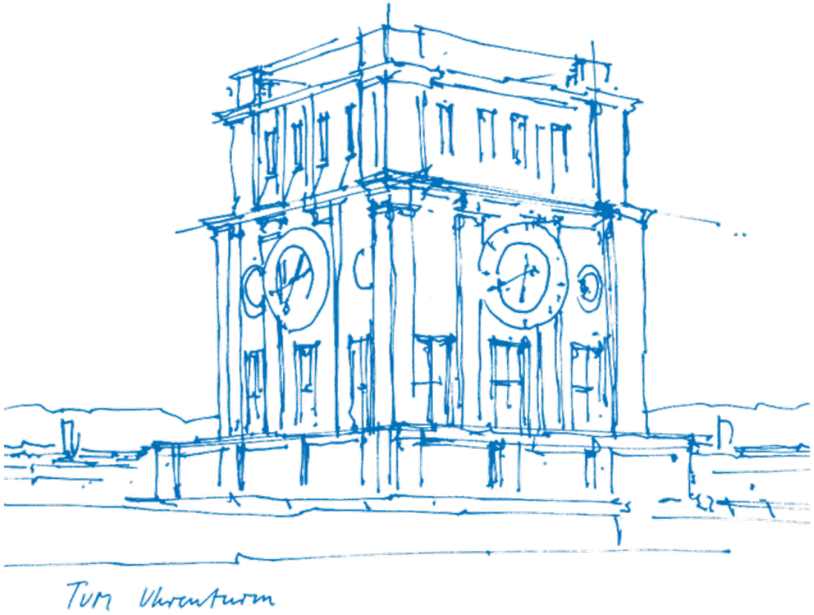
Consistent tests of independence via rank statistics

Mathias Drton

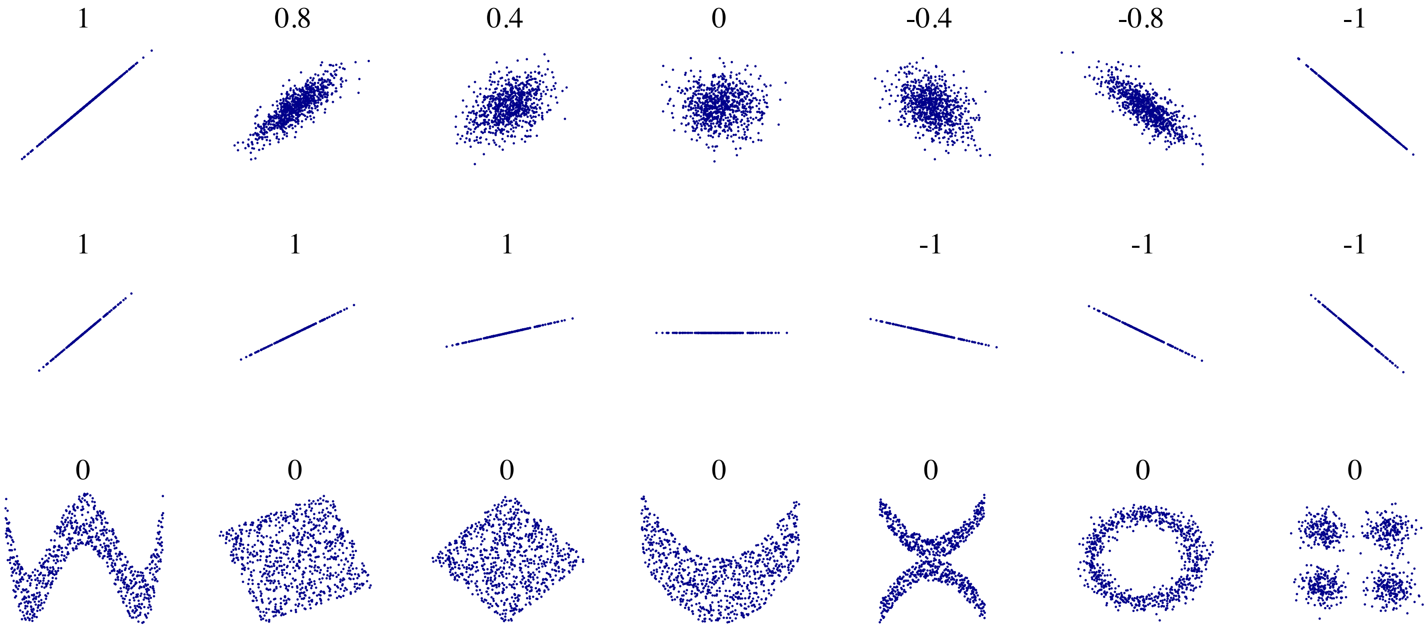
Department of Mathematics

Technical University of Munich (TUM)

(joint work with Hongjian Shi, Marc Hallin, Fang Han)



Values of Pearson correlation



Wikipedia: Correlation

Topic of the talk

- How to form statistical tests that detect general non-linear dependence?

  And why is this question interesting?

- How to form tests that 'work' for any type of continuous data?

  Considered solution: Use ranks of the observed values

- How to do this for vector-valued data?

  How to rank in dimension ≥ 2? Optimal transport perspective...

