Adaptive Multiscale Binary Expansion Tests for Independence

Yang Yang[1] Duo Zheng[2] Sandeep Jain[3] Kai Zhang[4] Ping-Shou Zhong[1]

[1]University of Illinois Chicago [2]Amazon [3]UIC Ophthalmology and Visual Science [4]University of North Carolina at Chapel Hill

wa-dCoBET: distribution-free, adaptive, scalable, and interpretable independence testing

Code Repository
github.com/yyang3388/Cobet

Problem

Goal: test independence for i.i.d. samples

$$\{(X_i, Y_i)\}_{i=1}^n, \qquad X \in \mathbb{R}^p,\ Y \in \mathbb{R}^q,$$

under complex nonlinear, non-monotone, and high-dimensional dependence.

Challenge | High dimension | Interpretability

Why it matters.

Central to causal discovery, feature screening, graphical models, and scientific data analysis.

Kernel/distance-based tests are powerful, but sensitivity depends on metric, bandwidth, or aggregation choices.

Limitations of current methods.

Classical tests (χ^2, Fisher's exact test, and extensions [1]) target univariate variables and become unreliable for multivariate X, Y: many categories are needed, leaving small expected cell counts under limited sample sizes [2, 3].

Distance-correlation weighting down-weights high-frequency components, reducing power for complex, non-linear, and local dependencies.

Main Theoretical Insight

Step 1: PIT. With marginal CDFs F_r, G_s of $X^{(r)}, Y^{(s)}$, define

$$U_i^{(r)} = 2F_r(X_i^{(r)}) - 1, \qquad V_i^{(s)} = 2G_s(Y_i^{(s)}) - 1,$$

so $U_i^{(r)}, V_i^{(s)} \in [-1, 1]$ are uniform and

$$X \perp Y \iff U \perp V, \qquad U, V \in [-1, 1]^d.$$

Step 2: binary expansion coefficients. Truncate the base-2 expansion at depth K:

$$U_i^{(r)} = \sum_{k=1}^{K} 2^{-k} A_{k,i}^{(r)}, \qquad V_i^{(s)} = \sum_{j=1}^{K} 2^{-j} B_{j,i}^{(s)},$$

with $A_{k,i}^{(r)}, B_{j,i}^{(s)} \in \{-1, +1\}$ (Bernoulli bits).

Step 3: interaction vectors. For $I = (I_1, \dots, I_p)$, $I_r \subseteq \{1, \dots, K\}$ (similarly J),

$$\mathcal{A}_{i,I} = \prod_{r=1}^{p} \prod_{k \in I_r} A_{k,i}^{(r)}, \qquad \mathcal{B}_{i,J} = \prod_{s=1}^{q} \prod_{j \in J_s} B_{j,i}^{(s)},$$

$$\vec{A}_i = (\mathcal{A}_{i,I})_{I \in \mathcal{I}_p^*} \in \mathbb{R}^{d_A}, \qquad \vec{B}_i = (\mathcal{B}_{i,J})_{J \in \mathcal{I}_q^*} \in \mathbb{R}^{d_B}.$$

Step 4: equivalence. With $m_{I,J} = \mathrm{Cov}(\mathcal{A}_{i,I}, \mathcal{B}_{i,J})$,

$$\boxed{X \perp Y \iff m_{I,J} = 0 \quad \forall (I, J)}$$

General Dependence Measure

Using characteristic functions $\varphi_{UV}, \varphi_U, \varphi_V$ of the transformed variables, define

$$\mathcal{V}^2(U, V) = \int_{\mathbb{R}^p}\int_{\mathbb{R}^q} |\varphi_{UV}(t, s) - \varphi_U(t)\varphi_V(s)|^2 \, w^2(t, s) \, dt \, ds.$$

Expanding in the binary interaction basis collapses $\mathcal{V}^2(U, V)$ into a bilinear form in $\Sigma_{AB} = \mathrm{Cov}(\vec{A}_i, \vec{B}_i)$:

$$\mathcal{V}^2(U, V) = \mathrm{tr}(\Sigma_{AB} W_B \Sigma_{BA} W_A),$$

so every weight $w^2(t, s)$ induces a matrix pair (W_A, W_B).

