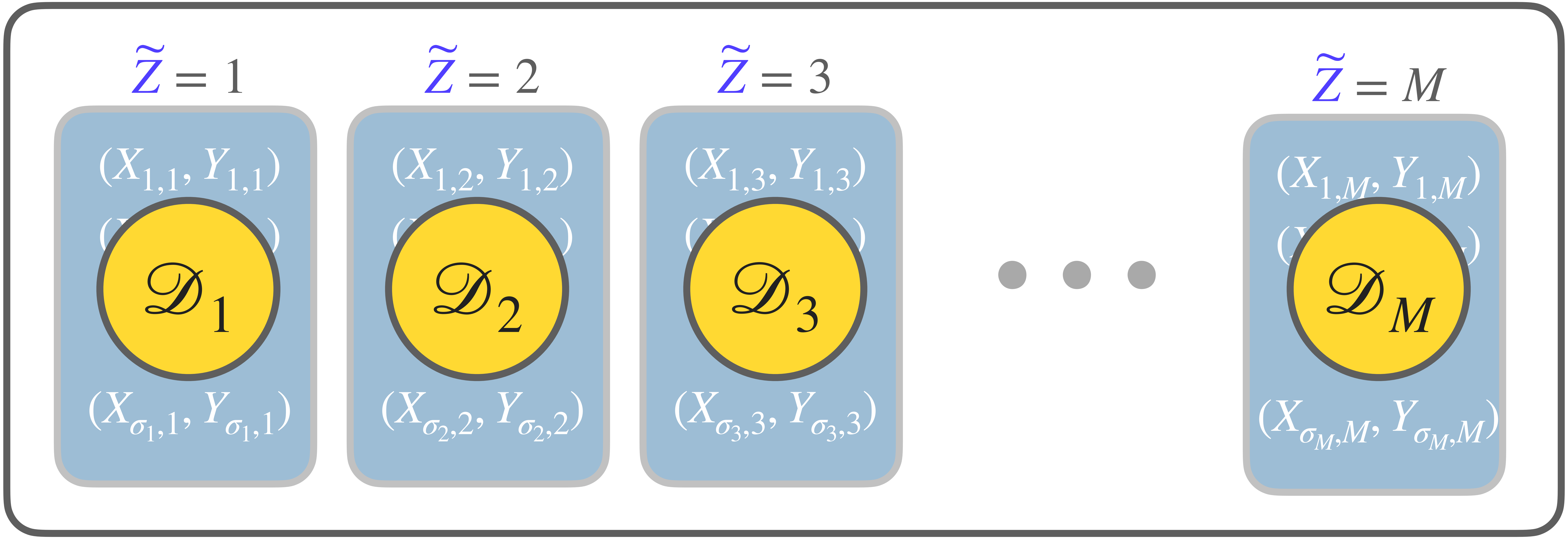
Local Permutation Tests

- **Partition** the domain of Z into $B_1, \ldots, B_M$ bins
- **Transform** $\{(X_i, Y_i, Z_i)\}_{i=1}^n \mapsto \{(X_i, Y_i, \widetilde{Z}_i)\}_{i=1}^n$ where $\widetilde{Z}_i = k$ if $Z_i \in B_k$



Local Permutation Tests

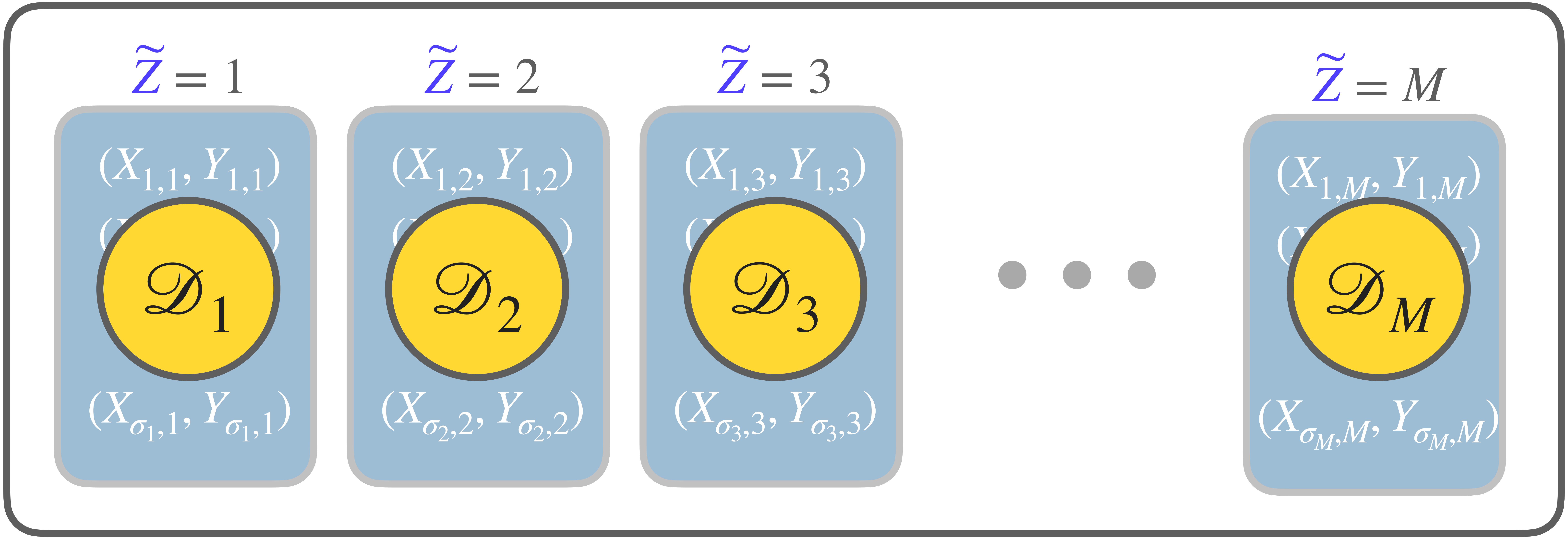
- **Compute** a binning-based test statistic $T := T(\mathscr{D}_1, \mathscr{D}_2, \ldots, \mathscr{D}_M)$



Local Permutation Tests

- **Compute** a binning-based test statistic $T := T(\mathscr{D}_1, \mathscr{D}_2, \ldots, \mathscr{D}_M)$

e.g., $T = \sum_{i=1}^{M} \chi^2(\mathscr{D}_i)$ or $T = \max_{1 \le i \le M} \chi^2(\mathscr{D}_i)$



Local Permutation Tests

$\widetilde{Z} = 1$

$(X_{1,1}, Y_{1,1})$
$(X_{2,1}, Y_{2,1})$
$\vdots$
$(X_{\sigma_1,1}, Y_{\sigma_1,1})$

$\widetilde{Z} = 2$

$(X_{1,2}, Y_{1,2})$
$(X_{2,2}, Y_{2,2})$
$\vdots$
$(X_{\sigma_2,2}, Y_{\sigma_2,2})$

$\widetilde{Z} = 3$

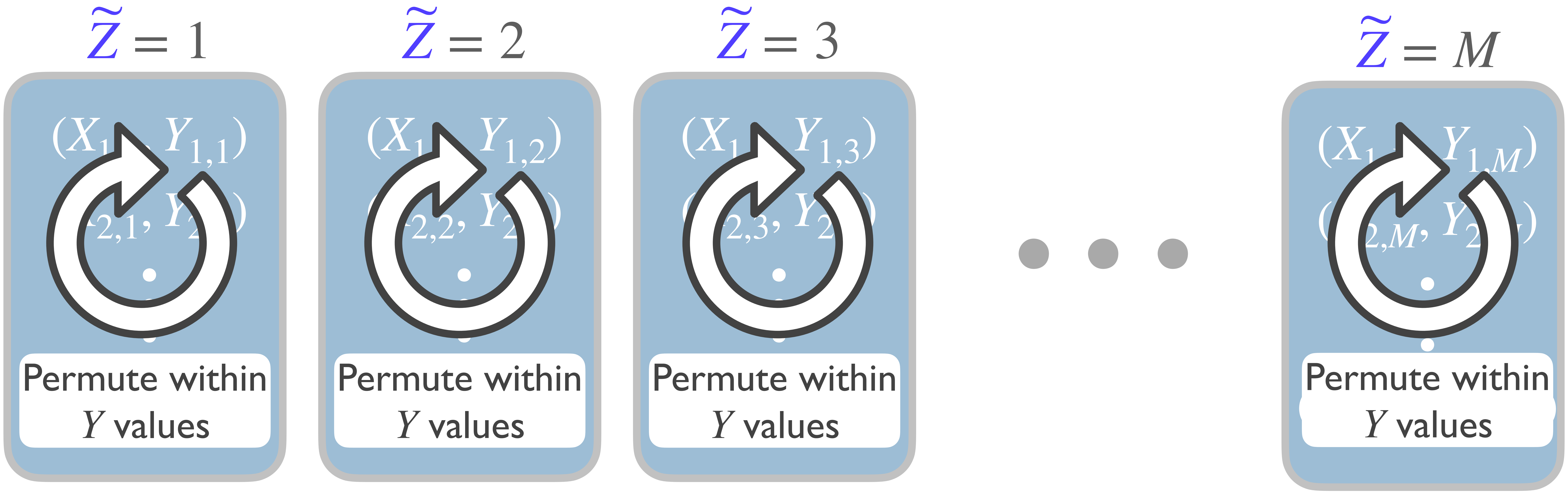
$(X_{1,3}, Y_{1,3})$
$(X_{2,3}, Y_{2,3})$
$\vdots$
$(X_{\sigma_3,3}, Y_{\sigma_3,3})$

$\cdots$

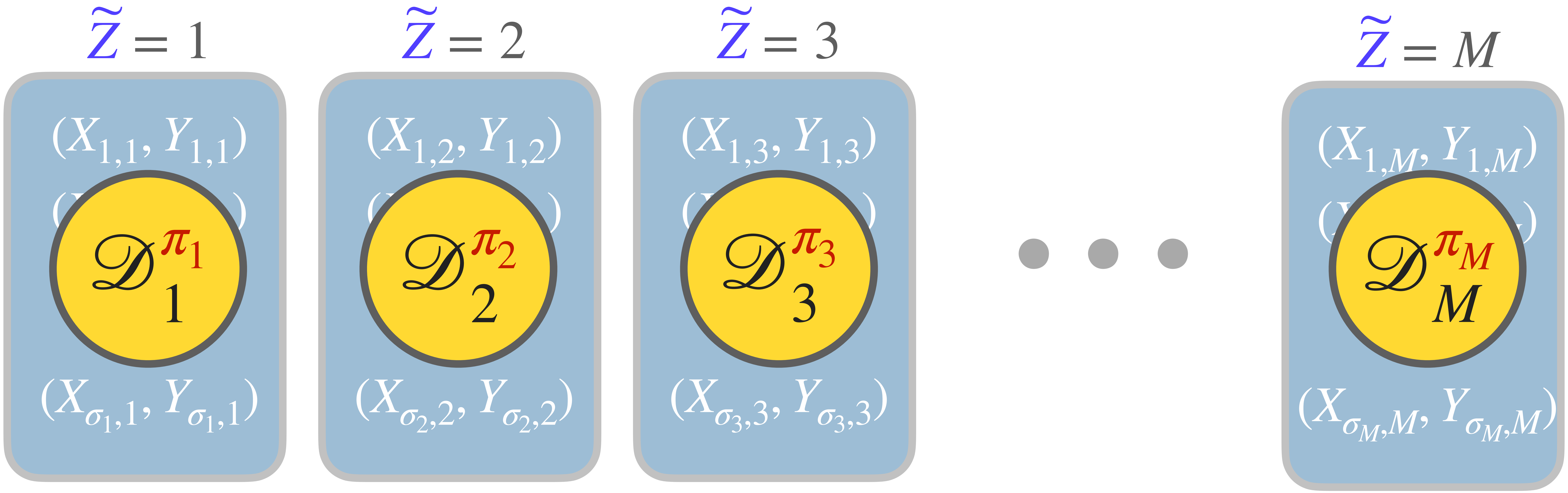
$\widetilde{Z} = M$

$(X_{1,M}, Y_{1,M})$
$(X_{2,M}, Y_{2,M})$
$\vdots$
$(X_{\sigma_M,M}, Y_{\sigma_M,M})$

Local Permutation Tests



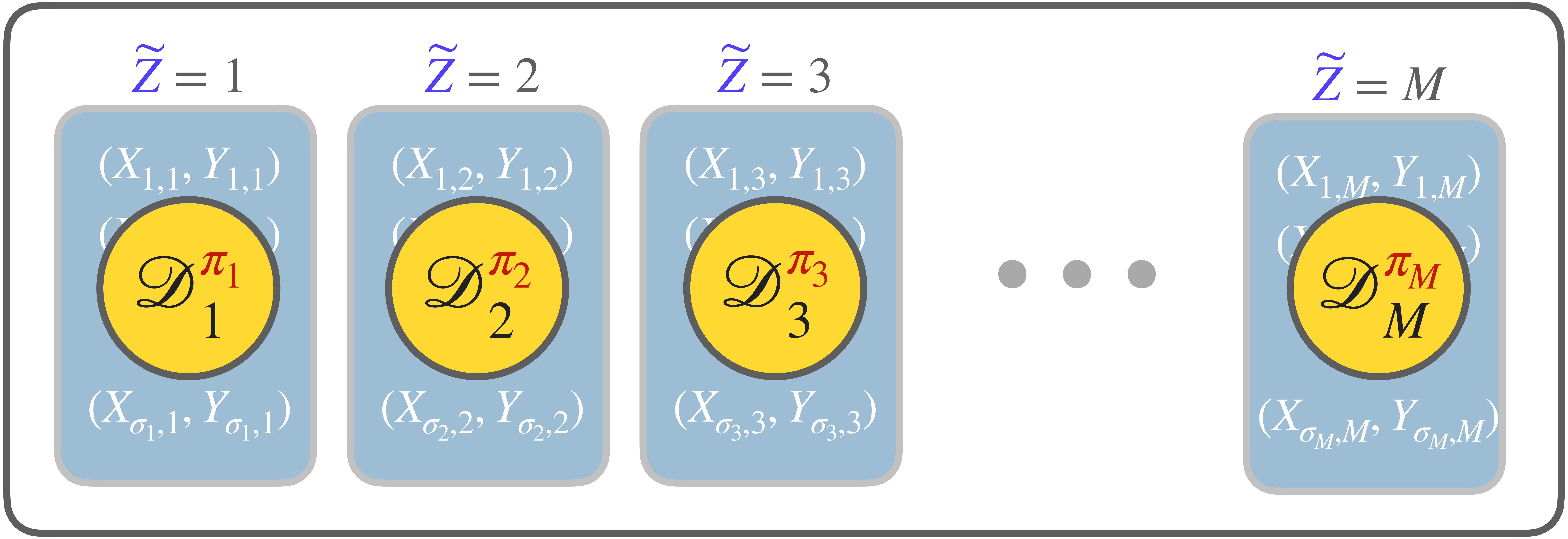
Local Permutation Tests



Local Permutation Tests

- **Compute** a binning-based test statistic $T^{\pi} := T(\mathscr{D}_1^{\pi_1}, \mathscr{D}_2^{\pi_2}, \ldots, \mathscr{D}_M^{\pi_M})$

e.g., $T^{\pi} = \sum_{i=1}^{M} \chi^2(\mathscr{D}_i^{\pi_1})$ or $T^{\pi} = \max_{1 \le i \le M} \chi^2(\mathscr{D}_i^{\pi_1})$



Algorithm: Local Permutation Tests

- **Compute** a binning-based test statistic $T := T(\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_M)$

For $i = 1, \ldots, K$

Permute Y values within bins

$T_i^\pi := T(\mathcal{D}_1^{\pi_1}, \mathcal{D}_2^{\pi_2}, \ldots, \mathcal{D}_M^{\pi_M})$

End

- **Reject** the null if $T > q_{1-\alpha}$

The $1 - \alpha$ quantile of $\{T, T_1^\pi, \ldots, T_K^\pi\}$

Algorithm: Local Permutation Tests

- **Compute** a binning-based test statistic $T := T(\mathscr{D}_1, \mathscr{D}_2, \ldots, \mathscr{D}_M)$

For $i = 1, \ldots, K$

Permute Y values within bins

$$T_i^{\pi} := T(\mathscr{D}_1^{\pi_1}, \mathscr{D}_2^{\pi_2}, \ldots, \mathscr{D}_M^{\pi_M})$$