1. Why insist on consistency?

Motivation from causal discovery

Fact: Conditional expectation $\mathbb{E}[Y|X]$ 'best' predicts Y as function of X and

$$\operatorname{Corr}\left[\, Y - \mathbb{E}[Y|X]\,,\; X\,\right] \;=\; \operatorname{Corr}\left[\, X - \mathbb{E}[X|Y]\,,\; Y\,\right] \;=\; 0.$$

Motivation from causal discovery

Fact: Conditional expectation $\mathbb{E}[Y|X]$ 'best' predicts Y as function of X and

$$\text{Corr}\,[\,Y - \mathbb{E}[Y|X]\,,\;X\,] \;=\; \text{Corr}\,[\,X - \mathbb{E}[X|Y]\,,\;Y\,] \;=\; 0.$$

- Causal model (additive noise model):

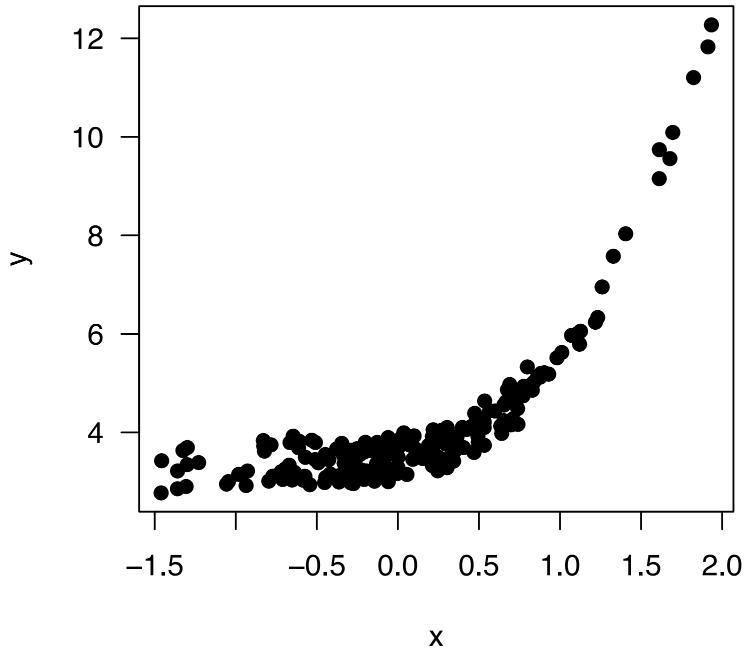
$$Y = f(X) + \epsilon \quad \text{with} \quad X \perp\!\!\!\perp \epsilon.$$

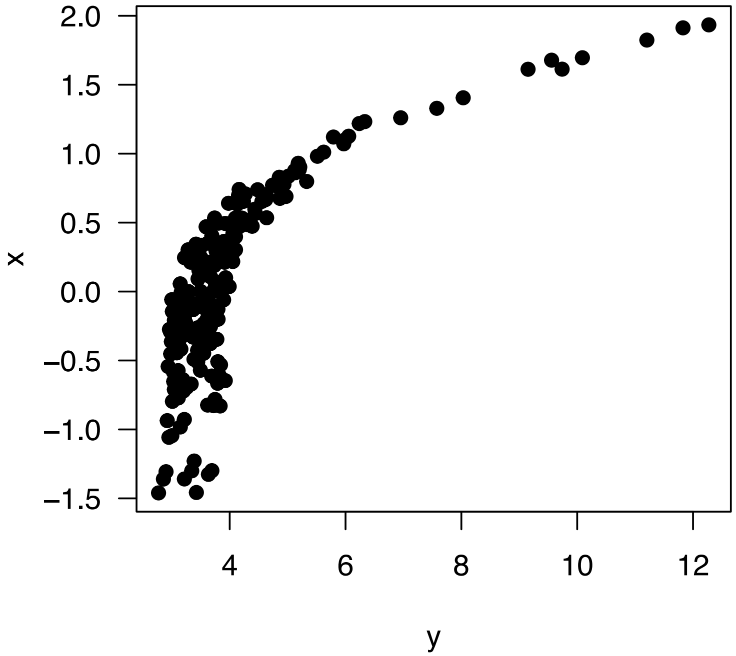
- In this model, $Y - \mathbb{E}[Y|X] = Y - f(X) = \epsilon$ is independent of X (not only uncorrelated).

- In contrast, for a general non-linear f, it holds that

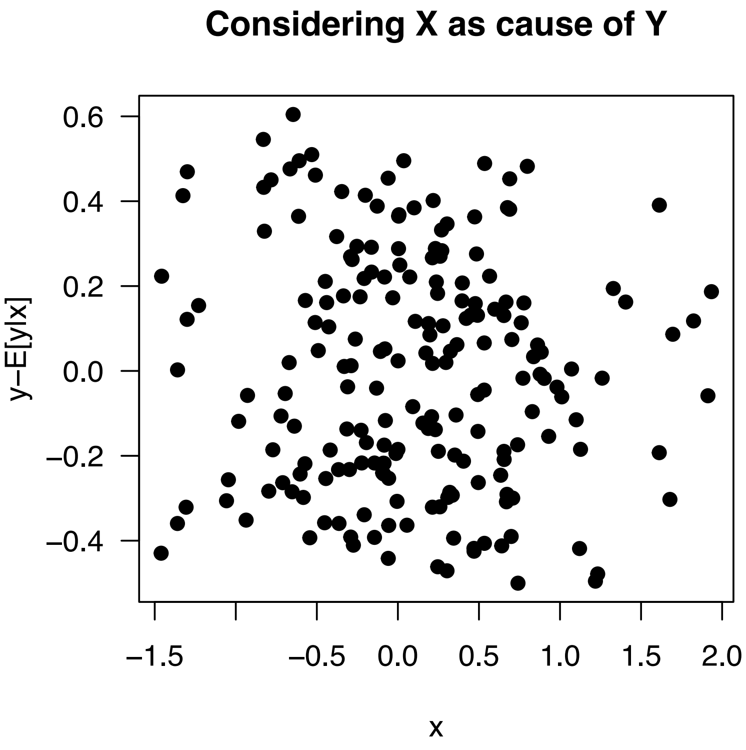
$$X - \mathbb{E}[X|Y] \quad \not\!\perp\!\!\!\perp \quad Y \qquad \text{(despite zero Pearson correlation).}$$