$w^2 \equiv 1 \Rightarrow$ **CoBET** | $w^2 =$ distance kernel $\Rightarrow$ **dCoBET**

CoBET: $W_A = I_{d_A}$, $W_B = I_{d_B}$, giving $\mathcal{V}_{I,K}^2 = \|\Sigma_{AB}\|_F^2$.

dCoBET: $w^2(t, s) = \{\|t\|_p^{p+1} \|s\|_q^{q+1}\}^{-1}$, giving $W_A = K^{(p)}$, $W_B = K^{(q)}$.

Contribution 1: binary expansion theory

A complete multiscale characterization of independence via binary interaction moments, converting a nonparametric problem into testing structured cross-covariances—no RKHS or kernel choice needed.

Low-order bits summarize coarse shifts; higher-order bits capture localized and tail dependence.

Interaction products allow dependence across coordinates and resolutions.

Test Statistic and Asymptotics

Estimate $\mathcal{V}_{W,K}^2(U, V)$ by the degree-4 U-statistic

$$T_n^{(W)} = T_{1n}^{(W)} - 2T_{2n}^{(W)} + T_{3n}^{(W)},$$

$$T_{1n}^{(W)} = \frac{1}{n(n-1)} \sum_{i \neq j} (\vec{A}_i^\top W_A \vec{A}_j)(\vec{B}_j^\top W_B \vec{B}_i),$$

with $T_{2n}^{(W)}, T_{3n}^{(W)}$ analogous sums over 3 and 4 distinct indices. This is unbiased, $\mathbb{E}\{T_n^{(W)}\} = \mathcal{V}_{W,K}^2(U, V)$, and under $H_0 : U \perp V$,

$$Z_n^{(W)} = \frac{T_n^{(W)}}{\sqrt{\mathrm{Var}(T_n^{(W)})}} \implies \mathcal{N}(0, 1),$$

with an analytic variance estimator.

Kernel Representation

Binary interaction vectors are exponentially large ($d_A = 2^{pK} - 1$, $d_B = 2^{qK} - 1$) but **never explicitly constructed**. For $i \neq j$,

$$\vec{A}_i^\top \vec{A}_j = \prod_{r=1}^{p} \prod_{k=1}^{K} \left(1 + A_{k,i}^{(r)} A_{k,j}^{(r)}\right) - 1,$$

and for a weight matrix W_A,

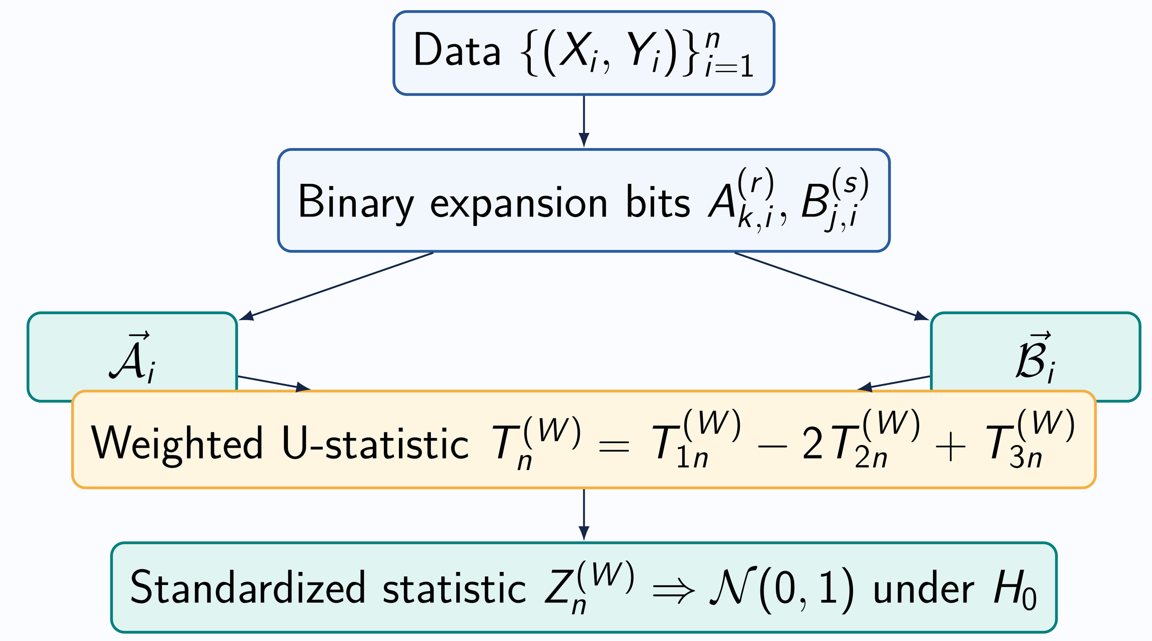
$$\vec{A}_i^\top W_A \vec{A}_j = \int_{\mathbb{R}^p} \mathcal{C}^2(t) \left[\prod_{r,k}\left(1 + A_{k,i}^{(r)} \pi_k^{(r)}(t_r)\right) - 1\right]\left[\prod_{r,k}\left(1 + A_{k,j}^{(r)} \pi_k^{(r)}(t_r)\right) - 1\right] w_0^2(t) \, dt,$$

