Entropy Regularized Optimal Transport Independence Criterion

Lang Liu

University of Washington

March 18, 2022

Team

Lang Liu

Soumik Pal

Zaid Harchaoui

Statistical Test of Independence

Problem:

- Let $(X, Y) \sim P_{XY}$ on $\mathcal{X} \times \mathcal{Y}$ with marginals P_X and P_Y.
- Let $\{(X_i, Y_i)\}_{i=1}^n$ be i.i.d. copies of (X, Y).

$$\mathbf{H}_0 : X \text{ and } Y \text{ are independent} \leftrightarrow \mathbf{H}_1 : X \text{ and } Y \text{ are dependent.}$$

Statistical Test of Independence

Problem:

- Let $(X, Y) \sim P_{XY}$ on $\mathcal{X} \times \mathcal{Y}$ with marginals P_X and P_Y.
- Let $\{(X_i, Y_i)\}_{i=1}^n$ be i.i.d. copies of (X, Y).

$$\mathbf{H}_0 : X \text{ and } Y \text{ are independent} \leftrightarrow \mathbf{H}_1 : X \text{ and } Y \text{ are dependent}.$$

Applications:

- Independent component analysis (Bach & Jordan '02, Gretton et al. '05).
- Causal inference (Pfister et al. '18, Chakraborty & Zhang '19).
- Self-supervised learning (Li et al. '21).

Statistical Test of Independence

Strategy:

- Define an independence criterion $T(X, Y)$ such that
  - $T(X, Y) \geq 0$,
  - $T(X, Y) = 0$ iff X and Y are independent.
- Estimate the criterion from data $T_n(X, Y)$.
- Select a *significance level* α and choose a *critical value* $t_n(\alpha)$.
- Reject $\mathbf{H}_0$ if $T_n(X, Y) > t_n(\alpha)$.

Statistical Test of Independence

Strategy:

- Define an independence criterion $T(X, Y)$ such that
  - $T(X, Y) \geq 0$,
  - $T(X, Y) = 0$ iff X and Y are independent.
- Estimate the criterion from data $T_n(X, Y)$.
- Select a *significance level* α and choose a *critical value* $t_n(\alpha)$.
- Reject $\mathbf{H}_0$ if $T_n(X, Y) > t_n(\alpha)$.
- Evaluation metrics:
  - Type I error rate $\mathbb{P}(\text{reject} \mid \mathbf{H}_0) \lesssim \alpha$.
  - Statistical power $\mathbb{P}(\text{reject} \mid \mathbf{H}_1)$ as larger as possible.

Statistical Test of Independence

Independence tests:

- Classical independence criterion (Hoeffding ’48, Kruskal ’58, Lehmann ’66)
  - Pearson’s correlation coefficient.
  - Spearman’s ρ.
  - Kendall’s τ.

Statistical Test of Independence

Independence tests:

- Classical independence criterion (Hoeffding '48, Kruskal '58, Lehmann '66)
  - Pearson's correlation coefficient.
  - Spearman's ρ.
  - Kendall's τ.
- Distance-based independence criterion.
  - Hilbert-Schmidt independence criterion (HSIC) (Gretton et al. '05).
  - Distance covariance (dCov) (Székely et al. '07).

Statistical Test of Independence

Independence tests:

- Classical independence criterion (Hoeffding '48, Kruskal '58, Lehmann '66)
  - Pearson's correlation coefficient.
  - Spearman's ρ.
  - Kendall's τ.
- Distance-based independence criterion.
  - Hilbert-Schmidt independence criterion (HSIC) (Gretton et al. '05).
  - Distance covariance (dCov) (Székely et al. '07).
- Optimal transport based independence criterion.
  - Wasserstein correlation coefficient (Wiesel '21, Mordant and Segers '21, Nies et al. '21).
  - Rank-based independence criterion (Shi et al. '20, Deb & Sen '21).

