

# Supplementary Material: Proofs and R Codes for Spacing-Based Dependence Measures

## 1 Appendix I: Proofs of Theorems and Lemmas

### 1.1 Proof of Theorem 1 (Asymptotic Distribution of $\hat{\tau}_S$ )

*Proof.* We treat the spacing-based Kendall's statistic  $\hat{\tau}_S$  as a  $U$ -statistic of order 2, defined as

$$\hat{\tau}_S = \frac{2}{(n-1)(n-2)} \sum_{1 \leq i < j \leq n-1} h((S_i^X, S_i^Y), (S_j^X, S_j^Y)),$$

where the kernel function  $h((s_1^X, s_1^Y), (s_2^X, s_2^Y)) = \text{sign}(s_1^X - s_2^X) \cdot \text{sign}(s_1^Y - s_2^Y)$ .

Here,  $S_i^X = X_{(i+1)} - X_{(i)}$  and  $S_i^Y = Y_{(i+1)} - Y_{(i)}$  denote the spacings between successive order statistics of the marginal samples  $X_{(1)}, \dots, X_{(n)}$  and  $Y_{(1)}, \dots, Y_{(n)}$ , respectively.

Assuming that the marginal distributions  $F_X$  and  $F_Y$  are continuous and that  $(X_i, Y_i)$  are i.i.d. under the null hypothesis of independence, the resulting spacings  $S_i^X$  and  $S_i^Y$  are independent across the two marginals. Furthermore, the asymptotic behavior of spacings in i.i.d. samples has been well studied; under mild regularity conditions, the vector of spacings is asymptotically i.i.d. as well.

Given that  $h(\cdot, \cdot)$  is a bounded, symmetric kernel function with finite variance, we invoke Hoeffding's central limit theorem for  $U$ -statistics (see [Hoeffding \(1948\)](#)) to establish the asymptotic normality:

$$\sqrt{n}(\hat{\tau}_S - \tau_S) \xrightarrow{d} \mathcal{N}(0, \sigma_{\tau_S}^2),$$

where  $\tau_S = \mathbb{E}[h((S_1^X, S_1^Y), (S_2^X, S_2^Y))]$  is the population version of the spacing-based Kendall's measure.

The asymptotic variance  $\sigma_{\tau_S}^2$  is determined by the first-order projection of the kernel:

$$\sigma_{\tau_S}^2 = 4 \cdot \text{Var} \left( \mathbb{E} [h((S_1^X, S_1^Y), (S_2^X, S_2^Y)) \mid (S_1^X, S_1^Y)] \right),$$

which captures the leading contribution to the variance in the  $U$ -statistic asymptotics. The factor of 4 arises from the standard normalization of  $U$ -statistics of order 2.

This completes the proof.  $\square$

### 1.2 Proof of Theorem 2

*Proof.* We consider the statistic  $\hat{\tau}_S^*$  defined based on the spacings of the samples  $\{X_i\}_{i=1}^n$  and  $\{Y_i\}_{i=1}^n$ , where  $X$  and  $Y$  are independent continuous random variables with absolutely continuous distributions.

#### Step 1: Centering and Scaling

Under the independence assumption,  $\hat{\tau}_S^*$  is a degenerate  $U$ -statistic of order 4 with kernel  $h_4$ , symmetric in its arguments. More explicitly,  $\hat{\tau}_S^*$  can be written as

$$\hat{\tau}_S^* = \binom{n}{4}^{-1} \sum_{1 \leq i_1 < i_2 < i_3 < i_4 \leq n} h_4((X_{i_1}, Y_{i_1}), \dots, (X_{i_4}, Y_{i_4})),$$

where  $h_4$  is the symmetric kernel function associated with the statistic.

Since  $X$  and  $Y$  are independent, the kernel  $h_4$  is degenerate of order 1 and 2, i.e.,

$$E[h_4(Z_1, Z_2, Z_3, Z_4) \mid Z_1] = 0, \quad \text{and} \quad E[h_4(Z_1, Z_2, Z_3, Z_4) \mid Z_1, Z_2] = 0,$$

where  $Z_i = (X_i, Y_i)$ .

