REVSTAT – Statistical Journal
Supplementary file

Received: January 2023 Revised: June 2024 Accepted: June 2024

Supplementary —Energy Distance and Kernel Mean Embeddings for Two-Sample Survival Testing with Applications in Immunotherapy Clinical Trials

Authors: Marcos Matabuena ✉
– CiTIUS, Centro Singular de Investigación en Tecnoloxías Intelixentes
Universidad de Santiago de Compostela,
(marcos.matabuena@usc.es)

Oscar Hernan Madrid Padilla
– Department of Statistics, University of California Los Angeles (oscar.madrid@stat.ucla.edu)

Appendix A Theoretical results

0.1. Asymptotic distribution

The asymptotic distribution of statistics under the null hypothesis will be established only for the case of maximum mean discrepancy (MMD). However, given the equivalence between the tests based on the kernel mean embeddings and the energy distance Sejdinovic et al. (2013) this is not restrictive.

To begin with, we consider the distribution functions P_0, P_1, Q_0 and Q_1 whose meaning was established in Section 1.1 along with the random samples $\{(X_{j,i}, \delta_{j,i})\}_{j=0,1;i=1,\dots,n_j}$.

Next, let us consider the embeddings $\mu_{P_0}(\cdot) = \int_0^{\tau_0} K(\cdot, x)\, dP_0(x) \in H_K$ and $\mu_{P_1}(\cdot) = \int_0^{\tau_1} K(\cdot, x) dP_1(x) \in H_K$, where H_K is the RKHS induced by the kernel K and τ_0, τ_1 are the maximum possible lifetimes defined at the beginning of Section 3.

Under the null hypothesis $P_0 = P_1$ and $\tau_0 = \tau_1$. Then, $P_0' = P_1'$ (see Section 3), and $\mu_{P_0'}(\cdot) = \mu_{P_1'}(\cdot) = \frac{1}{P_0(\tau_0)} \int_0^{\tau_0} K(\cdot, x)\, dP_0(x)$, where $\mu_{P_0'}$ denotes the kernel mean embedding Muandet et al. (2017) of the distribution P_0'.

Given arbitrary elements of random sample for each population $X_{0,i}, X_{0,j}$ $(i = 1, \cdots, n_0, j = 1, \cdots, n_0)$,

✉ Corresponding author

$X_{1,i'}, X_{1,j'}$ $(i' = 1, \cdots, n_1, j' = 1, \cdots, n_1)$, as $\mu_{P_0'} = \mu_{P_1'}$, we can replace $K(X_{0,i}, X_{0,j})$, $K(X_{1,i'}, X_{1,j'})$ and $K(X_{0,i}, X_{1,i'})$ with $K^*(X_{0,i}, X_{0,j})$, $K^*(X_{1,i'}, X_{1,j'})$ and $K^*(X_{0,i}, X_{1,i'})$, where $K^* : [0,\tau_0] \times [0,\tau_0] \to \mathbb{R}$ is defined as follows:

$$\begin{aligned} K^*(X_{0,i}, X_{0,j}) &:= \langle K(X_{0,i}, \cdot) - \mu_{P_0'}, K(X_{0,j}, \cdot) - \mu_{P_0'} \rangle = \\ & K(X_{0,i}, X_{0,j}) - \frac{1}{P(\tau_0)} \int_0^{\tau_0} K(X_{0,i}, x) dP_0(x) - \frac{1}{P(\tau_0)} \int_0^{\tau_0} K(X_{0,j}, x) dP_0(x) + \\ & \frac{1}{P_0(\tau_0)} \frac{1}{P_0(\tau_0)} \int_0^{\tau_0} K(x, x') dP_0(x) dP_0(x') = \\ & K(X_{0,i}, X_{0,j}) - E_{X \sim P_0'}(K(X_{0,i}, X)) - E_{X \sim P_0'}(K(X_{0,j}, X)) + E_{X \sim P_0', X' \sim P_0'}(K(X, X')). \end{aligned}$$

The previous translation does not change the value between $K(\cdot,\cdot)$ and $K^*(\cdot,\cdot)$. This gives the equivalent form of the empirical MMD $\tilde{\gamma}_K^2(P_0, P_1)$ (see equation (10)):

$$\begin{aligned} &(0.1) \\ &\tilde{\gamma}_K^2(P_0, P_1) = \frac{\sum_{i=1}^{n_0} \sum_{j \neq i}^{n_0} W_{i:n_0}^0 W_{j:n_0}^0 K^*(X_{0,(i:n_0)}, X_{0,(j:n_0)})}{\sum_{j=1}^{n_0} \sum_{j \neq i}^{n_0} W_{i:n_0}^0 W_{j:n_0}^0} + \frac{\sum_{i=1}^{n_1} \sum_{j \neq i}^{n_1} W_{i:n_1}^1 W_{j:n_1}^1 K^*(X_{1,(i:n_1)}, X_{1,(j:n_1)})}{\sum_{i=1}^{n_1} \sum_{j \neq i}^{n_1} W_{i:n_1}^1 W_{j:n_1}^1} \\ &\qquad - 2 \frac{\sum_{i=1}^{n_0} \sum_{j=1}^{n_1} W_{i:n_0}^0 W_{i:n_1}^1 K^*(X_{0,(i:n_0)}, X_{1,(j:n_1)})}{\sum_{i=1}^{n_0} \sum_{j=1}^{n_1} W_{i:n_0}^0 W_{i:n_1}^1}. \end{aligned}$$

Now, note that $K^*(\cdot,\cdot)$ is a degenerate kernel:

$$E_{X \sim P_0'} K^*(X, y)) = \left[E_X(K(X, y)) - E_{X,X'} K(X, X') - E_X(K(X, y)) + E_{X,X'} K(X, X') \right] = 0 \quad \forall y \in [0, \tau_0].$$

Consequently, in following terms:

$$\frac{\sum_{i=1}^{n_0} \sum_{j \neq i}^{n_0} W_{i:n_0}^0 W_{j:n_0}^0 K^*(X_{0,(i:n_0)}, X_{0,(j:n_0)})}{\sum_{j=1}^{n_0} \sum_{j \neq i}^{n_0} W_{i:n_0}^0 W_{j:n_0}^0} \text{ and } \frac{\sum_{i=1}^{n_1} \sum_{j \neq i}^{n_1} W_{i:n_1}^1 W_{j:n_1}^1 K^*(X_{1,(i:n_1)}, X_{1,(j:n_1)})}{\sum_{i=1}^{n_1} \sum_{j \neq i}^{n_1} W_{i:n_1}^1 W_{j:n_1}^1},$$

we can apply the limit theorems for U-statistics under right-censored data Bose and Sen (2002); Fernández and Rivera (2018). In particular, we will use the results Fernández and Rivera (2018) under the weakest conditions to use the theorems. Under the conditions assumed in Section 1.1 along with the Euclidean distance and kernel of Table 1, we can apply the theoretical results directly.