Entropy Regularized Optimal Transport Independence Criterion

ETIC—define $T(X, Y)$ by

$$\bar{S}_\lambda(P_{XY}, P_X \otimes P_Y) := S_\lambda(P_{XY}, P_X \otimes P_Y) - \frac{1}{2} S_\lambda(P_{XY}, P_{XY}) - \frac{1}{2} S_\lambda(P_X \otimes P_Y, P_X \otimes P_Y),$$

- S_λ *Sinkhorn distance* (Cuturi '13, Ferradans et al. '14)

$$\min_{\gamma \in \mathrm{CP}(P_{XY}, P_X \otimes P_Y)} \left[\int c((x, y), (x', y')) \mathrm{d}\gamma((x, y), (x', y')) + \lambda \mathrm{KL}(\gamma \| P_{XY} \otimes P_X \otimes P_Y) \right].$$

Entropy Regularized Optimal Transport Independence Criterion

ETIC—define $T(X, Y)$ by

$$\bar{S}_\lambda(P_{XY}, P_X \otimes P_Y) := S_\lambda(P_{XY}, P_X \otimes P_Y) - \frac{1}{2}S_\lambda(P_{XY}, P_{XY}) - \frac{1}{2}S_\lambda(P_X \otimes P_Y, P_X \otimes P_Y),$$

- S_λ *Sinkhorn distance* (Cuturi '13, Ferradans et al. '14)

$$\min_{\gamma \in \mathrm{CP}(P_{XY}, P_X \otimes P_Y)} \left[\int c((x,y),(x',y'))\mathrm{d}\gamma((x,y),(x',y')) + \lambda \mathrm{KL}(\gamma \| P_{XY} \otimes P_X \otimes P_Y) \right].$$

- $\bar{S}_\lambda$ *Sinkhorn divergence* (Ramdas et al. '17, Genevay et al. '18, Feydy et al. '19).

Entropy Regularized Optimal Transport Independence Criterion

- Consider an *additive cost*:

$$c\big((x, y), (x', y')\big) = c_1(x, x') + c_2(y, y').$$

- Define two Gibbs kernels

$$k_1(x, x') := \exp(-c_1(x, x')/\lambda) \quad \text{and} \quad k_2(y, y') := \exp(-c_2(y, y')/\lambda).$$

Theorem (Informal, LPH '22)

Assume that the cost c_i is Lipschitz continuous and the Gibbs kernel k_i is universal for $i \in \{1, 2\}$. Then the ETIC is a ***valid independence criterion****.*

Entropy Regularized Optimal Transport Independence Criterion

Example: weighted quadratic cost

$$c\big((x, y), (x', y')\big) := w_1 \left\|x - x'\right\|^2 + w_2 \left\|y - y'\right\|^2 .$$

The objective in $S_\lambda(P_{XY}, P_X \otimes P_Y)$ is

$$\int c\big((x, y), (x', y')\big) \mathrm{d}\gamma\big((x, y), (x', y')\big) + \lambda \mathrm{KL}(\gamma \| P_{XY} \otimes P_X \otimes P_Y)$$

Entropy Regularized Optimal Transport Independence Criterion

Example: weighted quadratic cost

$$c\big((x,y),(x',y')\big) := w_1 \left\|x - x'\right\|^2 + w_2 \left\|y - y'\right\|^2.$$

The objective in $S_\lambda(P_{XY}, P_X \otimes P_Y)$ is

$$\int c\big((x,y),(x',y')\big)\mathrm{d}\gamma\big((x,y),(x',y')\big) + \lambda \mathrm{KL}(\gamma \| P_{XY} \otimes P_X \otimes P_Y)$$
$$= \int \left(w_1 \left\|x - x'\right\|^2 + w_2 \left\|y - y'\right\|^2 \right) \mathrm{d}\gamma\big((x,y),(x',y')\big) + \lambda \mathrm{KL}(\gamma \| P_{XY} \otimes P_X \otimes P_Y)$$

Entropy Regularized Optimal Transport Independence Criterion

Example: weighted quadratic cost

$$c\big((x,y),(x',y')\big) := w_1 \left\|x - x'\right\|^2 + w_2 \left\|y - y'\right\|^2.$$