End

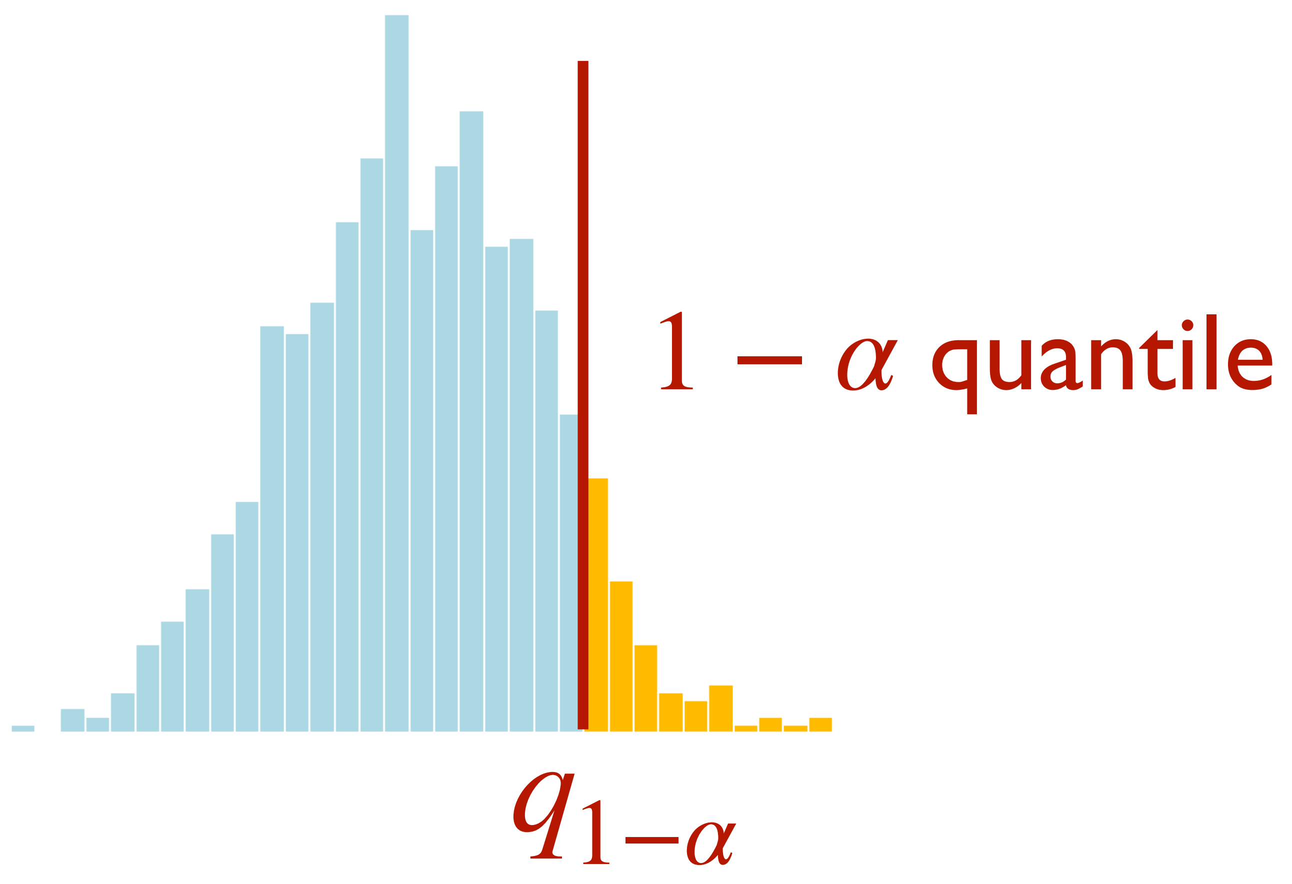
- **Reject** the null if $T > q_{1-\alpha}$

The $1 - \alpha$ **quantile of** $\{T, T_1^{\pi}, \ldots, T_K^{\pi}\}$

Algorithm: Local Permutation Tests

- **Compute** a binning-based test statistic $T := T(\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_M)$

$$\boxed{\begin{array}{l} \textit{For } i = 1,\ldots,K \\ \quad \text{Permute } Y \text{ values within bins} \\ \quad T_i^\pi := T(\mathcal{D}_1^{\pi_1}, \mathcal{D}_2^{\pi_2}, \ldots, \mathcal{D}_M^{\pi_M}) \\ \textit{End} \end{array}}$$



- **Reject** the null if $T > q_{1-\alpha}$

The $1-\alpha$ quantile of $\{T, T_1^\pi, \ldots, T_K^\pi\}$

Local Permutation Tests

Remark

- Unlike the two-sample and independence testing problem, there is **no guarantee** that the **local permutation test** is **valid** under the null

- In fact, **any valid CI test** has power ***at most*** its size α against **any alternative**! *(Shah & Peters 2020)*

Local Permutation Tests

Remark

- Unlike the two-sample and independence testing problem, there is **no guarantee** that the **local permutation test** is **valid** under the null

- In fact, **any valid CI test** has power ***at most*** its size α against **any alternative**! *(Shah & Peters 2020)*

Hardness Result for CI Testing

Theorem [Shah & Peters 2020]

- Let $\mathscr{P}_{\text{conti}}$ be the class of continuous distributions on $\mathbb{R}^d$, $\mathscr{P}_{0,\text{conti}}$ be the subset of $\mathscr{P}_{\text{conti}}$ such that $X \perp\!\!\!\perp Y \,|\, Z$, and $\mathscr{P}_{1,\text{conti}} = \mathscr{P}_{\text{conti}} \backslash \mathscr{P}_{0,\text{conti}}$.

Hardness Result for CI Testing

Distributions under H_0

Theorem [Shah & Peters 2020]

- Let $\mathscr{P}_{\text{conti}}$ be the **class of continuous distributions** on $\mathbb{R}^d$, $\mathscr{P}_{0,\text{conti}}$ be the subset of $\mathscr{P}_{\text{conti}}$ such that $X \perp\!\!\!\perp Y \mid Z$, and $\mathscr{P}_{1,\text{conti}} = \mathscr{P}_{\text{conti}} \backslash \mathscr{P}_{0,\text{conti}}$.

Distributions under H_1

Hardness Result for CI Testing

Theorem [Shah & Peters 2020]

- Let $\mathscr{P}_{\text{conti}}$ be the **class of continuous distributions** on $\mathbb{R}^d$, $\mathscr{P}_{0,\text{conti}}$ be the subset of $\mathscr{P}_{\text{conti}}$ such that $X \perp\!\!\!\perp Y \,|\, Z$, and $\mathscr{P}_{1,\text{conti}} = \mathscr{P}_{\text{conti}} \backslash \mathscr{P}_{0,\text{conti}}$.
- Suppose that a test ϕ satisfies

$$\sup_{P \in \mathscr{P}_{0,\text{conti}}} \mathbb{E}_P[\phi] \leq \alpha\,.$$

Hardness Result for CI Testing

Theorem [Shah & Peters 2020]

- Let $\mathscr{P}_{\mathrm{conti}}$ be the class of continuous distributions on $\mathbb{R}^d$, $\mathscr{P}_{0,\mathrm{conti}}$ be the subset of $\mathscr{P}_{\mathrm{conti}}$ such that $X \perp\!\!\!\perp Y \,|\, Z$, and $\mathscr{P}_{1,\mathrm{conti}} = \mathscr{P}_{\mathrm{conti}} \backslash \mathscr{P}_{0,\mathrm{conti}}$.
- Suppose that a test ϕ satisfies

$$\sup_{P \in \mathscr{P}_{0,\mathrm{conti}}} \mathbb{E}_P[\phi] \leq \alpha\,.$$

Then for **any** $P \in \mathscr{P}_{1,\mathrm{conti}}$, the power of ϕ is **bounded above** by

$$\mathbb{E}_P[\phi] \leq \alpha\,.$$

Intuition for the Hardness Result

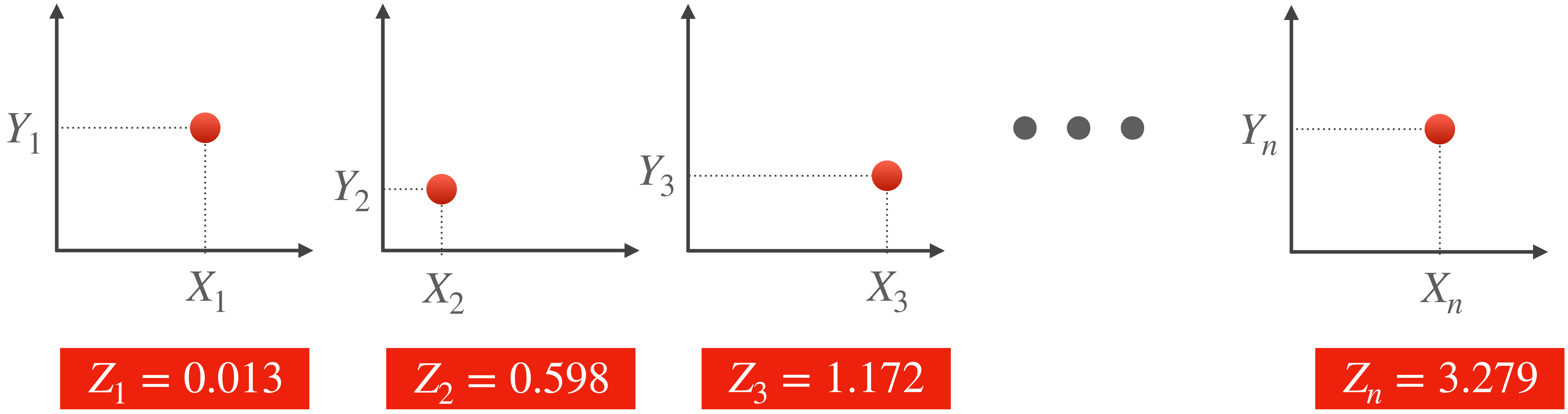
- When Z is continuous, we never observe the same value of Z

→ *The effective sample size is one!*

Intuition for the Hardness Result

- When Z is continuous, we never observe the same value of Z

  ➡ *The effective sample size is one!*



Question.

What about scenarios where Z follows a **discrete** or a **mixture** distribution? Is CI testing still **hard**?

New Hardness Result for CI Testing

Random sample from P_Z

Theorem [KNBW 2022]

- For an arbitrary $J \geq n(n-1)$, let $\rho_{J,P} = \mathbb{P}_P(Z_1, \ldots, Z_J$ are distinct)

No-collision probability

New Hardness Result for CI Testing

Theorem [KNBW 2022]

- For an arbitrary $J \geq n(n-1)$, let $\rho_{J,P} = \mathbb{P}_P(Z_1, \ldots, Z_J$ are distinct)
- Suppose that a test ϕ satisfies

$$\sup_{P \in \mathscr{P}_{0,\mathrm{disc}}} \mathbb{E}_P[\phi] \leq \alpha$$

Class of null distributions including discrete P_Z

New Hardness Result for CI Testing

Theorem [KNBW 2022]

- For an arbitrary $J \geq n(n-1)$, let $\rho_{J,P} = \mathbb{P}_P(Z_1, \ldots, Z_J \text{ are distinct})$
- Suppose that a test ϕ satisfies

$$\sup_{P \in \mathscr{P}_{0,\mathrm{disc}}} \mathbb{E}_P[\phi] \leq \alpha$$

Then for **any** $P \in \mathscr{P}_1$, the power of ϕ is **bounded above** by

$$\mathbb{E}_P[\phi] \leq \alpha \times \rho_{J,P} + (1 - \rho_{J,P}) + \frac{n(n-1)}{J}$$

New Hardness Result for CI Testing

Theorem [KNBW 2022]

- For an arbitrary $J \geq n(n-1)$, let $\rho_{J,P} = \mathbb{P}_P(Z_1, \ldots, Z_J$ are distinct)
- Suppose that a test ϕ satisfies

$$\sup_{P \in \mathscr{P}_{0,\mathrm{disc}}} \mathbb{E}_P[\phi] \leq$$

When $\rho_{J,P}$ is close to one, then the **power** becomes close to the size

Then for **any** $P \in \mathscr{P}_1$, the power of ϕ is **bounded above** by

$$\mathbb{E}_P[\phi] \leq \alpha \times \rho_{J,P} + (1 - \rho_{J,P}) + \frac{n(n-1)}{J}$$

New Hardness Result for CI Testing

Examples

- **Continuous** distributions ($\rho_{J,P} = 1$ for any $J \geq 2$)

Any **valid** CI test has the power **upper bounded** by

$$\mathbb{E}_P[\phi] \leq \alpha$$

New Hardness Result for CI Testing

Examples

- Continuous distributions ($\rho_{J,P} = 1$ for any $J \geq 2$)

  Any valid CI test has the power upper bounded by

$$\mathbb{E}_P[\phi] \leq \alpha$$

- **Discrete uniform** distribution on $\{1,2,\ldots,M\}$ with $n^4/M \to 0$

  Any **valid** CI test has the power **upper bounded** by

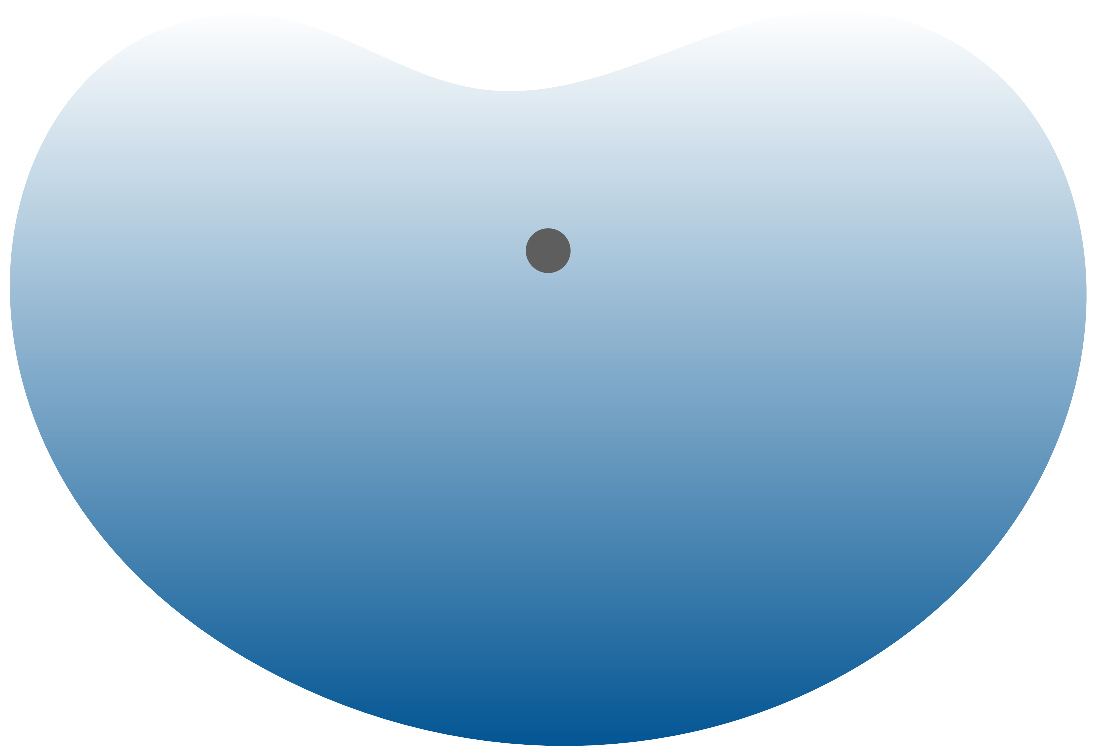
$$\mathbb{E}_P[\phi] \leq \alpha + o(1)$$

We need **assumptions** to make CI testing **feasible**
especially in continuous settings