By Corollary 2.9 Fernández and Rivera (2018), under the null hypothesis and $\tau_0 = \tau_1$, we have:

$$\frac{\sum_{i=1}^{n_0} \sum_{j \neq i}^{n_0} W_{i:n_0}^0 W_{j:n_0}^0 K^*(X_{0,(i:n_0)}, X_{0,(j:n_0)})}{\sum_{j=1}^{n_0} \sum_{j \neq i}^{n_0} W_{i:n_0}^0 W_{j:n_0}^0} \xrightarrow{D} c_1 + \psi$$

and

$$\frac{\sum_{i=1}^{n_1} \sum_{j \neq i}^{n_1} W_{i:n_1}^1 W_{j:n_1}^1 K^*(X_{1,(i:n_1)}, X_{1,(j:n_1)})}{\sum_{i=1}^{n_1} \sum_{j \neq i}^{n_1} W_{i:n_1}^1 W_{j:n_1}^1} \xrightarrow{D} c_2 + \psi$$

where $\psi = \sum_{i=1}^{\infty} \lambda_i (\epsilon_i^2 - 1)$, with ϵ_i $i.i.d$ standard normal random variables and c_1, c_2 are two constants specified in Fernández and Rivera (2018) that are not relevant for our purposes.

The structure of the previous limits coincides with the case without censoring in the degenerate case. More concretely, the limit is $c + \psi$ Korolyuk and Borovskich (1994) where c is a constant.

For the term:

$$2\frac{\sum_{i=1}^{n_0}\sum_{j=1}^{n_1} W_{i:n_0}^0 W_{i:n_1}^1 K^*(X_{0,(i:n_0)}, X_{1,(j:n_1)})}{\sum_{i=1}^{n_0}\sum_{j=1}^{n_1} W_{i:n_0}^0 W_{i:n_1}^1}$$

we have a U-statistic of two samples under right-censored data in the degenerate case. In this case, the limiting distribution is:

$$\sqrt{n_0 n_1}\frac{\sum_{i=1}^{n_0}\sum_{j=1}^{n_1} W_{i:n_0}^0 W_{i:n_1}^1 K^*(X_{0,(i:n_0)}, X_{1,(j:n_1)})}{\sum_{i=1}^{n_0}\sum_{j=1}^{n_1} W_{i:n_0}^0 W_{i:n_1}^1} \xrightarrow{D} \eta_\infty$$

$$\eta_\infty = \sum_{j=1}^{\infty} \lambda_j \tau_j \epsilon_j,$$

where $\{\tau_j\}_{j=1}^{\infty}$ and $\{\epsilon_j\}_{j=1}^{\infty}$ are two independent sequences of standard normal random variables.

0.2. Consistency against all alternatives

Theorem 0.1. *Let S, A be arbitrary spaces defined on $\mathbb{R}^+$, S contained in A, and let $\gamma(x, y)$ be a continuous, symmetric, real function on $A \times A$. Suppose X, X', Y, Y' are independent random variables, X, X' identically distributed, and Y, Y' are identically distributed. We suppose, moreover, that $\gamma(X, X')$, $\gamma(Y, Y')$, and $\gamma(X, Y)$ have finite expected values on A. Then*

$$2\frac{\int_S\int_S \gamma(x,y)\,dP(x)\,dQ(y)}{\int_S dP(x)\int_S dQ(y)} - \frac{\int_S\int_S \gamma(x,y)\,dP(x)\,dP(y)}{(\int_S dP(x))^2} - \frac{\int_S\int_S \gamma(x,y)\,dQ(x)\,dQ(y)}{(\int_S dQ(x))^2} \geq 0 \tag{0.2}$$

if and only if γ is negative-definite, where P and Q denote the distributions of X and Y respectively. If γ is strictly negative, then equality holds if and only if X and Y are identically distributed on S.

Proof: By Theorem 1 Székely and Rizzo (2005), it is verified:

$$2\int_A\int_A \gamma(x,y)\,dP(x)\,dQ(y) - \int_A\int_A \gamma(x,y)\,dP(x)\,dP(y) - \int_A\int_A \gamma(x,y)\,dQ(x)\,dQ(y) \geq 0, \tag{0.3}$$

if and only if γ is negative-definite. If γ is strictly negative, then equality holds if and only if X and Y are identically distributed on A.

Now, we define the following random variables on S: X^* and Y^* with distribution functions P' and Q' respectively, as follows:

$$dP'(x) = c_1 dP(x) \quad \text{and} \quad dQ'(x) = c_2 dQ(x),$$

where $c_1 = \frac{1}{\int_S dP(x)}$ and $c_2 = \frac{1}{\int_S dQ(x)}$, and we consider their copies $X^{*\prime}$ and $Y^{*\prime}$. Since $\gamma(X, X')$, $\gamma(Y, Y')$, and $\gamma(X, Y)$ have finite expected values in A, then $\gamma(X^*, X^{*\prime})$, $\gamma(Y^*, Y^{*\prime})$, and $\gamma(X^*, Y^*)$ have finite expected values in S. Moreover, let $\gamma(x, y)$ be a continuous, symmetric, real function in $S \times S$.

This leads to:

$$2c_1c_2 \int_S \int_S \gamma(x,y)\, dP(x)\, dQ(y) - c_1^2 \int_S \int_S \gamma(x,y)\, dP(x)\, dP(y) - c_2^2 \int_S \int_S \gamma(x,y)\, dQ(x)\, dQ(y) \geq 0 \tag{0.4}$$

if and only if γ is negative-definite, and

$$2c_1c_2 \int_S \int_S \gamma(x,y)\, dP(x)\, dQ(y) - c_1^2 \int_S \int_S \gamma(x,y)\, dP(x)\, dP(y) - c_2^2 \int_S \int_S \gamma(x,y)\, dQ(x)\, dQ(y) = 0 \tag{0.5}$$

if X^* and Y^* are identically distributed in S (with γ being strictly negative) or equivalently $P(t) = Q(t)$ for all $t \in S$. □