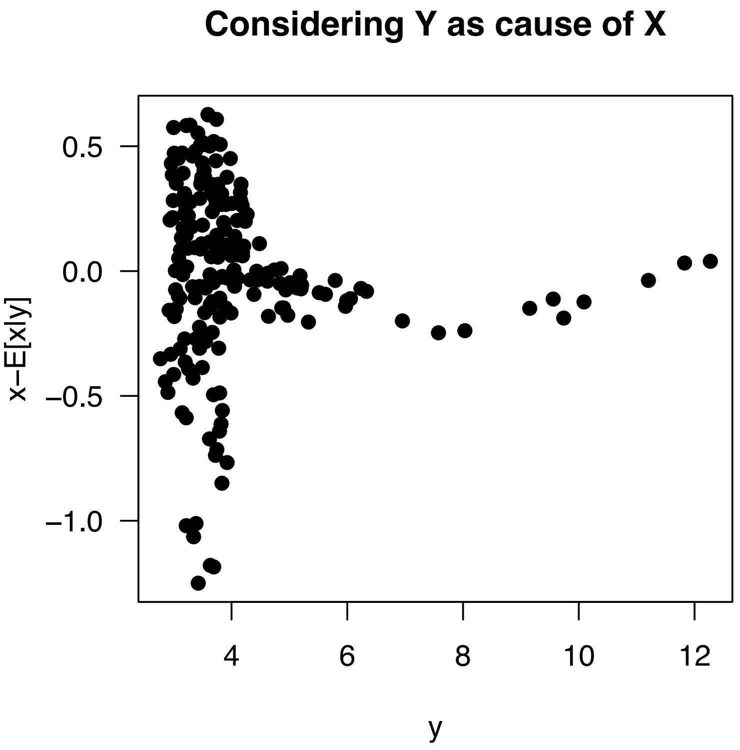
Causal discovery: Two plots of one bivariate dataset





Causal discovery: Subtracting effect of postulated cause





Causal discovery

```r
cor(x,ey)
## [1] -0.06502044
cor.test(x,ey)$p.value
## [1] 0.360332
```

```r
cor(y,ex)
## [1] -0.08028996
cor.test(y,ex)$p.value
## [1] 0.2584004
```

Causal discovery

```r
cor(x,ey)
```

```
## [1] -0.06502044
```

```r
cor.test(x,ey)$p.value
```

```
## [1] 0.360332
```

```r
cor(y,ex)
```

```
## [1] -0.08028996
```

```r
cor.test(y,ex)$p.value
```

```
## [1] 0.2584004
```

```r
tauStarTest(x,ey)
```

```
## Test Type: asymptotic continuous
## Input Length: 200
##
## Results:
##  t* value Asym. p-val
##   0.00081     0.26329
```

```
## Bootstrap p-value:  0.086
```

```r
tauStarTest(y,ex)
```

```
## Test Type: asymptotic continuous
## Input Length: 200
##
## Results:
##  t* value Asym. p-val
##   0.02626       1e-04
```

```
## Bootstrap p-value:  0.001
```

2. Rank correlations

Rank correlations

- If Pearson is out, then 'next best thing':

$$\text{Spearman's } \rho \quad \text{or} \quad \text{Kendall's } \tau.$$

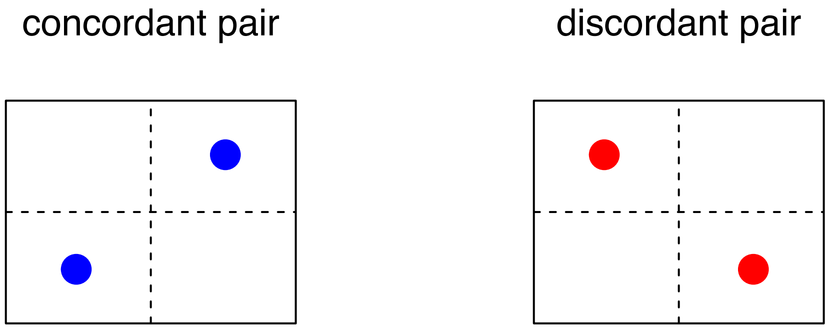
- Estimators $\hat{\rho}_n$ and $\hat{\tau}_n$ are functions of the ranks of $(X^{(1)}, \ldots, X^{(n)})$ and of $(Y^{(1)}, \ldots, Y^{(n)})$.
- The rank of $X^{(i)}$ is

$$R_i^X = \#\{j : X^{(j)} \leq X^{(i)}\}, \qquad i = 1, \ldots, n.$$

```
##         x ranks
##  -0.9804     1
##  -0.4679     2
##   1.7163     7
##   0.5131     4
##   0.4287     3
##   0.7011     5
##   0.7526     6
```

Kendall

- Correlation coefficient: $\hat{\tau}_n = \dfrac{\#\{\text{concordant pairs}\} - \#\{\text{discordant pairs}\}}{\binom{n}{2}}$
- Point configurations:



- Correlation measure: $\tau = \Pr(\text{concordant pair}) - \Pr(\text{discordant pair})$

Spearman

- Correlation coefficient: $\hat{\rho}_n =$ Pearson correlation of rank vectors

```
cor(x, y, method = "spearman")
## [1] 0.7575758
```

```
cor(rank(x), rank(y))
## [1] 0.7575758
```