**Step 2: Hoeffding Decomposition**

By the Hoeffding decomposition for U-statistics, the asymptotic distribution is governed by the second order projection

$$h_2(z_1, z_2) := E[h_4(z_1, z_2, Z_3, Z_4)].$$

Under independence, the first order terms vanish due to degeneracy, and the kernel is fully degenerate of order 2.

**Step 3: Spectral Decomposition of the Kernel**

The kernel  $h_2$  induces a Hilbert–Schmidt integral operator  $T$  on  $L^2(P_Z)$  defined by

$$(Tf)(z) = \int h_2(z, z')f(z') dP_Z(z'),$$

where  $P_Z$  is the joint distribution of  $Z = (X, Y)$ .

Since  $h_2$  is symmetric and square integrable, by Mercer’s theorem,  $T$  admits the spectral decomposition

$$h_2(z, z') = \sum_{j=1}^{\infty} \lambda_j \phi_j(z) \phi_j(z'),$$

where  $\{\phi_j\}$  is an orthonormal basis of  $L^2(P_Z)$  and  $\{\lambda_j\}$  are the eigenvalues of  $T$ .

**Step 4: Asymptotic Distribution**

By the theory of degenerate U-statistics (see e.g. Serfling (1980), Chapter 5), the asymptotic distribution of

$$(n-1)\hat{\tau}_S^* = \binom{n}{4} \hat{\tau}_S^* \cdot \frac{(n-1)}{\binom{n}{4}} \approx \binom{n}{2} \hat{\tau}_S^*,$$

after proper scaling, converges in distribution to

$$\sum_{j=1}^{\infty} \lambda_j (Z_j^2 - 1),$$

where  $Z_j \sim \mathcal{N}(0, 1)$  are independent standard normal random variables.

This follows from the fact that the degenerate U-statistic asymptotically behaves like a weighted sum of independent  $\chi_1^2 - 1$  variables, each corresponding to an eigenvalue of the kernel operator.

**Step 5: Regularity Conditions**

The regularity of the spacing distributions ensures the existence of bounded moments and that the kernel  $h_4$  is square integrable with respect to the joint distribution  $P_Z^4$ , which justifies the application of the Hilbert–Schmidt spectral theory and guarantees the convergence.

Hence, we conclude that

$$(n-1)\hat{\tau}_S^* \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j (Z_j^2 - 1).$$

This completes the proof. □

### 1.3 Proof of Lemma 1 (Non-contiguous Limit of $\hat{\tau}_S$ )

*Proof.* Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be i.i.d. observations from a bivariate distribution with continuous marginals and joint distribution function  $H(x, y)$ , which admits copula representation  $C(u, v)$  via Sklar's Theorem. Under the fixed (non-contiguous) alternative, we assume that the dependence structure between  $X$  and  $Y$  is stable and does not converge to independence as  $n \rightarrow \infty$ . In particular,  $C(u, v) \neq uv$ .

Define the spacing-based version of Kendall's tau:

$$\hat{\tau}_S = \frac{2}{(n-1)(n-2)} \sum_{1 \leq i < j \leq n-1} \text{sign}(S_i^X - S_j^X) \cdot \text{sign}(S_i^Y - S_j^Y),$$

where  $S_i^X = X_{(i+1)} - X_{(i)}$ , and similarly for  $S_i^Y$ .

Under regularity conditions (continuity and strict monotonicity of marginals), the spacings  $\{S_i^X\}, \{S_i^Y\}$  remain informative about the joint dependence structure induced by the copula  $C$ , even though the spacings themselves are not independent.

Since  $\hat{\tau}_S$  is a symmetric  $U$ -statistic of order 2, and the kernel

$$h((s_1^X, s_1^Y), (s_2^X, s_2^Y)) = \text{sign}(s_1^X - s_2^X) \cdot \text{sign}(s_1^Y - s_2^Y)$$

has finite expectation under the joint law of spacings derived from  $H(x, y)$ , the strong law of large numbers for  $U$ -statistics (see [Serfling \(1980\)](#)) yields:

$$\hat{\tau}_S \xrightarrow{a.s.} \tau_S^* := \mathbb{E}[\text{sign}(S_1^X - S_2^X) \cdot \text{sign}(S_1^Y - S_2^Y)].$$

Here,  $\tau_S^*$  is a deterministic constant that characterizes the strength of dependence under the fixed alternative. Since  $C(u, v) \neq uv$ , the joint distribution of the spacings is not a product measure, implying  $\tau_S^* \neq 0$  in general.