Theorem 0.2. *Let $X_{j,i} = \min(T_{j,i}, C_{j,i}) \sim$ i.i.d. $P_{c(j)}$ and $\delta_{j,i} = 1\{X_{j,i} = T_{j,i}\}$ for $j = 0, 1; i = 1, \dots, n_j$ with $P_{c(j)}$ for $j = 0, 1$ and under the conditions assumed in Section 1.1 imposed on the variables $T_{j,i} \sim$ i.i.d.P_j and $C_{j,i} \sim$ i.i.d.Q_j for $j = 0, 1; i = 1, \dots, n_j$. Then:*

$$\tilde{\epsilon}_\alpha(P_0, P_1) \overset{n_0,n_1\to\infty}{\to} \epsilon_{c(\alpha)}(P_0, P_1) = 2\frac{\int_0^{\tau_0}\int_0^{\tau_1} ||x-y||^\alpha\, dP_0^*(x)\, dP_1^*(y)}{\int_0^{\tau_0}\int_0^{\tau_1} dP_0^*(x)\, dP_1^*(y)} \tag{0.6}$$

$$-\frac{\int_0^{\tau_0}\int_0^{\tau_0} ||x-y||^\alpha\, dP_0^*(x)\, dP_0^*(y)}{\int_0^{\tau_0}\int_0^{\tau_0} dP_0^*(x)\, dP_0^*(y)} - \frac{\int_0^{\tau_1}\int_0^{\tau_1} ||x-y||^\alpha\, dP_1^*(x)\, dP_1^*(y)}{\int_0^{\tau_1}\int_0^{\tau_1} dP_1^*(x)\, dP_1^*(y)}, \tag{0.7}$$

$$\tilde{\gamma}_K(P_0, P_1) \overset{n_0,n_1\to\infty}{\to} \gamma_{c(K)}(P_0, P_1) = 2\frac{\int_0^{\tau_0}\int_0^{\tau_1} K(x,y)\, dP_0^*(x)\, dP_1^*(y)}{\int_0^{\tau_0}\int_0^{\tau_1} dP_0^*(x)\, dP_1^*(y)} \tag{0.8}$$

$$-\frac{\int_0^{\tau_0}\int_0^{\tau_0} K(x,y)\, dP_0^*(x)\, dP_0^*(y)}{\int_0^{\tau_0}\int_0^{\tau_0} dP_0^*(x)\, dP_0^*(y)} - \frac{\int_0^{\tau_1}\int_0^{\tau_1} K(x,y)\, dP_1^*(x)\, dP_1^*(y)}{\int_0^{\tau_1}\int_0^{\tau_1} dP_1^*(x)\, dP_1^*(y)}, \tag{0.9}$$

where

$$P_0^*(x) = \begin{cases} P_0(x) & \text{if } x < \tau_0 \\ P_0(\tau_0^-) + 1\{\tau_0 \in A^1\}P_0(\tau_0) & \text{if } x \geq \tau_0 \end{cases}$$

and

$$P_1^*(x) = \begin{cases} P_1(x) & \text{if } x < \tau_1 \\ P_1(\tau_1^-) + 1\{\tau_1 \in A^1\}P_1(\tau_1) & \text{if } x \geq \tau_1 \end{cases}$$

Here, $\tau_0 = \inf\{x : 1 - P_{c(0)}(x) = 0\}$, $\tau_1 = \inf\{x : 1 - P_{c(1)}(x) = 0\}$, $A^0 = \{x \in \mathbb{R} | P_{c(0)}\{x\} > 0\}$, and $A^1 = \{x \in \mathbb{R} | P_{c(1)}\{x\} > 0\}$.

Proof: The proof consists of repeatedly applying the strong laws of large numbers for U-statistics Kaplan-Meier with two samples Stute and Wang (1993), along with the convergence results for the U-statistic of degree two for randomly censored data Bose and Sen (1999).

According to Stute and Wang (1993):

$$\sum_{i=1}^{n_1}\sum_{j=1}^{n_1} W^0_{i:n_0} W^1_{j:n_1} h(X_{0,(i:n_0)}, X_{1,(j:n_1)}) \overset{n_0,n_1\to\infty}{\to} \int_0^{\tau_0}\int_0^{\tau_1} h(x,y)\, dP_0^*(x)\, dP_1^*(y)$$

where h is a given kernel of degree two such that

$$\int h(x,y)\, dP_0(x)\, dP_1(y) < \infty.$$

By hypothesis, $P_{c(j)}$ for $j = 0, 1$ is a continuous distribution function. Then, A^0 and A^1 are empty sets, and therefore $P_0^*(x) = P_0(x)$ for all $x \in [0, \tau_0]$ and $P_1^*(x) = P_1(x)$ for all $x \in [0, \tau_1]$.

Applying the previous result with $h(x, y) = 1$ to the following expressions, along with the properties of convergence in probability, we have:

$$\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_1} W^0_{i:n_0} W^1_{j:n_1} h(X_{0,(i:n_0)}, X_{1,(j:n_1)})}{\sum_{i=1}^{n_1}\sum_{j=1}^{n_1} W^0_{i:n_0} W^1_{j:n_1}} \overset{n_0,n_1\to\infty}{\to} \frac{\int_0^{\tau_0}\int_0^{\tau_1} h(x,y)\, dP_0^*(x)\, dP_1^*(y)}{\int_0^{\tau_0}\int_0^{\tau_1} dP_0^*(x)\, dP_1^*(y)}.$$

Using Theorem 1 of Bose and Sen (1999), it is also verified that

$$\frac{\sum_{i=1}^{n_0}\sum_{j\neq i}^{n_0} W^0_{i:n_0} W^0_{j:n_0} h(X_{0,(i:n_0)}, X_{0,(j:n_0)})}{\sum_{i=1}^{n_0}\sum_{j\neq i}^{n_0} W^0_{i:n_0} W^0_{j:n_0}} \overset{n_0\to\infty}{\to} \frac{\int_0^{\tau_0}\int_0^{\tau_0} h(x,y)\, dP_0^*(x)\, dP_0^*(y)}{\int_0^{\tau_0}\int_0^{\tau_0} dP_0^*(x)\, dP_0^*(y)},$$

and

$$\frac{\sum_{i=1}^{n_1}\sum_{j\neq i}^{n_1} W^1_{i:n_1} W^1_{j:n_1} h(X_{1,(i:n_1)}, X_{1,(j:n_1)})}{\sum_{i=1}^{n_1}\sum_{j\neq i}^{n_1} W^1_{i:n_1} W^1_{j:n_1}} \overset{n_1\to\infty}{\to} \frac{\int_0^{\tau_1}\int_0^{\tau_1} h(x,y)\, dP_1^*(x)\, dP_1^*(y)}{\int_0^{\tau_1}\int_0^{\tau_1} dP_1^*(x)\, dP_1^*(y)}.$$

Finally, taking $h(x, y)$ as $||x - y||^\alpha$ or $h(x, y) = K(x, y)$ and applying the properties of convergence in probability of the sum of two random variables, the desired result is obtained. □