Spearman

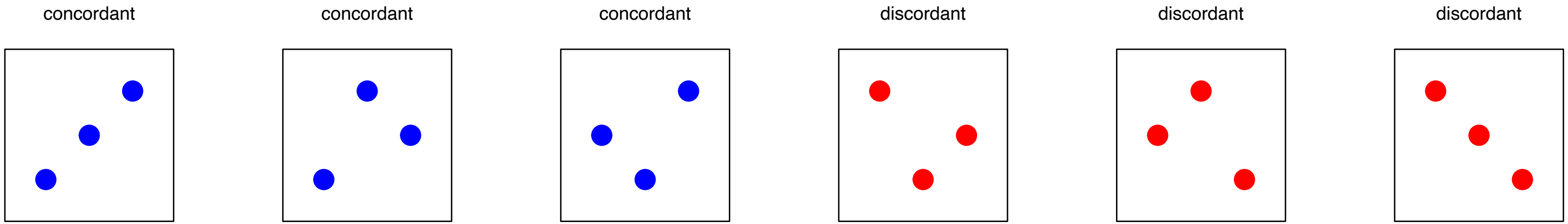
- Correlation coefficient: $\hat{\rho}_n =$ Pearson correlation of rank vectors

```
cor(x, y, method = "spearman")
## [1] 0.7575758
```

```
cor(rank(x), rank(y))
## [1] 0.7575758
```

- Correlation measure: $\rho = \Pr(\text{concordant triple}) - \Pr(\text{discordant triple})$

- Point configurations:



Rank correlations are great for independence testing

- Let $(X^{(1)}, Y^{(1)}), \dots, (X^{(n)}, Y^{(n)})$ be i.i.d. sample from continuous distribution $P^{(X,Y)}$.

- Fact:

  If X and Y independent, then two rank vectors are independent and uniformly distributed:

$$X \perp\!\!\!\perp Y \quad \Longrightarrow \quad \forall n \in \mathbb{N} : (R_1^X, \dots, R_n^X) \perp\!\!\!\perp (R_1^Y, \dots, R_n^Y) \text{ and}$$
$$(R_1^X, \dots, R_n^X), (R_1^Y, \dots, R_n^Y) \sim \text{Uniform}(\mathcal{S}_n)$$

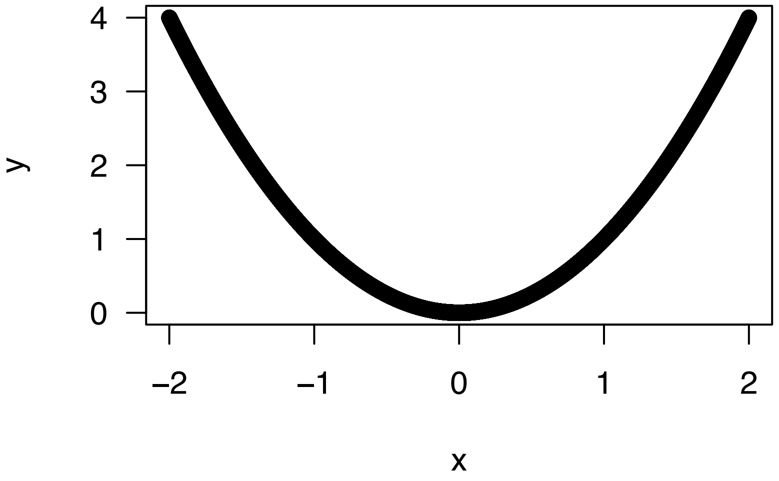
- Conclusion:

  Rank correlation coefficients are distribution-free under null hypothesis of independence.

  Calibrate independence tests via exact distributions (for small n) or uniformly valid asymptotic approximations (e.g., $\hat{\rho}_n$ and $\hat{\tau}_n$ are asymptotically normal).

Spearman/Kendall not consistent

Although ρ and τ have many uses, they do not give consistent tests of independence.




```r
cor(x, y, method = "spearman")

## [1] 6.648968e-05

library(pcaPP)
cor.fk(x, y)

## [1] 4.490223e-05

# 'fast' Kendall
```

Kendall:
naive $O(n^2)$; efficient $O(n \log(n))$

Hoeffding's problem

- We know

$$\boxed{X \perp\!\!\!\perp Y} \implies \forall n \in \mathbb{N} : \boxed{\begin{array}{l}(R_1^X, \ldots, R_n^X) \perp\!\!\!\perp (R_1^Y, \ldots, R_n^Y) \text{ and} \\ (R_1^X, \ldots, R_n^X), (R_1^Y, \ldots, R_n^Y) \sim \text{Uniform}(\mathcal{S}_n)\end{array}}$$

- Is there a converse?

- What is the smallest $n \equiv n_{\text{Hoeffding}}$ such that

$$\boxed{\begin{array}{l}(R_1^X, \ldots, R_n^X) \perp\!\!\!\perp (R_1^Y, \ldots, R_n^Y) \text{ and} \\ (R_1^X, \ldots, R_n^X), (R_1^Y, \ldots, R_n^Y) \sim \text{Uniform}(\mathcal{S}_n)\end{array}} \implies \boxed{X \perp\!\!\!\perp Y} \quad ?$$

- Certainly, $n_{\text{Hoeffding}} > 2$ ... e.g., $Y = X^2$ with $X \sim N(0, 1)$.

3. Consistent rank correlations

Hoeffding's D (1948)

$$D = \int (F(x,y) - F_X(x)F_Y(y))^2 \, \mathrm{d}F(x,y)$$

- Here, F is the joint distribution function and F_X, F_Y are the marginal d.f.
- D is consistent for absolutely continuous bivariate distributions: $D = 0 \iff X \perp\!\!\!\perp Y$.
- Unbiased estimator (U-statistic) based on ranks of 5-tuples of data points:

$$\widehat{D}_n = \frac{1}{\binom{n}{5}} \sum_{1 \le i_1 < i_2 < \cdots < i_5 \le n} h_D\left(\begin{pmatrix} X^{(i_1)} \\ Y^{(i_1)} \end{pmatrix}, \ldots, \begin{pmatrix} X^{(i_5)} \\ Y^{(i_5)} \end{pmatrix} \right).$$

For example, $\int F_X(x)^2 F_Y(y)^2 \, \mathrm{d}F(x,y) = \mathbb{E}[\mathbf{1}(X_1 \le X_5)\mathbf{1}(X_2 \le X_5)\mathbf{1}(Y_3 \le Y_5)\mathbf{1}(Y_4 \le Y_5)]$.

