Optimal rates for independence testing via permutation tests

Tom Berrett (joint with Ioannis Kontoyiannis and Richard Samworth)
University of Warwick

Department of Statistics, University of Washington

October 22nd, 2021

Independence testing

Independence testing is one of the most studied problems in statistics. Given data on a random pair (X, Y), we aim to test

$$H_0 : X \text{ is independent of } Y.$$

	Middle school or lower	High school	Bachelor's	Master's	PhD or higher
Never married	18	36	21	9	6
Married	12	36	45	36	21
Divorced	6	9	9	3	3
Widowed	3	9	9	6	3

Source: `https://www.spss-tutorials.com/chi-square-independence-test/`.

With discrete variables and $p_{jk} = \mathbb{P}(X = j, Y = k)$, we test

$$H_0 : p_{jk} = r_j q_k \quad \text{for some } (r_j), (q_k).$$

The χ^2 test

In such contingency tables, it is standard practice to use Pearson's χ^2 test, where we compare the test statistic

$$T = \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{(o_{jk} - e_{jk})^2}{e_{jk}}$$

to the quantiles of the $\chi^2_{(J-1)(K-1)}$ distribution.

j/k	1	2	3	4	5
1	o_{11}	o_{12}	o_{13}	o_{14}	o_{15}
2	o_{21}	o_{22}	o_{23}	o_{24}	o_{25}
3	o_{31}	o_{32}	o_{33}	o_{34}	o_{35}
4	o_{41}	o_{42}	o_{43}	o_{44}	o_{45}

$$e_{jk} = \frac{o_{j+} \cdot o_{+k}}{n}$$

Although very popular, this test has serious drawbacks.

Asymptotic null distributions

It is well known that, for a fixed null distribution P, we have $T \xrightarrow{\mathrm{d}} \chi^2_{(J-1)(K-1)}$, so that Pearson's test is (pointwise) asymptotically valid.

However, this convergence is not uniform. For a fixed $\lambda > 0$ set $p = \sqrt{\lambda/n}$ and consider the null distribution given by:

p^2	$p(1-p)$
$p(1-p)$	$(1-p)^2$

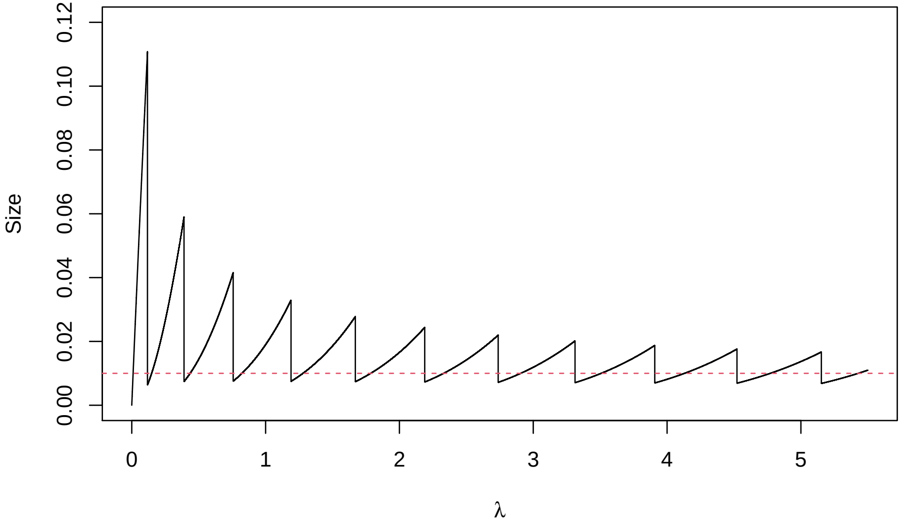
Here, we have

$$T \xrightarrow{\mathrm{d}} \frac{(Z-\lambda)^2}{\lambda},$$

where $Z \sim \mathrm{Poi}(\lambda)$.

Lack of Type I error control

If we compare T to the 99th quantile of the χ_1^2 distribution we get



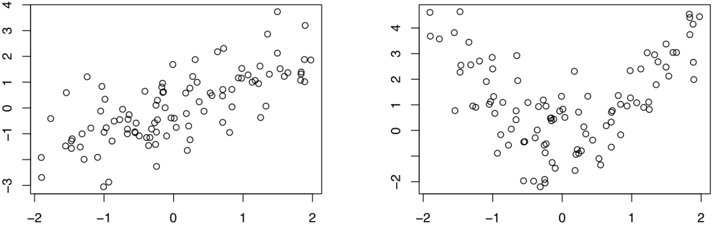
Continuous data

Beyond discrete data, there is a vast literature on testing independence of continuous variables.

Classical measures include:

- Pearson's correlation (e.g. Pearson, 1920);
- Kendall's tau (Kendall, 1938);
- Hoeffding's D (Hoeffding, 1948).

These are limited to linear or monotonic dependence, or bivariate settings.



Background

Modern datasets often exhibit complex dependence which is not well captured by these classical measures.

As a result, many new measures and tests have been proposed and studied recently:

- HSIC (Gretton et al., 2005; Sejdinovic et al., 2013; Pfister et al., 2018; Albert et al., 2019);
- Distance covariance (Székely, Rizzo and Bakirov, 2007; Székely and Rizzo, 2013);
- Nearest neighbour methods (B. and Samworth, 2019);
- Rank-based tests (Weihs et al., 2017; Shi, Drton and Han, 2019; Deb and Sen, 2019);
- Empirical copula processes (Kojadinovic and Holmes, 2009);
- Sample space partitioning (Gretton and Györfi, 2010; Heller et al., 2016).

