Journal of Statistics and Computer Science
Vol. 2, Issue 2, 2023, pp. 147-159
https://doi.org/10.47509/JSCS.2023.v02i02.06

Equivalence tests based on weighted L_2-distance between cumulative distribution functions

Vladimir Ostrovski

ERGO Group AG, ERGO-Platz 1, 40477 Düsseldorf

vladimir.ostrovski77@googlemail.com

Abstract

We introduce a new family of equivalence tests for a fully specified continuous distribution on $\mathbb{R}$. The tests are based on the weighted L_2-distance between cumulative distribution functions. The asymptotic distribution of the test statistic is derived using the functional delta method. The local asymptotic optimality of the proposed tests is shown. An easy-to-compute estimator for the asymptotic variance of the test statistic is provided. The tests can be carried out using the asymptotic approximation or the percentile-t bootstrap method. For the special case of the Anderson-Darling distance, a comprehensive simulation study of finite sample properties is performed. A practical method of finding appropriate values for the tolerance parameter is given.

MSC 2020: 62G10

Key words: equivalence test, Cramér-von Mises distance, Anderson-Darling distance, uniformity test, weighted CDF

Article History
Received : 26 October 2023; Revised : 19 November 2023, Accepted : 15 December 2023; Published : 30 December 2023

To cite this paper
Vladimir Ostrovski (2023). Equivalence tests based on weighted L_2- distance between cumulative distribution functions. *Journal of Statistics and Computer Science.* 2(2), 147-159.

1 Introduction

A common task in applied statistics is to evaluate whether observed data conform to a specified distribution. For this purpose, goodness-of-fit tests are very often performed in practice. Goodness-of-fit tests are tailored to establish lack of fit to a hypothetical distribution. Therefore, goodness-of-fit tests are considered inappropriate for showing that observed data are consistent with a given probability distribution, see Hodges and Lehmann (1954), Berger and Delampady (1987), Lindsay and Liu (2009) and more recently Rao and Lovric (2016).

Equivalence tests are designed specifically to show that the observed data are sufficiently close to a given probability distribution, see Wellek (2010) for an overview. We consider equivalence tests for univariate continuous probability distributions on $\mathbb{R}$ based on the weighted L_2-distance between cumulative distribution functions (CDFs). The observed data $X_1, \ldots, X_n$ are identically and independently distributed accordingly to an unknown continuous CDF F on $\mathbb{R}$, where $n \in \mathbb{N}$ denotes the sample size. Let G denote the CDF of the hypothetical univariate continuous distribution. The weighted L_2-distance between F and G is $d(F, G) = \int (F - G)^2 \, d\mu$, where μ is a non-negative continuous measure on $\mathbb{R}$. The distance $d(F, G)$ is very popular in the field of the goodness-of-fit testing, see Baringhaus, Ebner, and Henze (2017) for an overview. The CDF F can be efficiently estimated by the empirical CDF F_n.

Example 1. The famous Cramér-von Mises distance $d(F, G) = \int (F - G)^2 \, dG$ is a special case with $\mu = G$. The popular Anderson-Darling distance $d(F, G) = \int \frac{(F-G)^2}{G(1-G)} dG$ is another special case with $d\mu = \frac{1}{G(1-G)} dG$.

There are first applications of distance $d(F, G)$ for equivalence testing. Baringhaus and Henze (2017) and Ostrovski (2022) use Cramér-von Mises distance to test equivalence to a fully specified univariate continuous distribution. Baringhaus, Gaigall, and Thiele (2018) apply the Cramér-von Mises distance in equivalence tests for uniformity on an unknown interval so the interval boundaries need to be estimated. The equivalence test problem is $H_0 = \{d(F,G) \geq \varepsilon\}$ and $H_1 = \{d(F,G) < \varepsilon\}$, where $\varepsilon > 0$ is a tolerance parameter. If H_0 can be rejected for an appropriate value of ε then the true underlying CDF F is sufficiently close to the given CDF G. Using the empirical CDF F_n as a plug-in estimator of F, we obtain the test statistic

