An adaptive test based on Kendall's τ

for independence in high dimensions

Sam Olson

Synopsis of Paper

- Testing the mutual independence for high-dimensional data
- Known that L_2-type statistics have lower power under sparse cases
- Known that L_∞-type statistics have lower power under dense cases
- **Goal:** Develop an adaptive test based on Kendall's τ to work well in both situations
- Determine necessary assumptions and the asymptotic null distribution of the proposed statistic
- Assess how well the test does compared to other testing methods
- Results indicate the adaptive test performs well in either dense or sparse cases

What is Kendall's τ?

$$\tau = \frac{(\text{Count of concordant pairs}) - (\text{Count of discordant pairs})}{(\text{Number of pairs})}$$

- Any pair of observations (x_i, y_i) and (x_j, y_j), where $i < j$, are said to be *concordant* if the sort order of (x_i, x_j) and (y_i, y_j) agrees.
- That is, if either both $x_i > x_j$ and $y_i > y_j$ holds or both $x_i < x_j$ and $y_i < y_j$; otherwise they are said to be *discordant*.
- There are other types of Kendall's τ, e.g., τ_a, τ_b, and τ_c.

History of Kendall's τ and related concepts

- **1897** — (Fechner 1897) introduces the *method of signs* for succession-dependence.
- **1938** — (Kendall 1938) develops the τ rank correlation coefficient.
- **1958** — (Kruskal 1958) generalizes Kendall's ideas into a general nonparametric testing framework.
- **1958–1990s** — Applied to time series settings for tests of serial dependence. (El-Shaarawi and Niculescu 1992; Hamed 2011).
- **2020-2024** — Kendall's τ was used for "match rate analysis" to determine whether two separate processes replicate "close enough" healthcare metrics (my personal experience).
- **2024** — (Shi et al. 2024), the focus of this presentation, develop adaptive high-dimensional independence tests using Kendall's τ.
- **2025** — (Han, Ma, and Xie 2025) extend to a broader class of sum-of-powers tests.

In Detail: Shi et al. (2024): Problem Statement

$$H_0 : X_1, \ldots, X_d \text{ are mutually independent}$$

Testing full independence in **high dimensions** ($d \gg n$), where both *dense* and *sparse* alternatives may occur.

Why Kendall's τ?

- **Rank-based**, **distribution-free**, and robust to **heavy tails**.
- Works even when moments (e.g., variances) are infinite.
- Avoids dependence on the data-generation process (no parametric assumptions!)
- Each $\tau_{k\ell}$ measures pairwise monotonic dependence between X_k, X_ℓ.

Dense vs. Sparse Settings

Setting	Dependence Structure	Suitable Statistic
Dense	Many weak correlations	L_2-type (sum-type)
Sparse	Few strong correlations	L_∞-type (max-type)

Method (Overview) I

1. Compute pairwise Kendall's taus $\tau_{k\ell}$.
2. Construct two base statistics:

- S_τ (L_2-type):

$$S_\tau = \omega_2^{-1/2}\left(\sum_{k>\ell}\tau_{k\ell}^2 - \frac{d(d-1)}{2}\omega_1\right) \Rightarrow S_\tau \xrightarrow{d} N(0,1)$$

- M_τ (L_∞-type):

$$M_\tau = \omega_1^{-1}\max_{k<\ell}\tau_{k\ell}^2 - 4\ln d + \ln\ln d \Rightarrow M_\tau \xrightarrow{d} \text{Gumbel}$$

Where ω_1, ω_2 are constants reflecting the variance structure of pairwise Kendall's τ's under independence (H_0).

Method (Overview) II

3. Combine the two via the *minimum p-value approach* (an "adaptive test"):

$$C_\tau = \min\{1 - \Phi(S_\tau),\, 1 - F_{\text{Gumbel}}(M_\tau)\}$$

What is an Adaptive Test?

- A single procedure that **automatically adapts** to the dependence pattern.
- If data are **dense**, S_τ dominates; if **sparse**, M_τ dominates.
- C_τ effectively selects the stronger signal through p-value $= \min(p_{S_\tau}, p_{M_\tau})$.