Aims

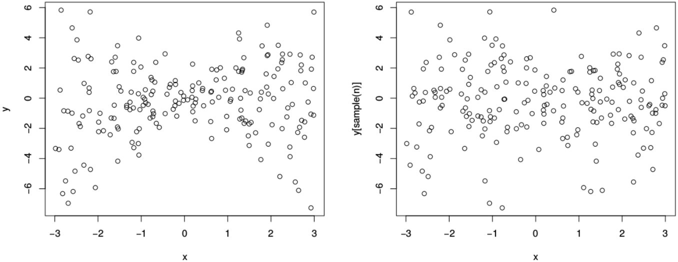
What do we want from an independence test?

- **Validity:** a test that reliably controls the Type I error across the entire null.
- **Power:** as large as possible among valid tests.

We will see that we can achieve both of these aims with *permutation tests*.

Permutation tests

A practical and popular approach for independence testing is to carry out a permutation test (e.g. Pitman, 1938; Fisher, 1935).

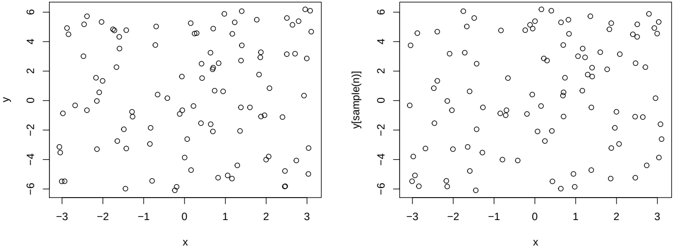


For any test statistic $T : \mathcal{X}^n \times \mathcal{Y}^n \to \mathbb{R}$ and i.i.d. uniformly random permutations $\pi_1, \ldots, \pi_B \in \mathcal{S}_n$ we can set $Y_i^{(b)} = Y_{\pi_b(i)}$ and calculate the p-value

$$p = \frac{1 + \sum_{b=1}^{B} \mathbb{1}\{T(\boldsymbol{X}, \boldsymbol{Y}^{(b)}) \geq T(\boldsymbol{X}, \boldsymbol{Y})\}}{1 + B}.$$

Size of permutation tests

Under H_0, the datasets $(\boldsymbol{X}, \boldsymbol{Y}), (\boldsymbol{X}, \boldsymbol{Y}^{(1)}), \ldots, (\boldsymbol{X}, \boldsymbol{Y}^{(B)})$ are exchangeable.



Hence, $T(\boldsymbol{X}, \boldsymbol{Y}), T(\boldsymbol{X}, \boldsymbol{Y}^{(1)}), \ldots, T(\boldsymbol{X}, \boldsymbol{Y}^{(B)})$ are exchangeable.

The rank of $T(\boldsymbol{X}, \boldsymbol{Y})$ among $T(\boldsymbol{X}, \boldsymbol{Y}), T(\boldsymbol{X}, \boldsymbol{Y}^{(1)}), \ldots, T(\boldsymbol{X}, \boldsymbol{Y}^{(B)})$ is uniformly distributed on $\{1, \ldots, B+1\}$, and so

$$\mathbb{P}(p \leq \alpha) \leq \frac{\lfloor \alpha(B+1) \rfloor}{B+1} \leq \alpha$$

for all $\alpha \in [0, 1]$.

Asymptotic power of permutation tests

Many studies of the power of permutation tests use the approach of Hoeffding (1952).

Theorem (Hoeffding (1952))

Say $(X_1, Y_1) \sim P_n$ for some sequence (P_n). Suppose that

$$\mathbb{P}(T(\boldsymbol{X}, \boldsymbol{Y}^{(1)}) \leq u, T(\boldsymbol{X}, \boldsymbol{Y}^{(2)}) \leq v) \to R(u)R(v)$$

and $\mathbb{P}(T(\boldsymbol{X}, \boldsymbol{Y}) \leq u) \to H(u)$ for distribution functions H, R. If H and R are continuous at $R^{-1}(1-\alpha)$ and R is strictly increasing at $R^{-1}(1-\alpha)$, then, if $B \to \infty$, we have

$$\mathbb{P}(p \leq \alpha) \to 1 - H\big(R^{-1}(1-\alpha)\big)$$

as $n \to \infty$.

This can be used for results on power against sequences of local alternatives.

Permutation tests

Permutation tests:

- require no assumptions for non-asymptotic Type I error control;
- can be used with any test statistic.

Despite the popularity of permutations tests, many open problems remain concerning their power properties. Existing works typically study pointwise asymptotics or relatively simple settings (e.g. Hoeffding, 1952; Romano, 1989; Lehmann and Romano, 2005; Albert et al., 2015).

We show that they can achieve minimax rate optimal power.

Outline of the rest of the talk

1. Problem statement and formalisation
2. U-statistic permutation tests
   - Discrete case
   - Continuous case
   - General upper bound
3. Lower bounds
4. Distributional results and power function

Problem statement

Given i.i.d. pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ taking values in $\mathcal{X} \times \mathcal{Y}$, we construct tests ψ_n of the null hypothesis of independence

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

Writing $\mathcal{P}_0$ for the set of all null distributions, we insist that

$$\psi_n \in \Psi_n(\alpha) := \Big\{\psi : (\mathcal{X} \times \mathcal{Y})^n \to [0,1] : \sup_{P \in \mathcal{P}_0} \mathbb{E}_P(\psi) \leq \alpha\Big\},$$

the set of (randomised) independence tests with size $\leq \alpha$.

Strength of dependence

For power results, we assume (X, Y) has density f w.r.t. $\mu = \mu_X \otimes \mu_Y$, such that

$$(\mathcal{X} \times \mathcal{Y}, \mu)$$

is a separable measure space.

Our upper bounds are general, and lower bounds match in settings such as

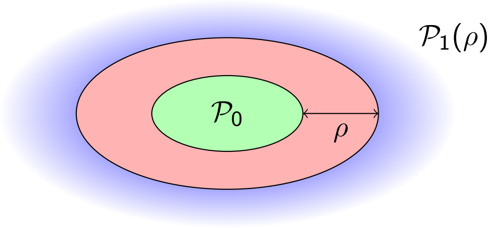
- Discrete data: $\mathcal{X} \times \mathcal{Y} = \{1, \ldots, J\} \times \{1, \ldots, K\}$ some $J, K \in \mathbb{N} \cup \{\infty\}$.
- Continuous data: $\mathcal{X} \times \mathcal{Y} = [0, 1]^{d_X + d_Y}$.
- Infinite-dimensional data: $\mathcal{X} \times \mathcal{Y} = [0, 1]^{\mathbb{N}}$.