The objective in $S_\lambda(P_{XY}, P_X \otimes P_Y)$ is

$$\begin{aligned}
&\int c\big((x,y),(x',y')\big)\mathrm{d}\gamma\big((x,y),(x',y')\big) + \lambda \mathrm{KL}(\gamma \| P_{XY} \otimes P_X \otimes P_Y) \\
&= \int \left(w_1 \left\|x - x'\right\|^2 + w_2 \left\|y - y'\right\|^2\right) \mathrm{d}\gamma\big((x,y),(x',y')\big) + \lambda \mathrm{KL}(\gamma \| P_{XY} \otimes P_X \otimes P_Y) \\
&= w_1 \int \left\|x - x'\right\|^2 \mathrm{d}\gamma_1(x,x') + w_2 \int \left\|y - y'\right\|^2 \mathrm{d}\gamma_2(y,y') + \lambda \mathrm{KL}(\gamma \| P_{XY} \otimes P_X \otimes P_Y).
\end{aligned}$$

Computational Aspects of ETIC

ETIC test statistic:

$$T_n(X, Y) = \bar{S}_\lambda(\hat{P}_{XY}, \hat{P}_X \otimes \hat{P}_Y),$$

where $\hat{P}_{XY} = \frac{1}{n}\sum_{i=1}^n \delta_{(X_i, Y_i)}$, $\hat{P}_X = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, and $\hat{P}_Y = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$.

Computational Aspects of ETIC

ETIC test statistic:

$$T_n(X, Y) = \bar{S}_\lambda(\hat{P}_{XY}, \hat{P}_X \otimes \hat{P}_Y),$$

where $\hat{P}_{XY} = \frac{1}{n}\sum_{i=1}^n \delta_{(X_i, Y_i)}$, $\hat{P}_X = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, and $\hat{P}_Y = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$.

- Naïve Sinkhorn algorithm: $\tilde{O}(n^4)$ time and $O(n^4)$ space.

	Cost matrix	**Computation per iteration**
Sinkhorn	$n^2 \times n^2$	$n^2 \times n^2$ and $n^2 \times 1$

Computational Aspects of ETIC

ETIC test statistic:

$$T_n(X, Y) = \bar{S}_\lambda(\hat{P}_{XY}, \hat{P}_X \otimes \hat{P}_Y),$$

where $\hat{P}_{XY} = \frac{1}{n}\sum_{i=1}^n \delta_{(X_i, Y_i)}$, $\hat{P}_X = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, and $\hat{P}_Y = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$.

- Naïve Sinkhorn algorithm: $\tilde{O}(n^4)$ time and $O(n^4)$ space.
- *Tensor Sinkhorn algorithm*: $\tilde{O}(n^3)$ time and $O(n^2)$ space.

	Cost matrix	**Computation per iteration**
Sinkhorn	$n^2 \times n^2$	$n^2 \times n^2$ and $n^2 \times 1$
Tensor Sinkhorn	Two $n \times n$	$n \times n$ and $n \times n$

Computational Aspects of ETIC

ETIC test statistic:

$$T_n(X, Y) = \bar{S}_\lambda(\hat{P}_{XY}, \hat{P}_X \otimes \hat{P}_Y),$$

where $\hat{P}_{XY} = \frac{1}{n}\sum_{i=1}^n \delta_{(X_i,Y_i)}$, $\hat{P}_X = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, and $\hat{P}_Y = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$.

- Naïve Sinkhorn algorithm: $\tilde{O}(n^4)$ time and $O(n^4)$ space.
- *Tensor Sinkhorn algorithm*: $\tilde{O}(n^3)$ time and $O(n^2)$ space.
- Tensor Sinkhorn with *random feature approximation*: $\tilde{O}(pn^2)$ time and $O(n^2)$ space.