Theoretical Results I

With all the following base assumptions:

- Finite 8th moment (for asymptotic normality of τ to hold)
- Dependence boundedness conditions (α-mixing type)
- $\frac{d}{n} \to \infty$ allowed

Then, **Under** H_0:

- S_τ and M_τ are **asymptotically independent** — key innovation.
- Therefore,

$$(S_\tau, M_\tau) \xrightarrow{d} (Z_1, Z_2) \quad \text{with } Z_1 \sim N(0,1),\ Z_2 \sim \text{Gumbel.}$$

Theoretical Results II

- Consequently,

$$C_\tau = \min\{1 - \Phi(Z_1),\, 1 - F(Z_2)\} \xrightarrow{d} W = \min(U_1, U_2),$$

where $U_1, U_2 \sim$ Uniform$(0,1)$.

$$\Rightarrow H(t) = 2t - t^2, \quad t \in [0,1].$$

Decision rule:
Reject H_0 if $C_\tau < 1 - \sqrt{1-\alpha}$.

Practical Implementation

- Two variants:
  - TC_τ: uses theoretical (asymptotic) critical values.
  - MC_τ: uses Monte Carlo–simulated critical values (finite-sample accurate).
- Distribution-free; efficient table lookup possible for (n, d).

Key Properties of the Adaptive Test

- Joint Asymptotic independence of S_τ, M_τ allows for an adaptive combination (hence the valid statistical test of independence)
- **Distribution-free:** No parametric or moment assumptions.
- **Robust:** Handles heavy-tailed or non-Gaussian data.
- **Adaptive:** Unified test for both dense and sparse dependence.
- **Asymptotic theory:**
  - $S_\tau \to N(0,1)$
  - $M_\tau \to$ Gumbel
  - S_τ, M_τ independent
  - $C_\tau \to W$ with $H(t) = 2t - t^2$

Results I

Under various settings, we compare the following methods:

- S_r: Pearson–correlation L_2-type test; best for *dense dependence*.
- TS_τ: Kendall's tau L_2-type test using *asymptotic* critical values.
- MS_τ: Same as TS_τ but with *Monte Carlo* critical values.
- M_r: Pearson–correlation L_∞-type test; best for *sparse dependence*.
- TM_τ: Kendall's tau L_∞-type test using *asymptotic* Gumbel limits.
- MM_τ: Same as TM_τ but with *Monte Carlo* critical values.
- TC_τ: *Adaptive* Kendall's tau test combining S_τ and M_τ (asymptotic).
- MC_τ: *Adaptive* Kendall's tau test combining S_τ and M_τ (Monte Carlo).
- PE_r: *Power-enhanced* Pearson test improving S_r under sparse cases.
- $U_{\min}$: *Adaptive U-statistic* test combining multiple orders via minimum p-value.

Results II

There are a number of tables for the various conditions the author's tested the statistical tests. The main focus is on the *Size* and *Power* (mainly the latter) of the statistical tests.

Overall:

- While one particular non-adaptive test may do best under a particular setting (dense/sparse),
- Both implementations of the adaptive test proposed does as good (is roughly as powerful) as the "best" method, but for **both** settings
- Highlighted portions of the tables that follow are the adaptive tests