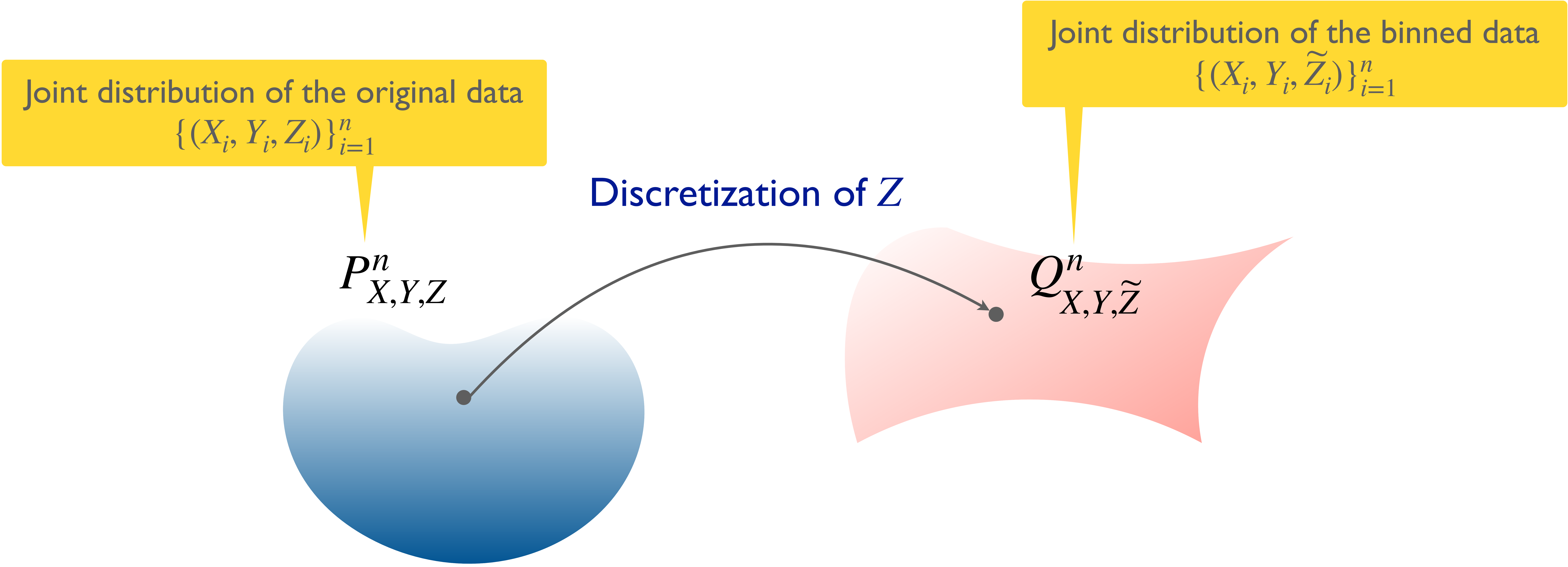
Validity of Local Permutation Tests

Joint distribution of the original data
$\{(X_i, Y_i, Z_i)\}_{i=1}^n$

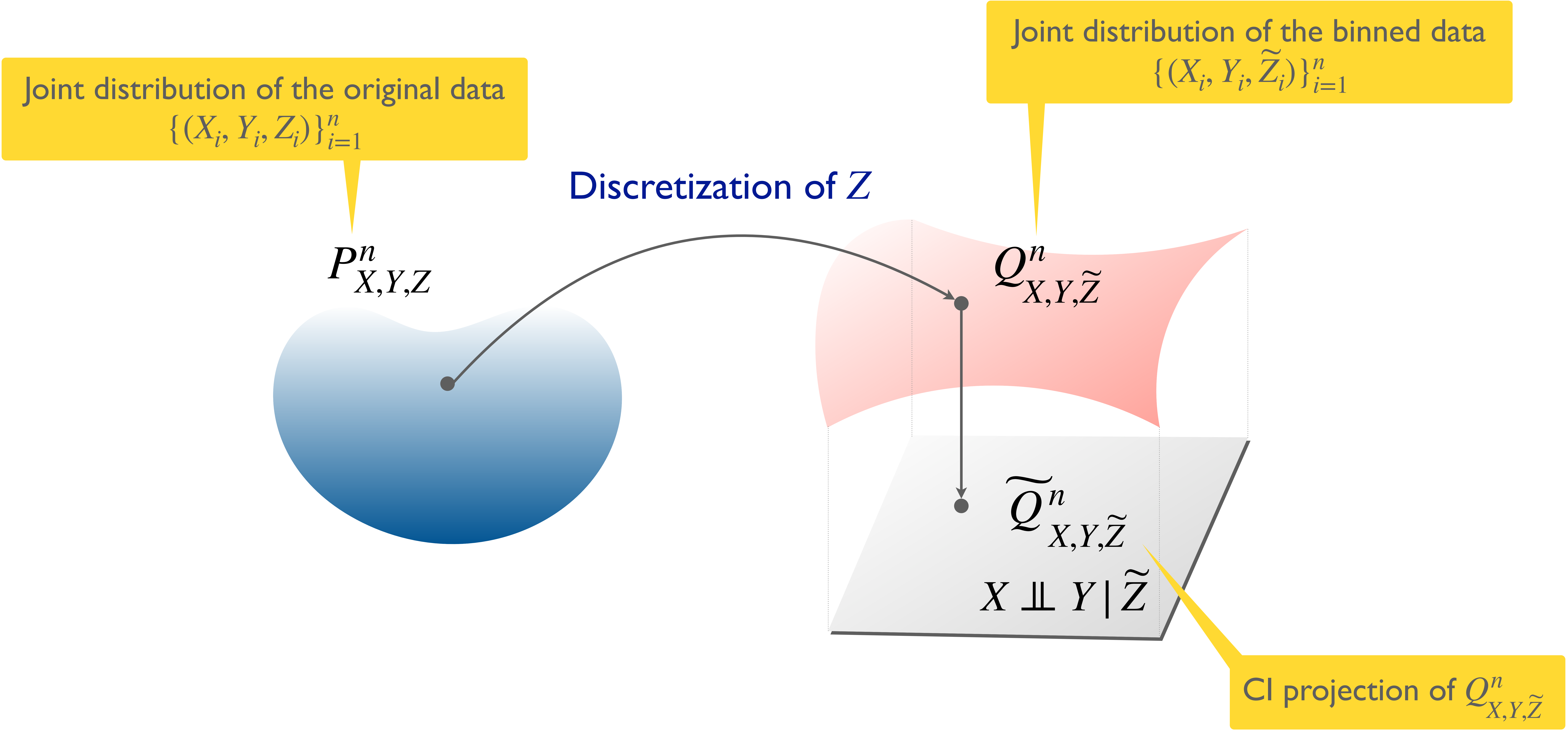
$P^n_{X,Y,Z}$



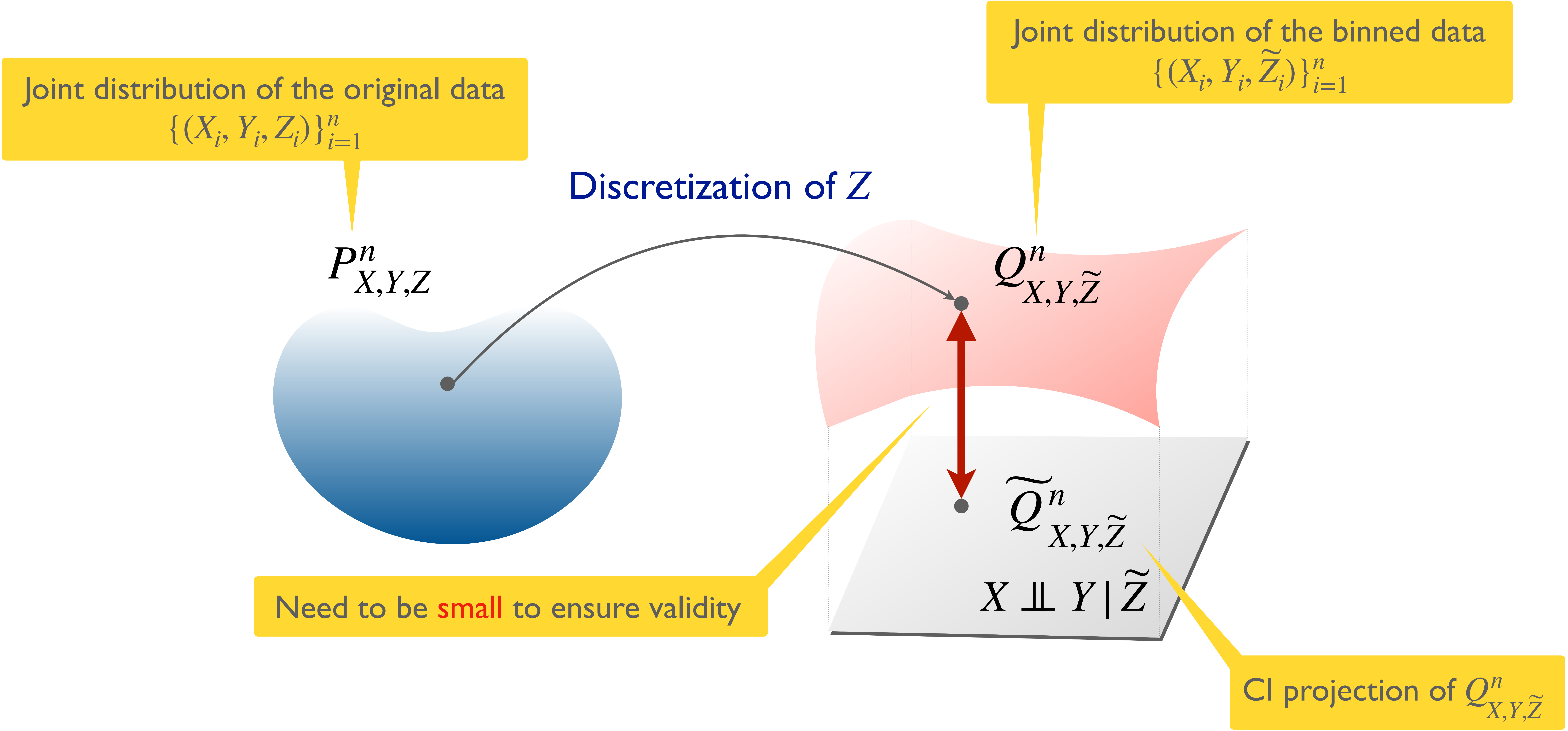
Validity of Local Permutation Tests



Validity of Local Permutation Tests



Validity of Local Permutation Tests



Validity of Local Permutation Tests

Lemma [KNBW 2022]

- Suppose that the distribution $P_{X,Y,Z}$ satisfies $X \perp\!\!\!\perp Y \mid Z$.

Validity of Local Permutation Tests

Lemma [KNBW 2022]

- Suppose that the distribution $P_{X,Y,Z}$ satisfies $X \perp\!\!\!\perp Y \,|\, Z$.
- Then for any $\alpha \in (0,1)$, the **Type I error** of the local permutation test is bounded above by

$$\mathbb{P}_{P^n_{X,Y,Z}}(T > q_{1-\alpha}) \leq \alpha + d_{\mathrm{TV}}\left(Q^n_{X,Y,\tilde{Z}}, \widetilde{Q}^n_{X,Y,\tilde{Z}}\right)$$

* Total variation distance

$$d_{\mathrm{TV}}(P,Q) = \sup_{A \in \mathcal{F}} |P(A) - Q(A)| = \frac{1}{2}\|P - Q\|_1$$

Validity of Local Permutation Tests

Lemma [KNBW 2022]

- Suppose that the distribution $P_{X,Y,Z}$ satisfies $X \perp\!\!\!\perp Y \mid Z$.
- Then for any $\alpha \in (0,1)$, the **Type I error** of the local permutation test is bounded above by

CI projection of $Q^n_{X,Y,\widetilde{Z}}$

$$\mathbb{P}_{P^n_{X,Y,Z}}(T > q_{1-\alpha}) \leq \alpha + d_{\mathrm{TV}}\big(Q^n_{X,Y,\widetilde{Z}}, \widetilde{Q}^n_{X,Y,\widetilde{Z}}\big)$$

Discretized Distribution

Validity of Local Permutation Tests

Lemma [KNBW 2022]

- Suppose that the distribution $P_{X,Y,Z}$ satisfies $X \perp\!\!\!\perp Y \mid Z$.
- Then for any $\alpha \in (0,1)$, the **Type I error** of the local permutation test is bounded above by

$$\mathbb{P}_{P^n_{X,Y,Z}}(T > q_{1-\alpha}) \leq \alpha + d_{\mathrm{TV}}\big(Q^n_{X,Y,\widetilde{Z}}, \widetilde{Q}^n_{X,Y,\widetilde{Z}}\big)$$

The bound is **universal** but **abstract**

Next goal: make this bound more **explicit** under smoothness conditions

Validity of Local Permutation Tests

Definition [Generalized Hellinger distance]

Given $\gamma \geq 1$, the **generalized Hellinger distance** with parameter γ between P and Q is defined as

$$d_\gamma(P, Q) = \left(\frac{1}{2} \int |p^{1/\gamma} - q^{1/\gamma}|^\gamma \, d\mu \right)^{1/\gamma}$$

Validity of Local Permutation Tests

Definition [Generalized Hellinger distance]

Given $\gamma \geq 1$, the **generalized Hellinger distance** with parameter γ between P and Q is defined as

$$d_\gamma(P, Q) = \left(\frac{1}{2} \int |p^{1/\gamma} - q^{1/\gamma}|^\gamma \, d\mu \right)^{1/\gamma}$$

For example

$$d_\gamma(P, Q) = \mathrm{TV}(P, Q) \quad \text{when } \gamma = 1$$

$$d_\gamma(P, Q) = \mathrm{Hellinger}(P, Q) \quad \text{when } \gamma = 2$$

Validity of Local Permutation Tests

Definition [γ-Hellinger Lipschitzness]

Let $\mathscr{P}_{0,\gamma}(L) \subset \mathscr{P}_0$ be the collection of null distributions such that for all $z, z' \in \mathscr{Z}$

$$d_\gamma(P_{X\,|\,Z=z}, P_{X\,|\,Z=z'}) \leq L\delta(z, z') \quad \text{and} \quad d_\gamma(P_{Y\,|\,Z=z}, P_{Y\,|\,Z=z'}) \leq L\delta(z, z')$$

where $\delta(z, z')$ is a distance between z and z' (e.g., Euclidean distance)

Validity of Local Permutation Tests

Definition [γ-Hellinger Lipschitzness]

Let $\mathscr{P}_{0,\gamma}(L) \subset \mathscr{P}_0$ be the collection of null distributions such that for all $z, z' \in \mathscr{Z}$

$$d_\gamma(P_{X \mid Z=z}, P_{X \mid Z=z'}) \leq L\delta(z, z') \quad \text{and} \quad d_\gamma(P_{Y \mid Z=z}, P_{Y \mid Z=z'}) \leq L\delta(z, z')$$

where $\delta(z, z')$ is a distance between z and z' (e.g., Euclidean distance)

In a nutshell

Both $P_{X \mid Z=z}$ and $P_{Y \mid Z=z}$ are **smooth** functions with respect to z

Validity of Local Permutation Tests

Theorem [KNBW 2022]

- Let h be the maximum diameter of bins $B_1, \ldots, B_M$

Validity of Local Permutation Tests

Theorem [KNBW 2022]

- Let h be the maximum diameter of bins $B_1, \ldots, B_M$
- For any $\alpha \in (0,1)$, the **Type I error** of the local permutation test over $\mathscr{P}_{0,\gamma}(L)$ is bounded above by

$$\sup_{P_{X,Y,Z} \in \mathscr{P}_{0,\gamma}(L)} \mathbb{P}_{P^n_{X,Y,Z}}(T > q_{1-\alpha}) \leq \begin{cases} \alpha + O\left(\sqrt{n} h^{\gamma}\right), & \text{if } \gamma \in [1,2] \\ \alpha + O\left(\sqrt{n} h^{2}\right), & \text{if } \gamma > 2 \end{cases}$$

Validity of Local Permutation Tests

$$\sup_{P_{X,Y,Z} \in \mathscr{P}_{0,\gamma}(L)} \mathbb{P}_{P^n_{X,Y,Z}}(T > q_{1-\alpha}) \leq \begin{cases} \alpha + O\big(\sqrt{n}h^{\gamma}\big), & \text{if } \gamma \in [1,2] \\ \alpha + O\big(\sqrt{n}h^{2}\big), & \text{if } \gamma > 2 \end{cases}$$

Remark

- This result holds for **any binning-based test statistic T**

Validity of Local Permutation Tests

$$\sup_{P_{X,Y,Z} \in \mathscr{P}_{0,\gamma}(L)} \mathbb{P}_{P^n_{X,Y,Z}}(T > q_{1-\alpha}) \leq \begin{cases} \alpha + O\big(\sqrt{n}h^{\gamma}\big), & \text{if } \gamma \in [1,2] \\ \alpha + O\big(\sqrt{n}h^{2}\big), & \text{if } \gamma > 2 \end{cases}$$

Remark

- This result holds for **any binning-based test statistic T**
- If $\sqrt{n}h^{\gamma} \to 0$, then the local permutation test is **asymptotically valid**

Validity of Local Permutation Tests

$$\sup_{P_{X,Y,Z} \in \mathscr{P}_{0,\gamma}(L)} \mathbb{P}_{P^n_{X,Y,Z}}(T > q_{1-\alpha}) \leq \begin{cases} \alpha + O\big(\sqrt{n}h^{\gamma}\big), & \text{if } \gamma \in [1,2] \\ \alpha + O\big(\sqrt{n}h^{2}\big), & \text{if } \gamma > 2 \end{cases}$$

Remark

- This result holds for **any binning-based test statistic T**
- If $\sqrt{n}h^{\gamma} \to 0$, then the local permutation test is **asymptotically valid**
- If $\sqrt{n}h^{\gamma} \to \infty$, then we can construct T such that **the type I error blows up**

Validity of Local Permutation Tests

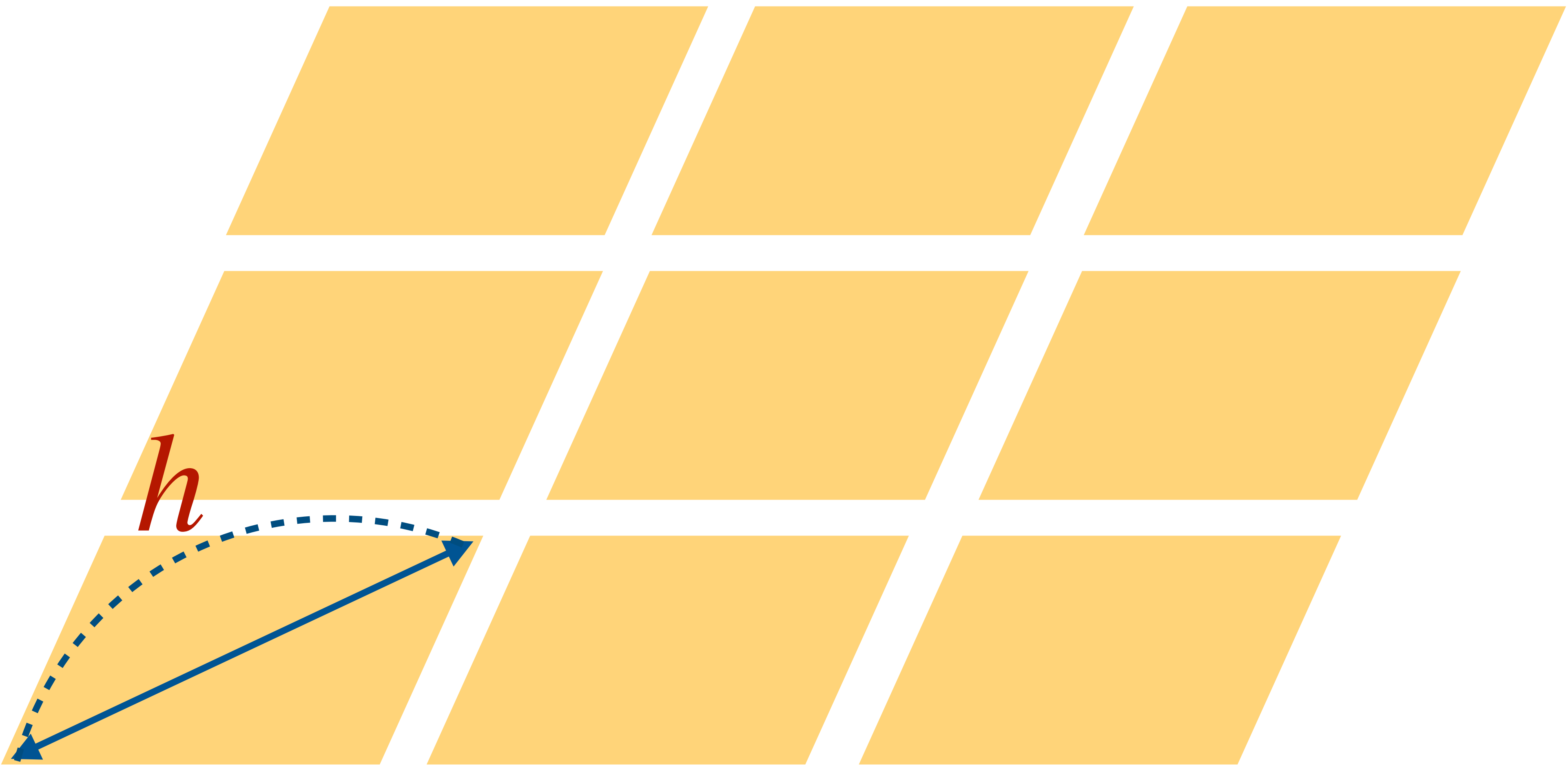
$$\sup_{P_{X,Y,Z} \in \mathscr{P}_{0,\gamma}(L)} \mathbb{P}_{P^n_{X,Y,Z}}(T > q_{1-\alpha}) \leq \begin{cases} \alpha + O\big(\sqrt{n}h^{\gamma}\big), & \text{if } \gamma \in [1,2] \\ \alpha + O\big(\sqrt{n}h^{2}\big), & \text{if } \gamma > 2 \end{cases}$$