$$T(F_n) = \sqrt{n}\,(d(F_n, G) - \varepsilon)$$

The distance $d(F_n, G)$ can be calculated by numerical integration. It can be transformed as follows to facilitate the computation:

$$d(F_n, G) = \int (F_n - G)^2 \, d\mu = \sum_{i=0}^{n} \int_{X_{(i)}}^{X_{(i+1)}} \left(\frac{i}{n} - G\right)^2 d\mu$$

where $X_{(1)} \leq \ldots \leq X_{(n)}$ are the order statistics, $X_{(0)} = -\infty$ and $X_{(n+1)} = +\infty$. The summands $\int_{X_{(i)}}^{X_{(i+1)}} \left(\frac{i}{n} - G\right)^2 d\mu$ have a closed-form expression in the case of the Cramér-von Mises and Anderson-Darling distances. The time-proven Cramér-von Mises test statistic is

$$nd(F_n, G) = \frac{1}{12n} + \sum_{i=1}^{n} \left(U_{(i)} - \frac{2i-1}{2n}\right)^2$$

where $U_i = G(X_i)$ for all i and $U_{(1)} \leq \ldots \leq U_{(n)}$ are the order statistics. Presented in a similar way, the Anderson-Darling test statistic is

$$nd(F_n, G) = -n - \sum_{i=1}^{n} \frac{2i-1}{n} \left[\ln\left(U_{(i)}\right) + \ln\left(1 - U_{(n+1-i)}\right)\right]$$

2 Asymptotic theory

In this section, the asymptotic distribution of the test statistic $T(F_n)$ is derived using the functional delta method. Then we show that the equivalence tests based on $T(F_n)$ are locally asymptotically most powerful (LAMP), see van der Vaart (1998, Chapter 25) for the theory of LAMP tests.

Proposition 2. *The statistical functional* $\kappa : L_2(\mu) \to \mathbb{R}, F \mapsto d(F,G)$ *is Hadamard differentiable everywhere. The derivative of* κ *at* $F \in L_2(\mu)$ *is the continuous linear function* $L_2(\mu) \to \mathbb{R}, h \mapsto \int 2(F-G)\, h d\mu$.

Proof. Let $(t_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ so that $t_n > 0$ for all $n \in \mathbb{N}$ and $t_n \to 0$ for $n \to \infty$. Let $(h_n)_{n\in\mathbb{N}}$ be a sequence in $L_2(\mu)$ so that $\int (h_n - h)^2 d\mu \to 0$ for $n \to \infty$. We show that

$$\begin{aligned} & \frac{1}{t_n}\left(\int (F + t_n h_n - G)^2 d\mu - \int (F-G)^2 d\mu\right) - 2\int h(F-G)\, d\mu \\ = \; & 2\int (h_n - h)(F-G)\, d\mu + t_n \int h_n^2 d\mu \end{aligned}$$

converges to 0 for $n \to \infty$. Applying the Cauchy-Schwartz inequality to the first summand, we obtain

$$\left(\int (h_n - h)(F-G)\, d\mu\right)^2 \leq \int (h_n - h)^2 d\mu \int (F-G)^2 d\mu \to 0$$

for $n \to \infty$. The second summand $t_n \int h_n^2 d\mu$ converges to 0 for $n \to \infty$ because $t_n \to 0$ and $\int h_n^2 d\mu \to \int h^2 d\mu$. □

Corollary 3. *Let* F *be the true underlying CDF of the observed data. If* $d(F,G) = \varepsilon$ *then the test statistic* $T(F_n)$ *converges weakly to* $2\int (F-G)\,\mathbb{G}_F d\mu$, *where* $\mathbb{G}_F$ *is a* F*-Brownian bridge process.*

Proof. By Donsker's theorem (see van der Vaart (1998, p. 266, Theorem 19.3)),

$\sqrt{n}(F_n - F)$ converges weakly to $\mathbb{G}_F$. Applying the functional delta method (see van der Vaart (1998, p. 287, Theorem 20.8), we obtain the asymptotic distribution of $T(F_n) = \sqrt{n}(d(F_n, G) - d(F, G))$. $\square$

Example 4. For the Cramér-von Mises distance, the test statistic $T(F_n)$ converges weakly to $2\int (F - G)\mathbb{G}_F dG$. In the case of the Anderson-Darling distance, $T(F_n)$ converges weakly to $2\int \frac{F-G}{G(1-G)}\mathbb{G}_F dG$.