	Cost matrix	**Computation per iteration**
Sinkhorn	$n^2 \times n^2$	$n^2 \times n^2$ and $n^2 \times 1$
Tensor Sinkhorn	Two $n \times n$	$n \times n$ and $n \times n$
Tensor Sinkhorn with RF	Two $n \times p$	$n \times n$ and $n \times p$

The Tensor Sinkhorn Algorithm

- A and B distributions on $\{x_i\}_{i=1}^n \times \{y_i\}_{i=1}^n$.
- $K_1 = \exp(-C_1/\lambda)$ and $K_2 = \exp(-C_2/\lambda)$.
- Compute $S_\lambda(A, B)$.

Algorithm 1 Tensor Sinkhorn Algorithm

1: **Input:** A, B, K_1, and K_2.
2: Initialize $U \leftarrow \mathbf{1_{n\times n}}$ and $V \leftarrow \mathbf{1_{n\times n}}$.
3: **while** not converge **do**
4: $\quad U \leftarrow A \oslash (K_1 V K_2^\top)$ and $V = B \oslash (K_1^\top U K_2)$.
5: **end while**
6: **Output:** $\langle \varepsilon \log U, A\rangle_{\mathbf{F}} + \langle \varepsilon \log V, B\rangle_{\mathbf{F}}$.

Gradient Backpropagation through ETIC

What if $\{X_i\}_{i=1}^n$ are complicated objects, e.g., images?

- Obtain feature representations $Z_i = f_\theta(X_i)$, e.g., neural networks.
- Choose the squared Euclidean distance on the feature embedding space.
- Learn a representation via

$$\sup_\theta T_n(f_\theta(X), Y).$$

- ETIC is amenable to gradient backpropagation.

Statistical Properties of ETIC

Theorem (LPH '22)

Assume that c is the ***weighted quadratic cost*** *and P_X and P_Y are supported on a* ***bounded domain with radius*** *R. Then we have, with probability at least $1-\delta$,*

$$|T_n(X,Y) - T(X,Y)| \leq C_d \left(\lambda + \frac{R^{5d+16}}{\lambda^{5d/2+7}} \sqrt{\log \frac{6}{\delta}} \right) \frac{1}{\sqrt{n}}.$$

Statistical Properties of ETIC

Theorem (LPH '22)

*Assume that c is the **weighted quadratic cost** and P_X and P_Y are supported on a **bounded domain with radius** R. Then we have, with probability at least $1-\delta$,*

$$|T_n(X,Y) - T(X,Y)| \leq C_d \left(\lambda + \frac{R^{5d+16}}{\lambda^{5d/2+7}} \sqrt{\log \frac{6}{\delta}} \right) \frac{1}{\sqrt{n}}.$$

Remark

- *Rate of convergence* $O(n^{-1/2})$.
- *The choice of* $\lambda = R^2$ *gives* $C_d \sqrt{\log(6/\delta)} R^2/\sqrt{n}$.
- $T(X,Y) = T_\lambda(X,Y) \to 0$ *as* $\lambda \to \infty$.

Statistical Properties of ETIC

Theorem (LPH '22)

*Assume that c is the **weighted quadratic cost** and P_X and P_Y are supported on a **bounded domain with radius** R. Then we have, with probability at least $1-\delta$,*

$$|T_n(X,Y) - T(X,Y)| \leq C_d \left(\lambda + \frac{R^{5d+16}}{\lambda^{5d/2+7}} \sqrt{\log \frac{6}{\delta}} \right) \frac{1}{\sqrt{n}}.$$

Remark

The above theorem implies that the power of the ETIC test is asymptotically one.