Thus, under a non-contiguous fixed alternative, the statistic  $\hat{\tau}_S$  converges almost surely to a non-zero limit that reflects the stable dependence in the underlying population. □

### 1.4 Proof of Theorem 3

Under the independence assumption and the existence of finite fourth moments of the spacings, the squared spacing-based distance covariance statistic  $\text{dCov}_S^2$  can be expressed as a degenerate  $U$ -statistic of order 2 with kernel

$$h_2((X, Y), (X', Y')) = |X - X'| \cdot |Y - Y'|.$$

By Hoeffding's decomposition, the kernel is degenerate under independence, and the associated Hilbert–Schmidt operator  $T$  defined by

$$(Tf)(z) = \int h_2(z, z') f(z') dP_Z(z'), \quad z = (x, y),$$

is symmetric and positive semi-definite.

By Mercer's theorem, this operator admits an eigen-decomposition:

$$h_2(z, z') = \sum_{j=1}^{\infty} \lambda_j \phi_j(z) \phi_j(z'),$$

where  $\{\lambda_j\}$  are the eigenvalues and  $\{\phi_j\}$  form an orthonormal basis in  $L^2(P_Z)$ .

Classical asymptotic theory for degenerate U-statistics (e.g., Serfling, 1980) then implies that

$$(n-1)\text{dCov}_S^2 \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j Z_j^2,$$

where  $\{Z_j\}$  are i.i.d. standard normal random variables.

The finite fourth moment condition ensures square integrability of the kernel  $h_2$  and the applicability of the Hilbert–Schmidt spectral theory, completing the proof.

### 1.5 Proof of Theorem 4

Suppose  $F_X$  belongs to the maximum domain of attraction  $D(G_\xi)$  of the extreme value distribution with index  $\xi$ .

By the theory of extreme value domains (cf. De Haan and Ferreira (2006)), the behavior of spacings derived from samples of  $F_X$  is governed by the tail parameter  $\xi$ . In particular, moments of the generalized spacing distribution depend continuously on  $\xi$ .

As  $\xi \rightarrow -\infty$  (corresponding to the Weibull domain with a finite upper endpoint), the spacings become increasingly concentrated, causing the variance of statistics based on these spacings to decrease and converge to zero.

Hence, the limiting variance of each spacing-based statistic is a function of  $\xi$ , which tends to zero as  $\xi \rightarrow -\infty$ .

This completes the proof.

### 1.6 Proof of Theorem 5

Under the contiguous alternative  $f_n(x, y)$ , the joint distribution deviates from independence at the  $1/\sqrt{n}$  rate. The spacing-based Kendall’s tau statistic  $\hat{\tau}_S$  is asymptotically linear and admits an influence function representation.

By applying a functional central limit theorem and standard U-statistic theory under contiguous alternatives (van der Vaart, 1998), we have

$$\sqrt{n}(\hat{\tau}_S - \tau_0) \xrightarrow{d} \mathcal{N}(\mu_\tau, \sigma_\tau^2),$$

where  $\tau_0 = 0$  under independence.

The asymptotic mean shift  $\mu_\tau$  captures the first-order effect of the alternative and is given by

$$\mu_\tau = \mathbb{E}[\text{sign}(S_1^X - S_2^X) \text{sign}(S_1^Y - S_2^Y) h(X_1, Y_1)],$$

where  $h$  encodes the local perturbation of the joint density.

This follows from the linearization of  $\hat{\tau}_S$  and the projection of the score function of the alternative onto the tangent space at independence. □

### 1.7 Proof of Lemma 2

Under a fixed, non-contiguous alternative with copula  $C \neq uv$ , the distribution of  $(X, Y)$  has a fixed dependence structure.