Next, we provide a short introduction to LAMP tests. The local asymptotic optimality is based on the notion of the locally smooth parametric submodels and the corresponding bounds on the asymptotic test power. Let $F_0 \in L_2(\mu)$ be a fixed CDF so that $d(F_0, G) = \varepsilon$. Let $[0, \delta] \to L_2(\mu), t \mapsto F_t$ be a parametric submodel for $\delta > 0$. Assume that the submodel is differentiable at F_0 in the mean square, see van der Vaart (1998, p. 362). We consider all asymptotic α-level tests for the equivalence test problem. Then for any sequence $\left(F_{t/\sqrt{n}}\right)_{n\in\mathbb{N}}$ and $t > 0$, there is an asymptotic upper bound on the power of asymptotic α-level tests at $F_{t/\sqrt{n}}$, see van der Vaart (1998, p. 384, Theorem 25.44). An asymptotic α-level test is LAMP at F_0 if the test power attains this upper bound for all submodels that are differentiable at F_0 in the mean square. Next, we derive the efficient influence function $\tilde{k}$ of the statistical functional κ. Then we show that the asymptotic distributions of the test statistic $T(F_n)$ and $S_n = \sqrt{n}\sum_{i=1}^{n} \tilde{k}(X_i)$ coincide. This implies that the asymptotic α-level tests based on the statistic $T(F_n)$ are LAMP at any CDF $F_0 \in L_2(\mu)$ with $d(F_0, G) = \varepsilon$.

Proposition 5. *The efficient influence function of the functional* κ *at* F_0 *is* $\tilde{\kappa}(x) = 2\int (F_0(s) - G(s))\left(1_{(-\infty,s]}(x) - F_0(s)\right) d\mu(s)$.

Proof. Let $[0, \delta] \to L_2(\mu), t \mapsto F_t$ be a parametric submodel for $\delta > 0$ that is differentiable at F_0 in the mean square with a tangent h. By van der Vaart (1998, p. 363, Lemma 25.14), the tangent h is in $L_2(F_0)$ and $\int h dF_0 = 0$.

Proposition 2 from Bickel, Klaassen, Ritov, and Wellner (1993, p. 457) implies $\frac{\partial}{\partial t} F_t(s) = \int h(x) 1_{(-\infty,s]}(x)\, dF_0(x)$ at $t = 0$. With this result we obtain $\frac{\partial}{\partial t}\kappa(F_t) = 2\int (F_0(s) - G(s)) \int h(x)\, 1_{(-\infty,s]}(x)\, dF_0(x)\, d\mu(s)$ at $t = 0$.
By Fubini's theorem, we conclude
$\frac{\partial}{\partial t}\kappa(F_t) = \int h(x) \left[2\int (F_0(s) - G(s))\, 1_{(-\infty,s]}(x)\, d\mu(s)\right] dF_0(x) = \int h\tilde{\kappa} dF_0$
at $t = 0$, where the last equality is due to
$\int h(x) \left[\int (F_0(s) - G(s))\, F_0(s)\, d\mu(s)\right] dF_0(x) = \int (F_0 - G)\, F_0 d\mu \int h dF_0 = 0$
Obviously, $\int \tilde{\kappa} dF_0$ equals zero. Due to $\left|1_{(-\infty,s]}(x) - F_0(s)\right| \leq 1$ and $F_0, G \in L_2(\mu) \subset L_1(\mu)$, we obtain $|\tilde{\kappa}| \leq 2\int |F_0 - G|\, d\mu < \infty$. Consequently, $\tilde{\kappa} \in L_2(F_0)$. Similar to Example 25.16 of van der Vaart (1998, p. 364), we can construct a submodel $[0,\delta] \to L_2(\mu), t \mapsto F_t$ for $\delta > 0$ that is differentiable at F_0 in the mean square with a tangent $\tilde{\kappa}$. Therefore, $\tilde{\kappa}$ is the efficient score function of k at F_0 by definition, see van der Vaart (1998, Section 25.4). □