Minimax framework

Given a suitable (e.g. smooth) class $\mathcal{P}_1$ of distributions, define the alternatives

$$\mathcal{P}_1(\rho) = \left\{ P \in \mathcal{P}_1 : D(P) \geq \rho^2 \right\}$$

$$D(f) = \int (f - f_X f_Y)^2 \, d\mu$$



For a test ψ_n and $\beta \in (0, 1-\alpha)$ we define the minimax separation radius

$$\rho^*_{n,\alpha,\beta}(\mathcal{P}_1, \psi_n) = \inf\left\{ \rho > 0 : \sup_{P \in \mathcal{P}_0} \mathbb{E}_P(\psi) + \sup_{P \in \mathcal{P}_1(\rho)} \mathbb{E}_P(1-\psi) \leq \alpha + \beta \right\}.$$

Minimax framework

We look for tests ψ_n^* such that

$$\rho^*_{n,\alpha,\beta}(\mathcal{P}_1, \psi_n^*) \asymp \inf_{\psi_n \in \Psi_n(\alpha)} \rho^*_{n,\alpha,\beta}(\mathcal{P}_1, \psi_n),$$

and will call such tests *rate optimal*.

- Minimax rate optimal tests were found for certain univariate problems by Ingster (1989), Ermakov (1990) and Ingster (1996).
- In the multivariate setting, Albert et al. (2019) finds minimax rates with a test using an oracle critical value.
- In independent work, Kim, Balakrishnan and Wasserman (2021) proved the minimax rate optimality of HSIC-based permutation tests.

Hardness result

In typical nonparametric problems, uniform power results are impossible without some such assumptions.

Theorem (Continuous case)

Suppose $\mathcal{X} \times \mathcal{Y} = [0,1]^{d_X + d_Y}$ *with* μ_X, μ_Y *Lebesgue measures, and let* $\psi_n : (\mathcal{X} \times \mathcal{Y})^n \to [0,1]$ *be a test such that*

$$\int \psi_n \, d\mu^{\otimes n} \leq \alpha.$$

Then, for any $\epsilon > 0$ *and* $\rho \in (0, 2^{-1/2}]$*, there exists a density* f *such that* $D(f) = \rho^2$ *and*

$$\mathbb{E}_f(\psi_n) \leq \alpha + \epsilon.$$

Given a test, we can always find a type of alternative it has poor power against.

Smoothness

Letting $(p_{jk})_{j\in\mathcal{J},k\in\mathcal{K}} = (p_j^X p_k^Y)_{j\in\mathcal{J},k\in\mathcal{K}}$ be an orthonormal basis for $L^2(\mu)$, for $f \in L^2(\mu)$ define

$$a_{jk} = \int p_{jk} f \, d\mu, \quad a_{j\bullet} = \int p_j^X f_X \, d\mu_X, \quad a_{\bullet k} = \int p_k^Y f_Y \, d\mu_Y$$

and, given $\theta \in [0,\infty]^{\mathcal{J}\times\mathcal{K}}$, further define

$$S_\theta(f) = \sum_{j\in\mathcal{J},k\in\mathcal{K}} \theta_{jk}^2 \{a_{jk}(f) - a_{j\bullet}(f)a_{\bullet k}(f)\}^2.$$

We will consider classes of the form

$$\mathcal{P}_1 = \big\{P : P \text{ has density } f \text{ with } S_\theta(f) \le r^2\big\}.$$

1 Problem statement and formalisation

2 U-statistic permutation tests
- Discrete case
- Continuous case
- General upper bound

3 Lower bounds

4 Distributional results and power function

Test statistic

Now that we have formalised the problem, we search for suitable tests.

Our test statistic will be an estimator of the dependence measure

$$D(f) = \int (f - f_X f_Y)^2 \, d\mu = \sum_{j \in \mathcal{J}, k \in \mathcal{K}} (a_{jk} - a_{j\bullet} a_{\bullet k})^2.$$

For a subset $\mathcal{M} \subseteq \mathcal{J} \times \mathcal{K}$ we write

$$\begin{aligned} & h\big((x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)\big) \\ & = \sum_{(j,k) \in \mathcal{M}} \{p_{jk}(x_1, y_1) p_{jk}(x_2, y_2) - 2 p_{jk}(x_1, y_1) p_{jk}(x_2, y_3) \\ & \qquad + p_{jk}(x_1, y_2) p_{jk}(x_3, y_4)\} \end{aligned}$$

and

$$\hat{D}_n = \frac{1}{4! \binom{n}{4}} \sum_{\substack{i_1, i_2, i_3, i_4 \\ \text{distinct}}} h\big((X_{i_1}, Y_{i_1}), (X_{i_2}, Y_{i_2}), (X_{i_3}, Y_{i_3}), (X_{i_4}, Y_{i_4})\big).$$

Permutation test

For $b = 1, \dots, B$ we generate independent uniformly random permutations $\pi_1, \dots, \pi_B$ and set

$$\hat{D}_n^{(b)} = \frac{1}{4!\binom{n}{4}} \sum_{\substack{i_1, i_2, i_3, i_4 \\ \text{distinct}}} h\big((X_{i_1}, Y_{\pi_b(i_1)}), (X_{i_2}, Y_{\pi_b(i_2)}), (X_{i_3}, Y_{\pi_b(i_3)}), (X_{i_4}, Y_{\pi_b(i_4)})\big).$$

Calculate the p-value

$$p = \frac{1 + \sum_{b=1}^{B} \mathbb{1}\{\hat{D}_n^{(b)} \geq \hat{D}_n\}}{1 + B}$$

and reject H_0 if and only if $p \leq \alpha$.