- *Under $\mathbf{H}_0$, $T(X,Y) = 0$ and thus the critical value $t_n(\alpha)$ should be of order $O(n^{-1/2})$.*
- *Under $\mathbf{H}_1$, $T(X,Y) > 0$ and thus $T_n(X,Y)$ will alway exceed $t_n(\alpha)$ as $n \to \infty$.*

Statistical Properties of ETIC

Theorem (LPH '22)

Assume that c is the weighted quadratic cost and P_X and P_Y are ***sub-Gaussian with parameter*** σ^2. *Then we have*

$$\mathbb{E}\,|T_n(X,Y) - T(X,Y)| \le C_d \left(\lambda + \frac{\sigma^{\lceil 5d/2 \rceil + 6}}{\lambda^{\lceil 5d/4 \rceil + 2}} \right) \frac{1}{\sqrt{n}}.$$

Statistical Properties of ETIC

Theorem (LPH '22)

Assume that c is the weighted quadratic cost and P_X and P_Y are ***sub-Gaussian with parameter*** σ^2. *Then we have*

$$\mathbb{E}\,|T_n(X, Y) - T(X, Y)| \le C_d \left(\lambda + \frac{\sigma^{\lceil 5d/2 \rceil + 6}}{\lambda^{\lceil 5d/4 \rceil + 2}} \right) \frac{1}{\sqrt{n}}.$$

Remark

When $\lambda := \lambda_n = o(1)$ *is chosen such that* $\lambda_n = \omega(n^{-1/(\lceil 5d/2 \rceil + 4)})$, *we have*

$$T_n(X, Y) \to_{\mathbf{L}^1} \mathrm{W}_2^2(P_{XY}, P_X \otimes P_Y).$$

Independence Testing on Bilingual Text

Bilingual text

- Parallel European Parliament corpus (Koehn '05).
- Randomly select $n = 64$ English documents and a paragraph in each document.
  - **Dependent sample**: (English paragraph, same paragraph in French).
  - **Independent sample**: (English paragraph, random paragraph in French).
  - **Partially dependent sample**: (English paragraph, random paragraph in the same document in French).
- Feature embedding of dimension 768 with LaBSE (Feng et al. '20).

Independence Testing on Bilingual Text

Independence tests

- ETIC with a weighted quadratic cost inducing Gibbs kernels

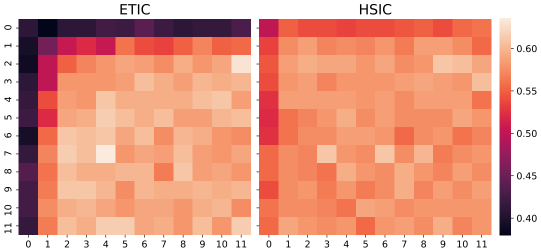
$$k_1(x, x') := \exp\{-\left\|x - x'\right\|^2 / \sigma_1\} \quad \text{and} \quad k_2(y, y') := \exp\{-\left\|y - y'\right\|^2 / \sigma_2\}.$$

- HSIC with the same kernels.
- Median heuristic: $\sigma_1 = r_1 M_x$ and $\sigma_2 = r_2 M_y$ with $r_1, r_2 \in [0.25, 4]$.

Independence Testing on Bilingual Text

- All tests have power one on the dependent sample.
- All tests have desired type I error rate on the independent sample.

ETIC outperforms HSIC for many r_1 and r_2 on the partially dependent sample.

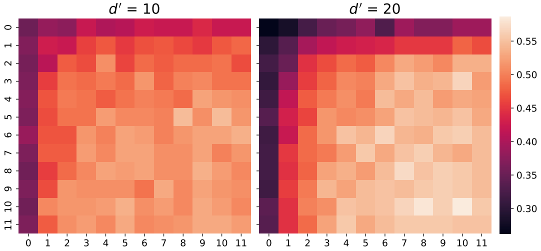


Independence Testing on Bilingual Text

ETIC with random feature approximation

- Principal component analysis (PCA) to reduce the dimension to $d' \in \{10, 20\}$.
- Use $p' = 700$ random features.

ETIC-RF with PCA performs similarly with enough random features.



Conclusion

- A new independence criterion ETIC and the associated test.
- An efficient algorithm to compute empirical ETIC.
- Amenable to gradient backpropagation.
- Finite-sample guarantees for its statistical properties.
- Higher power with a large range of hyperparameters.

Thank you!