Proposition 6. *Let $\sigma_n^2 = \sigma_n^2(X_1, \ldots, X_n)$ be a consistent estimator of the asymptotic variance of the test statistic $T(F_n)$. Let $(c_n)_{n\in\mathbb{N}}$ be a sequence of critical values so that $c_n \to q_\alpha$, where q_α denotes the lower α-quantile of the standard normal distribution. Then the test, that rejects H_0 if $T(F_n) \leq q_\alpha \sigma_n$, is LAMP at any F_0 with $d(F_0, G) = \varepsilon$.*

Proof. By Theorem 25.44 and Lemma 25.45 of van der Vaart (1998, p. 384), it is sufficient to show that the asymptotic distributions of $T(F_n)$ and S_n are equal under F_0, where $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{k}(X_i) = 2\int (F_0 - G) \sqrt{n}(F_n - F_0)\, d\mu$. By Donsker's theorem, $\sqrt{n}(F_n - F)$ converges weakly to $\mathbb{G}_{F_0}$. The continuous mapping theorem implies that S_n converges weakly to $2\int (F_0 - G)\, \mathbb{G}_{F_0} d\mu$. The asymptotic distribution of T_n is given in Corollary 3. □

3 Asymptotic and bootstrap-based tests

In this section, we derive an estimator for the asymptotic variance of the test statistic. The asymptotic and bootstrap-based tests are described and the asymptotic optimality of these tests is shown.

Proposition 7. *Suppose that* $d\mu = wdH$, *where* $H : \mathbb{R} \to \mathbb{R}$ *is a continuous CDF and* $w : \mathbb{R} \to \mathbb{R}$ *is a continuous function. Then the distribution of* $2\int (F - G)\, \mathbb{G}_F d\mu$ *is normal with mean* 0 *and variance*
$\sigma^2(F) = 4\int\int K(s,t,F)\, dH(s)\, dH(t)$, *where*
$K(x,y,F) = (F(x) - G(x))\, w(x)\, (F(y) - G(y))\, w(y)\, (F(x \wedge y) - F(x)F(y))$.

Proof. Using the identity $H^{-1} \circ H(x) = x$, we obtain
$2\int (F-G)\, \mathbb{G}_F d\mu = 2\int_0^1 \left(F \circ H^{-1} - G \circ H^{-1}\right) w \circ H^{-1} \mathbb{G}_{F\circ H^{-1}} d\lambda$, where λ denotes Lebesgue measure. The process $\left(F \circ H^{-1} - G \circ H^{-1}\right) w \circ H^{-1} \mathbb{G}_{F\circ H^{-1}}$ is Gaussian with zero means and covariance function $K\left(H^{-1}(s), H^{-1}(t), F\right)$. By Shorack and Wellner (2009, p. 42, Proposition 2.2.1), the distribution of $2\int (F-G)\, \mathbb{G}_F d\mu$ is normal with mean 0 and variance
$4\int_0^1\int_0^1 K\left(H^{-1}(s), H^{-1}(t), F\right) ds dt = 4\int\int K(s,t,F)\, dH(s)\, dH(t)$. □

Corollary 8. *Let* $f_k = \int_{X_{(k)}}^{X_{(k+1)}} \left(\frac{k}{n} - G(t)\right) w(t)\, dH(t)$ *for* $k = 0, \ldots, n$, *where* $X_{(0)} = -\infty$ *and* $X_{(n+1)} = +\infty$. *Under assumptions of Proposition 7, the estimator of the asymptotic variance of the test statistic*

$$\sigma^2(F_n) = 4\sum_{k=0}^{n}\sum_{l=0}^{n}\left(\frac{k \wedge l}{n} - \frac{kl}{n^2}\right) f_k f_l \tag{3.1}$$

is consistent almost surely.