Theorem 0.3. *Let $X_{j,i} = \min(T_{j,i}, C_{j,i}) \sim i.i.d.P_{c(j)}$ and $\delta_{j,i} = 1\{X_{j,i} = T_{j,i}\}$ for $j = 0, 1; i = 1, \ldots, n_j$ with $P_{c(j)}$ for $j = 0, 1$. Suppose also that the conditions stated in Section 1.1 hold for the random variables $T_{j,i} \sim i.i.d.P_j$ and $C_{j,i} \sim i.i.d.Q_j$ for $j = 0, 1; i = 1, \ldots, n_j$. Further assume that $\tau_0 = \tau_1$ or the support of the distribution functions P_0 and P_1 is contained in the intervals $[0, \tau_0]$ and $[0, \tau_1]$, respectively. Then, for testing the null $H_0 : P_0(t) = P_1(t)\ \forall t \in [0, \tau_1]$, the statistics $T_{\tilde\epsilon_\alpha}$ and $T_{\tilde\gamma^2_K}$ determine tests that are consistent against all fixed alternatives with continuous random variables.*

Proof: We assume without any restrictions that P_0 and P_1 have the same support (otherwise it is enough to extend the probability measure with less support to the higher one). If $\tau_0 = \tau_1$, we can apply Theorems 0.1 and 0.2 and then we have it guaranteed that:

$$\lim_{n_0\to\infty, n_1\to\infty} \tilde\epsilon_\alpha(P_0, P_1) = 2\frac{\int_0^{\tau_0}\int_0^{\tau_1} ||x-y||^\alpha\, dP_0^*(x)\, dP_1^*(y)}{\int_0^{\tau_0}\int_0^{\tau_1} dP_0^*(x)\, dP_1^*(y)} - \frac{\int_0^{\tau_0}\int_0^{\tau_0} ||x-y||^\alpha\, dP_0^*(x)\, dP_0^*(y)}{\int_0^{\tau_0}\int_0^{\tau_0} dP_0^*(x)\, dP_0^*(y)} - \frac{\int_0^{\tau_1}\int_0^{\tau_1} ||x-y||^\alpha\, dP_1^*(x)\, dP_1^*(y)}{\int_0^{\tau_1}\int_0^{\tau_1} dP_1^*(x)\, dP_1^*(y)} \geq 0 \tag{0.10}$$

$$\lim_{n_0\to\infty,n_1\to\infty} \tilde{\gamma}_K(P_0,P_1) = \frac{\int_0^{\tau_0}\int_0^{\tau_0} K(x,y)\,dP_0^*(x)\,dP_0^*(y)}{\int_0^{\tau_0}\int_0^{\tau_0} dP_0^*(x)\,dP_0^*(y)} + \frac{\int_0^{\tau_1}\int_0^{\tau_1} K(x,y)\,dP_1^*(x)\,dP_1^*(y)}{\int_0^{\tau_1}\int_0^{\tau_1} dP_1^*(x)\,dP_1^*(y)} - 2\frac{\int_0^{\tau_0}\int_0^{\tau_1} K(x,y)\,dP_0^*(x)\,dP_1^*(y)}{\int_0^{\tau_0}\int_0^{\tau_1} dP_0^*(x)\,dP_1^*(y)} \geq 0 \tag{0.11}$$

Furthermore, (0.10) and (0.11) are equal to zero if and only if $P_0(t) = P_1(t)$ for all $t \in [0,\tau_1]$.

Suppose $\exists t \in [0,\tau_1]$ such that $P_0(t) \neq P_1(t)$. Then we have strict inequality in (0.10) and (0.11), so with probability one:

$$\lim_{n_0\to\infty,n_1\to\infty} P(\tilde{\epsilon}_\alpha(P_0,P_1) = c_{\epsilon_\alpha} > 0) = 1 \quad \text{and} \quad \lim_{n_0\to\infty,n_1\to\infty} P(\tilde{\gamma}_K(P_0,P_1) = c_K > 0) = 1.$$

According to the theory of degenerate U-statistics Korolyuk and Borovskich (1994), under the null hypothesis, there exist constants c_{α_1} and c_{α_2} satisfying:

$$\lim_{n\to\infty} P\left(\frac{n_0n_1}{n_0+n_1}\hat{\epsilon}_\alpha(P_0,P_1) > c_{\alpha_1}\right) = \alpha \quad \text{and} \quad \lim_{n\to\infty} P\left(\frac{n_0n_1}{n_0+n_1}\hat{\gamma}_K(P_0,P_1) > c_{\alpha_2}\right) = \alpha.$$

Under the alternative hypothesis:

$$\lim_{n\to\infty} P\left(\frac{n_0n_1}{n_0+n_1}\hat{\epsilon}_\alpha(P_0,P_1) > c_{\alpha_1}\right) = 1 \quad \text{and} \quad \lim_{n\to\infty} P\left(\frac{n_0n_1}{n_0+n_1}\hat{\gamma}_K(P_0,P_1) > c_{\alpha_2}\right) = 1$$

since $n\hat{\epsilon}_\alpha(P_0,P_1) \to \infty$ and $n\hat{\gamma}_K(P_0,P_1)$ with probability one as $n \to \infty$.

In the case of $\tau_0 \neq \tau_1$, the support of the distribution functions P_0 and P_1 is contained in the intervals $[0,\tau_0]$ and $[0,\tau_1]$. In this situation, the normalization constants are 1, and the previous argument holds true. □

Appendix B: V-statistics as a Distance Between Samples

We will now establish that the statistics defined in (11)–(12) behave like distances between the elements of the sample $\{(X_{j,i},\delta_{j,i})\}_{j=0,1;i=1,\dots,n_j}$ defined in Section 1.1.

Given two arbitrary samples $A := \{(X_{j,i},\delta_{j,i})\}_{j=0;i=1,\dots,n_0}$ and $B := \{(X_{j,i},\delta_{j,i})\}_{j=1;i=1,\dots,n_1}$, a function $d: (\mathbb{R}^+ \times \{0,1\})^{n_0} \times (\mathbb{R}^+ \times \{0,1\})^{n_1} \to \mathbb{R}$ between A and B is a distance if:

- $d(A,B) \geq 0$ and $d(A,B) = 0$ iff $A = B$.
- $d(A,B) = d(B,A)$.

Moreover, given an arbitrary sample C, it is verified that:

- $d(A,B) \leq d(A,C) + d(B,C)$.