By the law of large numbers and the continuous mapping theorem, the spacing-based Kendall’s tau statistic  $\hat{\tau}_S$ , being a consistent estimator of the population Kendall’s tau defined through spacings, satisfies

$$\hat{\tau}_S \xrightarrow{p} \tau_S^\infty \neq 0,$$

where  $\tau_S^\infty$  is the limiting population-level measure of dependence induced by the copula  $C$ .

This limit is strictly nonzero since  $C \neq uv$  implies dependence.

## 1.8 Proof of Theorem 6 (Asymptotic Normality of $\hat{\tau}_S^*$ under Contiguous Alternatives)

*Proof.* Let  $\hat{\tau}_S^*$  denote the modified spacing-based Kendall's tau statistic, which can be written as a  $V$ -statistic of order 4 with symmetric kernel:

$$h((X_1, Y_1), \dots, (X_4, Y_4)) = \text{sign}(S_1^X - S_2^X) \cdot \text{sign}(S_3^Y - S_4^Y),$$

where  $S_i^X = X_{(i+1)} - X_{(i)}$ , and similarly for  $S_j^Y$ , with suitable indexing and ordering for the spacings.

Under the sequence of contiguous local alternatives, the joint density is modeled as:

$$f_n(x, y) = f_X(x)f_Y(y) \left( 1 + \frac{\delta}{\sqrt{n}} h(x, y) \right),$$

where  $h(x, y)$  is a square-integrable mean-zero function with respect to the product measure  $f_X(x)f_Y(y)$ , and  $\delta \in \mathbb{R}$  governs the strength of dependence.

Due to the structure of local alternatives (see [van der Vaart \(1998\)](#), Chapter 3), the expectation of the kernel under  $f_n$  differs from that under the null by a term of order  $O(1/\sqrt{n})$ . Thus, we can write:

$$\mathbb{E}_{f_n}[h(Z_1, Z_2, Z_3, Z_4)] = \mathbb{E}_{H_0}[h(Z_1, Z_2, Z_3, Z_4)] + \frac{\mu_{\tau^*}}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right),$$

where  $Z_i = (X_i, Y_i)$  and  $\mu_{\tau^*} = \mathbb{E}[h(Z_1, Z_2, Z_3, Z_4) \cdot \psi(Z_1, Z_2, Z_3, Z_4)]$ , with  $\psi$  derived from the score function corresponding to the parametric submodel  $f_n$ .

This deviation in expectation leads to a shift in the mean of the  $V$ -statistic, while the asymptotic variance remains that under the null due to the Le Cam third lemma.

Hence, by applying the asymptotic theory of  $V$ -statistics under contiguous alternatives (see [Serfling \(1980\)](#), Theorem 5.5.2), we obtain:

$$\sqrt{n}(\hat{\tau}_S^* - \tau_0^*) \xrightarrow{d} \mathcal{N}(\mu_{\tau^*}, \sigma_{\tau^*}^2),$$

where  $\tau_0^*$  is the null value (typically 0 under independence), and  $\sigma_{\tau^*}^2$  is the asymptotic variance under the null, computed via Hoeffding's projection. □

## 1.9 Proof of Lemma 3 (Non-Contiguous Limit of $\hat{\tau}_S^*$ )

*Proof.* Consider the case where the joint distribution of  $(X, Y)$  has a fixed copula  $C \neq uv$ , representing persistent dependence as  $n \rightarrow \infty$ .

The statistic  $\hat{\tau}_S^*$  is a  $V$ -statistic of order 4 with a symmetric kernel  $h_4$  that measures concordance among quadruples of observations based on marginal spacings.

Under the fixed alternative  $C$ , the kernel  $h_4$  has a non-degenerate expectation:

$$\tau_S^{*\infty} := \mathbb{E}_C[h_4((S_1^X, S_1^Y), (S_2^X, S_2^Y), (S_3^X, S_3^Y), (S_4^X, S_4^Y))] \in (0, 1),$$

which quantifies the population-level degree of 4-tuple concordance induced by the dependence structure of  $C$ .