Proof. The function $F \mapsto \sigma^2(F)$ is continuous by Proposition 7 and $F_n \to F$ almost surely for $n \to \infty$ by Glivenco-Cantelly theorem. Therefore, $\sigma^2(F_n) \to$

$\sigma^2(F)$ almost surely by the continuous mapping theorem. Observe that

$$\sigma^2(F_n) = \sum_{k=0}^{n}\sum_{l=0}^{n} 4\int_{X_{(k)}}^{X_{(k+1)}}\int_{X_{(l)}}^{X_{(l+1)}} K(s,t,F_n)\,dH(s)\,dH(t)$$

Then the closed formula (3.1) is derived by straightforward algebra and Fubini's theorem. □

Example 9. Let $U_{(k)} = G\left(X_{(k)}\right)$ for all $k = 1,\ldots,n$. Let $U_{(0)} = 0$ and $U_{(n+1)} = 1$. In the case of the Cramér-von Mises distance, we get a closed-form solution for f_k as follows

$f_k = \int_{X_{(k)}}^{X_{(k+1)}}\left(\frac{k}{n} - G(t)\right)dG(t) = \int_{U_{(k)}}^{U_{(k+1)}}\left(\frac{k}{n} - x\right)dx = \frac{k}{n}x - x^2\Big|_{U_{(k)}}^{U_{(k+1)}}$ using the substitution $x = G(t)$. For the Anderson-Darling distance we obtain

$f_k = \int_{X_{(k)}}^{X_{(k+1)}}\left(\frac{k}{n} - G(t)\right)dG(t) = \frac{k}{n}\ln\left(\frac{x}{1-x}\right) + \ln(1-x)\Big|_{U_{(k)}}^{U_{(k+1)}}$

The asymptotic test rejects H_0 if $T(F_n) \le q_\alpha \sigma(F_n)$, where q_α is the lower α-quantile of the standard normal distribution. The minimum tolerance parameter, for which the asymptotic test can reject H_0, is $\varepsilon_{\min}(F_n) = d(F_n, G) - \frac{1}{\sqrt{n}} q_\alpha \sigma(F_n)$. Under assumptions of Proposition 7, the asymptotic test is LAMP by Proposition 6.

Alternatively, the variance of the test statistic $T(F_n)$ can be estimated by empirical bootstrap. Let $\hat{\sigma}_n^2$ denote the usual bootstrap estimator of the variance of $T(F_n)$. The estimator $\hat{\sigma}_n^2$ can be computed by sampling with replacement from the original observations to any degree of accuracy. The bootstrap estimator $\hat{\sigma}_n^2$ replaces $\sigma^2(F_n)$ in the asymptotic test. Otherwise, everything remains the same. The estimator $\hat{\sigma}_n^2$ is consistent by van der Vaart (1998, p. 333, Theorem 23.7) if the asymptotic variance of the test statistic $T(F_n)$ is finite. Then the asymptotic test that uses the bootstrap estimator $\hat{\sigma}_n^2$ instead of $\sigma^2(F_n)$ is LAMP by Proposition 6.