with $\mathcal{C}^2(t) = \prod_{r,k} \cos^2(t_r/2^k)$, $\pi_k^{(r)}(t_r) = \tan(t_r/2^k)$.

Reduces $O(2^{pK})$ to $O(pK)$

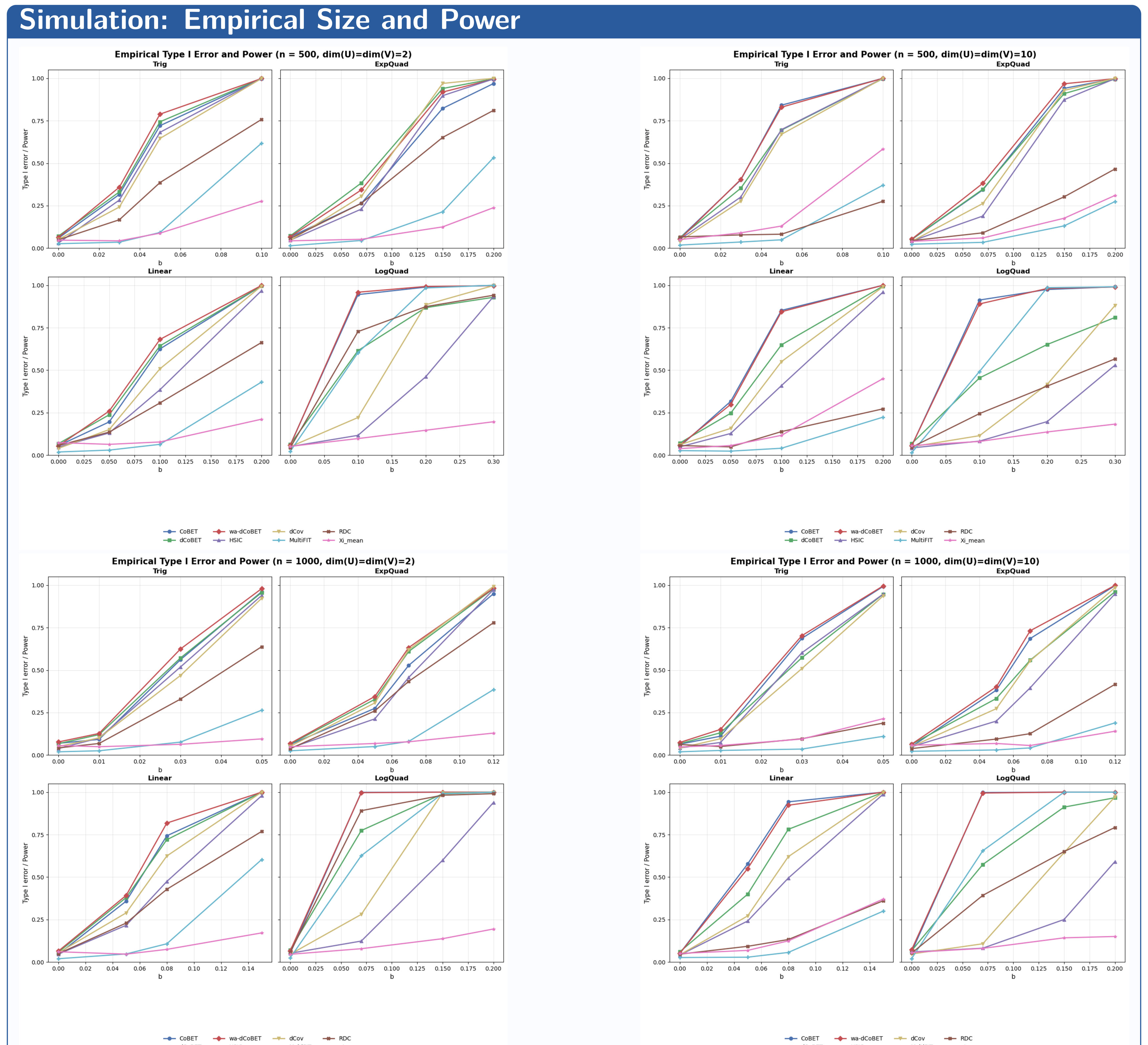
Consequence: $T_n^{(W)}$ is computable in closed form directly from the observed bits, without ever forming $\vec{A}_i \in \mathbb{R}^{d_A}$ or $\vec{B}_i \in \mathbb{R}^{d_B}$.