- Existence of this estimator implies $n_{\text{Hoeffding}} \le 5$.

Blum-Kiefer-Rosenblatt's R (1961)

$$R = \int (F(x,y) - F_X(x)F_Y(y))^2 \, \mathrm{d}F_X(x)F_Y(y)$$

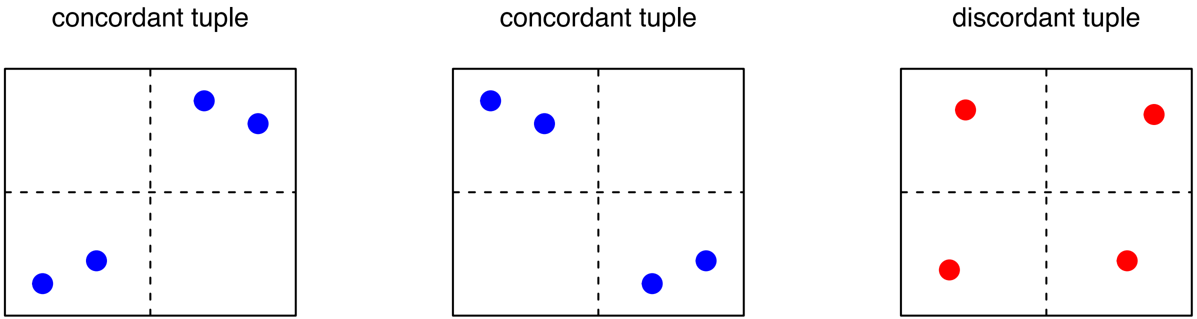
- Consistent for any bivariate distribution:

$$R = 0 \iff X \perp\!\!\!\perp Y.$$

- Admits unbiased estimator $\widehat{R}_n$ based on ranks of 6-tuples.
- Under independence, asymptotic equivalence: $\widehat{D}_n - \widehat{R}_n = o_p(1)$.

Bergsma-Dassios' τ^* (2014)

$$\tau^* = \mathbb{E}[\hat{\tau}_n^*], \quad \hat{\tau}_n^* = \frac{1}{\binom{n}{4}} \left(\#\{\text{concordant 4-tuples}\} - \#\{\text{discordant tuples}\} \right)$$



τ^* known to be consistent for dependence in continuous and discrete distributions.

Thus, $n_{\text{Hoeffding}} \leq 4$. Little known work of Yanagimoto (1970) gives $n_{\text{Hoeffding}} = 4$

Computation

- Naive counting of tuples:

$$
\begin{aligned}
&O(n^5) \text{ for } \widehat{D}_n,\\
&O(n^6) \text{ for } \widehat{R}_n,\\
&O(n^4) \text{ for } \widehat{\tau}^*_n.
\end{aligned}
$$

- But Hoeffding cleverly shows that his $\widehat{D}_n$ can be evaluated in $O(n \log n)$ time.
- For $\widehat{\tau}^*$, newest algorithm also achieves $O(n \log n)$ (Even-Zohar and Leng, 2021).

  (R package: `independence`)
- A relation of Yanagimoto (1970) implies $O(n \log n)$ computation also for BKR's $\widehat{R}_n$

  (D., Han, Shi, 2020).

Asymptotics

Theorem

If X and Y are independent continuous random variables, then

$$n\widehat{D}_n,\ n\widehat{R}_n,\ \frac{n}{36}\hat{\tau}_n^* \xrightarrow{d} \frac{1}{\pi^4}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{i^2 j^2}(\xi_{ij}^2 - 1)$$

where ξ_{ij}, $i, j \geq 1$, are i.i.d. $N(0,1)$ random variables.

- Under independence, degenerate U-statistics of order 2 (Hoeffding, 1948; Nandy, Weihs, D., 2016; D., Han, Shi, 2020).
- R package to compute asymptotic p-values for independence test: `tauStar` (Luca Weihs).

Chatterjee's new coefficient of correlation (2021)

Research continues...

Chatterjee (2021) proposed a rank correlation coefficient ξ_n that amazingly is

- computable in $O(n \log n)$,
- consistently estimates a dependence measure ξ with

$$\xi = 0 \iff X \perp\!\!\!\perp Y \qquad \text{(consistent for dependence), and}$$

$$\xi = 1 \iff \exists f : Y = f(X) \qquad \text{(consistent for functional dependence),}$$

- asymptotically normal.

Chatterjee's new coefficient of correlation (2021)

Research continues...

Chatterjee (2021) proposed a rank correlation coefficient ξ_n that amazingly is

- computable in $O(n \log n)$,
- consistently estimates a dependence measure ξ with

$$\xi = 0 \iff X \perp\!\!\!\perp Y \qquad \text{(consistent for dependence), and}$$

$$\xi = 1 \iff \exists f : Y = f(X) \qquad \text{(consistent for functional dependence),}$$

- asymptotically normal.