Discrete case

In the discrete case $\mathcal{X} \times \mathcal{Y} = \{1, \ldots, J\} \times \{1, \ldots, K\}$, the test statistic takes a simpler form.

$$\hat{D}_n = \frac{1}{n(n-3)} \sum_{j=1}^{J} \sum_{k=1}^{K} (o_{jk} - e_{jk})^2 - \frac{4}{n(n-2)(n-3)} \sum_{j=1}^{J} \sum_{k=1}^{K} o_{jk} e_{jk} + \ldots,$$

where we omit terms only depending on o_{j+} and o_{+k}.

Using Patefield's algorithm we can generate the permuted tables according to

$$\mathbb{P}\big((o_{jk}^{(1)}) = (n_{jk}) | (o_{jk})\big) = \frac{(\prod_{j=1}^{J} o_{j+}!)(\prod_{k=1}^{K} o_{+k}!)}{n \prod_{j=1}^{J} \prod_{k=1}^{K} n_{jk}!}.$$

Discrete case

We can take

$$\mathcal{P}_1(\rho) = \left\{ P : D(f) = \sum_{j=1}^{J} \sum_{k=1}^{K} \{f(j,k) - f_X(j) f_Y(k)\}^2 \geq \rho^2 \right\}.$$

Let ψ_n be the permutation test described above.

Theorem

Fix $\alpha, \beta \in (0,1)$ such that $\alpha + \beta < 1$. Then there exists $C = C(\alpha, \beta) \in (0, \infty)$ such that

$$\sup_{P \in \mathcal{P}_0} \mathbb{E}_P(\psi_n) + \sup_{P \in \mathcal{P}_1(Cn^{-1/2})} \mathbb{E}_P(1 - \psi_n) \leq \alpha + \beta$$

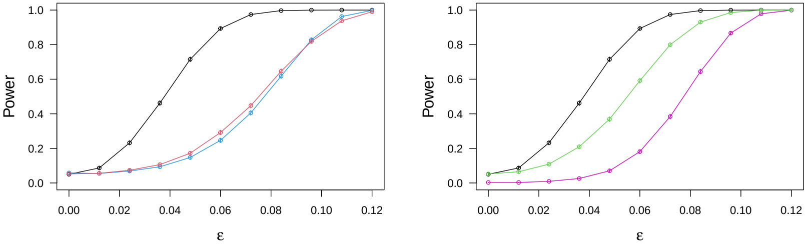
for all $n \in \mathbb{N}$, i.e. $\rho^*_{n,\alpha,\beta}(\mathcal{P}_1, \psi_n) \lesssim n^{-1/2}$.

In fact, $\hat{D}_n$ is the minimum variance unbiased estimator of D.

Sparse alternative

Consider the table with $J = 5, K = 8$ and with cell probabilities

$$p_{ij} = \frac{2^{-(j+k)}}{(1-2^{-J})(1-2^{-K})} + \epsilon\big(\mathbb{1}_{\{j=k=1\}} + \mathbb{1}_{\{j=k=2\}} - \mathbb{1}_{\{j=1,k=2\}} - \mathbb{1}_{\{j=2,k=1\}}\big).$$



Power curves of the USP test (black), compared with Pearson's test (left) and the G-test (right) ($n = 100$, $\alpha = 0.05$). Chi-squared quantile versions of these other tests are in blue (left) and purple (right); permutation versions are in red (left) and green (right).

Continuous case

When $\mathcal{X} = [0,1]^{d_X}$ and $\mathcal{Y} = [0,1]^{d_Y}$ we let (p_{jk}) be the Fourier basis and take

$$\mathcal{P}_1(\rho) = \Big\{ P : \sum_{j,k} (|j|^{s_X} \vee |k|^{s_Y}) \{a_{jk} - a_{j\bullet} a_{\bullet k}\}^2 \leq r^2, \int (f - f_X f_Y)^2 \geq \rho^2$$

$$\max(\|f\|_\infty, \|f_X\|_\infty, \|f_Y\|_\infty) \leq A \Big\}.$$

Let ψ_n be the permutation test described above.

Theorem

There exists $C = C(\alpha, \beta, s_X, s_Y, d_X, d_Y, A) \in (0, \infty)$ such that

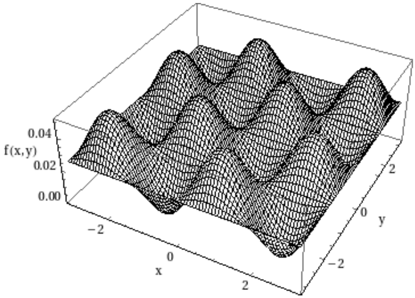
$$\rho^*_{n,\alpha,\beta}(\mathcal{P}_1, \psi_n) \leq C \Big(\frac{r^{d/(2s)}}{n} \Big)^{2s/(4s+d)},$$

where $d := d_X + d_Y$ and $s := d/(d_X/s_X + d_Y/s_Y)$.

Continuous data

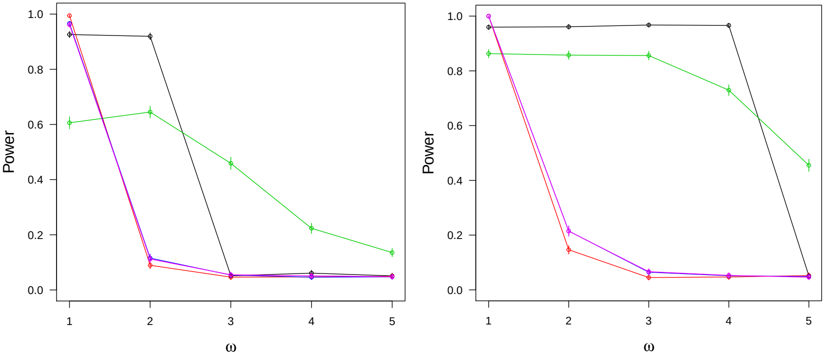
We consider data in $[0,1]^2$ with

$$f_\omega(x,y) = 1 + \sin(2\pi\omega x)\sin(2\pi\omega y).$$



Continuous data

$$f_\omega(x, y) = 1 + \sin(2\pi\omega x)\sin(2\pi\omega y).$$



Estimated power functions in the Sobolev example for our U-statistic permutation test (black) with $M = 2, n = 100$ (left) and $M = 4, n = 200$ (right), HSIC (red), distance covariance (blue), copula (purple) and MINTav (green). Error bars show two standard errors; other parameters: $\alpha = 0.05$, $B = 99$.