Graphical Illustration of the Test Statistic



Choice of W determines the test: $W_A = I_{d_A}, W_B = I_{d_B} \Rightarrow$ CoBET; $W_A = K^{(p)}, W_B = K^{(q)} \Rightarrow$ dCoBET; adaptive mixture $\Rightarrow$ wa-dCoBET.

Simulation: Empirical Size and Power



Rejection probability at $n = 500, 1000$, $\dim(U) = \dim(V) \in \{2, 10\}$, vs. HSIC, dCov, RDC, MultiFIT, MMD-DUAL.

Type I error stays near 5% at $b = 0$ across sample sizes and dimensions.

wa-dCoBET's advantage grows with both n and dimension.

Contribution 2: adaptive testing

wa-dCoBET blends identity and distance-inspired weights to improve power across heterogeneous alternatives.

Rather than committing to one dependence geometry, the procedure estimates which component has the stronger foldwise signal and forms an aggregated statistic.

Main Takeaways

Representation: independence $\Leftrightarrow$ vanishing binary interaction cross-covariances.

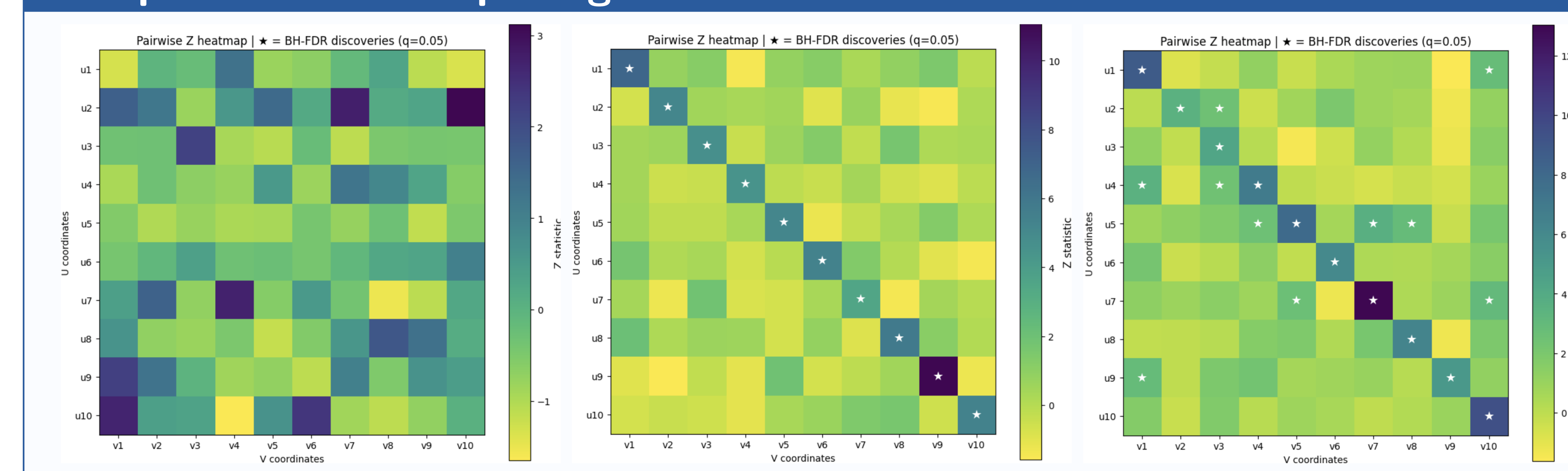
Scalability: kernel representation avoids exponential feature construction.

Adaptivity: SNR voting combines CoBET and dCoBET.

Theory: analytic null calibration removes costly resampling.

Interpretability: heatmaps locate dependence structure after rejection.