Results III

n	50				100			
d	50	100	200	400	50	100	200	400
Model 1								
S_r	0.042	0.055	0.048	0.053	0.047	0.044	0.047	0.049
TS_τ	0.044	0.053	0.049	0.049	0.050	0.043	0.053	0.053
MS_τ	0.046	0.057	0.052	0.051	0.056	0.045	0.055	0.055
M_r	0.013	0.007	0.001	0.001	0.021	0.020	0.013	0.009
TM_τ	0.029	0.028	0.018	0.013	0.029	0.027	0.027	0.033
MM_τ	0.044	0.063	0.052	0.051	0.041	0.047	0.044	0.052
TC_τ	0.037	0.037	0.031	0.029	0.040	0.036	0.037	0.044
MC_τ	0.042	0.056	0.047	0.040	0.049	0.048	0.056	0.053
PE_r	0.168	0.135	0.080	0.073	0.068	0.058	0.053	0.051
$U_{\min}$	0.060	0.073	0.065	0.072	0.062	0.060	0.061	0.055
Model 2								
S_r	0.418	0.439	0.432	0.440	0.578	0.568	0.577	0.574
TS_τ	0.040	0.057	0.054	0.044	0.047	0.053	0.049	0.045
MS_τ	0.043	0.057	0.056	0.047	0.051	0.054	0.053	0.045
M_r	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TM_τ	0.024	0.020	0.016	0.014	0.032	0.036	0.037	0.028
MM_τ	0.041	0.056	0.052	0.040	0.054	0.054	0.058	0.051
TC_τ	0.038	0.040	0.038	0.033	0.044	0.043	0.049	0.035
MC_τ	0.045	0.055	0.052	0.045	0.052	0.055	0.072	0.043
PE_r	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$U_{\min}$	NA	NA	NA	NA	NA	NA	NA	NA
Model 3								
S_r	0.056	0.057	0.061	0.059	0.051	0.062	0.051	0.059
TS_τ	0.048	0.045	0.047	0.056	0.047	0.049	0.046	0.048
MS_τ	0.049	0.049	0.050	0.057	0.052	0.052	0.049	0.049
M_r	0.091	0.156	0.225	0.360	0.141	0.264	0.493	0.765
TM_τ	0.033	0.020	0.016	0.017	0.034	0.026	0.030	0.030
MM_τ	0.052	0.045	0.053	0.042	0.055	0.043	0.051	0.052
TC_τ	0.044	0.031	0.033	0.034	0.041	0.041	0.042	0.031
MC_τ	0.047	0.046	0.052	0.048	0.053	0.049	0.063	0.041
PE_r	0.387	0.448	0.564	0.731	0.198	0.284	0.427	0.677
$U_{\min}$	NA	0.057	NA	NA	0.046	0.056	0.053	NA

Figure 1: Empirical sizes of tests

Results IV

n	50				100			
d	50	100	200	400	50	100	200	400
Model 4								
S_r	0.434	0.918	0.999	1.000	0.178	0.651	0.993	1.000
TS_τ	0.375	0.876	0.998	1.000	0.158	0.574	0.986	1.000
MS_τ	0.362	0.873	0.998	1.000	0.155	0.577	0.986	1.000
M_r	0.015	0.018	0.008	0.003	0.036	0.026	0.021	0.024
TM_τ	0.036	0.044	0.040	0.041	0.040	0.044	0.046	0.053
MM_τ	0.071	0.099	0.120	0.113	0.063	0.069	0.079	0.094
TC_τ	0.380	0.878	0.999	1.000	0.168	0.582	0.988	1.000
MC_τ	0.393	0.894	0.999	1.000	0.192	0.621	0.989	1.000
PE_r	0.510	0.925	0.999	1.000	0.207	0.654	0.994	1.000
$U_{\min}$	0.999	1.000	1.000	1.000	0.993	1.000	1.000	1.000
Model 5								
S_r	0.891	0.927	0.956	0.971	0.866	0.914	0.947	0.972
TS_τ	0.856	0.952	0.990	1.000	0.752	0.887	0.976	0.995
MS_τ	0.853	0.951	0.990	0.999	0.748	0.888	0.977	0.995
M_r	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TM_τ	0.426	0.514	0.594	0.722	0.308	0.407	0.530	0.684
MM_τ	0.545	0.691	0.823	0.907	0.377	0.492	0.639	0.801
TC_τ	0.888	0.967	0.994	1.000	0.794	0.908	0.987	0.997
MC_τ	0.896	0.973	0.997	1.000	0.819	0.923	0.992	0.998
PE_r	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$U_{\min}$	NA	NA	NA	NA	NA	NA	NA	NA

Figure 2: Empirical powers of tests in dense cases.