General upper bounds

Both previous upper bounds (and more), are consequences of the following general upper bound, holding when we have

$$\mathcal{P}_1 := \Big\{ P : S_\theta(f) = \sum_{j\in\mathcal{J},k\in\mathcal{K}} \theta_{jk}^2 \{a_{jk}(f) - a_{j\bullet}(f)a_{\bullet k}(f)\}^2 \le r^2$$

$$\max(\|f\|_\infty, \|f_X\|_\infty, \|f_Y\|_\infty) \le A \Big\}.$$

Let ψ_n be our general permutation test.

Theorem

Fix $\alpha, \beta \in (0,1)$ such that $\alpha + \beta < 1$. Then there exist $C, C' > 0$, depending only on α, β, θ and A, such that when $nr^2 \ge C'$ we have

$$\rho^*_{n,\alpha,\beta}(\mathcal{P}_1, \psi_n) \le C \inf_{\emptyset \ne \mathcal{M} \subseteq \mathcal{J}\times\mathcal{K}} \max\left\{ \frac{r}{\inf\{\theta_{jk} : (j,k) \notin \mathcal{M}\}}, \frac{|\mathcal{M}|^{1/4} \wedge \|h\|_\infty^{1/2}}{n^{1/2}} \right\}.$$

Upper bounds

We must choose $\mathcal{M}$ to balance the first two terms in the maximum.

Let $\omega : \mathbb{N} \to \mathcal{J} \times \mathcal{K}$ be such that $\theta_{\omega(1)} \leq \theta_{\omega(2)} \leq \ldots$, and write

$$m_0(t) = \min\{m \in \mathbb{N} : m^{1/2}\theta^2_{\omega(m)} > t\}.$$

Corollary

$$\rho^*(n, \alpha, \beta, \xi) \leq C \inf_{m \in \mathbb{N}} \max\left\{\frac{r}{\theta_{\omega(m)}}, \frac{m^{1/4}}{n^{1/2}}\right\} \leq C\frac{m_0(nr^2)^{1/4}}{n^{1/2}}.$$

Actually, we can choose $\mathcal{M}$ adaptively and achieve rate

$$C\frac{\log^{1/4} n}{n^{1/2}} m_0^{1/4}\left(\frac{nr^2}{\log^{1/2} n}\right).$$

1 Problem statement and formalisation

2 U-statistic permutation tests
- Discrete case
- Continuous case
- General upper bound

3 Lower bounds

4 Distributional results and power function

Lower bounds

We provide minimax lower bounds in the case $\mathcal{X} \times \mathcal{Y} = [0,1]^{d_X + d_Y}$.

We lower bound a smaller notion of minimax risk given by

$$\tilde{\mathcal{R}}(n, \xi, \rho) = \inf_{\psi \in \Psi_n(1)} \left\{ \mathbb{E}_{P_0}(\psi) + \sup_{P \in \mathcal{P}_1(\rho)} \mathbb{E}_P(1 - \psi) \right\},$$

where $P_0 = \mathrm{Unif}([0,1]^{d_X + d_Y})$.

For $\gamma \in (0,1)$ write

$$\tilde{\rho}^*(n, \gamma, \xi) = \inf\{\rho > 0 : \mathcal{R}(n, \xi, \rho) \leq \gamma\}.$$

We have $\tilde{\rho}^*(n, \alpha + \beta, \xi) \leq \rho^*(n, \alpha, \beta, \xi)$. When our upper and lower bounds match the risks $\mathcal{R}$ and $\tilde{\mathcal{R}}$ are equivalent (up to constants).

General lemma

Lemma

Let $(a_{jk})_{j\in\mathcal{J}\setminus\{j_0\},k\in\mathcal{K}\setminus\{k_0\}}$ *be a deterministic square-summable array of real numbers, let* $(\xi_{jk})_{j\in\mathcal{J}\setminus\{j_0\},k\in\mathcal{K}\setminus\{k_0\}}$ *be an i.i.d. array of Rademacher random variables, and define a random element of* $L^2(\mu)$ *by*

$$q(x,y) := 1 + \sum_{j\in\mathcal{J}\setminus\{j_0\},k\in\mathcal{K}\setminus\{k_0\}} a_{jk}\xi_{jk}p_{jk}(x,y).$$

If $q \geq 0$ *then* q *is a density. Let* $f \stackrel{d}{=} q|\{q \geq 0\}$ *and write* $\mathbb{E}\mathbb{P}_f^{\otimes n}$ *for the mixture distribution on* $(\mathcal{X}\times\mathcal{Y})^n$*. Then*