By the strong law of large numbers for  $V$ -statistics (see [Serfling, 1980](#)), we have almost sure convergence:

$$\hat{\tau}_S^* \xrightarrow{a.s.} \tau_S^{*\infty}.$$

Since  $\tau_S^{*\infty}$  is strictly positive under dependence, this establishes the lemma. □

### 1.10 Proof of Theorem 7 (Asymptotic Normality of $\text{dCov}_S^2$ )

*Proof.* Recall the spacing-based distance covariance statistic is defined as a V-statistic of order 2:

$$\text{dCov}_S^2 = \frac{1}{(n-1)^2} \sum_{i,j=1}^{n-1} A_{ij}^* B_{ij}^*,$$

where

$$A_{ij} = |S_i^X - S_j^X|, \quad B_{ij} = |S_i^Y - S_j^Y|,$$

and  $A_{ij}^*, B_{ij}^*$  denote the double-centered versions to remove location effects, ensuring the statistic is centered.

Under the contiguous alternative model:

$$f_n(x, y) = f_X(x) f_Y(y) \left( 1 + \frac{\delta}{\sqrt{n}} h(x, y) \right),$$

the joint distribution perturbs the null hypothesis of independence at rate  $1/\sqrt{n}$ .

By applying standard asymptotic theory for V-statistics with degenerate kernels under the null and non-degenerate first-order projections under contiguous alternatives (see [Serfling \(1980\)](#); see also [Székely and Rizzo \(2009\)](#)), the asymptotic distribution of  $\text{dCov}_S^2$  satisfies

$$\sqrt{n} (\text{dCov}_S^2 - \delta) \xrightarrow{d} \mathcal{N}(\mu_d, \sigma_d^2),$$

where  $\delta = 0$  under independence, and the mean shift

$$\mu_d = \text{Cov}_h(|S_1^X - S_2^X|, |S_1^Y - S_2^Y|) = \mathbb{E}_h[A_{12} B_{12}] - \mathbb{E}[A_{12}] \mathbb{E}[B_{12}]$$

is the covariance induced by the perturbation function  $h$ .

The variance  $\sigma_d^2$  can be expressed in terms of the projections of the kernel and the variance of the underlying spacing variables, ensuring asymptotic normality.

This establishes the asymptotic normality of  $\text{dCov}_S^2$  under contiguous alternatives.  $\square$

### 1.11 Proof of Lemma 4 (Non-Contiguous Limit of $\text{dCov}_S^2$ )

*Proof.* Under a fixed alternative hypothesis, the joint distribution of  $(X, Y)$  is governed by a copula  $C \neq uv$  that induces dependence independent of the sample size  $n$ . Consequently, the marginal spacings  $S_i^X$  and  $S_i^Y$  inherit this dependence structure.

By the law of large numbers for V-statistics with a fixed kernel, we have

$$\text{dCov}_S^2 = \frac{1}{(n-1)^2} \sum_{i,j=1}^{n-1} A_{ij}^* B_{ij}^* \xrightarrow{p} \mathbb{E}[A_{12}^* B_{12}^*].$$

Recall that the double-centering removes marginal means, so

$$\mathbb{E}[A_{12}^* B_{12}^*] = \mathbb{E}[|S_1^X - S_2^X| |S_1^Y - S_2^Y|] - \mathbb{E}[|S_1^X - S_2^X|] \cdot \mathbb{E}[|S_1^Y - S_2^Y|] = \delta^*.$$

Since the copula is not independence,  $\delta^*$  is strictly positive, reflecting persistent second-order dependence across spacings.

