IMS Collections
Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková
Vol. 7 (2010) 70–74

DOI: 10.1214/10-IMSCOLL707

On a paradoxical property of the Kolmogorov–Smirnov two-sample test

Alexander Y. Gordon[1] **and Lev B. Klebanov**[2]

University of North Carolina at Charlotte and Charles University at Prague

Abstract: The two-sample Kolmogorov–Smirnov test can lose power as the size of one sample grows while the size of the other sample remains constant. In this case, a paradoxical situation takes place: the use of additional observations weakens the ability of the test to reject the null hypothesis when it is false.

1. Biasedness of the Kolmogorov goodness-of-fit test

We start with partially known results on biasedness of the Kolmogorov goodness-of-fit test (see [1]).

Let us recall some definitions. Suppose that $X_1, \dots, X_n$ are independent and identically distributed (i.i.d.) random variables (observations) with (unknown) distribution function (d.f.) F. Based on the observations, one needs to test the hypothesis

$$H_0 : \quad F = F_0,$$

where F_0 is a fixed d.f.

Definition 1.1. *For a specific alternative hypothesis, a test is said to be unbiased if the probability of rejecting the null hypothesis*
(a) is greater than or equal to the significance level when the alternative is true, and
(b) is less than or equal to the significance level when the null hypothesis is true (i. e. the test is of the α level).
A test is said to be biased for an alternative hypothesis, if (a) is not true while (b) remains true (i. e. for this alternative test remains to be of level α).

Below we will consider a test with the following properties:

1. For a distance d in the space of d.f.'s we reject the null hypothesis H_0 if

$$d(G_n, F_0) > \delta_\alpha,$$

where G_n is a sample d.f. of $X_1, \dots, X_n$ and δ_α satisfies the inequality

$$\mathbb{P}\{d(G_n, F_0) > \delta_\alpha\} \le \alpha. \tag{1.1}$$

[1]Department of Mathematics and Statistics, University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC 28223, USA, e-mail: aygordon@gmail.com

[2]Department of Probability and Statistics Charles University, Sokolovská 83, Prague, 18675, Czech Republic, e-mail: levkl@gmail.com

AMS 2000 subject classifications: Primary 62G10

Keywords and phrases: Kolmogorov goodness-of-fit test, Kolmogorov–Smirnov two-sample test, unbiasedness

2. The test is distribution free, i. e., the probability

$$\mathbb{P}_F\{d(G_n, F) > \delta_\alpha\}$$

does not depend on the continuous d.f. F.

We call such tests *distance-based*.

Denote by $\mathcal{B}(F, \delta)$ an closed ball of radius $\delta > 0$ centered at F in the metric space of all d.f.'s with the distance d.

Let F_0 be a continuous d.f. and let δ_α be defined to satisfy (1.1).

Theorem 1.1. *Suppose that for some $\alpha > 0$ there exists a continuous d.f. F_a such that*

$$\mathcal{B}(F_a, \delta_\alpha) \subset \mathcal{B}(F_0, \delta_\alpha), \tag{1.2}$$

and

$$\mathbb{P}_{F_a}\{G_n \in B(F_0, \delta_\alpha) \setminus B(F_a, \delta_\alpha)\} > 0. \tag{1.3}$$

Then the distance-based test is biased for the alternative F_a.

Proof. Let $X_1, \dots, X_n$ be a sample from F_a and G_n be the corresponding sample d.f. Then

$$\mathbb{P}_{F_a}\{G_n \in \mathcal{B}(F_a, \delta_\alpha)\} \geq 1 - \alpha.$$

In view of (1.2) and (1.3) we have

$$\mathbb{P}_{F_a}\{G_n \in \mathcal{B}(F_0, \delta_\alpha)\} > 1 - \alpha,$$

that is

$$\mathbb{P}_{F_a}\{d(G_n, F_0) > \delta_\alpha\} < \alpha.$$

□

Note that Theorem 1.1 is not a consequence of the result [2], because the alternative distribution in [2] is an n-dimensional distribution, and therefore, the observations $X_1, \dots, X_n$ are not i.i.d. random variables.