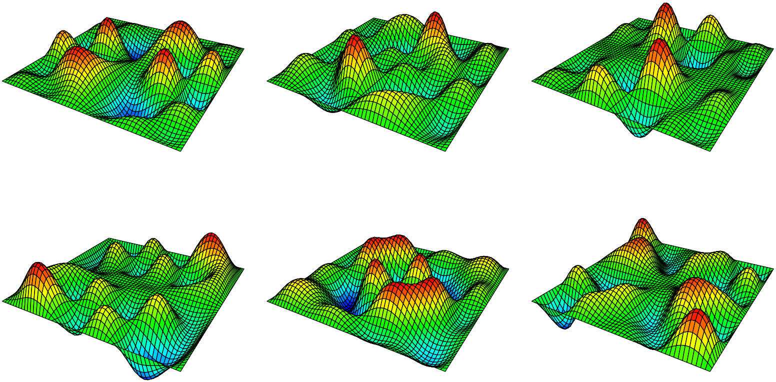
$$d_{\mathrm{TV}}^2\left(\mathbb{P}_{p_0}^{\otimes n}, \mathbb{E}\mathbb{P}_f^{\otimes n}\right) \leq \frac{\exp\left(\frac{(n+1)^2}{2}\sum_{j\in\mathcal{J}\setminus\{j_0\},k\in\mathcal{K}\setminus\{k_0\}} a_{jk}^4\right)}{4\mathbb{P}(q\geq 0)^2} - \frac{1}{4}.$$

When f takes values in $\mathcal{F}_\xi(\rho)$ then we have

$$\tilde{\mathcal{R}}(n,\xi,\rho) \geq 1 - d_{\mathrm{TV}}\left(\mathbb{P}_{p_0}^{\otimes n}, (\mathbb{E}\mathbb{P}_f)^{\otimes n}\right).$$

Fourier lower bounds

$$q(x,y) := 1 + \sum_{j \in \mathcal{J} \setminus \{j_0\}, k \in \mathcal{K} \setminus \{k_0\}} a_{jk} \xi_{jk} p_{jk}(x,y).$$



Realisations of q when $\mathcal{X} \times \mathcal{Y} = [0,1]^2$ and we use the Fourier basis.

Fourier lower bounds

With the Fourier basis on $[0,1]^d$ we may use empirical process techniques to bound

$$\mathbb{P}(q < 0) = \mathbb{P}\Big(\sup_{x,y} \sum_{j,k} a_{jk}\xi_{jk}p_{jk}(x,y) > 1\Big).$$

We show that when $nr^2 \geq 2$ and $(r^d/n^{2s})^{1/(4s+d)} \lesssim 1/\log^{1/2}(n)$ we have

$$\tilde{\rho}^*(n,\gamma,\xi) \gtrsim \left(\frac{r^{d/(2s)}}{n}\right)^{2s/(4s+d)},$$

to match our upper bound in (n, r).

1. Problem statement and formalisation

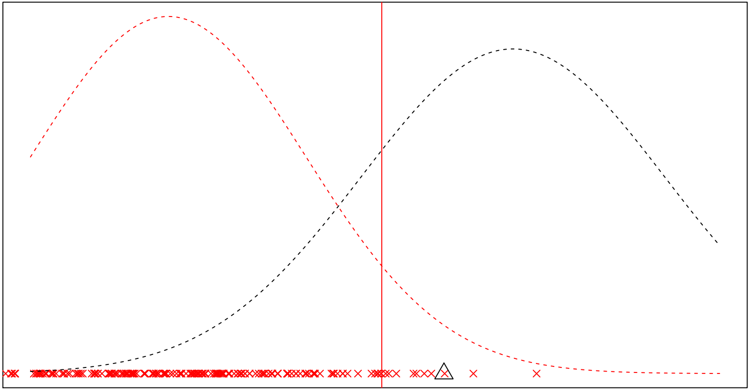
2. U-statistic permutation tests
   - Discrete case
   - Continuous case
   - General upper bound

3. Lower bounds

4. Distributional results and power function

Approximate power function

Upper bounds so far are based on bounds on the mean and variance of permuted U-statistics.



With a more sophisticated analysis we can give more detailed results.

Approximate power function

We can show that $\hat{D}_n, \hat{D}_n^{(1)}, \ldots, \hat{D}_n^{(B)}$ are approximately normally distributed and approximately independent for local alternatives.

Hence, with $j = B + 1 - \lceil \alpha(B+1) \rceil$,

$$\begin{aligned} \mathbb{P}(\text{Reject}) &= \mathbb{P}\big(\hat{D}_n > j\text{th largest } \hat{D}_n^{(b)}\big) \\ &\approx \mathbb{E}\bar{\Phi}\big(W_{(j)} - \mathbb{E}\hat{D}_n/\text{sd}(\hat{D}_n)\big), \end{aligned}$$

where $W_{(j)}$ is the jth order statistic of a standard normal sample.

Degenerate U-statistics

Under H_0, our statistic

$$\hat{D}_n = \frac{1}{4!\binom{n}{4}} \sum_{\substack{i_1,i_2,i_3,i_4 \\ \text{distinct}}} h\big((X_{i_1}, Y_{i_1}), (X_{i_2}, Y_{i_2}), (X_{i_3}, Y_{i_3}), (X_{i_4}, Y_{i_4})\big).$$

is degenerate, meaning that

$$\mathbb{E}\big\{h\big((x_1, y_1), (X_2, Y_2), (X_3, Y_3), (X_4, Y_4)\big)\big\} = 0$$

for all x_1, y_1.

In fact we may approximate

$$\hat{D}_n \approx \binom{n}{2}^{-1} \sum_{i_1 < i_2} h_2\big((X_{i_1}, Y_{i_1}), (X_{i_2}, Y_{i_2})\big)$$

for appropriate h_2.

Approximate normality of U-statistics

Suppose

$$U = \binom{n}{2}^{-1} \sum_{i_1 < i_2} h(Z_{i_1}, Z_{i_2})$$

with h symmetric, $\mathbb{E}\{h(Z_1, Z_2)^2\} < \infty$ and $\mathbb{E}h(z, Z) = 0$ for all z.