Remark

- This result holds for **any binning-based test statistic T**
- If $\sqrt{n}h^{\gamma} \to 0$, then the local permutation test is **asymptotically valid**
- If $\sqrt{n}h^{\gamma} \to \infty$, then we can construct T such that **the type I error blows up**
- We need to take h to be small to ensure **type I error control**

Trade-off between Type I error and Type II error

Small h

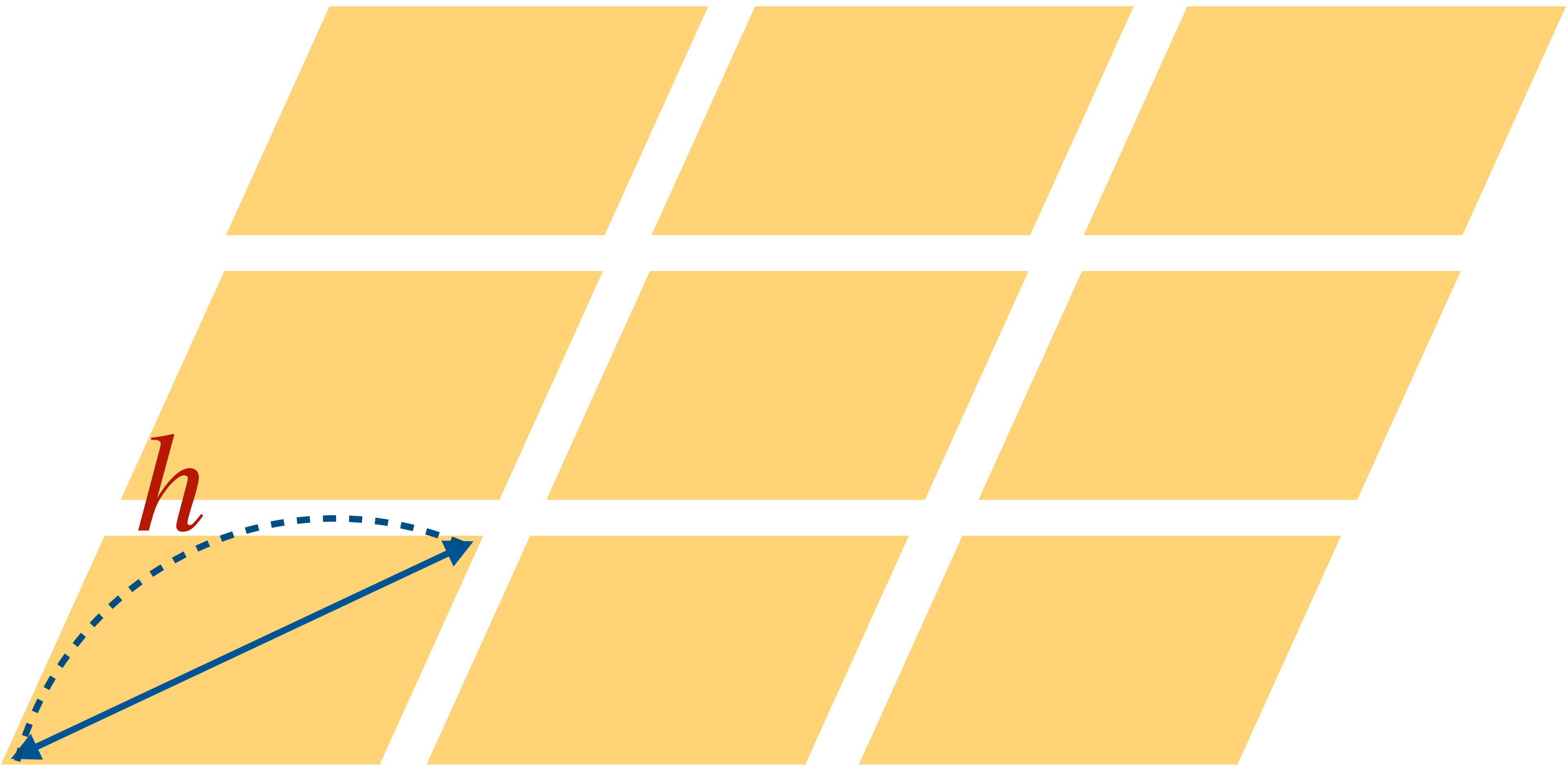


Large h

Trade-off between Type I error and Type II error

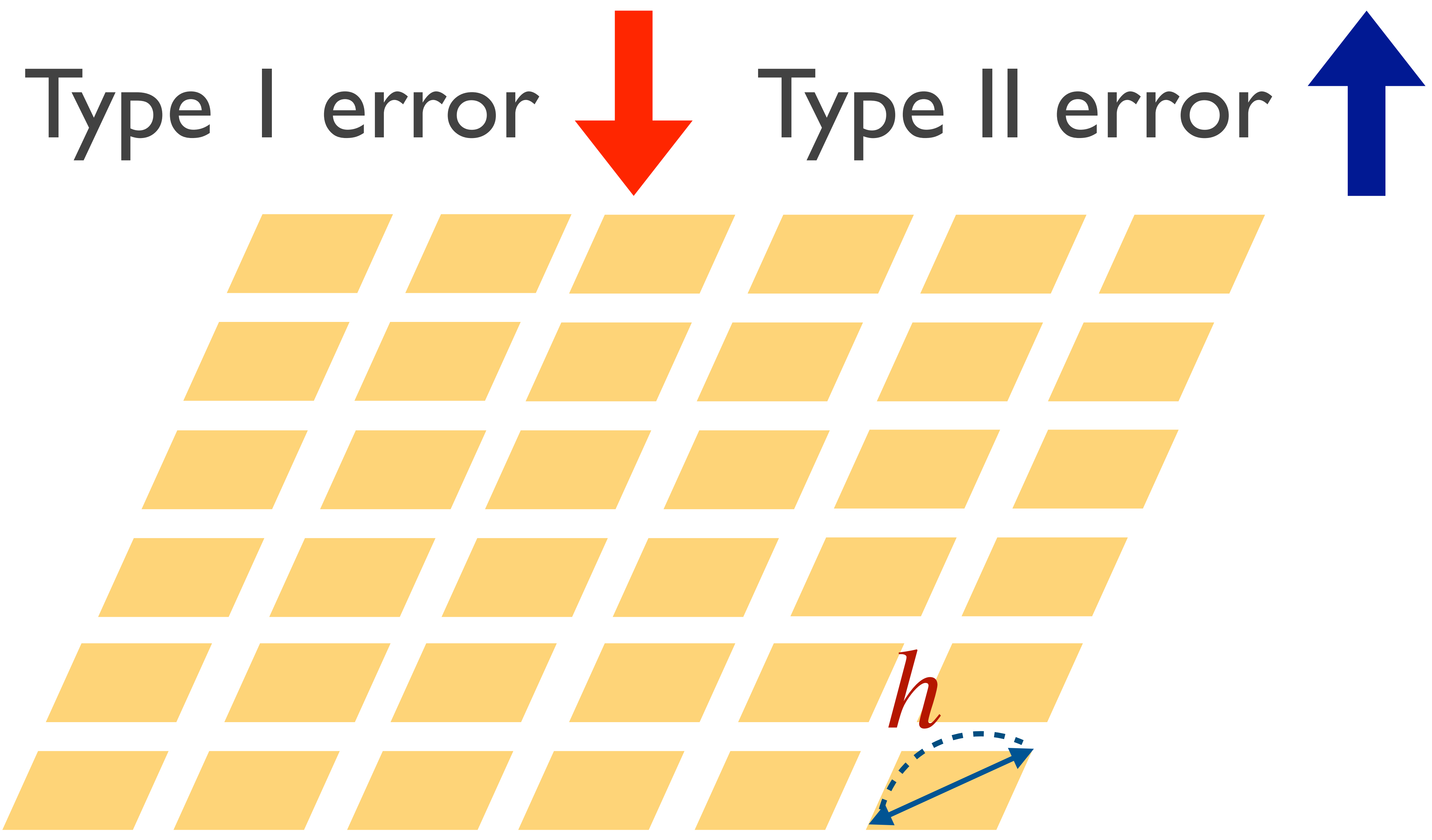
Type I error ↓ Type II error ↑

Small h

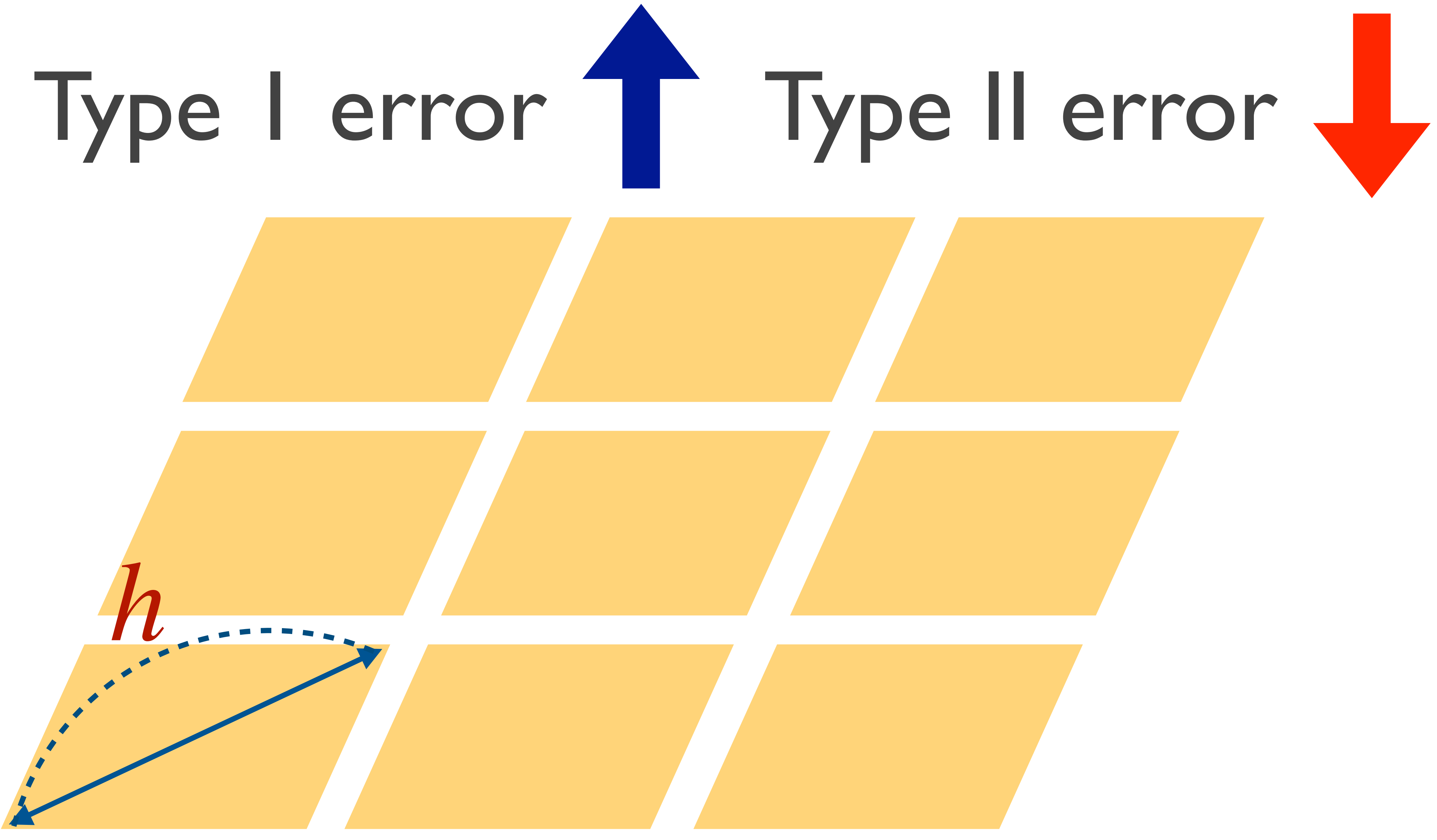


Large h

Trade-off between Type I error and Type II error

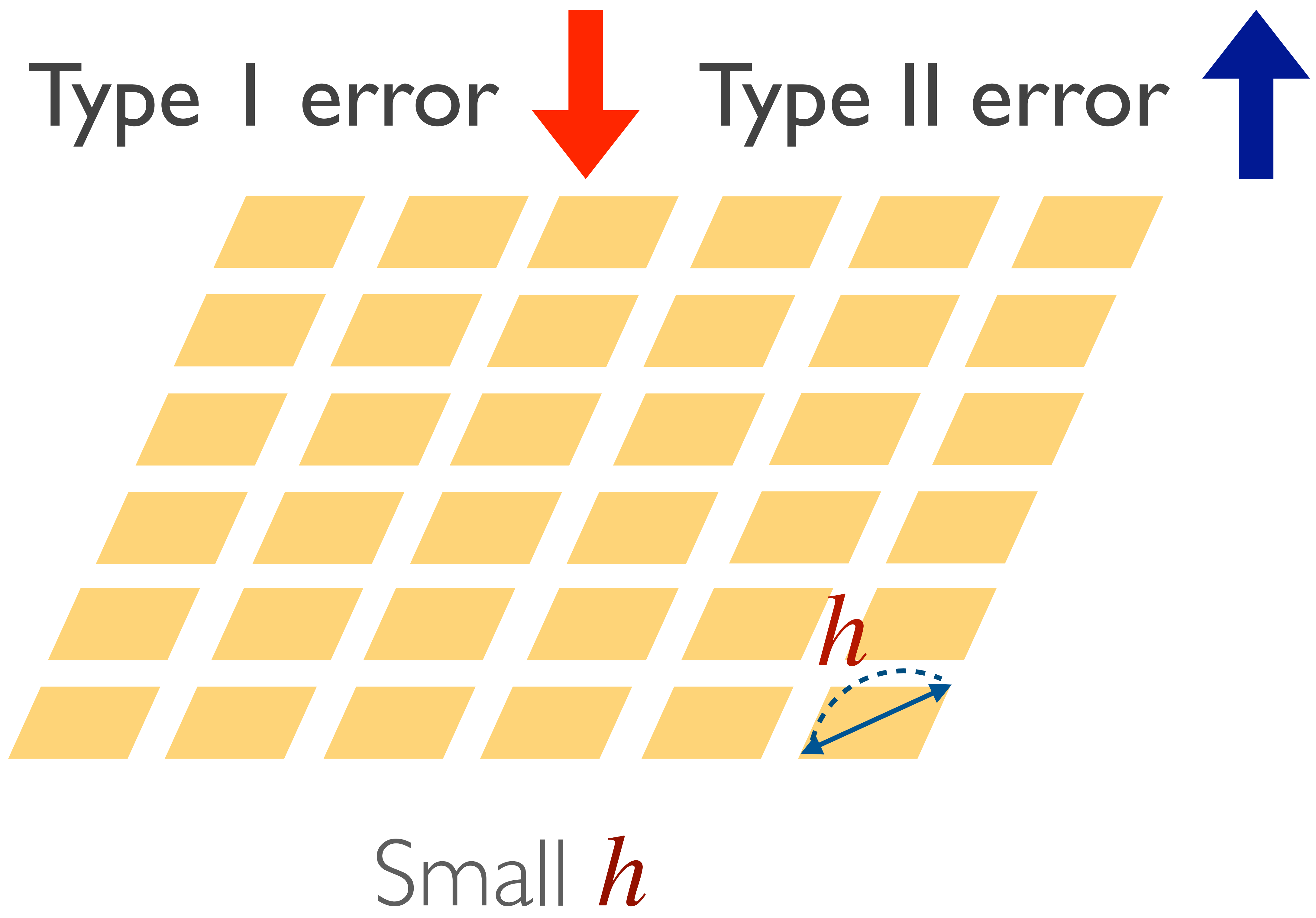


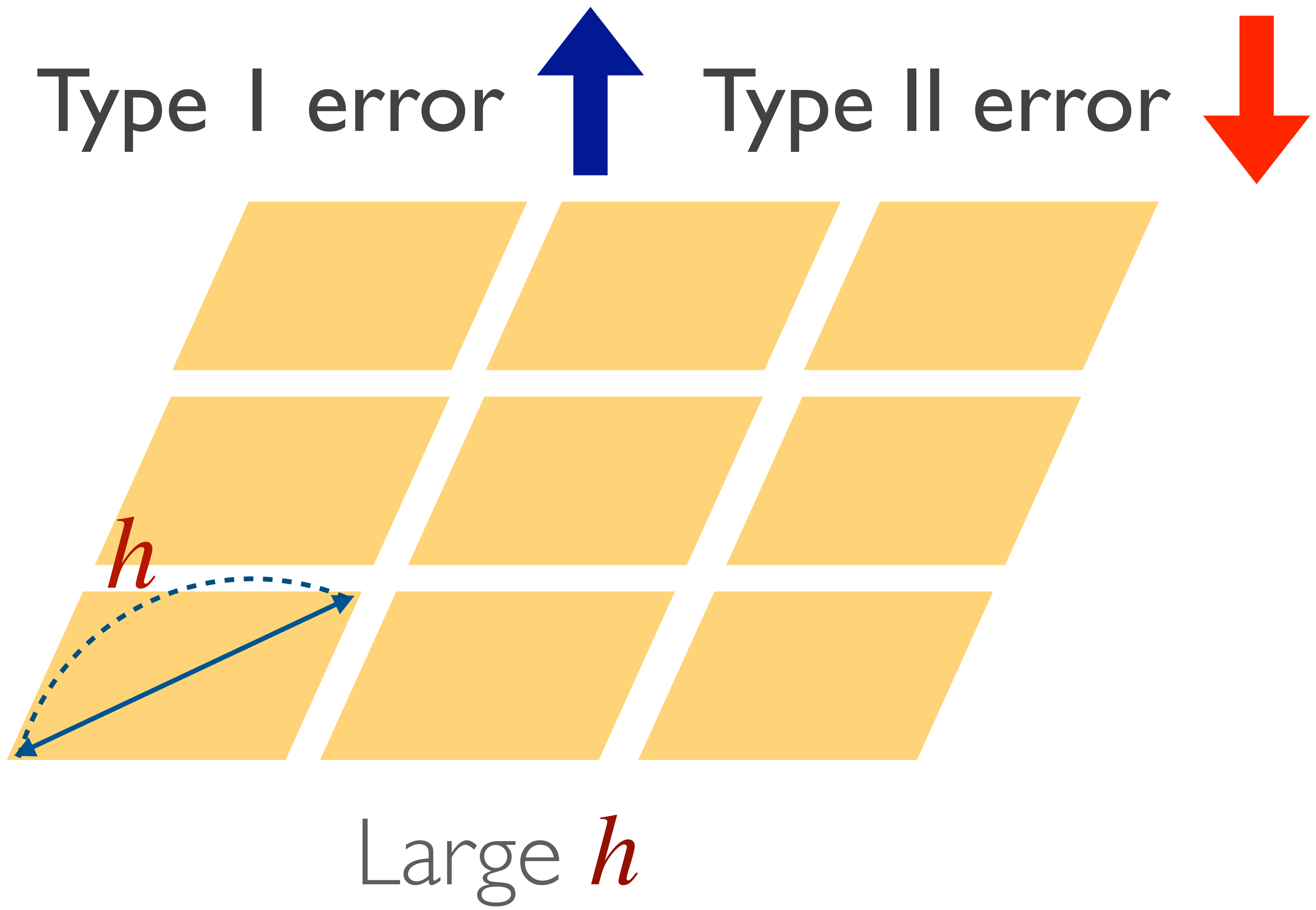
Small h



Large h

Trade-off between Type I error and Type II error





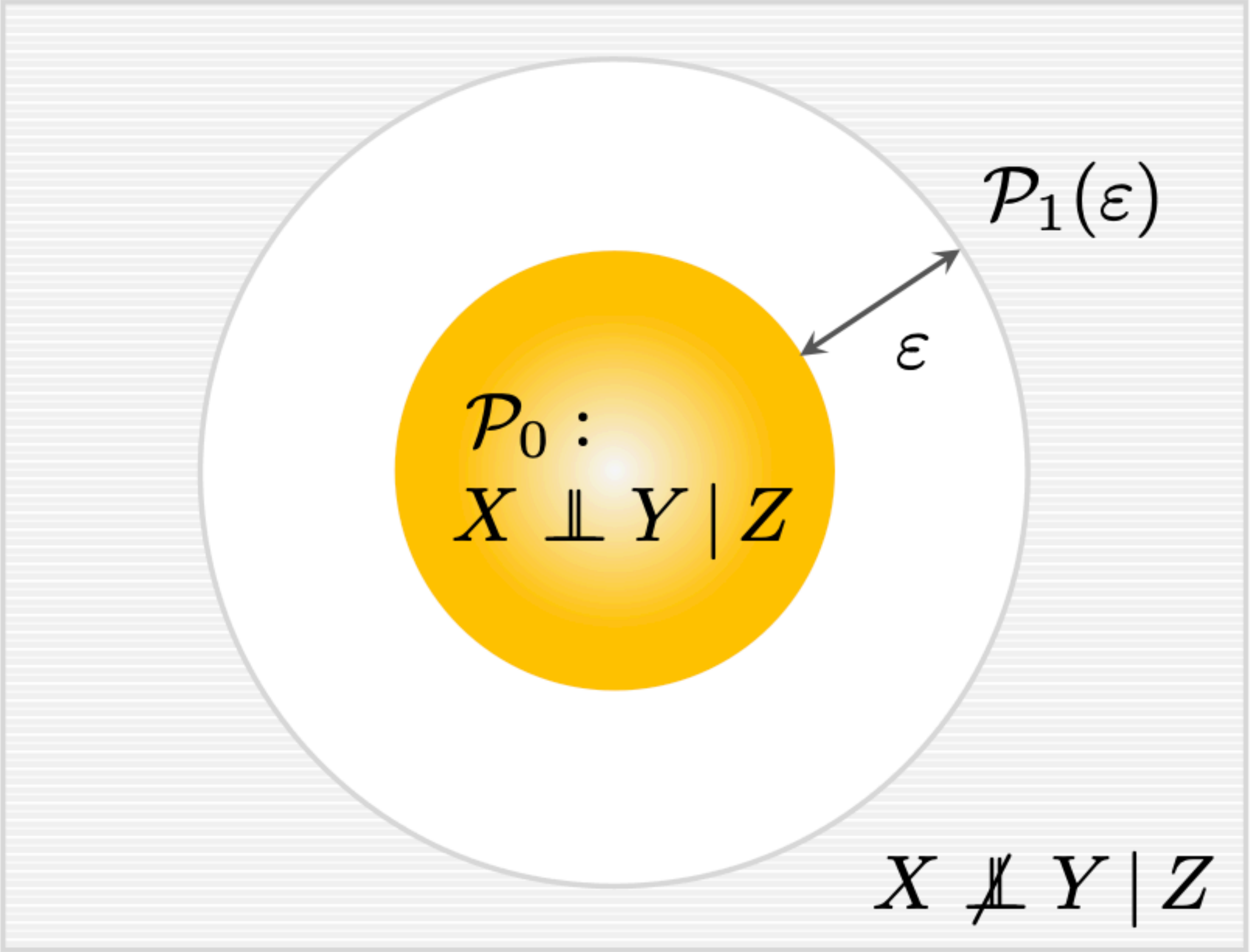
We need to balance this trade-off

Local permutation tests can achieve a certain **minimax optimality** by choosing h carefully

Detour: Minimax Testing Framework

- $\mathrm{Risk}(\epsilon) = \inf_{\phi \in \Phi} \left\{ \underbrace{\sup_{P \in \mathscr{P}_0} \mathbb{E}_P[\phi]}_{\text{Worst-case type I error}} + \underbrace{\sup_{P \in \mathscr{P}_1(\epsilon)} \mathbb{E}_P[1 - \phi]}_{\text{Worst-case type II error}} \right\}$