Unfortunately, independence test based on ξ_n has suboptimal power (Shi, D. & Han, 2022)

(cannot detect signals of order $\frac{1}{\sqrt{n}}$, but see Lin & Han, 2021, for possible improvements)

4. Multivariate independence

Testing multivariate independence

- Consider abs. continuous random vectors

$$\boldsymbol{X} \in \mathbb{R}^p, \quad \boldsymbol{Y} \in \mathbb{R}^q.$$

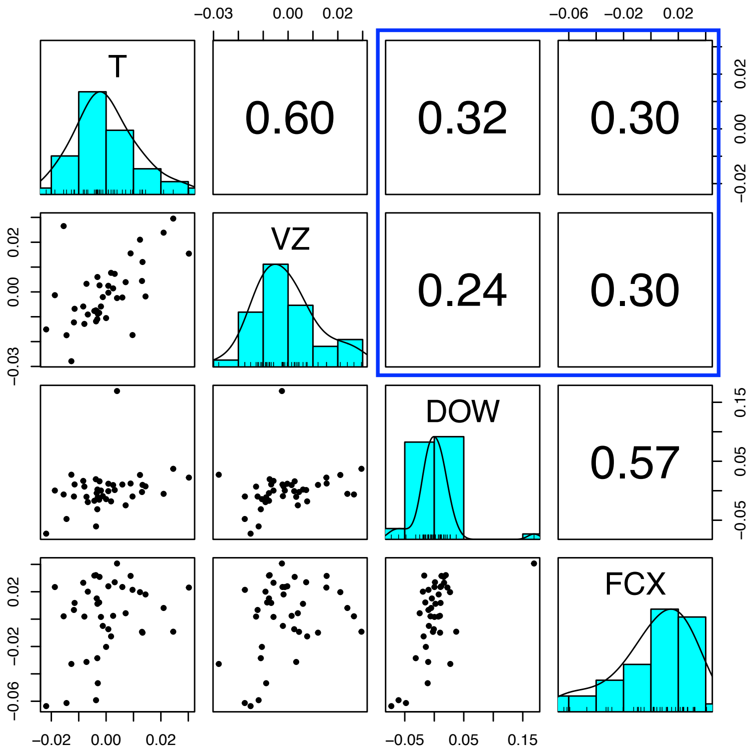
- Test independence

$$H_0 : \boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \quad \text{vs.} \quad H_1 : \boldsymbol{X} \not\perp\!\!\!\perp \boldsymbol{Y}$$

  based on an i.i.d. sample

$$\begin{pmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{Y}^{(1)} \end{pmatrix}, \ldots, \begin{pmatrix} \boldsymbol{X}^{(n)} \\ \boldsymbol{Y}^{(n)} \end{pmatrix} \sim P^{(\boldsymbol{X},\boldsymbol{Y})}.$$

- Goal: consistent and distribution-free test with local power...



Distance covariance

- Consistent measures of dependence have been developed also for general dimension.
- Distance covariance (Szekely et al., 2007) has become a popular option:

$$\mathrm{dCov}_n^2(\boldsymbol{X}, \boldsymbol{Y}) \;=\; \frac{1}{n^2} \sum_{ij} d_{ij,0}^X \cdot d_{ij,0}^X$$

  with $d_{ij,0}^X$ and $d_{ij,0}^Y$ recentered versions of pairwise distances $d_{ij}^X = \|\boldsymbol{X}^{(i)} - \boldsymbol{X}^{(j)}\|$ and $d_{ij}^Y = \|\boldsymbol{Y}^{(i)} - \boldsymbol{Y}^{(j)}\|$.

- For $\boldsymbol{X}$ and $\boldsymbol{Y}$ with finite first moments, this estimates

$$\mathrm{dCov}^2(\boldsymbol{X}, \boldsymbol{Y}) = \int_{\mathbb{R}^{p+q}} [\phi_{\boldsymbol{XY}}(s, t) - \phi_{\boldsymbol{X}}(s)\phi_{\boldsymbol{Y}}(t)]^2 \, w(s, t) \; ds \, dt$$

  which is zero iff $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y}$.

Distance covariance of ranks

- In the spirit of Spearman's ρ, we can apply distance covariance to vectors of ranks:
  - Compute ranks of each variable=coordinate of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. So for $\boldsymbol{X}$:

$$\begin{pmatrix} X_1^{(1)} & \dots & X_1^{(n)} \\ X_2^{(1)} & \dots & X_2^{(n)} \\ \vdots & & \vdots \\ X_p^{(1)} & \dots & X_p^{(n)} \end{pmatrix} \xmapsto{\text{row-wise marginal ranks}} \begin{pmatrix} R_1^{X_1} & \dots & R_n^{X_1} \\ R_1^{X_2} & \dots & R_n^{X_2} \\ \vdots & & \vdots \\ R_1^{X_p} & \dots & R_n^{X_p} \end{pmatrix}$$

  - Apply distance covariance to the resulting vectors of marginal ranks.

- Interestingly, in 1D ($p = q = 1$), this gives an estimate of BKR's R (Shi, D., Han, 2022).

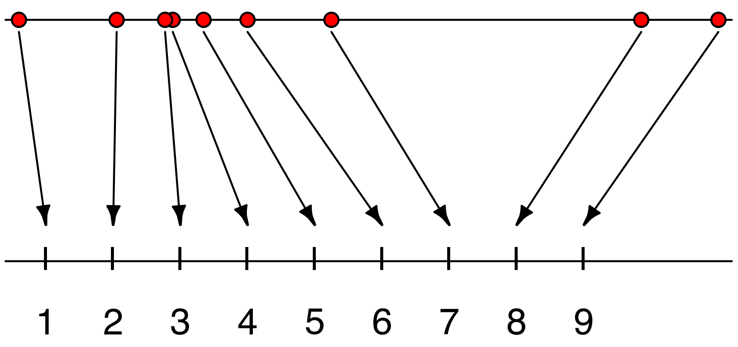
- However, in higher dimension, marginal ranks not distribution-free...