In order to improve the finite sample performance, we consider the bootstrap based test that implements the percentile-t bootstrap method, see van der Vaart (1998, Chapter 25) for details. Let $\hat{F_n}$ denote the empirical CDF of the usual bootstrap sample with replacement. The percentile-t bootstrap method approximates the unknown true distribution of the test statistic $T(F_n)$ by the conditional distribution of $\left(T\left(\hat{F_n}\right) - T(F_n)\right)/\sigma\left(\hat{F_n}\right)$ given F_n, which can be easily simulated by sampling with replacement. Let $\hat{q}_\alpha(F_n)$ denote the lower empirical α-quantile of $\left(T\left(\hat{F_n}\right) - T(F_n)\right)/\sigma\left(\hat{F_n}\right)$ given F_n. The percentile-t bootstrap test rejects H_0 if $T(F_n) \leq \hat{q}_\alpha(F_n)\,\sigma(F_n)$. Under assumptions of Proposition 7, the quantile $\hat{q}_\alpha(F_n)$ converges to q_α by van der Vaart (1998, p. 333, Theorem 23.7 and Theorem 23.9). Therefore, the percentile-t bootstrap test is LAMP by Proposition 6. The minimum tolerance parameter $\varepsilon_{\min}$, for which the percentile-t bootstrap test can reject H_0, equals $d(F_n, G) - \frac{1}{\sqrt{n}}\hat{q}_\alpha(F_n)\,\sigma(F_n)$.

4 Simulation Study

In this section, the finite sample properties of the proposed equivalence tests are studied by means of simulation. We consider the Anderson-Darling distance only because an extensive simulation study of the equivalence tests based on the Cramér-von Mises distance can be found in Ostrovski (2022). Let U denote the CDF of the uniform distribution on $[0,1]$. For the Anderson-Darling distance, it is sufficient to consider the case $G = U$ only because the identity

$$d(F,G) = \int \frac{(F-G)^2}{G(1-G)}dG = \int \frac{\left(F \circ G^{-1} - U\right)^2}{U(1-U)}dU$$

holds for any continuous CDF G. The tests are implemented in R and the complete simulation study is performed in R Studio. The source code is freely available under `https://github.com/TestingEquivalence/EquivalenceAD`. We use the following shorthand notation for the equivalence tests considered:

AT is the asymptotic test which uses the estimator $\sigma^2(F_n)$ of the asymptotic variance of the test statistic.

ATBV is the asymptotic test which uses the bootstrap estimator $\hat{\sigma}_n^2$ of the variance of the test statistic.

PTBT is the test based on the percentile-t bootstrap method.

All tests are performed at the nominal level 0.05. The test power is calculated using 1000 simulations in all cases. The number of bootstrap samples is 200 for ATBV and PTBT.

4.1 Test power at U and appropriate values of the tolerance parameter ε

To gain insight into how to determine appropriate values for the tolerance parameter ε, we calculate the test power at U. Table 1 displays the value of the tolerance parameter ε as a function of the test power at U for different sample sizes n. For a fixed test power, the corresponding values of ε are similar for AT, ATBV and PTBT. For a fixed sample size n, the test power at U decreases as ε is reduced.

Given a specific sample size, an appropriate value of the tolerance parameter ε can be determined by fixing the minimum acceptable test power at U. For example, the test power at U should be at least 0.9 and the sample size is 200. Then the appropriate value of ε is 0.025, see Table 1.

4.2 Type I error rates

The test power is computed at a number of different boundary points of H_0 to evaluate type I error rates. The CDF F of any boundary point under consideration is constructed as a linear combination $F = wB + (1-w)U$ similar

Table 1: Tolerance parameter ε as a function of the test power.

Test power	n	AT	ATBV	PTBT
0.9	50	0.091	0.096	0.097
	100	0.047	0.049	0.051
	200	0.023	0.025	0.025
0.8	50	0.070	0.078	0.062
	100	0.036	0.040	0.032
	200	0.018	0.020	0.017
0.7	50	0.057	0.066	0.042
	100	0.030	0.034	0.022
	200	0.015	0.017	0.011

to Baringhaus and Henze (2017) and Ostrovski (2022), where $w \in [0,1]$ and B is such CDF that $d(B,U) > \varepsilon$. CDFs B are selected from the extensive literature on goodness-of-fit tests for uniformity and are also considered in Ostrovski (2022). The beta distributions with different parameters and Stephens alternatives are used as CDFs B, see Table 2 for details. The Stephens alternatives consist of three different parametric CDFs A(k), B(k) and C(k), where the parameter k>0 controls the shape of CDF, see Stephens (1974) for details and exact formulas of CDFs. The test power at the different boundary points is summarized in Table 2 for the sample size 200. Similar simulation results are obtained for sample sizes 50 and 100. The tolerance parameter ε equals 0.025, so that the test power at U is over 0.9 for AT, ATBV and PTBT, see Table 1.