Thus,

$$\text{dCov}_S^2 \xrightarrow{p} \delta^* \neq 0.$$

$\square$

## 2 Appendix II: R Codes

```
### --- Libraries ---
install.packages(c("copula", "energy", "mvtnorm"), dependencies = TRUE)
library(copula)
library(energy)
library(mvtnorm)

### --- Frechet Generator Functions ---
rfrechet <- function(n, shape = 1) {
  u <- runif(n)
  return((1 / (-log(u)))^(1 / shape))
}

qfrechet <- function(p, shape = 1) {
  return((1 / (-log(p)))^(1 / shape))
}

### --- Marginal Spacing Calculator ---
get_spacings <- function(x) {
  sort(x)[-1] - sort(x)[-length(x)]
}

### --- Spacing-Based Kendall's Tau ---
tau_spacing <- function(x, y) {
  sx <- get_spacings(x)
  sy <- get_spacings(y)
  n <- length(sx)
  sum_val <- 0
  for (i in 1:(n - 1)) {
    for (j in (i + 1):n) {
      sum_val <- sum_val + sign(sx[i] - sx[j]) * sign(sy[i] - sy[j])
    }
  }
  return(2 * sum_val / (n * (n - 1)))
}

### --- Spacing-Based Bergsma-Dassios Tau* ---
tau_star_spacing <- function(x, y) {
  sx <- get_spacings(x)
  sy <- get_spacings(y)
  n <- length(sx)
  if (n < 4) return(NA) # not defined for fewer than 4 observations
  combs <- combn(n, 4)
  tau_val <- 0
  for (k in 1:ncol(combs)) {
    idx <- combs[, k]
    h <- function(s) {
      a <- s[1]; b <- s[2]; c <- s[3]; d <- s[4]
      sign((a - b) * (c - d)) * sign((a - c) * (b - d))
    }
  }
```

```

    tau_val <- tau_val + h(sx[idx]) * h(sy[idx])
  }
  return(tau_val / choose(n, 4))
}

### --- Spacing-Based Distance Covariance ---
dcov_spacing <- function(x, y) {
  sx <- get_spacings(x)
  sy <- get_spacings(y)
  A <- abs(outer(sx, sx, "-"))
  B <- abs(outer(sy, sy, "-"))

  A_centered <- A - rowMeans(A) - colMeans(A) + mean(A)
  B_centered <- B - rowMeans(B) - colMeans(B) + mean(B)

  return(mean(A_centered * B_centered))
}

### --- Power Simulation Function ---
sim_power <- function(n = 100, R = 500, alpha = 1.5, copula_type = "gumbel",
                      alt = c("noncontig", "contig"), stat = c("tau", "tau_star", "dcov"),
                      delta = 0.01) {
  alt <- match.arg(alt)
  stat <- match.arg(stat)

  powers <- numeric(R)

  if (copula_type == "gumbel") {
    cop <- gumbelCopula(param = 2, dim = 2)
  }

  for (r in 1:R) {
    if (alt == "noncontig") {
      u <- rCopula(n, cop)
      X <- qfrechet(u[,1], shape = alpha)
      Y <- qfrechet(u[,2], shape = alpha)
    } else {
      X <- rfrechet(n, shape = alpha)
      Y <- rfrechet(n, shape = alpha)
      h_xy <- delta * sin(X * Y) / sqrt(n)
      Y <- Y + h_xy
    }

    powers[r] <- switch(stat,
                        "tau" = tau_spacing(X, Y),
                        "tau_star" = tau_star_spacing(X, Y),
                        "dcov" = dcov_spacing(X, Y))
  }

  # Null distribution
  null_vals <- replicate(R, {