A set of all possible tests



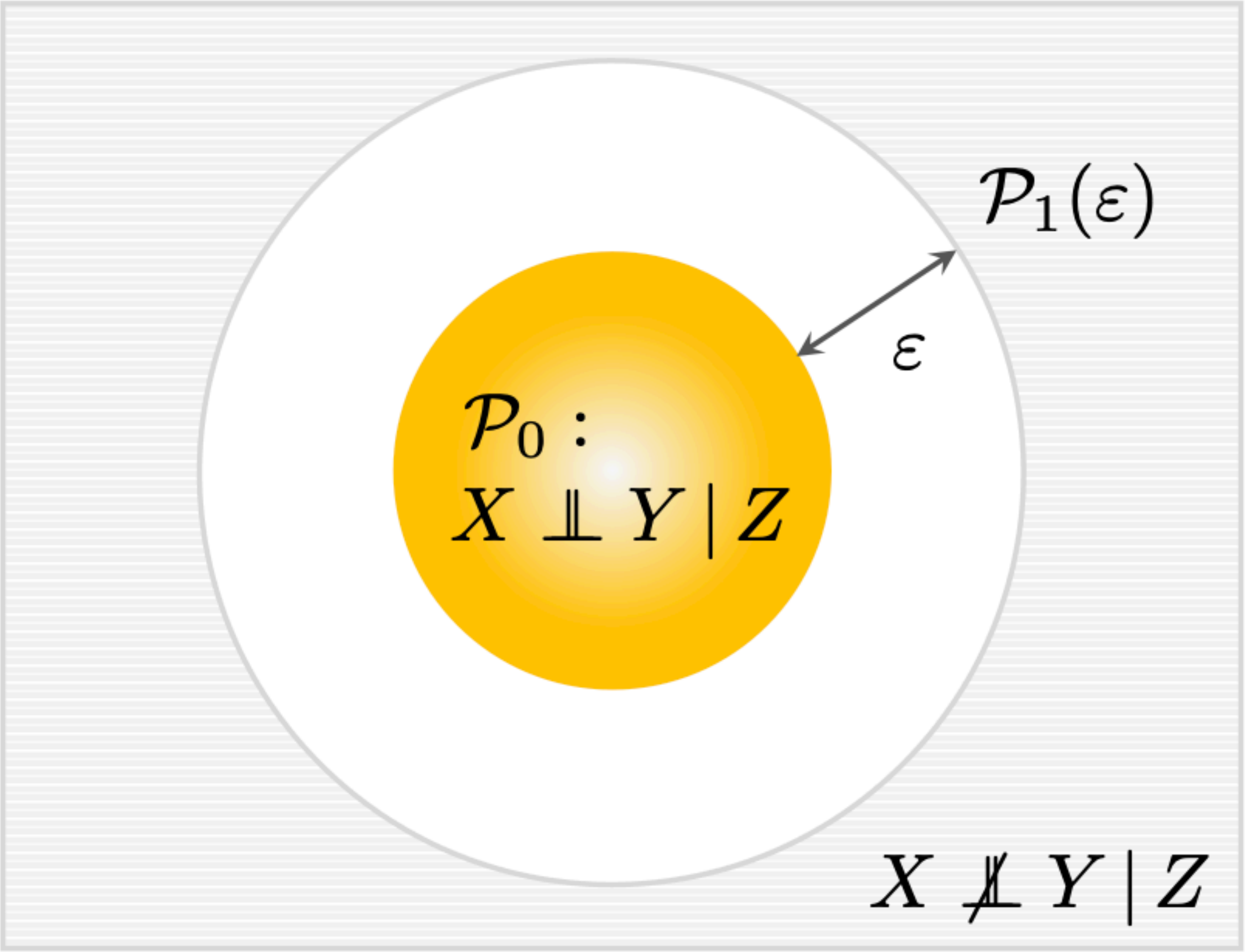
Detour: Minimax Testing Framework

- $\mathrm{Risk}(\epsilon) = \inf_{\phi \in \Phi} \left\{ \sup_{P \in \mathscr{P}_0} \mathbb{E}_P[\phi] + \sup_{P \in \mathscr{P}_1(\epsilon)} \mathbb{E}_P[1-\phi] \right\}$

Worst-case type I error   Worst-case type II error

A set of all possible tests

- $\epsilon^\star = \inf\left\{\epsilon : \mathrm{Risk}(\epsilon) \leq 0.1\right\}$



Detour: Minimax Testing Framework

- $\mathrm{Risk}(\epsilon) = \inf_{\phi \in \Phi} \left\{ \sup_{P \in \mathscr{P}_0} \mathbb{E}_P[\phi] + \sup_{P \in \mathscr{P}_1(\epsilon)} \mathbb{E}_P[1-\phi] \right\}$

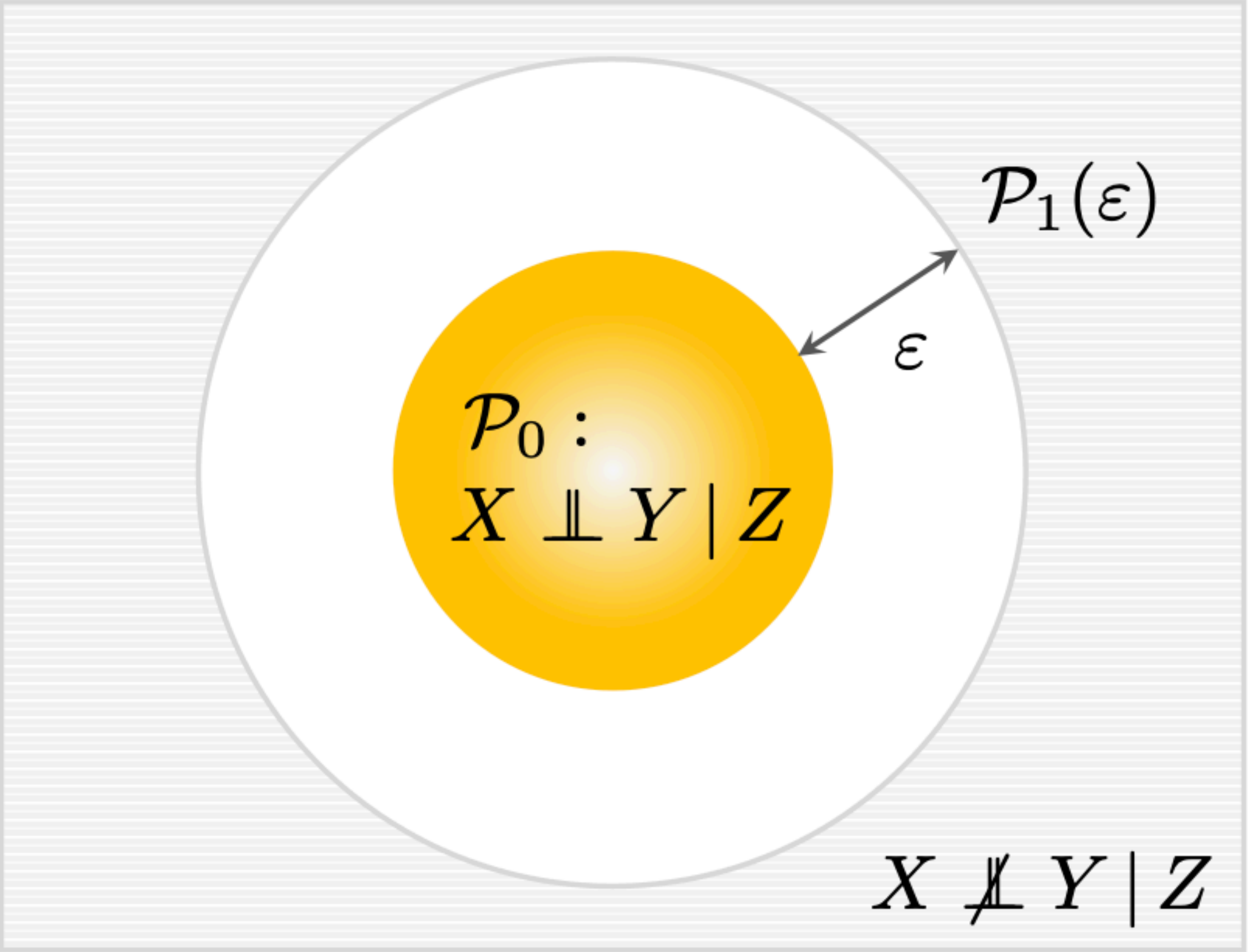
  Worst-case type I error   Worst-case type II error

  A set of all possible tests

- $\epsilon^\star = \inf\left\{\epsilon : \mathrm{Risk}(\epsilon) \le 0.1\right\}$

- We say that ϕ is **minimax optimal** if

$$\sup_{P \in \mathscr{P}_0} \mathbb{E}_P[\phi] + \sup_{P \in \mathscr{P}_1(\epsilon^\star)} \mathbb{E}_P[1-\phi] \le 0.1$$



Detour: Minimax Testing Framework

- $\mathrm{Risk}(\epsilon) = \inf_{\phi \in \Phi} \left\{ \sup_{P \in \mathscr{P}_0} \mathbb{E}_P[\phi] + \sup_{P \in \mathscr{P}_1(\epsilon)} \mathbb{E}_P[1 - \phi] \right\}$