Classical asymptotic theory says that

$$\binom{n}{2}^{1/2} U \xrightarrow{d} W = \sum_{j=1}^{\infty} \lambda_j (W_j^2 - 1)$$

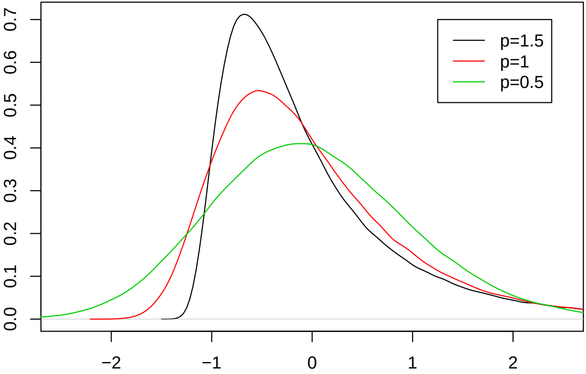
for i.i.d. standard Gaussians (W_j) and a square-summable sequence (λ_j).

When $h = h_n$ this limiting distribution may no longer be appropriate. If the (λ_j) become more diffuse then a normal approximation may hold.

Approximate normality of U-statistics

For each $q \in (2,3]$ there exists $C_q > 0$ depending only on q such that

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(W \le 2^{1/2}\|\lambda\|_2 x\right) - \Phi(x) \right| \le \min\left\{1,\, C_q \frac{\|\lambda\|_q}{\|\lambda\|_2}\right\}$$



The above plot shows the pdf when $\lambda_j = j^{-p}\mathbb{1}_{\{j\le 1000\}}$.

Approximate normality of U-statistics

When

$$U = \binom{n}{2}^{-1/2} \sum_{i_1 < i_2} h(Z_{i_1}, Z_{i_2}),$$

write $g(x, y) := \mathbb{E}\{h(x, Z)h(y, Z)\}$.

Proposition (Döbler and Peccati (2019), Theorem 3.3)

If $\mathbb{E}\{h(Z_1, Z_2)^2\} = 1$, $\mathbb{E}h(z, Z) = 0$ for all z, then there exists a universal constant C such that

$$d_{\mathrm{W}}(U, W)^2 \leq C \max\left[\mathbb{E}\{g(Z_1, Z_2)^2\}, \frac{\mathbb{E}\{h(Z_1, Z_2)^4\}}{n}\right],$$

where $W \sim N(0, 1)$, and d_{W} is the 1-Wasserstein distance.

Approximate normality of permuted U-statistics

We have a similar result for permuted U-statistics of the form

$$U^{(1)} = \binom{n}{2}^{-1/2} \sum_{i_1 < i_2} h\big((X_{i_1}, Y_{\pi(i_1)}), (X_{i_2}, Y_{\pi(i_2)})\big)$$

where π is a uniformly random permutation and where h additionally satisfies

$$\mathbb{E}h\big((x,y),(x',Y_1)\big) = \mathbb{E}h\big((x,y),(X_1,y')\big) = 0$$

for all $x, x' \in \mathcal{X}$ and $y, y' \in \mathcal{Y}$.

Proposition

There exists a universal constant $C > 0$ such that

$$d_{\mathrm{W}}(U^{(1)}, W)^2 \leq C \max\Big[\mathbb{E}\big|\mathbb{E}\{h\big((X_1,Y_2),(X_3,Y_1)\big)|X_3,Y_2\}\big|,$$
$$\mathbb{E}\{g\big((X_1,Y_2),(X_3,Y_4)\big)^2\}, \frac{1}{n}\max_{\sigma \in \mathcal{S}_4} \mathbb{E}\{h\big((X_1,Y_{\sigma(1)}),(X_2,Y_{\sigma(2)})\big)^4\}\Big].$$

Power function

In the $d_X = d_Y = 1$ Fourier problem, the signal strength is given by

$$\frac{\mathbb{E}(\hat{D}_n - \hat{D}_n^{(1)})}{\mathrm{Var}^{1/2}(\hat{D}_n^{(1)})} \sim \frac{\binom{n}{2}^{1/2} \sum_{(j,k)\in\mathcal{M}} |a_{jk} - a_{j\bullet}a_{\bullet k}|^2}{\sigma_{M,X}\sigma_{M,Y}} =: \Delta_f \,,$$

where $\sigma^2_{M,X} := \sum_{j=-2M}^{2M}(2M+1-j)|a_{j\bullet}|^2 \asymp M\|f_X\|_2^2$ and M is the maximum frequency our statistic looks at.

The quality of the normal approximation is controlled by the (small) quantity

$$\delta_* := \max\left\{ \frac{\Delta_f^{1/2} \vee 1}{M^{1/2}}, D^{1/4}, \left(\frac{M^2}{n}\right)^{1/2}, \frac{A_{M,X}A_{M,Y}}{M} \right\}^{1/3},$$

where $A_{M,X} := \sum_{j=-2M}^{2M} |a_{j\bullet}| = o(M^{1/2})$.

Power function

Theorem

Consider f with $\|f\|_\infty < \infty$, let $\alpha \in (0,1)$ and let $B \in \mathbb{N}$. With $s = \lceil \alpha(B+1) \rceil - 1$, let $\mathrm{B}_{B-s,s+1} \sim \mathrm{Beta}(B-s, s+1)$. Then there exists $C = C(\|f\|_\infty, \alpha) > 0$ such that

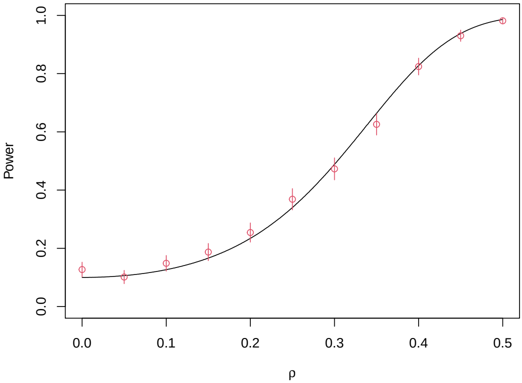
$$\begin{aligned} &\left|\mathbb{P}_f(p \le \alpha) - \mathbb{E}\bar{\Phi}\left(\Phi^{-1}(\mathrm{B}_{B-s,s+1}) - \Delta_f\right)\right| \\ &\qquad \le \epsilon_{n,M,B}(f) := C \min\{B^{4/3}\delta_*, (B^{-1/3} \vee \delta_*^{1/3})\}. \end{aligned}$$

Simulations

With $n = 300, M = 7, \alpha = 0.1, B = 99$ and

$$f(x,y) = 1 + \rho \sin(2\pi x)\sin(2\pi y)$$

for $\rho = 0, 0.05, 0.1, \ldots, 0.5$.