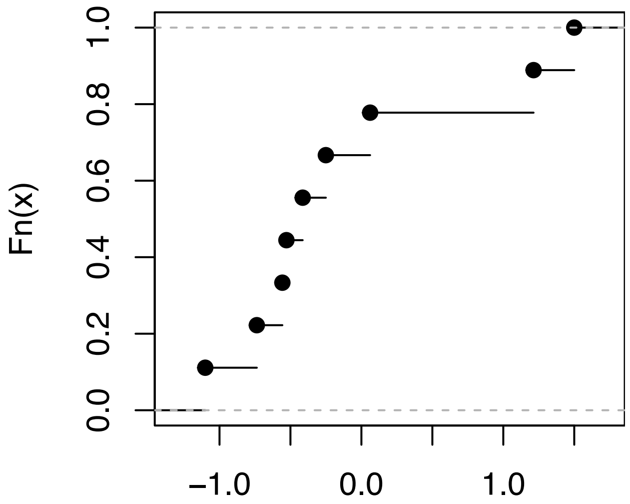
- Solution: New concept of multivariate ranks via optimal transport (Hallin et al., 2017,...)

Ranks and distribution function: Univariate case

Empirical distribution function computes the ranks:

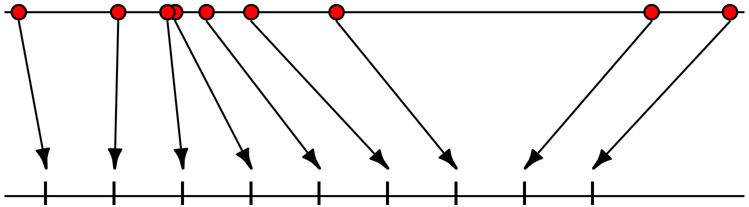
$$\widehat{F}_X^{(n)}(t) = \frac{1}{n}\#\{j : X^{(j)} \leq t\} \quad \Longrightarrow \quad \widehat{F}_X^{(n)}(X^{(i)}) \; = \; \frac{1}{n} R_i^X$$





Transport perspective

- Empirical distribution function $\widehat{F}_X^{(n)}$ solves an assignment problem: data $\rightarrow$ equidistant grid:



- Population d.f. $F_X(t) = \Pr(X \le t)$ is characterized as nondecreasing fct. that pushes P^X to Unif$(0,1)$.

Recall: If X has cdf F_X then $F_X(X) \sim$ Uniform$(0,1)$.

- Hallin's center-outward perspective (signed ranks):

$$2F_X - 1,\ 2\widehat{F}_X^{(n)} - 1 \text{ map to ``unit ball'' } (-1,1).$$

- Generalize to higher dimension via spherical uniform distribution U_p on unit ball $\{\boldsymbol{x} \in \mathbb{R}^p : \|\boldsymbol{x}\|_2 < 1\}$.

General dimension: Center-outward ranks and signs

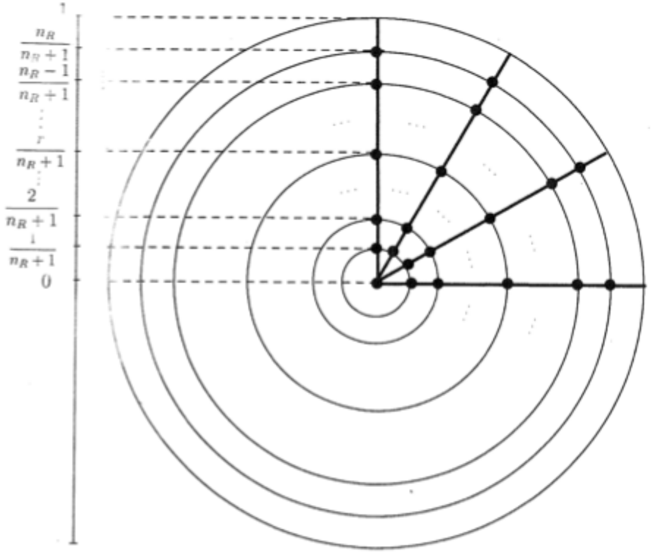
- Center-outward distribution function $\mathbf{F}_{\boldsymbol{X},\pm}$ is an optimal transport map that pushes $P^{\boldsymbol{X}}$ to U_p.

  If $\mathbb{E}[\|\boldsymbol{X}\|_2^2] < \infty$, then $\mathbf{F}_{\boldsymbol{X},\pm} = \arg\min_T \mathbb{E}\left[\|T(\boldsymbol{X}) - \boldsymbol{X}\|_2^2\right]$ subject to $P^{T(\boldsymbol{X})} = U_p$.

- Empirical center-outward d.f. $\widehat{\mathbf{F}}^{(n)}_{\boldsymbol{X},\pm}$ solves a linear sum assignment problem:

$$\widehat{\mathbf{F}}^{(n)}_{\boldsymbol{X},\pm} := \arg\min_{T_n} \sum_{i=1}^{n} \|\boldsymbol{X}^{(i)} - T_n(\boldsymbol{X}^{(i)})\|^2,$$

  where T_n assigns data points to a (suitably uniform) grid in unit ball.



Distribution-free consistent test

- Reject for large values of the test statistic:

$$W_n \; := \; \mathrm{dCov}_n^2\left(\hat{\mathbf{F}}^{(n)}_{\boldsymbol{X},\pm}(\boldsymbol{X}^{(i)})_{i=1}^n, \hat{\mathbf{F}}^{(n)}_{\boldsymbol{Y},\pm}(\boldsymbol{Y}^{(i)})_{i=1}^n\right)$$

- Test is consistent:
  a) $W_n \longrightarrow_{a.s.} \mathrm{dCov}^2\left(\mathbf{F}_{\boldsymbol{X},\pm}(\boldsymbol{X}), \mathbf{F}_{\boldsymbol{Y},\pm}(\boldsymbol{Y})\right)$, and

  b) $\mathbf{F}_{\boldsymbol{X},\pm}$ has an inverse that pushes back to $\boldsymbol{X}$.