Worst-case type I error

Worst-case type II error

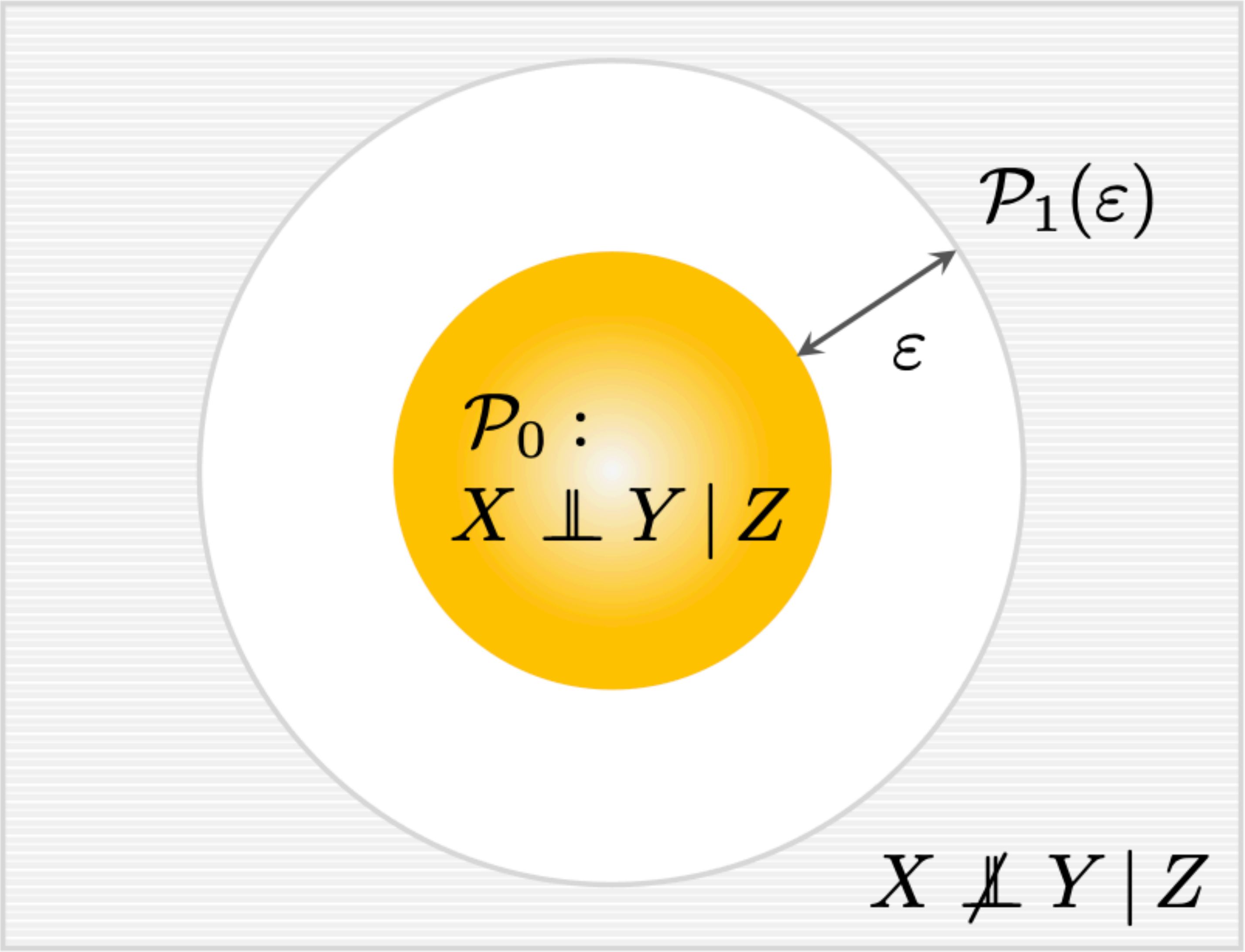
A set of all possible tests

- $\epsilon^\star = \inf\left\{\epsilon : \mathrm{Risk}(\epsilon) \leq 0.1\right\}$

- We say that ϕ is minimax "rate" optimal if

$$\sup_{P \in \mathscr{P}_0} \mathbb{E}_P[\phi] + \sup_{P \in \mathscr{P}_1(\epsilon)} \mathbb{E}_P[1 - \phi] \leq 0.1$$

whenever $\epsilon \asymp \epsilon^\star$



Optimality results of Neykov et al. (2021)

Let $(X, Y, Z) \in [0,1]^3$

Optimality results of Neykov et al. (2021)

Let $(X, Y, Z) \in [0,1]^3$

Class of null distributions

$\mathscr{P}_{0,\mathrm{NBL}}$: TV or χ^2 smooth (Z) + $X \perp\!\!\!\perp Y \mid Z$

Optimality results of Neykov et al. (2021)

Let $(X, Y, Z) \in [0,1]^3$

Class of null distributions

$$\mathscr{P}_{0,\mathrm{NBL}}\text{: TV or } \chi^2 \text{ smooth } (Z) + X \perp\!\!\!\perp Y \mid Z$$

Class of alternative distributions

$$\mathscr{P}_{1,\mathrm{NBL}}(\epsilon)\text{: TV smooth } (Z) + s\text{-Hölder smooth } (X, Y) + \inf_{P \in \mathscr{P}_0} d_{TV}(P, Q) \geq \epsilon$$

Optimality results of Neykov et al. (2021)

Let $(X, Y, Z) \in [0,1]^3$

$H_0 : P \in \mathscr{P}_{0,\mathrm{NBL}}$ versus

$H_1 : P \in \mathscr{P}_{1,\mathrm{NBL}}(\epsilon)$

Class of null distributions

$\mathscr{P}_{0,\mathrm{NBL}}$: TV or χ^2 smooth (Z) + $X \perp\!\!\!\perp Y \mid Z$

Class of alternative distributions

$\mathscr{P}_{1,\mathrm{NBL}}(\epsilon)$: TV smooth (Z) + s-Hölder smooth (X, Y) + $\inf_{P \in \mathscr{P}_0} d_{TV}(P, Q) \geq \epsilon$

Neykov et al. (2021) prove that the **minimax testing rate** for this problem is

$$\epsilon_n^\star \asymp n^{-\frac{2s}{5s+2}}$$

In our work [KNBW 2022]

- We show that the same test statistic calibrated by the **local permutation method** achieves the **same minimax rate** $\epsilon_n^\star \asymp n^{-\frac{2s}{5s+2}}$

- **However**, the optimality can be obtained over a slightly **smaller class of null distributions** (**Hellinger smoothness** instead of **TV smoothness**)

$\Longrightarrow$ **Double-binning method**

In our work [KNBW 2022]

- We show that the same test statistic calibrated by the **local permutation method** achieves the **same minimax rate** $\epsilon_n^\star \asymp n^{-\frac{2s}{5s+2}}$

- **However**, the optimality can be obtained over a slightly **smaller class of null distributions** (**Hellinger smoothness** instead of **TV smoothness**)

$\Longrightarrow$ Double-binning method

In our work [KNBW 2022]

- We show that the same test statistic calibrated by the **local permutation method** achieves the **same minimax rate** $\epsilon_n^\star \asymp n^{-\frac{2s}{5s+2}}$

- **However**, the optimality can be obtained over a slightly **smaller class of null distributions** (**Hellinger smoothness** instead of **TV smoothness**)

$\Longrightarrow$ Double-binning method

Summary

- We presented a **theoretical foundation** of local permutation tests
- We revisited the **hardness of CI testing** and showed that the hardness is determined by "**no-collision**" probability
- We illustrated that the local permutation test with a careful binning scheme can lead to **minimax optimality**

Summary

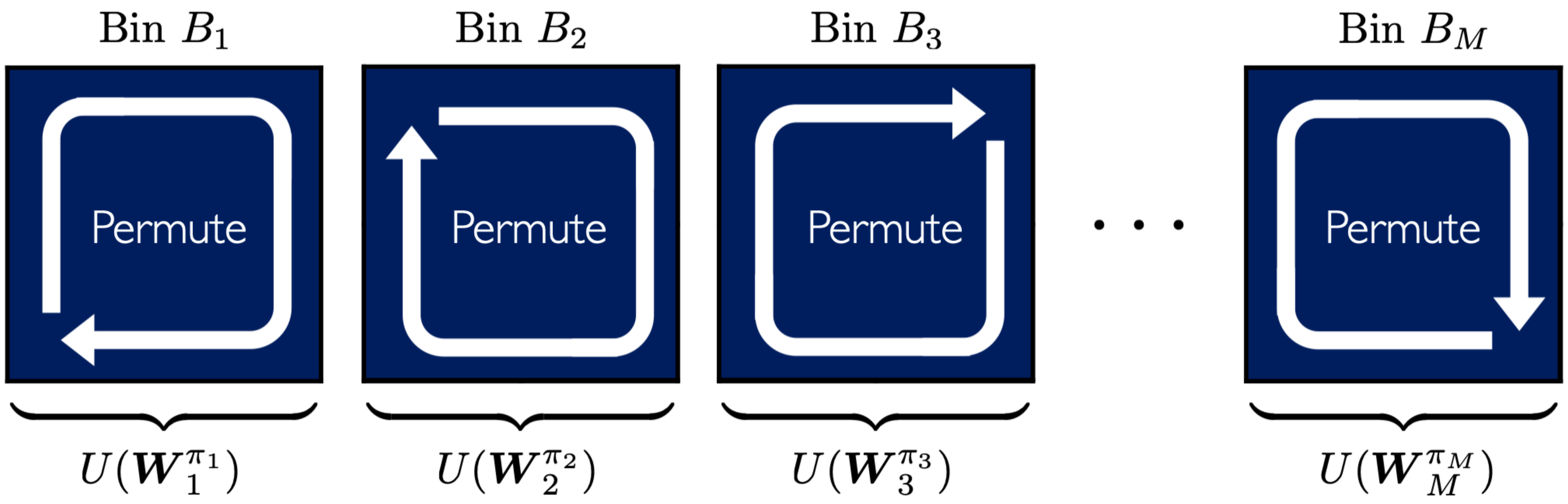
- We presented a **theoretical foundation** of local permutation tests
- We revisited the **hardness of CI testing** and showed that the hardness is determined by "**no-collision**" probability
- We illustrated that the local permutation test with a careful binning scheme can lead to **minimax optimality**

Summary

- We presented a **theoretical foundation** of local permutation tests
- We revisited the **hardness of CI testing** and showed that the hardness is determined by "**no-collision**" probability
- We illustrated that the local permutation test with a careful binning scheme can lead to **minimax optimality**

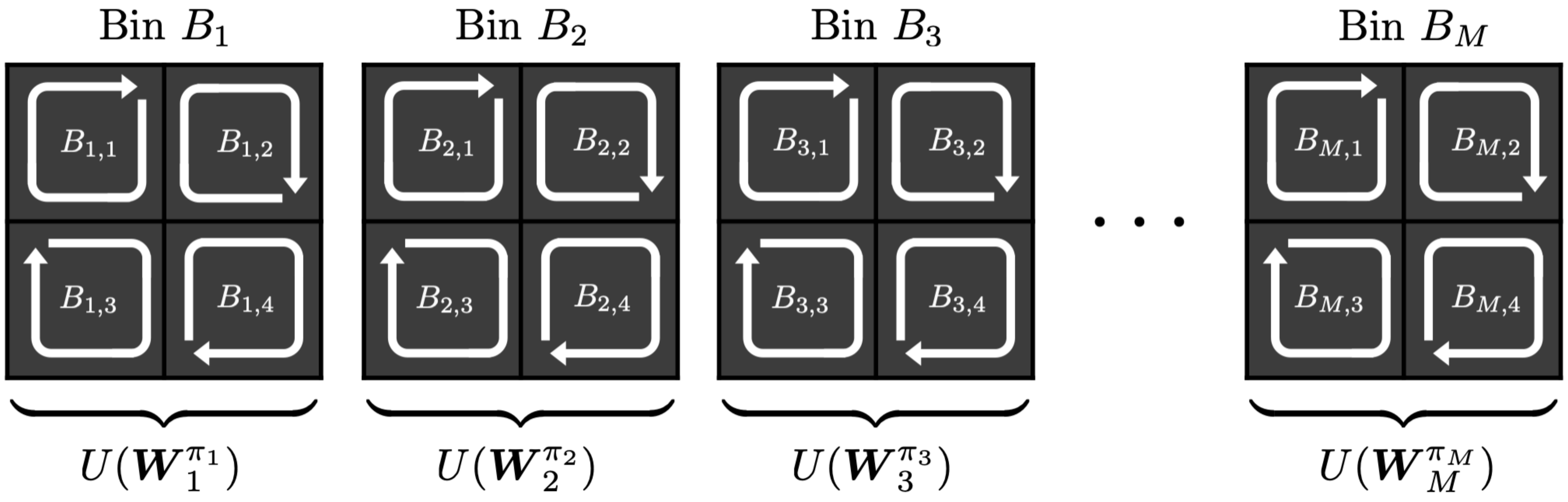
Thank you!

Double-binning method



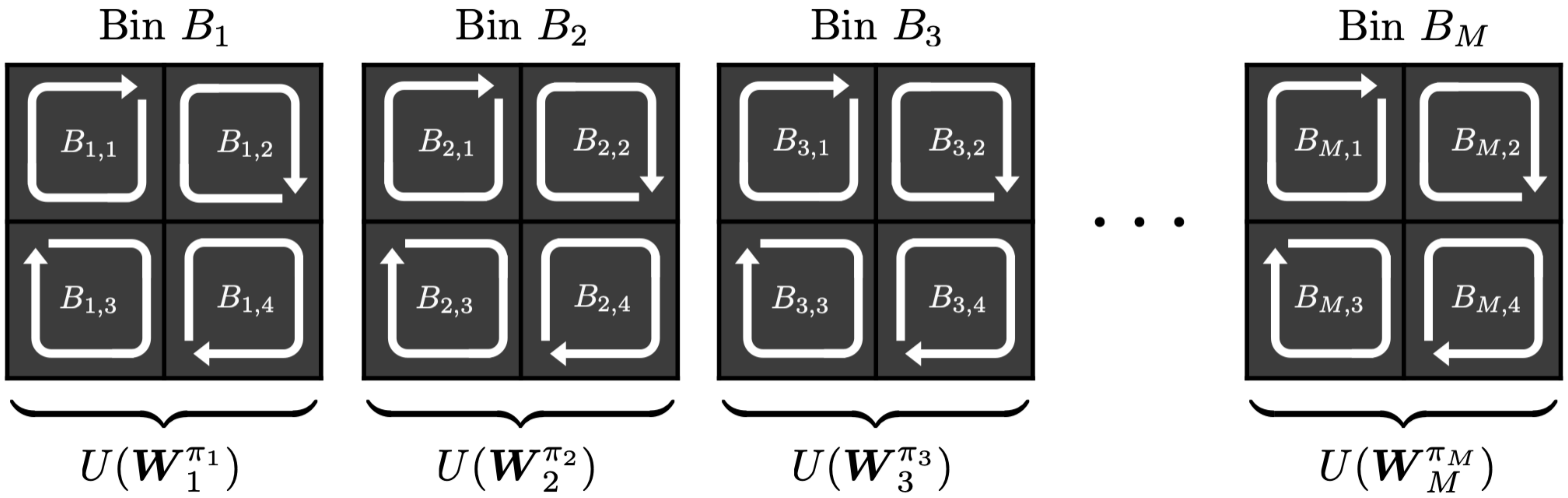
- Previously, we **permuted** a sample and **computed** a test statistic within **the same bin**

Double-binning method



- In **double-binning**, we consider bins of **two** distinct resolutions
  - **Permute** the observations within the **finer bins**
  - **Compute** test statistics over the **coarser bins**

Double-binning method



Decrease the type I error

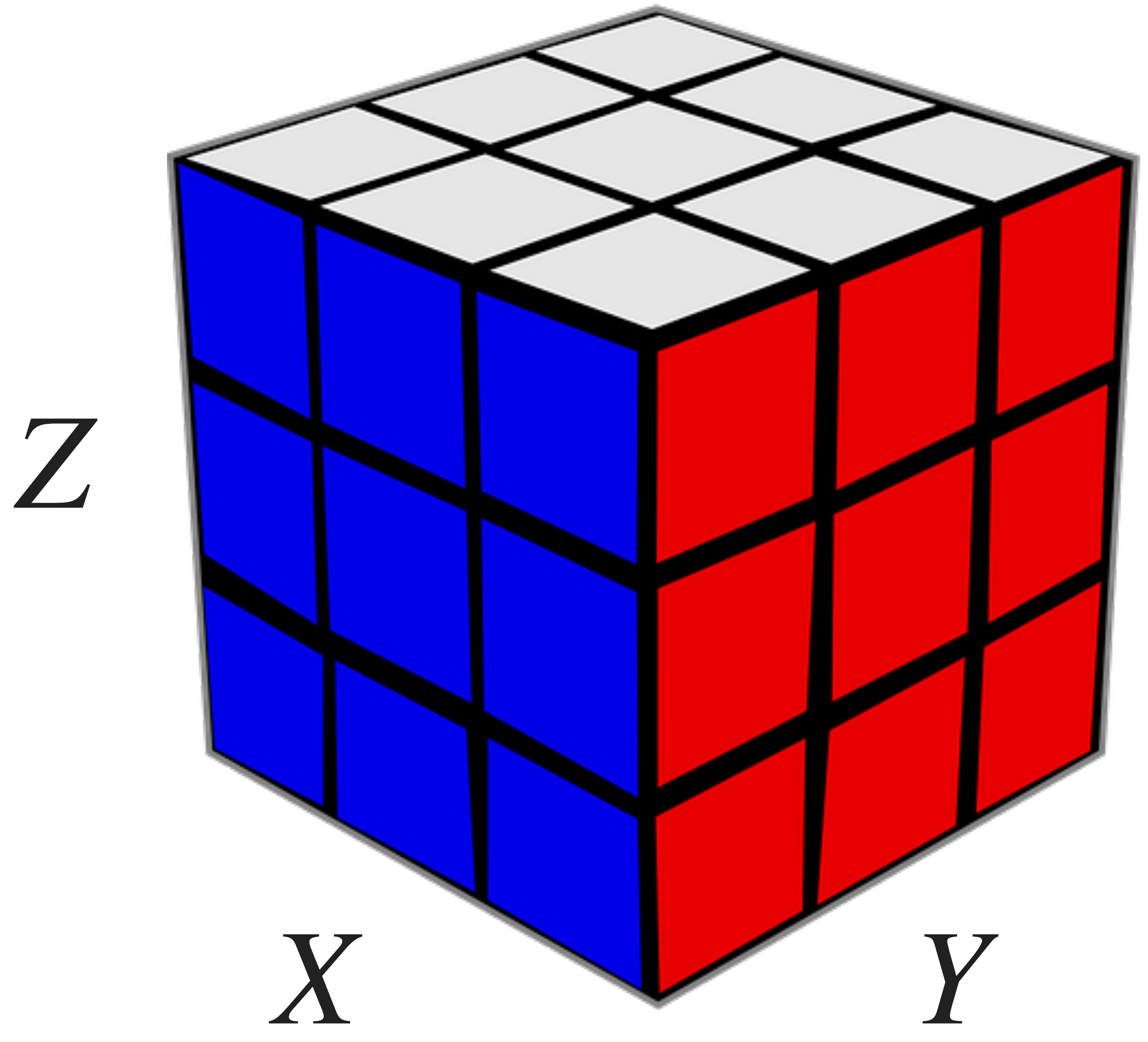
- In **double-binning**, we consider bins of **two** distinct resolutions
  - **Permute** the observations within the **finer bins**
  - **Compute** test statistics over the **coarser bins**