The population version of energy distance and maximum mean discrepancy with appropriate distances/kernel (for example, with Euclidean distance) verify those conditions with any pair of probability measures with finite moments of order 2. In parallel, considering the weights

$$W_i^0 = \frac{W_{i:n_0}^0}{\sum_{i=1}^{n_0} W_{i:n_0}^0} \quad \text{and} \quad W_j^1 = \frac{W_{j:n_1}^1}{\sum_{i=1}^{n_1} W_{j:n_1}^1} \quad (i = 1, \dots, n_0) \quad (j = 1, \dots, n_1),$$

we have

$$W_i^0 \geq 0, \quad W_j^1 \geq 0 \quad (i = 1, \dots, n_0) \quad (j = 1, \dots, n_1), \quad \sum_{i=1}^{n_0} W_i^0 = 1 \quad \text{and} \quad \sum_{j=1}^{n_1} W_j^1 = 1.$$

Now, we consider the probability measures P_0^*, P_1^* induced by the probabilities $(W_1^0, \dots, W_{n_0}^0)$, $(W_1^1, \dots, W_{n_1}^1)$ whose values are $(X_{01}, \dots, X_{0n_0})$ and $(X_{11}, \dots, X_{1n_1})$ respectively. It is trivially verified that the energy distance and the maximum mean discrepancy between P_0^* and P_1^* are well defined. By definition,

$$\epsilon_\alpha(P^*, Q^*) = 2E||X - Y||^\alpha - E||X - X^{'}||^\alpha - E||Y - Y^{'}||^\alpha \tag{0.12}$$

where $X, X' \sim^{i.i.d.} P^*$ and $Y, Y' \sim^{i.i.d.} Q^*$. Replacing 0.12 with the population-defined quantities,

(0.13)

$$\epsilon_\alpha(P_0^*, P_1^*) = 2\frac{\sum_{i=1}^{n_0} \sum_{j=1}^{n_1} W_{i:n_0}^0 W_{i:n_1}^1 ||X_{0(i:n_0)} - X_{1(j:n_1)}||^\alpha}{\sum_{i=1}^{n_0} \sum_{j=1}^{n_1} W_{i:n_0}^0 W_{j:n_1}^1} - \frac{\sum_{i=1}^{n_0} \sum_{j\neq i}^{n_0} W_{i:n_0}^0 W_{j:n_0}^0 ||X_{0(i:n_0)} - X_{0(j:n_0)}||^\alpha}{\sum_{i=1}^{n_0} \sum_{j=i}^{n_0} W_{i:n_0}^0 W_{j:n_0}^0} - \frac{\sum_{i=1}^{n_1} \sum_{j\neq i}^{n_1} W_{i:n_1}^1 W_{j:n_1}^1 ||X_{1(i:n_1)} - X_{1(j:n_1)}||^\alpha}{\sum_{i=1}^{n_1} \sum_{j=i}^{n_1} W_{i:n_1}^1 W_{j:n_1}^1}.$$

Therefore, equations 11 and 12 (see paper) always take values greater than or equal to zero. This is given if and only if $(W_1^0, \dots, W_{n_0}^0) = (W_1^1, \dots, W_{n_1}^1)$ and $(X_{01}, \dots, X_{0n_0}) = (X_{11}, \dots, X_{1n_1})$. This also implies $(\delta_{01}, \dots, \delta_{0n_0}) = (\delta_{11}, \dots, \delta_{1n_1})$.

Note that it is well known that the U-statistics do not verify that property in the general case. The same is true in the case of censorship present.

Appendix C: Statistics in the Multivariate Case

Let us now consider the construction of the statistics of energy distance and maximum mean discrepancy in the multivariate case. In this case, there is a lifetime $T \in \mathbb{R}^+$ with possible censorship and a vector of covariates $S \in \mathbb{R}^{p-1}$ without censorship. Possible practical applications of the above include the comparison of the equality of distribution according to the lifetime of individuals and certain clinical variables of patients, independence testing (Shen & Vogelstein, 2019), or change-point detection problems.

Let $H_{j,i} = (T_{j,i}, S_{j,i}) \sim P_j$ $(j = 0, 1; i = 1, \dots, n_j)$ and censoring times $C_{j,i} \sim Q_j$ $(j = 0, 1; i = 1, \dots, n_j)$, with distribution P_j defined as a subset of $\mathbb{R}^+ \times \mathbb{R}^{p-1}$ and the distributions Q_j on $\mathbb{R}^+$ $(j = 0, 1)$.

Here, the index j represents a population, and the index i a particular sample within the corresponding population. Moreover, the random variables $(T_{0,1}, S_{0,1}), \ldots, (T_{0,n_0}, S_{0,n_0}), \ldots, (T_{1,1}, S_{1,1}), \ldots, (T_{1,n_1}, S_{1,n_1}), C_{0,1}, \ldots, C_{n_0,n_0}, C_{1,n_1}, \ldots,$ are assumed to be independent of each other. In practice, only the random variables $(X_{j,i} = \min(T_{j,i}, C_{j,i}), S_{j,i})$ and $\delta_{j,i} = 1\{X_{j,i} = T_{j,i}\}$ $(j = 0, 1; i = 1, \ldots, n_j)$ are observed.

On the basis of the observed data $\{(X_{j,i}, S_{j,i}, \delta_{j,i})\}_{j=0,1;i=1,\ldots,n_j}$ we must approximate the distances $\epsilon_\alpha(P_0, P_1)$, $\gamma_K^2(P_0, P_1)$. In this case, we can use the Kaplan-Meier estimator in the presence of covariates (Stute, 1993; Gerds & Schumacher, 2006).

(0.14)

$$\hat{\epsilon}_\alpha(P_0, P_1) = 2\frac{\sum_{i=1}^{n_0}\sum_{j=1}^{n_1} W_{i:n_0}^0 W_{i:n_1}^1 ||H_{0,(i:n_0)} - H_{1,(j:n_1)}||^\alpha}{\sum_{i=1}^{n_0}\sum_{j=1}^{n_1} W_{i:n_0}^0 W_{j:n_1}^1} - \frac{\sum_{i=1}^{n_0}\sum_{j\neq i}^{n_0} W_{i:n_0}^0 W_{j:n_0}^0 ||H_{0,(i:n_0)} - H_{0,(j:n_0)}||^\alpha}{\sum_{i=1}^{n_0}\sum_{j\neq i}^{n_0} W_{i:n_0}^0 W_{j:n_0}^0} - \frac{\sum_{i=1}^{n_1}\sum_{j\neq i}^{n_1} W_{i:n_1}^1 W_{j:n_1}^1 ||H_{1,(i:n_1)} - H_{1,(j:n_1)}||^\alpha}{\sum_{i=1}^{n_1}\sum_{j\neq i}^{n_1} W_{i:n_1}^1 W_{j:n_1}^1}$$

(U-statistic α-energy distance under right censoring)

$$\hat{\gamma}_K^2(P_0, P_1) = \frac{\sum_{i=1}^{n_0}\sum_{j\neq i}^{n_0} W_{i:n_0}^0 W_{j:n_0}^0 K(H_{0,(i:n_0)}, H_{0,(j:n_0)})}{\sum_{i=1}^{n_0}\sum_{j\neq i}^{n_0} W_{i:n_0}^0 W_{j:n_0}^0} + \frac{\sum_{i=1}^{n_1}\sum_{j\neq i}^{n_1} W_{i:n_1}^1 W_{j:n_1}^1 K(H_{1,(i:n_1)}, H_{1,(j:n_1)})}{\sum_{i=1}^{n_1}\sum_{j\neq i}^{n_1} W_{i:n_1}^1 W_{j:n_1}^1} - 2\frac{\sum_{i=1}^{n_0}\sum_{j=1}^{n_1} W_{i:n_0}^0 W_{i:n_1}^1 K(H_{0,(i:n_0)}, H_{1,(j:n_1)})}{\sum_{i=1}^{n_0}\sum_{j=1}^{n_1} W_{i:n_0}^0 W_{i:n_1}^1}$$