```



```

X0 <- rfrechet(n, shape = alpha)
Y0 <- rfrechet(n, shape = alpha)
switch(stat,
  "tau" = tau_spacing(X0, Y0),
  "tau_star" = tau_star_spacing(X0, Y0),
  "dcov" = dcov_spacing(X0, Y0))
})

crit_val <- quantile(null_vals, 0.95)
power_est <- mean(powers > crit_val)
return(power_est)
}

### --- Main Power Analysis Over Alpha ---
alphas <- c(0.5, 0.8, 1.0, 1.2, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0)
results <- data.frame(
  alpha = alphas,
  tau_power_noncontig = NA,
  tau_star_power_noncontig = NA,
  dcov_power_noncontig = NA,
  tau_power_contig = NA,
  tau_star_power_contig = NA,
  dcov_power_contig = NA
)

for (i in seq_along(alphas)) {
  a <- alphas[i]
  cat("Simulating for alpha =", a, "\n")

  results$tau_power_noncontig[i] <- sim_power(alpha = a, alt = "noncontig", stat = "tau")
  results$tau_star_power_noncontig[i] <- sim_power(alpha = a, alt = "noncontig", stat = "tau_star")
  results$dcov_power_noncontig[i] <- sim_power(alpha = a, alt = "noncontig", stat = "dcov")

  results$tau_power_contig[i] <- sim_power(alpha = a, alt = "contig", stat = "tau")
  results$tau_star_power_contig[i] <- sim_power(alpha = a, alt = "contig", stat = "tau_star")
  results$dcov_power_contig[i] <- sim_power(alpha = a, alt = "contig", stat = "dcov")
}

print(results)
library(ggplot2)
library(reshape2)

df_long <- melt(results, id.vars = "alpha",
  variable.name = "Measure",
  value.name = "Power")

df_long$Type <- ifelse(grepl("noncontig", df_long$Measure), "Non-Contiguous", "Contiguous")
df_long$Statistic <- gsub("_power_(noncontig|contig)", "", df_long$Measure)

ggplot(df_long, aes(x = alpha, y = Power, color = Statistic, linetype = Type)) +
  geom_line(size = 1) +

```

```

labs(title = "Empirical Power vs Tail Parameter (alpha)",
      x = expression(alpha),
      y = "Empirical Power") +
theme_minimal(base_size = 14) +
scale_color_brewer(palette = "Dark2") +
ylim(0, 1)

# Sample R code snippet for simulation and power calculation

set.seed(123)

# Parameters
n <- 200
alpha_values <- c(1.1, 1.5, 2, 3)
power_results <-
data.frame(alpha = alpha_values, tau_S = NA, tau_S_star = NA, dCor_S = NA)

# Function to simulate Fréchet marginals
rfrechet <- function(n, alpha) {
  u <- runif(n)
  (1 / u)^(1/alpha)
}

for (i in seq_along(alpha_values)) {
  alpha <- alpha_values[i]

  # Simulate X and Y from Fréchet with dependence (example with linear correlation)
  X <- rfrechet(n, alpha)
  Y <- 0.5 * X + sqrt(1 - 0.5^2) * rfrechet(n, alpha)

  # Compute spacing-based statistics (placeholders)
  tau_S_val <- spacing_kendall_tau(X, Y)      # user-defined function
  tau_S_star_val <- spacing_tau_star(X, Y)    # user-defined function
  dCor_S_val <- spacing_distance_correlation(X, Y) # user-defined function

  # Store results
  power_results$tau_S[i] <- tau_S_val
  power_results$tau_S_star[i] <- tau_S_star_val
  power_results$dCor_S[i] <- dCor_S_val
}

print(power_results)

# Set seed for reproducibility
set.seed(123)

#####
# Simulation Setup
#####

n <- 200

```

```

alpha_values <- c(1.1, 1.5, 2, 3)
results <- data.frame(
  alpha = alpha_values,
  power_tau_S = NA,
  power_tau_S_star = NA,
  power_dCor_S = NA
)

# Simulate Fréchet marginals
rfrechet <- function(n, alpha) {
  u <- runif(n)
  (1 / u)^(1/alpha)
}

# Placeholder for spacing-based statistics functions (user to define)
spacing_kendall_tau <- function(x, y) {
  # Compute spacing-based Kendall's tau estimate
  # ...
  return(runif(1)) # dummy return value
}

spacing_tau_star <- function(x, y) {
  # Compute spacing-based Bergsma-Dassios tau* estimate
  # ...
  return(runif(1)) # dummy return value
}

spacing_distance_correlation <- function(x, y) {
  # Compute spacing-based distance correlation
  # ...
  return(runif(1)) # dummy return value
}

#####
# Power Computation Loop
#####

for (i in seq_along(alpha_values)) {
  alpha <- alpha_values[i]

  # Generate dependent sample (example with linear dependence)
  X <- rfrechet(n, alpha)
  Y <- 0.5 * X + sqrt(1 - 0.5^2) * rfrechet(n, alpha)

  # Calculate spacing-based statistics
  tau_S_val <- spacing_kendall_tau(X, Y)
  tau_S_star_val <- spacing_tau_star(X, Y)
  dCor_S_val <- spacing_distance_correlation(X, Y)

  # Store results
  results$power_tau_S[i] <- tau_S_val