Increase the power

Optimality results of Neykov et al. (2021)

Procedure

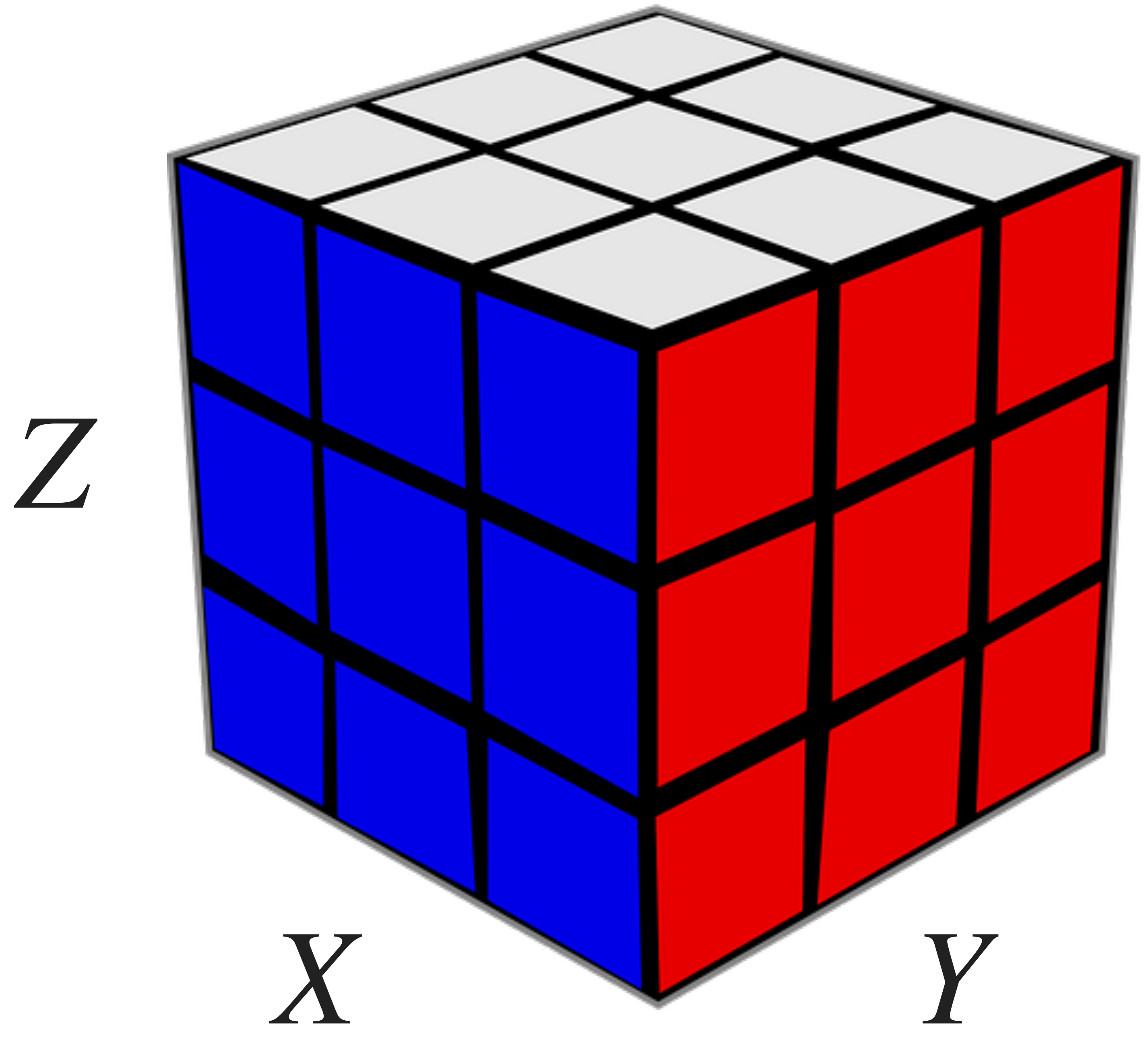
- Discretize (X, Y, Z) into $d_X \times d_Y \times d_Z$ bins



Optimality results of Neykov et al. (2021)

Procedure

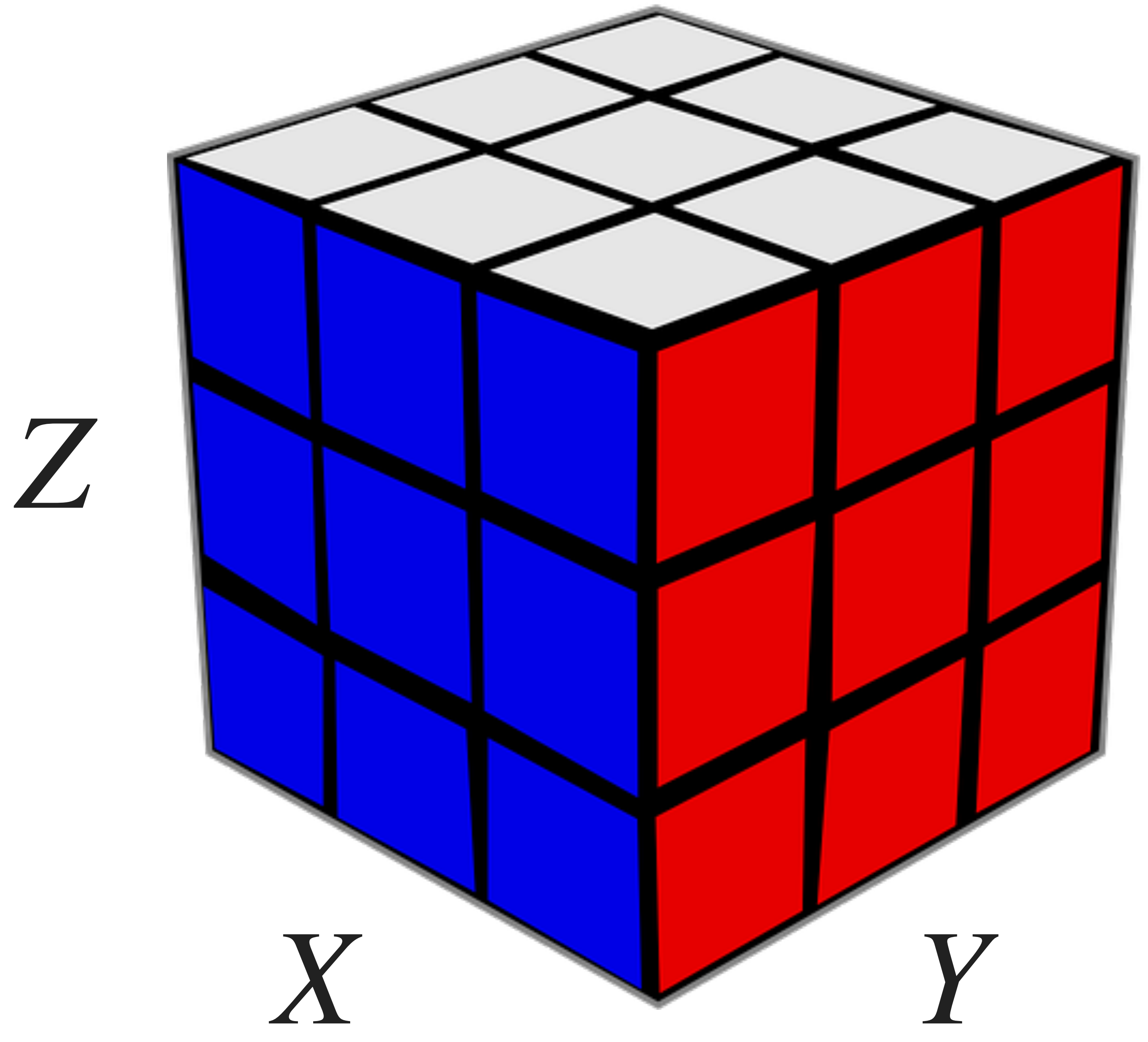
- Discretize (X, Y, Z) into $d_X \times d_Y \times d_Z$ bins
- Set $d_X = d_Y = \lceil n^{\frac{2}{5s+2}} \rceil$ and $d_Z = \lceil n^{\frac{2s}{5s+2}} \rceil$



Optimality results of Neykov et al. (2021)

Procedure

- Discretize (X, Y, Z) into $d_X \times d_Y \times d_Z$ bins
- Set $d_X = d_Y = \lceil n^{\frac{2}{5s+2}} \rceil$ and $d_Z = \lceil n^{\frac{2s}{5s+2}} \rceil$
- Compute the test statistic T proposed by Canonne et al. (2018) for **discrete CI testing**



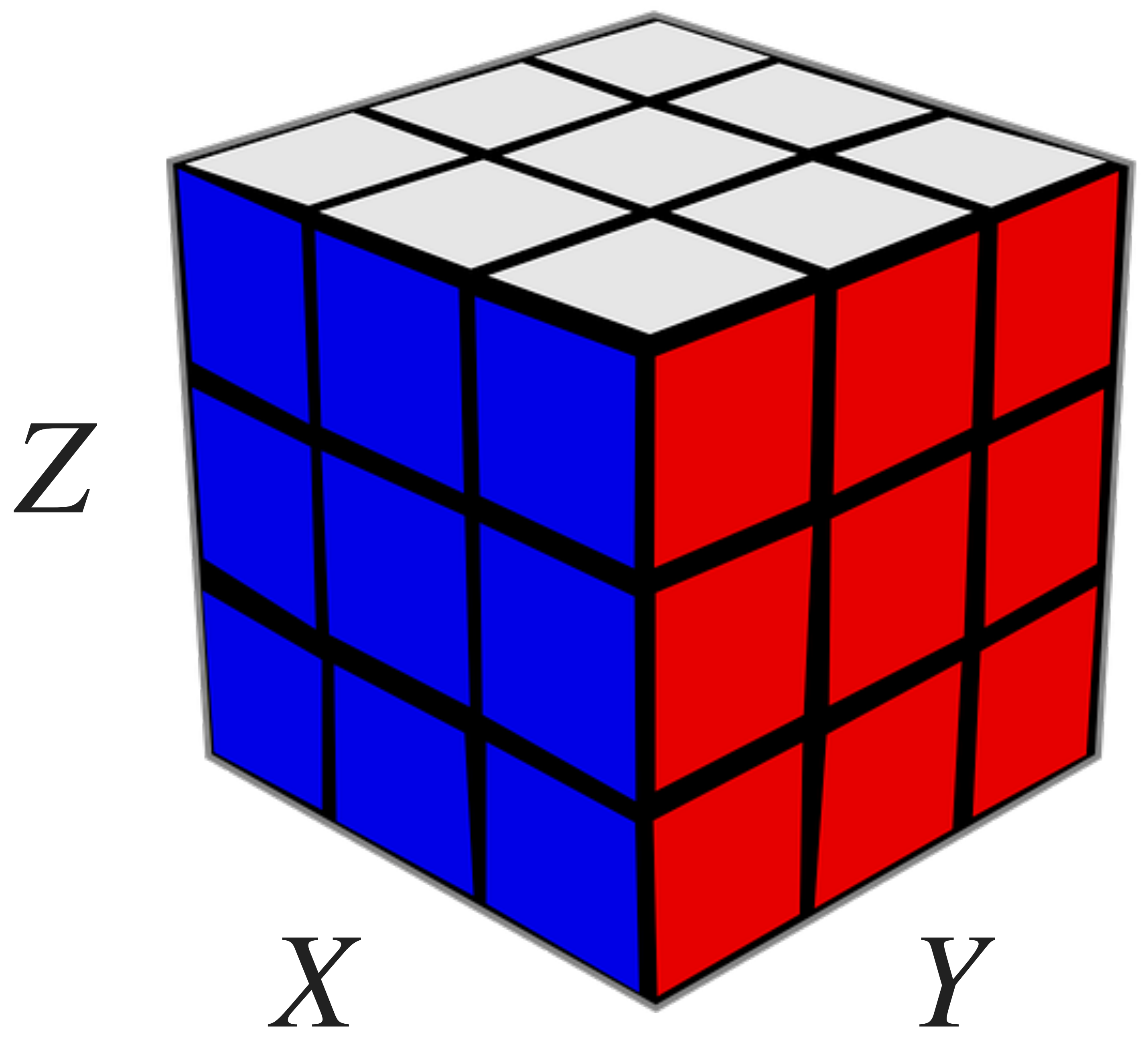
$$T = \sum_{i=1}^{d_Z} w_i U_i(\mathscr{D}_i)$$

Weighted sum of U-statistics

Optimality results of Neykov et al. (2021)

Procedure

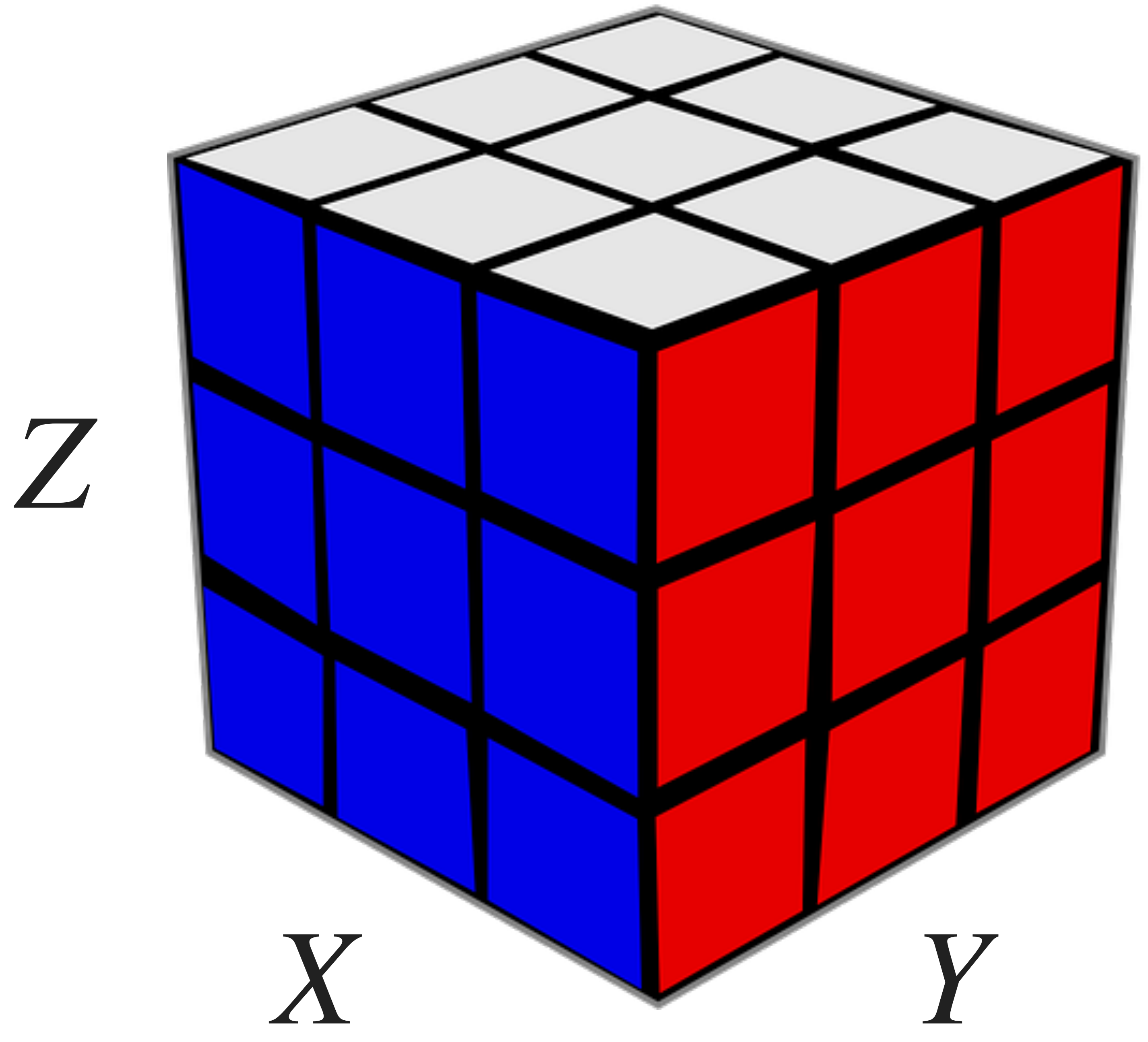
- Discretize (X, Y, Z) into $d_X \times d_Y \times d_Z$ bins
- Set $d_X = d_Y = \lceil n^{\frac{2}{5s+2}} \rceil$ and $d_Z = \lceil n^{\frac{2s}{5s+2}} \rceil$
- Compute the test statistic T proposed by Canonne et al. (2018) for **discrete CI testing**
- **Reject** the null when $T > Cd_Z^{1/2}$ for a sufficiently large C