(U-statistic kernel method under right censoring)

Analogously, we can define V-statistics as follows:

$$\hat{\epsilon}_\alpha(P_0, P_1) = 2\frac{\sum_{i=1}^{n_0}\sum_{j=1}^{n_1} W_{i:n_0}^0 W_{i:n_1}^1 ||H_{0,(i:n_0)} - H_{1,(j:n_1)}||^\alpha}{\sum_{i=1}^{n_0}\sum_{j=1}^{n_1} W_{i:n_0}^0 W_{j:n_1}^1} - \frac{\sum_{i=1}^{n_0}\sum_{j=1}^{n_0} W_{i:n_0}^0 W_{j:n_0}^0 ||H_{0,(i:n_0)} - H_{0,(j:n_0)}||^\alpha}{\sum_{i=1}^{n_0}\sum_{j=1}^{n_0} W_{i:n_0}^0 W_{j:n_0}^0} - \frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_1} W_{i:n_1}^1 W_{j:n_1}^1 ||H_{1,(i:n_1)} - H_{1,(j:n_1)}||^\alpha}{\sum_{i=1}^{n_1}\sum_{j=1}^{n_1} W_{i:n_1}^1 W_{j:n_1}^1}$$

(V-statistic α-energy distance under right censoring)

$$\hat{\gamma}_K^2(P_0, P_1) = \frac{\sum_{i=1}^{n_0}\sum_{j=1}^{n_0} W_{i:n_0}^0 W_{j:n_0}^0 K(H_{0,(i:n_0)}, H_{0,(j:n_0)})}{\sum_{j=1}^{n_0}\sum_{j=1}^{n_0} W_{i:n_0}^0 W_{j:n_0}^0} + \frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_1} W_{i:n_1}^1 W_{j:n_1}^1 K(H_{1,(i:n_1)}, H_{1,(j:n_1)})}{\sum_{i=1}^{n_1}\sum_{j=1}^{n_1} W_{i:n_1}^1 W_{j:n_1}^1} - 2\frac{\sum_{i=1}^{n_0}\sum_{j=1}^{n_1} W_{i:n_0}^0 W_{i:n_1}^1 K(H_{0,(i:n_0)}, H_{1,(j:n_1)})}{\sum_{i=1}^{n_0}\sum_{j=1}^{n_1} W_{i:n_0}^0 W_{i:n_1}^1}$$

(V-statistic kernel method under right censoring)

where

(0.15)
$$W_{i:n_0}^0 = \frac{\delta_{0,(i:n_0)}}{n_0 - i + 1} \prod_{j=1}^{i-1} \left[\frac{n_0 - j}{n_0 - j + 1}\right]^{\delta_{0,(j:n_0)}} \quad (i = 1, \ldots, n_0)$$

and

$$W^1_{i:n_1} = \frac{\delta_{1,(i:n_1)}}{n_1 - i + 1} \prod_{j=1}^{i-1} \left[\frac{n_1 - j}{n_1 - j + 1} \right]^{\delta_{1,(j:n_1)}} \quad (i = 1, \ldots, n_1) \tag{0.16}$$

It can be seen that this estimator is asymptotically efficient with the hypothesis of independence assumed between lifetimes and censorship times (Gerds & Schumacher, 2006). However, this situation is unrealistic in practice. Instead, T and C are often imposed to be conditionally independent given S (Fan & Gijbels, 1994).

Given the equivalence between the weights of the Kaplan-Meier estimator and the inverse-probability-of-censoring weighted average (Satten & Datta, 2001), a natural generalization for modeling dependent censorship is to calculate weights as follows:

$$W^0_{i:n_0} = \frac{\delta_{0(i:n_0)}}{n_0 \hat{P}(C_0 > X_{0,(i:n_0)} | S = S_{0,(i:n_0)})} \quad (i = 1, \ldots, n_0) \tag{0.17}$$

and

$$W^1_{i:n_1} = \frac{\delta_{1(i:n_1)}}{n_1 \hat{P}(C_1 > X_{1,(i:n_1)} | S = S_{1,(i:n_i)})} \quad (i = 1, \ldots, n_1) \tag{0.18}$$

The previous conditional probability of the censorship variable of each population can be estimated, for example, using the Cox model. In a one- or two-dimensional space, an alternative option is to use a non-parametric approach with the Beran estimator (the smoothed conditional Kaplan-Meier estimator). From the theoretical point of view, in the case of dependent censorship, the estimators with inverse-probability-of-censoring weighted average have the disadvantage that they are not asymptotically efficient

Appendix E: Null Hypothesis Results

Table 1: Proportion of p-values less than or equal to 0.05 for Exponential distribution under the null hypothesis

Method				Logrank	Gehan	Tarone	Peto	Fleming $\rho = 1, \gamma = 1$
	n_1	n_2	Censoring rate	$\hat{p}$	$\hat{p}$	$\hat{p}$	$\hat{p}$	$\hat{p}$
Exp(1)	20	20	0.1	0.050	0.046	0.046	0.044	0.058
Exp(1)	50	50	0.1	0.060	0.066	0.064	0.064	0.056
Exp(1.5)	20	20	0.1	0.066	0.058	0.062	0.058	0.062
Exp(1.5)	50	50	0.1	0.062	0.056	0.050	0.056	0.054
Exp(1)	20	20	0.3	0.058	0.054	0.056	0.056	0.056
Exp(1)	50	50	0.3	0.050	0.052	0.054	0.058	0.046
Exp(1.5)	20	20	0.3	0.054	0.054	0.050	0.052	0.052
Exp(1.5)	50	50	0.3	0.066	0.068	0.066	0.068	0.056

Table 2: Proportion of p-values less than or equal to 0.05 for Gamma distribution under the null hypothesis

Method				Logrank	Gehan	Tarone	Peto	Fleming $\rho=1,\gamma=1$
	n_1	n_2	Censoring rate	$\hat{p}$	$\hat{p}$	$\hat{p}$	$\hat{p}$	$\hat{p}$
Gamma(1,1)	20	20	0.3	0.054	0.054	0.056	0.058	0.054
Gamma(1,1)	50	50	0.1	0.038	0.038	0.030	0.032	0.050
Gamma(1.5,1.5)	20	20	0.1	0.046	0.048	0.048	0.048	0.062
Gamma(1.5,1.5)	50	50	0.1	0.046	0.044	0.046	0.044	0.050
Gamma(1,1)	20	20	0.3	0.056	0.060	0.058	0.052	0.066
Gamma(1,1)	50	50	0.3	0.054	0.052	0.058	0.050	0.046
Gamma(1.5,1.5)	20	20	0.3	0.058	0.060	0.064	0.062	0.066
Gamma(1.5,1.5)	50	50	0.3	0.050	0.062	0.060	0.062	0.050