Results V

n	50				100			
d	50	100	200	400	50	100	200	400
Model 6								
S_r	0.053	0.060	0.054	0.055	0.082	0.069	0.056	0.054
TS_τ	0.050	0.058	0.051	0.050	0.077	0.062	0.059	0.054
MS_τ	0.048	0.059	0.054	0.048	0.077	0.064	0.061	0.055
M_r	0.201	0.307	0.504	0.757	0.845	0.963	0.999	1.000
TM_τ	0.210	0.329	0.533	0.793	0.786	0.936	0.996	1.000
MM_τ	0.260	0.425	0.645	0.861	0.809	0.944	0.997	1.000
TC_τ	0.182	0.281	0.473	0.746	0.734	0.918	0.993	1.000
MC_τ	0.194	0.323	0.531	0.767	0.754	0.926	0.995	1.000
PE_r	0.492	0.605	0.760	0.926	0.860	0.956	0.997	1.000
$U_{\min}$	0.233	0.289	0.371	0.347	0.763	0.874	0.946	0.976
Model 7								
S_r	0.433	0.435	0.431	0.436	0.577	0.568	0.578	0.575
TS_τ	0.086	0.077	0.057	0.052	0.075	0.057	0.057	0.043
MS_τ	0.081	0.076	0.056	0.049	0.073	0.059	0.062	0.045
M_r	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TM_τ	0.806	0.869	0.924	0.951	0.646	0.727	0.783	0.836
MM_τ	0.834	0.904	0.952	0.967	0.687	0.760	0.820	0.870
TC_τ	0.755	0.833	0.895	0.933	0.592	0.682	0.755	0.798
MC_τ	0.769	0.853	0.918	0.941	0.615	0.704	0.778	0.812
PE_r	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$U_{\min}$	NA	NA	NA	NA	NA	NA	NA	NA

Figure 3: Empirical powers of tests in sparse cases.

Results VI

n	50				100			
d	50	100	200	400	50	100	200	400
$\rho = 0.02$								
TS_τ	0.078	0.185	0.400	0.782	0.168	0.413	0.822	0.993
MS_τ	0.072	0.179	0.399	0.776	0.168	0.420	0.828	0.993
TM_τ	0.033	0.024	0.014	0.018	0.045	0.054	0.045	0.029
MM_τ	0.056	0.062	0.055	0.057	0.073	0.077	0.074	0.057
TC_τ	0.093	0.191	0.404	0.783	0.179	0.430	0.827	0.993
MC_τ	0.101	0.225	0.450	0.815	0.201	0.461	0.856	0.994
$\rho = 0.04$								
TS_τ	0.387	0.764	0.972	0.998	0.809	0.990	1.000	1.000
MS_τ	0.375	0.759	0.972	0.998	0.803	0.990	1.000	1.000
TM_τ	0.043	0.044	0.022	0.029	0.089	0.082	0.087	0.101
MM_τ	0.080	0.108	0.081	0.074	0.126	0.130	0.141	0.160
TC_τ	0.395	0.766	0.974	0.998	0.814	0.991	1.000	1.000
MC_τ	0.415	0.796	0.981	0.998	0.826	0.991	1.000	1.000
$\rho = 0.06$								
TS_τ	0.781	0.981	0.998	1.000	0.993	1.000	1.000	1.000
MS_τ	0.772	0.980	0.998	1.000	0.993	1.000	1.000	1.000
TM_τ	0.061	0.065	0.055	0.048	0.142	0.183	0.172	0.177
MM_τ	0.110	0.128	0.149	0.132	0.211	0.257	0.270	0.279
TC_τ	0.786	0.981	0.998	1.000	0.993	1.000	1.000	1.000
MC_τ	0.799	0.983	0.998	1.000	0.995	1.000	1.000	1.000
$\rho = 0.08$								
TS_τ	0.958	0.998	1.000	1.000	1.000	1.000	1.000	1.000
MS_τ	0.957	0.998	1.000	1.000	1.000	1.000	1.000	1.000
TM_τ	0.125	0.106	0.091	0.074	0.283	0.299	0.356	0.376
MM_τ	0.182	0.201	0.245	0.176	0.379	0.402	0.486	0.527
TC_τ	0.961	0.998	1.000	1.000	1.000	1.000	1.000	1.000
MC_τ	0.964	0.998	1.000	1.000	1.000	1.000	1.000	1.000

Figure 4: Empirical powers under various strengths of dependence in dense cases.