Optimality results of Neykov et al. (2021)

Procedure

- Discretize (X, Y, Z) into $d_X \times d_Y \times d_Z$ bins
- Set $d_X = d_Y = \lceil n^{\frac{2}{5s+2}} \rceil$ and $d_Z = \lceil n^{\frac{2s}{5s+2}} \rceil$
- Compute the test statistic T proposed by Canonne et al. (2018) for **discrete CI testing**



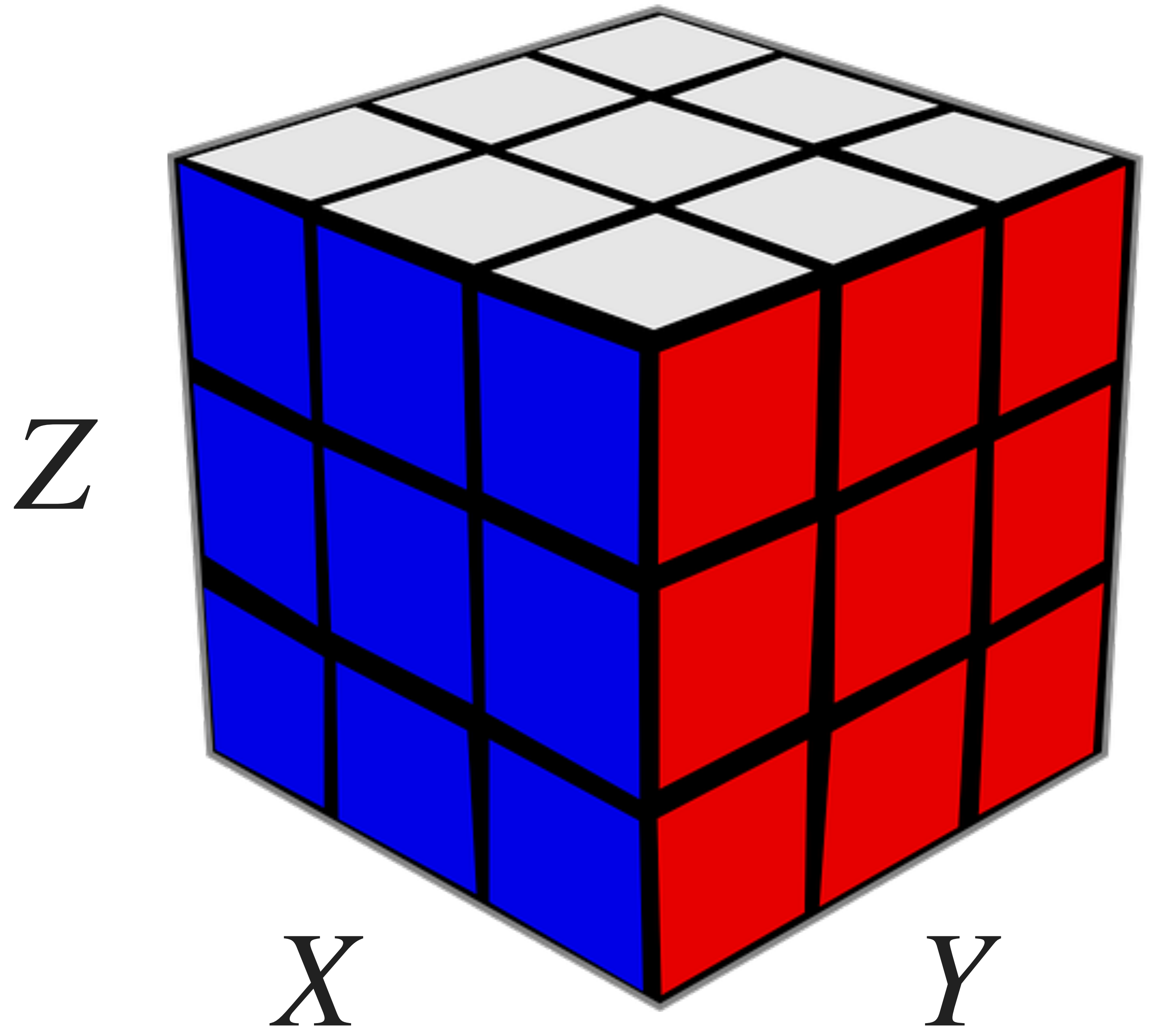
- **Reject** the null when $T > Cd_Z^{1/2}$ for a sufficiently large C

- It achieves the **minimax optimal rate**

Optimality results of Neykov et al. (2021)

Procedure

- Discretize (X, Y, Z) into $d_X \times d_Y \times d_Z$ bins
- Set $d_X = d_Y = \lceil n^{\frac{2}{5s+2}} \rceil$ and $d_Z = \lceil n^{\frac{2s}{5s+2}} \rceil$
- Compute the test statistic T proposed by Canonne et al. (2018) for **discrete CI testing**



- **Reject** the null when $T > Cd_Z^{1/2}$ for a sufficiently large C

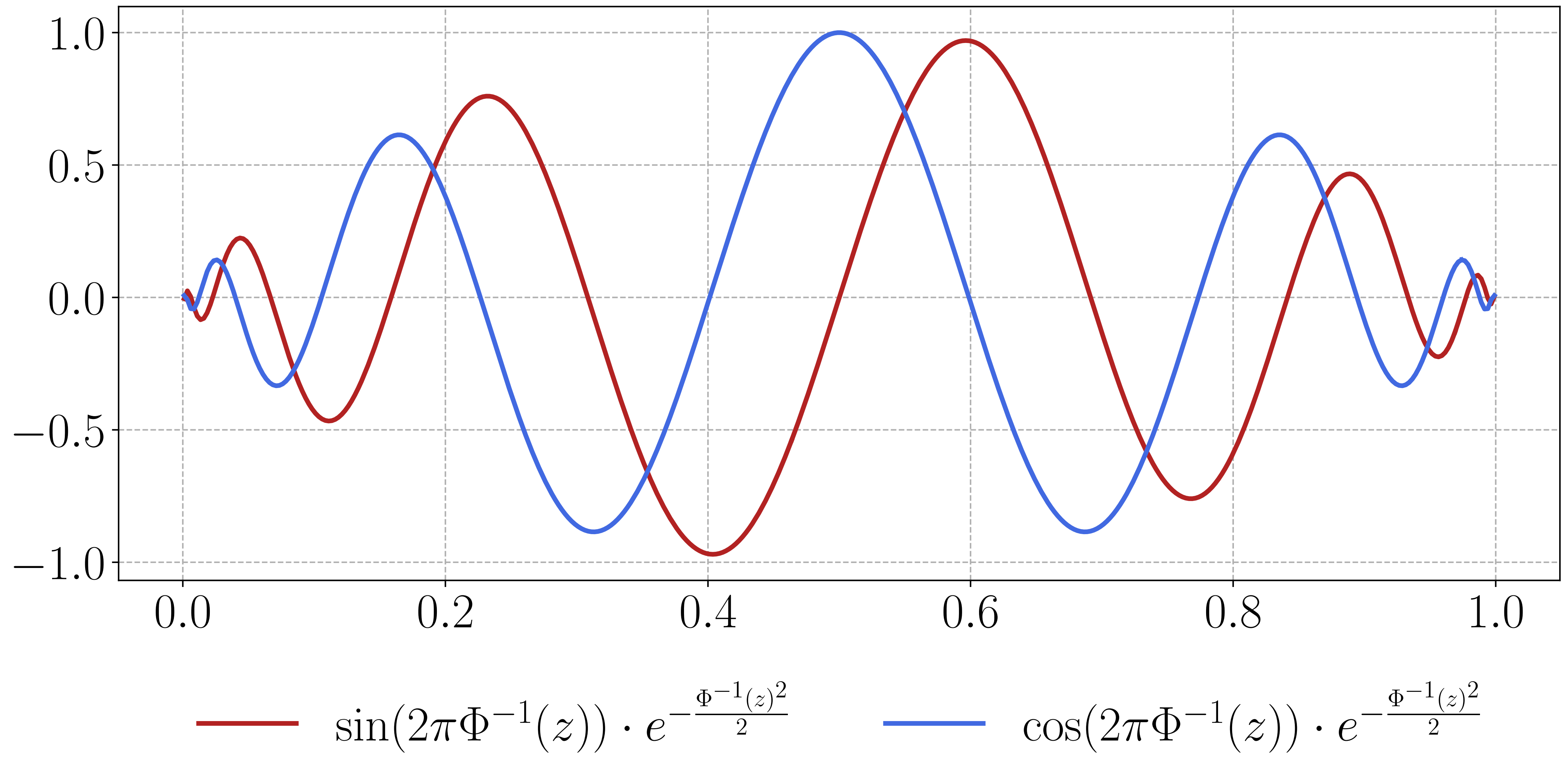
- It achieves the **minimax optimal rate**
- **However**, the unspecified number C makes this procedure **impractical**

Simulation Results

Simulation settings

- Under the null

  - $Z \sim \text{Uniform}[0,1]$
  - $X = \sin\{2\pi\Phi^{-1}(Z)\}e^{-\Phi^{-1}(Z)^2/2} + 0.3\varepsilon_1$
  - $Y = \cos\{2\pi\Phi^{-1}(Z)\}e^{-\Phi^{-1}(Z)^2/2} + 0.3\varepsilon_1$
  - $\varepsilon_1, \varepsilon_2 \overset{i.i.d.}{\sim} N(0,1)$



Simulation settings

Under the null

- $Z \sim \text{Uniform}[0,1]$
- $X = \sin\{2\pi\Phi^{-1}(Z)\}e^{-\Phi^{-1}(Z)^2/2} + 0.3\varepsilon_1$
- $Y = \cos\{2\pi\Phi^{-1}(Z)\}e^{-\Phi^{-1}(Z)^2/2} + 0.3\varepsilon_1$
- $\varepsilon_1, \varepsilon_2 \overset{i.i.d.}{\sim} N(0,1)$

Under the alternative

- $Z \sim \text{Uniform}[0,1]$
- $X = \sin\{2\pi\Phi^{-1}(Z)\}e^{-\Phi^{-1}(Z)^2/2} + 0.3\varepsilon_1$
- $Y = 0.15X + \cos\{2\pi\Phi^{-1}(Z)\}e^{-\Phi^{-1}(Z)^2/2} + 0.3\varepsilon_1$
- $\varepsilon_1, \varepsilon_2 \overset{i.i.d.}{\sim} N(0,1)$

Simulation settings

- Under the null
  - $Z \sim \text{Uniform}[0,1]$
  - $X = \sin\{2\pi\Phi^{-1}(Z)\}e^{-\Phi^{-1}(Z)^2/2} + 0.3\varepsilon_1$
  - $Y = \cos\{2\pi\Phi^{-1}(Z)\}e^{-\Phi^{-1}(Z)^2/2} + 0.3\varepsilon_1$
  - $\varepsilon_1, \varepsilon_2 \overset{i.i.d.}{\sim} N(0,1)$

- Under the alternative
  - $Z \sim \text{Uniform}[0,1]$
  - $X = \sin\{2\pi\Phi^{-1}(Z)\}e^{-\Phi^{-1}(Z)^2/2} + 0.3\varepsilon_1$
  - $Y = 0.15X + \cos\{2\pi\Phi^{-1}(Z)\}e^{-\Phi^{-1}(Z)^2/2} + 0.3\varepsilon_1$
  - $\varepsilon_1, \varepsilon_2 \overset{i.i.d.}{\sim} N(0,1)$

- Test statistic: $T = \sum_{i=1}^{M} U_i(W_i)$

$U_i(W_i)$: a test statistic for independence between X and Y

Simulation settings

- Under the null
  - $Z \sim \text{Uniform}[0,1]$
  - $X = \sin\{2\pi\Phi^{-1}(Z)\}e^{-\Phi^{-1}(Z)^2/2} + 0.3\varepsilon_1$
  - $Y = \cos\{2\pi\Phi^{-1}(Z)\}e^{-\Phi^{-1}(Z)^2/2} + 0.3\varepsilon_1$
  - $\varepsilon_1, \varepsilon_2 \overset{i.i.d.}{\sim} N(0,1)$

- Under the alternative
  - $Z \sim \text{Uniform}[0,1]$
  - $X = \sin\{2\pi\Phi^{-1}(Z)\}e^{-\Phi^{-1}(Z)^2/2} + 0.3\varepsilon_1$
  - $Y = 0.15X + \cos\{2\pi\Phi^{-1}(Z)\}e^{-\Phi^{-1}(Z)^2/2} + 0.3\varepsilon_1$
  - $\varepsilon_1, \varepsilon_2 \overset{i.i.d.}{\sim} N(0,1)$

- Test statistic: $T = \sum_{i=1}^{M} U_i(W_i)$

- Sample size: $n = 1000$

$U_i(W_i)$: a test statistic for independence between X and Y

Type I error vs. Number of bins

Significance level
$\alpha = 0.05$

The number of bins

Test Statistics	5	10	15	20	25	30	35
HSIC	1.000	0.443	0.105	0.055	0.057	0.048	0.040
Dcov	1.000	0.668	0.159	0.074	0.050	0.043	0.036
Cov	1.000	0.755	0.185	0.078	0.070	0.049	0.051
XiCov	0.986	0.103	0.058	0.043	0.065	0.058	0.053
Tau	1.000	0.679	0.158	0.073	0.066	0.054	0.048
TauStar	1.000	0.623	0.144	0.065	0.061	0.051	0.049
MIT	0.991	0.121	0.066	0.057	0.062	0.056	0.049

As # **of bins** **increases** (i.e., maximum diameter decreases), the **type I error** **decreases** consistently

Type I error vs. Number of bins

Significance level
$\alpha = 0.05$

The number of bins

Test Statistics	5	10	15	20	25	30	35
HSIC	1.000	0.443	0.105	0.055	0.057	0.048	0.040
Dcov	1.000	0.668	0.159	0.074	0.050	0.043	0.036
Cov	1.000	0.755	0.185	0.078	0.070	0.049	0.051
XiCov	0.986	0.103	0.058	0.043	0.065	0.058	0.053
Tau	1.000	0.679	0.158	0.073	0.066	0.054	0.048
TauStar	1.000	0.623	0.144	0.065	0.061	0.051	0.049
MIT	0.991	0.121	0.066	0.057	0.062	0.056	0.049

As # **of bins** **increases** (i.e., maximum diameter decreases), the **type I error** **decreases** consistently

Power vs. Number of bins

Significance level
$\alpha = 0.05$

The number of bins

Test Statistics	5	10	15	20	25	30	35
HSIC	1.000	0.996	0.870	0.653	0.501	0.410	0.369
Dcov	1.000	1.000	0.982	0.900	0.782	0.696	0.612
Cov	1.000	1.000	0.998	0.955	0.878	0.801	0.734
XiCov	1.000	0.558	0.270	0.211	0.196	0.160	0.154
Tau	1.000	1.000	0.989	0.908	0.805	0.714	0.659
TauStar	1.000	1.000	0.974	0.856	0.722	0.604	0.537
MIT	1.000	0.627	0.356	0.294	0.226	0.192	0.196

As # **of bins** **increases** (i.e., maximum diameter decreases),
the **power** **decreases** consistently

Power vs. Number of bins

Significance level
$\alpha = 0.05$

The number of bins

Test Statistics	5	10	15	20	25	30	35
HSIC	1.000	0.996	0.870	0.653	0.501	0.410	0.369
Dcov	1.000	1.000	0.982	0.900	0.782	0.696	0.612
Cov	1.000	1.000	0.998	0.955	0.878	0.801	0.734
XiCov	1.000	0.558	0.270	0.211	0.196	0.160	0.154
Tau	1.000	1.000	0.989	0.908	0.805	0.714	0.659
TauStar	1.000	1.000	0.974	0.856	0.722	0.604	0.537
MIT	1.000	0.627	0.356	0.294	0.226	0.192	0.196

As # **of bins** **increases** (i.e., maximum diameter decreases), the **power** **decreases** consistently

Double-binning approach: Type I error

Significance level
$\alpha = 0.05$

The number of bins

Test Statistics

	5	10	15	20	25	30	35	Double-Binning
HSIC	1.000	0.443	0.105	0.055	0.057	0.048	0.040	0.055
Dcov	1.000	0.668	0.159	0.074	0.050	0.043	0.036	0.065
Cov	1.000	0.755	0.185	0.078	0.070	0.049	0.051	0.077
XiCov	0.986	0.103	0.058	0.043	0.065	0.058	0.053	0.047
Tau	1.000	0.679	0.158	0.073	0.066	0.054	0.048	0.057
TauStar	1.000	0.623	0.144	0.065	0.061	0.051	0.049	0.062
MIT	0.991	0.121	0.066	0.057	0.062	0.056	0.049	0.066

# of coarser bin: $\lceil n^{0.3} \rceil$
# of finer bin: $\lceil n^{0.3} \rceil \times \lceil n^{0.3} \rceil$

Double-binning approach: Power

Significance level
$\alpha = 0.05$

The number of bins

Test Statistics	5	10	15	20	25	30	35	Double-Binning
HSIC	X	X	X	0.653	0.501	0.410	0.369	0.928
Dcov	X	X	X	0.900	0.782	0.696	0.612	0.987
Cov	X	X	X	0.955	0.878	0.801	0.734	0.988
XiCov	X	X	X	0.211	0.196	0.160	0.154	0.411
Tau	X	X	X	0.908	0.805	0.714	0.659	0.987
TauStar	X	X	X	0.856	0.722	0.604	0.537	0.976
MIT	X	X	X	0.294	0.226	0.192	0.196	0.480

(X) They do not control the type I error;
Hence ignored in power comparisons

# of coarser bin: $\lceil n^{0.3} \rceil$
# of finer bin: $\lceil n^{0.3} \rceil \times \lceil n^{0.3} \rceil$