Method				Energy distance $\alpha=1$	Kernel Gaussian $\sigma=1$	Kernel Laplacian $\sigma=1$
	n_1	n_2	Censoring rate	$\hat{p}$	$\hat{p}$	$\hat{p}$
Gamma(1,1)	20	20	0.3	0.058	0.052	0.060
Gamma(1,1)	50	50	0.1	0.044	0.042	0.040
Gamma(1.5,1.5)	20	20	0.1	0.062	0.060	0.060
Gamma(1.5,1.5)	50	50	0.1	0.050	0.054	0.052
Gamma(1,1)	20	20	0.3	0.058	0.064	0.062
Gamma(1,1)	50	50	0.3	0.058	0.056	0.062
Gamma(1.5,1.5)	20	20	0.3	0.068	0.070	0.056
Gamma(1.5,1.5)	50	50	0.3	0.056	0.056	0.068

Table 3: Proportion of p-values less than or equal to 0.05 for Lognormal distribution under the null hypothesis

Method				Logrank	Gehan	Tarone	Peto	Fleming $\rho=1,\gamma=1$
	n_1	n_2	Censoring rate	$\hat{p}$	$\hat{p}$	$\hat{p}$	$\hat{p}$	$\hat{p}$
Lognormal(0,0.5)	20	20	0.1	0.052	0.044	0.044	0.042	0.048
Lognormal(0,0.5)	50	50	0.1	0.040	0.034	0.040	0.036	0.040
Lognormal(0,0.25)	20	20	0.1	0.062	0.078	0.076	0.080	0.054
Lognormal(0,0.25)	50	50	0.1	0.036	0.044	0.044	0.040	0.038
Lognormal(0,0.5)	20	20	0.3	0.042	0.052	0.040	0.048	0.050
Lognormal(0,0.5)	50	50	0.3	0.078	0.082	0.078	0.082	0.066
Lognormal(0,0.25)	20	20	0.3	0.050	0.056	0.058	0.054	0.042
Lognormal(0,0.25)	50	50	0.3	0.046	0.060	0.046	0.058	0.048

Method				Energy distance $\alpha=1$	Kernel Gaussian $\sigma=1$	Kernel Laplacian $\sigma=1$
	n_1	n_2	Censoring rate	$\hat{p}$	$\hat{p}$	$\hat{p}$
Lognormal(0,0.5)	20	20	0.1	0.050	0.052	0.054
Lognormal(0,0.5)	50	50	0.1	0.040	0.038	0.040
Lognormal(0,0.25)	20	20	0.1	0.084	0.076	0.080
Lognormal(0,0.25)	50	50	0.1	0.038	0.040	0.044
Lognormal(0,0.5)	20	20	0.3	0.046	0.050	0.046
Lognormal(0,0.5)	50	50	0.3	0.072	0.074	0.074
Lognormal(0,0.25)	20	20	0.3	0.056	0.060	0.054
Lognormal(0,0.25)	50	50	0.3	0.044	0.040	0.052

Table 4: Empirical mean and standard deviation of p-values for Exponential distribution under the null hypothesis

Method				Logrank	Gehan	Tarone	Peto	Fleming $\rho = 1, \gamma = 1$
	n_1	n_2	Censoring rate	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$
Exp(1)	50	50	0.1	0.492 ± 0.293	0.489 ± 0.295	0.478 ± 0.280	0.486 ± 0.292	0.490 ± 0.295
Exp(1.5)	20	20	0.1	0.486 ± 0.289	0.475 ± 0.291	0.481 ± 0.288	0.458 ± 0.288	
Exp(1.5)	50	50	0.1	0.492 ± 0.299	0.501 ± 0.295	0.492 ± 0.295	0.498 ± 0.295	0.479 ± 0.293
Exp(1)	20	20	0.3	0.495 ± 0.284	0.502 ± 0.297	0.500 ± 0.294	0.499 ± 0.295	0.507 ± 0.286
Exp(1)	50	50	0.3	0.503 ± 0.296	0.486 ± 0.297	0.491 ± 0.296	0.486 ± 0.297	0.502 ± 0.287
Exp(1.5)	20	20	0.3	0.497 ± 0.295	0.499 ± 0.289	0.493 ± 0.285	0.495 ± 0.286	0.501 ± 0.292
Exp(1.5)	50	50	0.3	0.496 ± 0.298	0.492 ± 0.294	0.495 ± 0.299	0.492 ± 0.294	0.500 ± 0.299

Method				Energy distance $\alpha = 1$	Kernel Gaussian $\sigma = 1$	Kernel Laplacian $\sigma = 1$
	n_1	n_2	Censoring rate	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$
Exp(1)	50	50	0.1	0.482 ± 0.293	0.481 ± 0.298	0.478 ± 0.294
Exp(1.5)	20	20	0.1	0.482 ± 0.287	0.490 ± 0.285	0.493 ± 0.288
Exp(1.5)	50	50	0.1	0.482 ± 0.295	0.485 ± 0.293	0.481 ± 0.289
Exp(1)	20	20	0.3	0.508 ± 0.288	0.503 ± 0.285	0.506 ± 0.287
Exp(1)	50	50	0.3	0.494 ± 0.297	0.493 ± 0.297	0.495 ± 0.297
Exp(1.5)	20	20	0.3	0.500 ± 0.290	0.492 ± 0.284	0.506 ± 0.292
Exp(1.5)	50	50	0.3	0.489 ± 0.301	0.489 ± 0.301	0.490 ± 0.302

Table 5: Empirical mean and standard deviation of p-values for Gamma distribution under the null hypothesis