Results VII

n	50				100			
d	50	100	200	400	50	100	200	400
$\rho = 0.6$								
TS_τ	0.056	0.064	0.054	0.042	0.111	0.078	0.051	0.062
MS_τ	0.057	0.062	0.055	0.044	0.108	0.079	0.055	0.059
TM_τ	0.571	0.408	0.271	0.174	0.990	0.973	0.952	0.891
MM_τ	0.636	0.511	0.399	0.274	0.993	0.979	0.957	0.911
TC_τ	0.512	0.363	0.238	0.155	0.984	0.962	0.926	0.866
MC_τ	0.534	0.399	0.287	0.179	0.986	0.965	0.942	0.875
$\rho = 0.7$								
TS_τ	0.085	0.070	0.055	0.045	0.204	0.095	0.055	0.058
MS_τ	0.077	0.070	0.056	0.045	0.203	0.097	0.055	0.062
TM_τ	0.876	0.828	0.698	0.561	1.000	1.000	0.999	0.997
MM_τ	0.902	0.875	0.806	0.651	1.000	1.000	0.999	0.998
TC_τ	0.902	0.875	0.806	0.651	1.000	1.000	0.999	0.998
MC_τ	0.860	0.803	0.690	0.535	1.000	1.000	0.999	0.997
$\rho = 0.8$								
TS_τ	0.129	0.087	0.060	0.045	0.356	0.122	0.062	0.059
MS_τ	0.117	0.080	0.061	0.044	0.354	0.126	0.066	0.061
TM_τ	0.992	0.988	0.973	0.951	1.000	1.000	1.000	1.000
MM_τ	0.995	0.996	0.988	0.969	1.000	1.000	1.000	1.000
TC_τ	0.987	0.983	0.960	0.929	1.000	1.000	1.000	1.000
MC_τ	0.987	0.987	0.974	0.942	1.000	1.000	1.000	1.000
$\rho = 0.9$								
TS_τ	0.197	0.097	0.062	0.050	0.621	0.201	0.082	0.065
MS_τ	0.187	0.096	0.064	0.046	0.616	0.204	0.089	0.065
TM_τ	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MM_τ	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TC_τ	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MC_τ	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Figure 5: Empirical powers under various strengths of dependence in sparse cases.

Applications

- The paper also applies the method to two “real-world” examples:
  - **Welding (4 vars, n=40):** rank-based rejects null hypothesis; Pearson fails to reject.
  - **Biochemical (8 vars, n=12):** adaptive detects group differences.
- I also implemented the method myself (and validated against the author's own method)
  - Applied (mainly for fun) to a Consulting Case
  - Apriori believed the covariate to be independent, and they were (at least according to the adaptive test)!

Conclusion

- Kendall's τ connects classic nonparametric tests to modern high-dimension inference.
- Rank-based adaptive tests are practical and robust; 2025 work generalizes to sum-of-powers (Han, Ma, and Xie 2025).
- We should consider this test when we know we have high-dimensional data, but don't know whether it is dense or sparse.
- Allows us to test for independence beyond the typical "data collection assessment" and "variable relations graphs"

References I

El-Shaarawi, A. H., and Stefan P. Niculescu. 1992. "On Kendall's Tau as a Test of Trend in Time Series Data." *Environmetrics* 3 (4): 385–411.

Fechner, Gustav Theodor. 1897. *Kollektivmasslehre*. Leipzig: Verlag von Wilhelm Engelmann. https://www.google.com/books/edition/Kollektivmasslehre/bgQZAAAAMAAJ?hl=en.

Hamed, K. H. 2011. "The Distribution of Kendall's Tau for Testing the Significance of Cross-Correlation in Persistent Data." *Hydrological Sciences Journal* 56 (5): 841–53. https://doi.org/10.1080/02626667.2011.586948.

Han, Lijuan, Yun Ma, and Junshan Xie. 2025. "An Adaptive Test of the Independence of High-Dimensional Data Based on Kendall Rank Correlation Coefficient." *Journal of Nonparametric Statistics* 37 (3): 632–56. https://doi.org/10.1080/10485252.2024.2435852.

Kendall, M. G. 1938. "A New Measure of Rank Correlation." *Biometrika* 30 (1/2): 81–93. https://www.jstor.org/stable/2332226.

Kruskal, William H. 1958. "Ordinal Measures of Association." *Journal of the American Statistical Association* 53 (284): 814–61. https://doi.org/10.1080/01621459.1958.10501481.

Shi, Xiangyu, Yuanyuan Jiang, Jiang Du, and Zhuqing Miao. 2024. "An Adaptive Test Based on Kendall's Tau for Independence in High Dimensions." *Journal of Nonparametric Statistics* 36 (4): 1064–87. https://doi.org/10.1080/10485252.2023.2296521.