Paper (AISTATS 2022): arxiv.org/abs/2112.15265

Appendix

The Schrödinger Bridge Problem and Entropy Regularized OT

The Schrödinger bridge (SCB) problem

- Schrödinger's lazy gas experiment (Schrödinger '32).
- The SCB problem in continuum (Föllmer '88, Léonard '12).
- Survey on SCB (Léonard '14, Chen et al. '21).

Discrete entropy regularized optimal transport (EOT)

- Discrete EOT (Cuturi '13, Ferradans et al. '14).
- Limit laws (Bigot et al. '19, Klatt et al. '20).
- Finite-sample bounds (Genevay et al. '19, Mena and Weed '19).

Discrete Schrödinger bridge

- Discrete SCB for a particular cost (Pal and Wong '20).
- Discrete SCB for general costs (HLP '20).

Properties of ETIC

Proposition (LPH '22)

Let c be a continuous cost function. If either c is bounded or P_{XY} and $P_X \otimes P_Y$ have compact support, it holds that

$$T_\lambda(X, Y) \to \begin{cases} 0 & \text{if } c = c_1 \oplus c_2 \\ -\frac{1}{2} HSIC_{c_1, c_2}(X, Y) & \text{if } c = c_1 \otimes c_2, \end{cases} \quad \text{as } \lambda \to \infty.$$

Moreover, if both P_{XY} and $P_X \otimes P_Y$ are densities (or discrete measures), then

$$T_\lambda(X, Y) \to C_{OT}(P_{XY}, P_X \otimes P_Y), \quad \text{as } \lambda \to 0.$$

The Tensor Sinkhorn Algorithm

- A and B distributions on $\{x_i\}_{i=1}^n \times \{y_i\}_{i=1}^n$.
- $K_1 = \exp(-C_1/\lambda)$ and $K_2 = \exp(-C_2/\lambda)$.
- Compute $S_\lambda(A, B)$.

Algorithm 1 Tensor Sinkhorn Algorithm

1: **Input:** A, B, K_1, and K_2.
2: Initialize $U \leftarrow \mathbf{1_{n\times n}}$ and $V \leftarrow \mathbf{1_{n\times n}}$.
3: **while** not converge **do**
4: $\quad U \leftarrow A \oslash (K_1 V K_2^\top)$ and $V = B \oslash (K_1^\top U K_2)$.
5: **end while**
6: **Output:** $\langle \varepsilon \log U, A\rangle_{\mathbf{F}} + \langle \varepsilon \log V, B\rangle_{\mathbf{F}}$.

The Tensor Sinkhorn Algorithm

Tensor Sinkhorn algorithm

$$U = A \oslash (K_1 V K_2^\top)$$

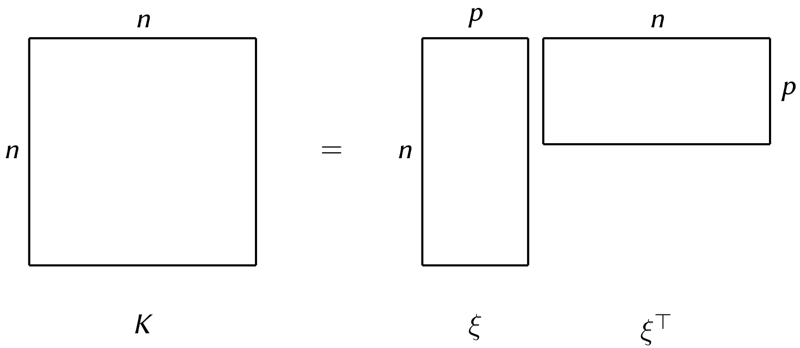
$$U = A \oslash \left[\begin{pmatrix} (K_2)_{1,1} & \dots & (K_1)_{1,n} \\ \vdots & & \vdots \\ (K_1)_{n,1} & \dots & (K_1)_{n,n} \end{pmatrix} \begin{pmatrix} V_{1,1} & \dots & V_{1,n} \\ \vdots & & \vdots \\ V_{n,1} & \dots & V_{n,n} \end{pmatrix} \begin{pmatrix} (K_2)_{1,1} & \dots & (K_2)_{n,1} \\ \vdots & & \vdots \\ (K_2)_{1,n} & \dots & (K_2)_{n,n} \end{pmatrix} \right].$$