Method				Logrank	Gehan	Tarone	Peto	Fleming $\rho = 1, \gamma = 1$
	n_1	n_2	Censoring rate	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$
Gamma(1,1)	20	20	0.1	0.491 ± 0.284	0.510 ± 0.294	0.498 ± 0.288	0.506 ± 0.292	0.493 ± 0.282
Gamma(1,1)	50	50	0.1	0.505 ± 0.292	0.508 ± 0.287	0.505 ± 0.290	0.508 ± 0.288	0.502 ± 0.290
Gamma(1.5,1.5)	20	20	0.1	0.515 ± 0.299	0.520 ± 0.290	0.519 ± 0.289	0.522 ± 0.291	
Gamma(1.5,1.5)	50	50	0.1	0.493 ± 0.291	0.505 ± 0.289	0.509 ± 0.289	0.506 ± 0.288	0.505 ± 0.291
Gamma(1,1)	20	20	0.3	0.484 ± 0.288	0.475 ± 0.289	0.477 ± 0.297	0.467 ± 0.288	0.464 ± 0.288
Gamma(1,1)	50	50	0.3	0.485 ± 0.292	0.513 ± 0.300	0.498 ± 0.293	0.511 ± 0.300	0.474 ± 0.287
Gamma(1.5,1.5)	20	20	0.3	0.484 ± 0.297	0.499 ± 0.294	0.490 ± 0.295	0.494 ± 0.292	0.484 ± 0.294
Gamma(1.5,1.5)	50	50	0.3	0.509 ± 0.289	0.490 ± 0.291	0.493 ± 0.288	0.489 ± 0.292	0.514 ± 0.288

Method				Energy distance $\alpha = 1$	Kernel Gaussian $\sigma = 1$	Kernel Laplacian $\sigma = 1$
	n_1	n_2	Censoring rate	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$
Gamma(1,1)	20	20	0.1	0.501 ± 0.294	0.512 ± 0.297	0.508 ± 0.296
Gamma(1,1)	50	50	0.1	0.503 ± 0.291	0.512 ± 0.292	0.508 ± 0.288
Gamma(1.5,1.5)	20	20	0.1	0.519 ± 0.295	0.515 ± 0.301	0.516 ± 0.295
Gamma(1.5,1.5)	50	50	0.1	0.499 ± 0.290	0.493 ± 0.291	0.495 ± 0.292
Gamma(1,1)	20	20	0.3	0.477 ± 0.288	0.479 ± 0.289	0.484 ± 0.287
Gamma(1,1)	50	50	0.3	0.489 ± 0.293	0.497 ± 0.296	0.497 ± 0.293
Gamma(1.5,1.5)	20	20	0.3	0.491 ± 0.293	0.493 ± 0.294	0.494 ± 0.294
Gamma(1.5,1.5)	50	50	0.3	0.495 ± 0.295	0.492 ± 0.293	0.490 ± 0.295

Table 6: Empirical mean and standard deviation of p-values for Lognormal distribution under the null hypothesis

Method				Logrank	Gehan	Tarone	Peto	Fleming $\rho = 1, \gamma = 1$
	n_1	n_2	Censoring rate	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$
Lognormal(0,0.5)	20	20	0.1	0.472 ± 0.279	0.477 ± 0.287	0.470 ± 0.285	0.473 ± 0.286	0.483 ± 0.287
Lognormal(0,0.5)	50	50	0.1	0.508 ± 0.283	0.504 ± 0.279	0.508 ± 0.286	0.504 ± 0.281	0.515 ± 0.296
Lognormal(0,0.25)	20	20	0.1	0.484 ± 0.291	0.476 ± 0.295	0.473 ± 0.291	0.471 ± 0.293	0.487 ± 0.290
Lognormal(0,0.25)	50	50	0.1	0.517 ± 0.292	0.523 ± 0.291	0.522 ± 0.293	0.522 ± 0.291	0.506 ± 0.284
Lognormal(0,0.5)	20	20	0.3	0.495 ± 0.285	0.489 ± 0.288	0.488 ± 0.287	0.485 ± 0.286	0.516 ± 0.289
Lognormal(0,0.5)	50	50	0.3	0.476 ± 0.296	0.473 ± 0.293	0.468 ± 0.287	0.470 ± 0.292	0.487 ± 0.296
Lognormal(0,0.25)	20	20	0.3	0.516 ± 0.293	0.526 ± 0.306	0.524 ± 0.303	0.524 ± 0.306	0.518 ± 0.286
Lognormal(0,0.25)	50	50	0.3	0.491 ± 0.289	0.500 ± 0.296	0.496 ± 0.295	0.498 ± 0.295	0.494 ± 0.289

Method				Energy distance $\alpha = 1$	Kernel Gaussian $\sigma = 1$	Kernel Laplacian $\sigma = 1$
	n_1	n_2	Censoring rate	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$	$\overline{x} \pm \sigma$
Lognormal(0,0.5)	20	20	0.1	0.490 ± 0.287	0.490 ± 0.283	0.493 ± 0.287
Lognormal(0,0.5)	50	50	0.1	0.503 ± 0.283	0.500 ± 0.283	0.500 ± 0.283
Lognormal(0,0.25)	20	20	0.1	0.481 ± 0.294	0.481 ± 0.294	0.482 ± 0.296
Lognormal(0,0.25)	50	50	0.1	0.517 ± 0.289	0.517 ± 0.291	0.516 ± 0.288
Lognormal(0,0.5)	20	20	0.3	0.495 ± 0.288	0.495 ± 0.287	0.497 ± 0.287
Lognormal(0,0.5)	50	50	0.3	0.482 ± 0.293	0.482 ± 0.294	0.488 ± 0.297
Lognormal(0,0.25)	20	20	0.3	0.522 ± 0.293	0.526 ± 0.298	0.526 ± 0.298
Lognormal(0,0.25)	50	50	0.3	0.504 ± 0.291	0.501 ± 0.296	0.502 ± 0.296

Appendix F: Additional Content

The Bessel functions of the second kind $\Gamma(\cdot)$ (see Table 1) are solutions of the Bessel differential equations that have a singularity at $x = 0$. Bessel's differential equations are defined as follows:

$$x^2 \frac{d^2\Gamma}{dx^2} + x\frac{d\Gamma}{dx} + \left(x^2 - \alpha^2\right)\Gamma = 0$$

for an arbitrary complex number α, the order of the Bessel function.

REFERENCES

Bose, A. and Sen, A. (1999). The strong law of large numbers for kaplan–meier u-statistics. *Journal of Theoretical Probability*, 12(1):181–200.

Bose, A. and Sen, A. (2002). Asymptotic distribution of the kaplan–meier u-statistics. *Journal of Multivariate Analysis*, 83(1):84 – 123.

Fernández, T. and Rivera, N. (2018). Kaplan-meier v and u-statistics.

Korolyuk, V. S. and Borovskich, Y. V. (1994). *Theory of U-statistics*, volume 273. Springer Science & Business Media.

Muandet, K., Fukumizu, K., Sriperumbudur, B., Schölkopf, B., et al. (2017). Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends® in Machine Learning*, 10(1-2):1–141.

Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013). Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *The Annals of Statistics*, pages 2263–2291.

Stute, W. and Wang, J.-L. (1993). Multi-sample u-statistics for censored data. *Scandinavian journal of statistics*, pages 369–374.

Székely, G. J. and Rizzo, M. L. (2005). A new test for multivariate normality. *Journal of Multivariate Analysis*, 93(1):58–80.