- Test is distribution-free as under H_0:

$$\left[\hat{\mathbf{F}}^{(n)}_{\boldsymbol{X},\pm}(\boldsymbol{X}^{(i)})\right]_{i=1}^n \perp\!\!\!\perp \left[\hat{\mathbf{F}}^{(n)}_{\boldsymbol{Y},\pm}(\boldsymbol{Y}^{(i)})\right]_{i=1}^n, \qquad \left[\hat{\mathbf{F}}^{(n)}_{\boldsymbol{X},\pm}(\boldsymbol{X}^{(i)})\right]_{i=1}^n, \left[\hat{\mathbf{F}}^{(n)}_{\boldsymbol{Y},\pm}(\boldsymbol{Y}^{(i)})\right]_{i=1}^n \sim \text{Uniform on the grid.}$$

Asymptotic distribution and local power

Theorem

If $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *are independent, then*

$$nW_n \longrightarrow_d \sum_{v=1}^{\infty} \lambda_v(\xi_v^2 - 1),$$

where the ξ_v *are i.i.d.* $N(0,1)$ *r.v., and the* λ_v *depend on the dimensions* p *and* q*.*

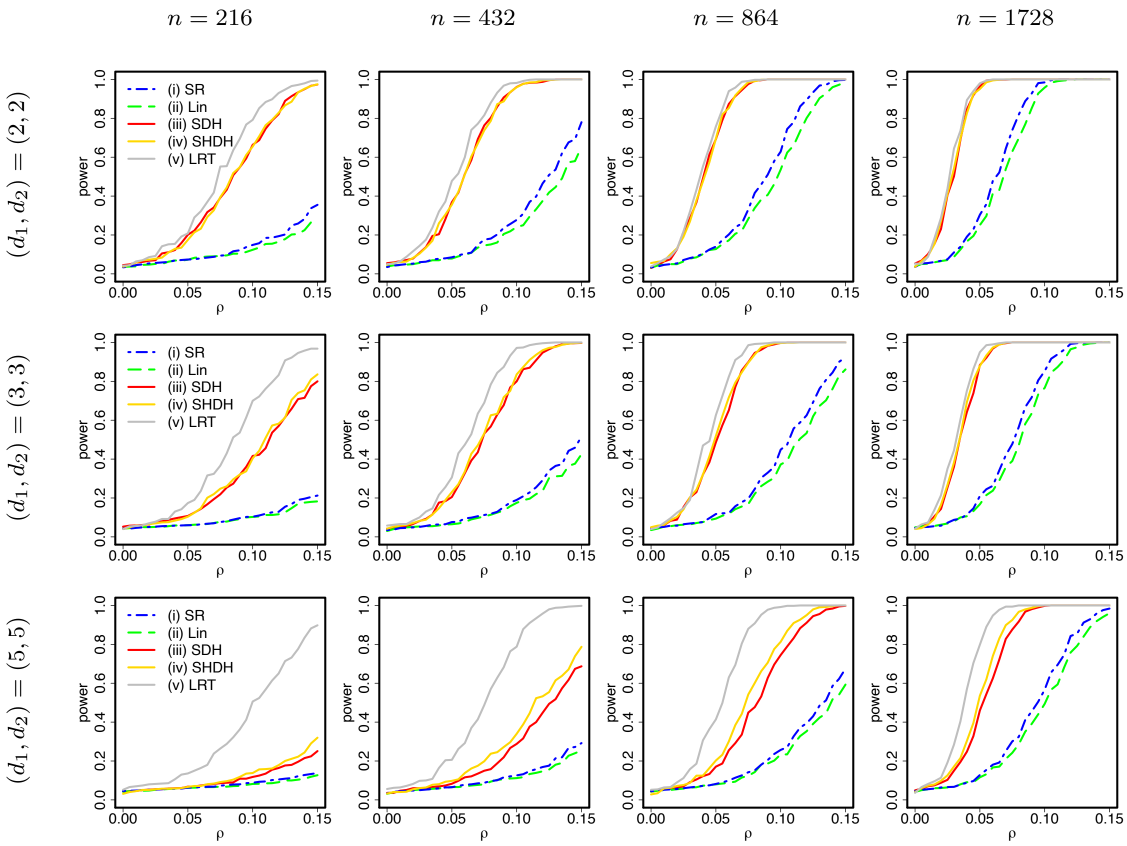
- Proof gives a Hajek representation theorem that shows that $W_n - \widetilde{W}_n = o_p(1)$ for

$$\boxed{\widetilde{W}_n \;:=\; \mathrm{dCov}_n^2\Big(\big[\mathbf{F}_{\boldsymbol{X},\pm}(\boldsymbol{X}^{(i)})\big]_{i=1}^n,\ \big[\mathbf{F}_{\boldsymbol{Y},\pm}(\boldsymbol{Y}^{(i)})\big]_{i=1}^n\Big)}.$$

- Local power analysis (LeCam's third lemma in a degenerate U-statistics setting) shows that test is powerful wrt. quadratic mean differentiable alternatives at a $1/\sqrt{n}$ distance

  (e.g., Gaussian with cross-covariance of order $1/\sqrt{n}$).

One instance of a simulation (Gaussian example)



Conclusion

- Some applications really require consistent tests of independence.
- Consistent rank correlations can be computed efficiently.
- Center-outward ranks and signs provide a distribution-free generalization to higher dimensions.
- Some reading:

Chatterjee (2022).
A survey of some recent developments in measures of association.

Shi, Drton, Han (2022).
On the power of Chatterjee's rank correlation. Biometrika
Distribution-free consistent independence tests via center-outward ranks and signs. JASA

Shi, Hallin, Drton, Han (2022).
On universally consistent and fully distribution-free rank tests of vector independence. Ann. Statist.