Sinkhorn algorithm

$$\mathrm{Vec}(U) = \mathrm{Vec}(A) \oslash \mathrm{Vec}(K_1 V K_2^\top) = \mathrm{Vec}(A) \oslash ((K_2 \otimes K_1)\mathrm{Vec}(V))$$

$$u = a \oslash \left[\begin{pmatrix} (K_2)_{1,1} K_1 & \dots & (K_2)_{1,n} K_1 \\ \vdots & \dots & \vdots \\ (K_2)_{n,1} K_1 & \dots & (K_2)_{n,n} K_1 \end{pmatrix} v \right].$$

Random Feature Approximation



ETIC with Random Features

- Consider Gibbs kernels of the form

$$k_1(x,x') = \int \varphi(x,u)^\top \varphi(x',u)d\rho_1(u) \quad \text{and} \quad k_2(y,y') = \int \psi(y,v)^\top \psi(y',v)d\rho_2(v).$$

- Obtain an i.i.d. sample $\boldsymbol{u} := \{u_k\}_{k=1}^p$ and approximate $k_1(x,x')$ by

$$k_{1,\boldsymbol{u}}(x,x') := \frac{1}{p}\sum_{k=1}^p \varphi(x,u_k)^\top \varphi(x',u_k).$$

- Obtain an i.i.d. sample $\boldsymbol{v} := \{v_k\}_{k=1}^p$ and approximate $k_2(y,y')$ by

$$k_{2,\boldsymbol{v}}(y,y') := \frac{1}{p}\sum_{k=1}^p \psi(y,v_k)^\top \psi(y',v_k).$$

ETIC with Random Features

Approximate $c((x, y), (x', y'))$ by

$$c_{\boldsymbol{u},\boldsymbol{v}}((x, y), (x', y')) := -\lambda \log k_{1,\boldsymbol{u}}(x, x') - \lambda \log k_{2,\boldsymbol{v}}(y, y').$$

Proposition (LPH ’22)

Let $p = \Omega(\tau^{-2} \log (n/\delta))$. Under appropriate assumptions, it holds that, with probability at least $1 - \delta$,

$$\left| S_{\lambda, c_{\boldsymbol{u},\boldsymbol{v}}}(A, B) - S_{\lambda, c}(A, B) \right| \leq \tau.$$

ETIC-Based Tests

The ETIC test with regularization parameter λ:

$$\psi(\alpha) := \mathbb{1}\{T_{n,\lambda}(X, Y) > t_{n,\lambda}(\alpha)\},$$

where α is the significance level and $H_{n,\lambda}(\alpha)$ is the critical value.

The adaptive ETIC test:

$$\psi_a(\alpha) := \mathbb{1}\left\{\max_{\lambda\in\Lambda} \bar{T}_{n,\lambda}(X, Y) > t_{n,\Lambda}(\alpha)\right\}$$

- Λ a finite set of regularization parameters.
- $\bar{T}_{n,\lambda}(X, Y) = [T_{n,\lambda}(X, Y) - \mathbb{E}[T_{n,\lambda}(X, Y)]]/\mathrm{Sd}(T_{n,\lambda}(X, Y))$ the studentized version.

Statistical Properties of ETIC

Empirical Sinkhorn divergence

$$\bar{S}_\lambda\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{U_i}, \frac{1}{n}\sum_{i=1}^{n}\delta_{V_i}\right).$$

Empirical ETIC

$$\bar{S}_\lambda\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{(X_i, Y_i)}, \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{(X_i, Y_j)}\right).$$

	First marginnal	**Second marginal**	**Are two marginals independent?**
SD	Sum of i.i.d. point masses	Sum of i.i.d. point masses	Yes
ETIC	Sum of i.i.d. point masses	**Sum of dependent point masses**	**No**