Here $\sigma^2_{M,X} = \sigma^2_{M,Y} = 2M + 1$ and $A_{M,X} = A_{M,Y} = 1$.

Conclusion

- We have provided a general upper bound on the minimax separation for independence testing by considering the power of a permutation test.
- We have given matching lower bounds, which therefore establish that our permutation test is optimal in certain cases.
- Based on new approximate normality results for U-statistics calculated on permuted data sets we prove more detailed power results.

Thank you!

Berrett, T. B., Kontoyiannis, I. and Samworth, R. J. (2021) Optimal rates for independence testing via U-statistic permutation tests. *Ann. Statist.*, to appear.

Berrett, T. B. and Samworth R. J. (2021) USP: an independence test that improves on Pearson's chi-squared and the G-test. `arXiv:2101.10880`

References

Albert, M., Bouret, Y., Fromont, M. and Reynaud-Bouret, P. (2015) Bootstrap and permutation tests of independence for point processes. *Ann. Statist.*, **43**, 2537–64.

Agresti, A. (1996) *An Introduction to Categorical Data Analysis* (1st edition). Wiley Series in Probability and Statistics.

B., and Samworth, R. J. (2019) Nonparametric independence testing via mutual information. *Biometrika, to appear.*

Deb, N. and Sen, B. (2019). Multivariate rank-based distribution-free nonparametric testing using measure transportation. *Available at* `arXiv:1909.08733`.

Döbler, C. and Peccati, G. (2019) Quantitative CLTs for symmetric U-statistics. *Electron. J. Probab.*, **24**, 1–43.

Ermakov, M. S. (1990) Asymptotically minimax tests for nonparametric hypotheses concerning the distribution density. *J. Sov. Math.*, **52**, 2891–2898.

Fisher, R. A. (1935) *The Design of Experiments* (1st Ed.). Oliver and Boyd, Edinburgh.

Gretton A., Bousquet O., Smola A. and Schölkopf B. (2005) Measuring Statistical Dependence with Hilbert-Schmidt Norms. *Algorithmic Learning Theory*, 63–77.

Gretton, A. and Györfi, L. (2010). Consistent nonparametric tests of independence. *J. Mach. Learn. Res.*, **11**, 1391–423.

Kendall, M. G. (1938) A new measure of rank correlation. *Biometrika*, **30**, 81–93.

Kim, I., Balakrishnan, S. and Wasserman, L. (2021) Minimax optimality of permutation tests. *Ann. Statist., to appear*.

Kojadinovic, I. and Holmes, M. (2009) Tests of independence among continuous random vectors based on Cramér–von Mises functionals of the empirical copula process. *J. Multivariate Anal.*, **100**, 1137–54.

References

Heller, R., Heller, Y., Kaufman, S., Brill, B. and Gorfine, M. (2016) Consistent distribution-free K-sample and independence tests for univariate random variables. *J. Mach. Learn. Res.*, **17**, 1–54.

Hall, P. (1984) Central limit theorem for integrated square error multivariate nonparametric density estimators. *J. Multivar. Anal.*, **14**, 1–16.

Hoeffding, W. (1948) A non-parametric test of independence. *Ann. Math. Statist.*, **19**, 546–57.

Hoeffding, W. (1952) The large-sample power of tests based on permutations of observations. *Ann. Math. Statist*, **23**, 169–192.

Ingster, Y. I. (1989) Asymptotic minimax testing of independence hypothesis. *J. Sov. Math.*, **44**, 466–476.

Ingster, Y. I. (1996) Minimax testing of the hypothesis of independence for ellipsoids in ℓ_p. *J. Math. Sci.*, **81**, 2406–2420.

Lehmann, E. L. and Romano, J. P. (2005) *Testing Statistical Hypotheses* (3rd Ed.). Springer, New York.

Albert, M., Laurent, B., Marrel, A. and Meynaoui, A. (2019) Adaptive test of independence based on HSIC measures. *Available at* `arXiv:1902.06441`.

Pitman, E. J. G. (1938) Significance tests which may be applied to samples from any populations: III. The analysis of variance test. *Biometrika*, **29**, 322–335.

Pearson, K. (1920) Notes on the history of correlation. *Biometrika*, **13**, 25–45.

Pfister, N., Bühlmann, P., Schölkopf, B. and Peters, J. (2018) Kernel-based tests for joint independence. *J. R. Statist. Soc. B.*, **80**, 5–31.

References

Romano, J. P. (1989) Bootstrap and randomization tests of some nonparametric hypotheses. *Ann. Statist.*, **17**, 141–159.

Sejdinovic, D., Sriperumbudur, B., Gretton, A. and Fukumizu, K. (2013) Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.*, **41**, 2263–2291.

Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.

Shi, H., Drton, M., and Han, F. (2019) Distribution-free consistent independence tests via Hallin's multivariate rank. *Available at* `arXiv:1909.10024`.

Song, L., Smola, A., Gretton, A., Bedo, J. and Borgwardt, K. (2012) Feature selection via dependence maximization. *J. Mach. Learn. Res.*, **13**, 1393–1434.

Sullivan III, M. (2017) *Statistics: Informed Decisions using Data* (5th edition). Pearson.

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

Székely, G. J. and Rizzo, M. L. (2013) The distance correlation t-test of independence in high dimension. *J. Multivariate Anal.*, **117**, 193–213.

Weihs, L., Drton, M. and Meinshausen, N. (2017) SymRC: Estimating symmetric rank covariances. `https://github.com/Lucaweihs/SymRC`.