Interpretable Heatmap Diagnostics



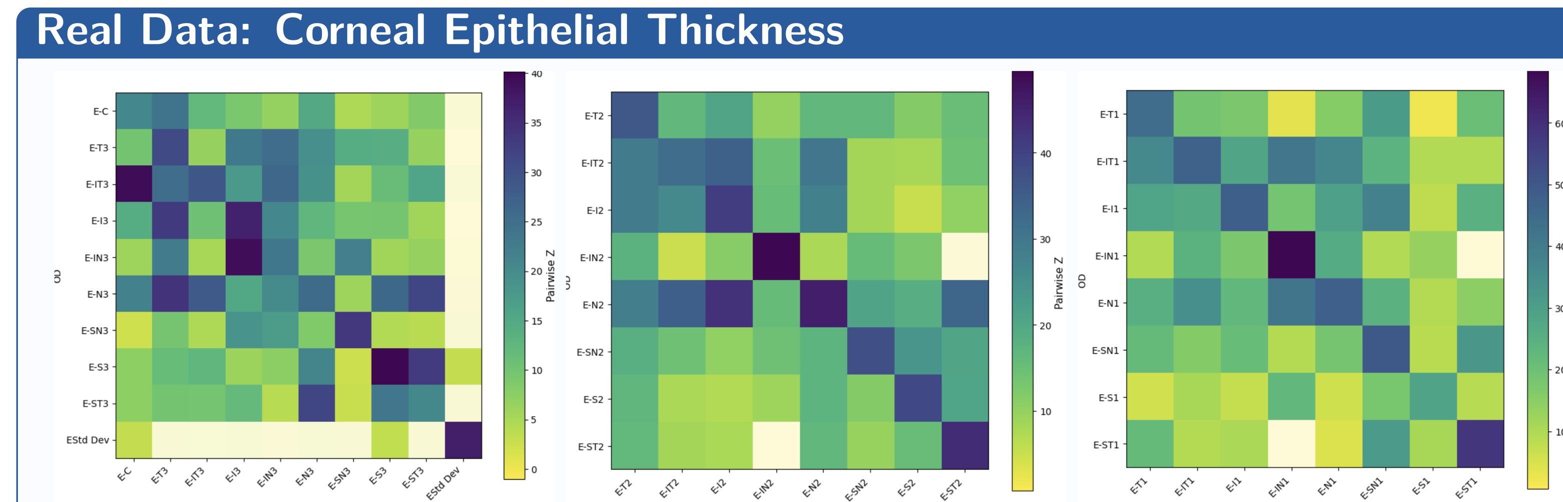
Null | Matched-coordinate | Copula-induced

Binary-expansion heatmaps show *where* dependence occurs after rejection.

Sparse hotspots indicate localized dependence; diagonal structure reflects matched-coordinate interactions.

Off-diagonal blocks reveal cross-variable relationships.

Real Data: Corneal Epithelial Thickness



Left: spatial thickness map for one eye, with regions defined by angular sectors and concentric rings. Right two: the three concentric groupings used for analysis.

Data.

189 subjects; epithelial thickness measured in 25 predefined corneal regions for both eyes (OD and OS) via anterior segment optical coherence tomography (AS-OCT).

Each observation is a spatial thickness map, with regions corresponding to angular sectors and concentric rings.

Analysis design.

Goal: dependence *between* eyes across anatomically corresponding regions.

To balance geometric fidelity and dimensionality, the 25 regions are grouped into three concentric sections analyzed separately: (i) central + inner ring ($p = q = 10$), (ii) middle ring ($p = q = 8$), (iii) outer ring ($p = q = 8$).

For each section, the OD and OS thickness vectors are treated as X and Y, respectively—e.g., for the central + inner region, $X, Y \in \mathbb{R}^{10}$.

Method	Result	p-value
wa-dCoBET	$Z = 5.749$	4.47×10^{-9}
HSIC	–	4.99×10^{-4}
dCov	–	9.9×10^{-4}

Strong diagonal patterns indicate bilateral symmetry; central off-diagonal signals suggest stronger regional coherence near the visual axis.

Impact

The proposed framework offers a scalable and interpretable alternative to kernel-based independence tests for high-dimensional machine learning, biostatistics, causal discovery, and scientific data analysis.

References

[1] A. Agresti. *Categorical Data Analysis*, 3rd ed. Wiley, 2013.

G. J. Székely, M. L. Rizzo, N. K. Bakirov. Measuring and testing dependence by correlation of distances. *Ann. Statist.*, 35(6):2769–2794, 2007.

A. Gretton, O. Bousquet, A. Smola, B. Schölkopf. Measuring statistical dependence with Hilbert-Schmidt norms. *ALT*, 2005.