```

```

    results$power_tau_S_star[i] <- tau_S_star_val
    results$power_dCor_S[i] <- dCor_S_val
}

print(results)

#####
# Real Data Example: Airfoil Self-Noise
#####

# Load data (assuming CSV file in working directory)
airfoil_data <- read.csv("airfoil_self_noise.csv")

# Select variables: Sound Pressure Level (target), Frequency, Angle of Attack, etc.
target <- airfoil_data$Sound.pressure.level
freq <- airfoil_data$Frequency
angle <- airfoil_data$Angle.of.attack
chord <- airfoil_data$Chord.length
velocity <- airfoil_data$Free.stream.velocity
displacement <- airfoil_data$Suction.side.displacement.thickness

# Compute classical and spacing-based dependence measures (placeholder)
compute_all_dependence <- function(x, y) {
  list(
    kendall_tau = cor(x, y, method = "kendall"),
    tau_S = spacing_kendall_tau(x, y),
    bergsma_dassios_tau_star = bergsma_dassios_tau_star(x, y), # placeholder
    tau_S_star = spacing_tau_star(x, y),
    dCor = energy::dcor(x, y),
    dCor_S = spacing_distance_correlation(x, y)
  )
}

dependence_results <- data.frame(
  Predictor = c("Frequency", "Angle of attack", "Chord length",
    "Free-stream velocity", "Displacement thickness"),
  Kendall_tau = NA,
  Tau_S = NA,
  Tau_star = NA,
  Tau_S_star = NA,
  dCor = NA,
  dCor_S = NA
)

vars <- list(freq, angle, chord, velocity, displacement)

for (i in seq_along(vars)) {
  deps <- compute_all_dependence(target, vars[[i]])
  dependence_results[i, 2:7] <- unlist(deps)
}

```

```
print(dependence_results)
```

## References

- Asuncion, A., & Newman, D. J. (2007). UCI Machine Learning Repository. *University of California, Irvine, School of Information and Computer Sciences*. Available at: <https://archive.ics.uci.edu/ml>
- Bergsma, W., & Dassios, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli*, 20(2), 1006–1028. Available at: <https://projecteuclid.org/euclid.bj/1392037549>
- Bickel, P. J., & Lehmann, E. L. (1975). Descriptive statistics for nonparametric models. III. Dispersion. *Annals of Statistics*, 3(6), 1133–1169. Available at: <https://projecteuclid.org/euclid.aos/1176343289>
- De Haan, L., & Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer.
- Genest, C., & Favre, A. C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4), 347–368. Available at: [https://ascelibrary.org/doi/10.1061/\(ASCE\)1084-0699\(2007\)12:4\(347\)](https://ascelibrary.org/doi/10.1061/(ASCE)1084-0699(2007)12:4(347))
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19(3), 293–325. Available at: <https://projecteuclid.org/euclid.aoms/1177730196>
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2), 81–93.
- Mikosch, T. (2006). Copulas: Tales and facts. *Extremes*, 9(1), 3–20. Available at: <https://link.springer.com/article/10.1007/s10687-006-0003-5>
- Nagaraja, H. N., Bharath, L. S., & Rao, J. N. K. (2017). Spacings around an order statistic. *Journal of Statistical Theory and Practice*, 11, 372–383. Available at: <https://www.tandfonline.com/doi/full/10.1080/15598608.2016.1247014>
- Resnick, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer. Available at: <https://link.springer.com/book/10.1007/978-0-387-69036-0>
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. ISBN 978-0471090672
- Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics*, 35(6), 2769–2794. Available at: <https://projecteuclid.org/euclid.aos/1194461726>
- Székely, G. J., & Rizzo, M. L. (2009). Brownian distance covariance. *Annals of Applied Statistics*, 3(4), 1236–1265. Available at: <https://projecteuclid.org/euclid.aos/1267453943>
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge. ISBN 978-0521784504