The test power of all three tests varies considerably from point to point and deviates significantly from the nominal value 0.05. AT is not conservative at many boundary points and the power of AT is the most variable. PTBT has some non-conservative tendencies, as the power of PTBT is often above the nominal level. The power of ATBV is closer to the nominal level at most boundary points compared to AT and PTBT. Therefore, ATBV performs best compared to AT and PTBT, although it is not conservative at some boundary points considered.

Table 2: Test power at the boundary points of H_0. The tolerance parameter is $\varepsilon = 0.025$.

CDF B	AT	ATBV	PTBT	CDF B	AT	ATBV	PTBT
Beta(0.5, 0.5)	0.026	0.016	0.063	A(2.5)	0.070	0.066	0.057
Beta(0.5, 1.0)	0.053	0.044	0.059	A(3)	0.070	0.058	0.055
Beta(0.5, 1.5)	0.090	0.065	0.077	B(0.25)	0.041	0.023	0.074
Beta(0.5, 2.0)	0.107	0.087	0.105	B(0.5)	0.021	0.015	0.038
Beta(1, 1.5)	0.068	0.061	0.056	B(2)	0.031	0.023	0.060
Beta(1, 2)	0.096	0.087	0.076	B(2.5)	0.025	0.022	0.058
Beta(1.5, 2)	0.057	0.046	0.067	B(3)	0.031	0.026	0.065
Beta(2, 2)	0.030	0.024	0.059	C(0.25)	0.036	0.028	0.059
A(0.25)	0.090	0.057	0.102	C(0.5)	0.043	0.030	0.079
A(0.5)	0.081	0.059	0.081	C(2)	0.017	0.005	0.047
A(1.5)	0.077	0.068	0.061	C(2.5)	0.018	0.012	0.033
A(2)	0.083	0.070	0.064	C(3)	0.035	0.022	0.064

References

Baringhaus, L., Ebner, B., and Henze, N. (2017). The limit distribution of weighted l^2-goodness-of-fit statistics under fixed alternatives, with applications. *Annals of the Institute of Statistical Mathematics*, *69*(5), 969–995.

Baringhaus, L., Gaigall, D., and Thiele, J. P. (2018). Statistical inference for L_2-distances to uniformity. *Computational Statistics*, *33*(4), 1863–1896.

Baringhaus, L., and Henze, N. (2017). Cramér-von Mises distance: probabilistic interpretation, confidence intervals, and neighbourhood-of-model validation. *Journal of Nonparametric Statistics*, *29*(2), 167-188.

Berger, J. O., and Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, *2*(3), 317 – 335.

Bickel, P., Klaassen, C., Ritov, Y., and Wellner, J. (1993). *Efficient and adaptive estimation for semiparametric models.* Johns Hopkins University Press, Baltimore.

Hodges, J., and Lehmann, E. (1954). Testing the approximate validity of statistical hypotheses. *J. Roy. Statist. Soc. Ser. B*, *16*(2), 261–268.

Lindsay, B., and Liu, J. (2009). Model Assessment Tools for a Model False World. *Statistical Science*, *24*(3), 303 – 318.

Ostrovski, V. (2022). On application of the Cramér-von Mises distance for equivalence testing. *Journal of Statistics and Computer Science*, *1*(1), 1–9.

Rao, C. R., and Lovric, M. M. (2016). Testing point null hypothesis of a normal mean and the truth: 21st century perspective. *Journal of Modern Applied Statistical Methods*, *15*(2), 2-21.

Shorack, G., and Wellner, J. (2009). *Empirical processes with applications to statistics.* Society for Industrial and Applied Mathematics.

Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, *69*(347), 730–737.

van der Vaart, A. (1998). *Asymptotic statistics.* Cambridge University.

Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority.* CRC Press.