Consider now the Kolmogorov goodness-of-fit test. Clearly, it is a distance-based test for the uniform distance

$$d(F, G) = \sup_x |F(x) - G(x)|. \tag{1.4}$$

Let us show that there are F_0 and F_a such that (1.2) holds. Without loss of generality we may choose

$$F_0(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

For a fixed n, we define δ_α so that (1.1) is true.

The ball $\mathcal{B}(F_0, \delta_\alpha)$ with $\delta_\alpha = 0.2$ is shown in Figure 1. Its center – the function F_0 – is shown in black, while the lower and upper "boundaries" of the ball are shown in gray.

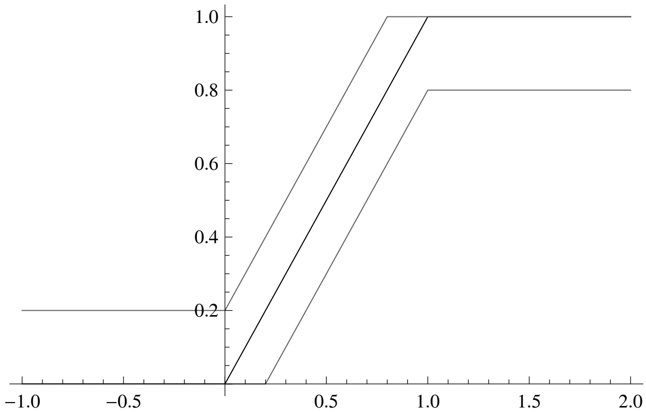

FIG 1. *The ball* $\mathcal{B}(F_0, \delta_\alpha)$.

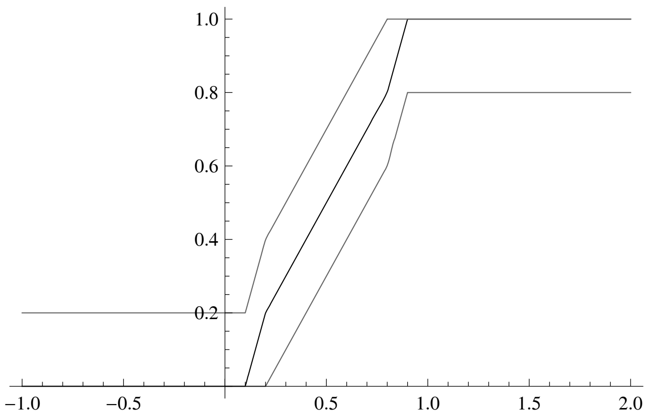

FIG 2. *The ball* $\mathcal{B}(F_a, \delta_\alpha)$.

Consider now the following d.f.:

$$F_a(x) = \begin{cases} 0, & x < \delta_\alpha/2, \\ 2x - \delta_\alpha, & \delta_\alpha/2 \le x < \delta_\alpha, \\ x, & \delta_\alpha \le x < 1 - \delta_\alpha, \\ 2x - (1 - \delta_\alpha), & 1 - \delta_\alpha \le x < 1 - \delta_\alpha/2, \\ 1, & x \ge 1 - \delta_\alpha/2. \end{cases}$$

Comparing Figures 1 and 2, we see that $\mathcal{B}(F_a, \delta_\alpha) \subset \mathcal{B}(F_0, \delta_\alpha)$, and therefore Kolmogorov test is biased for alternative F_a.

2. Biasedness of the Kolmogorov–Smirnov two-sample test for substantially different sizes of the samples and the paradox

Let us turn to two-sample problem. Suppose that we have two samples $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$, where all observations are independent. We also suppose that all

X_i's have the same d.f. F and all Y_j's – the same d.f. G. We suppose that both F and G are continuous functions. The null hypothesis is now $H_0: F = G$. It is clear that, without loss of generality, we may assume

$$G(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \le x < 1, \\ 1, & x \ge 1. \end{cases} \tag{2.5}$$

In addition, we suppose that

$$\mathtt{supp} F \subset [0,1] \text{ and } F \text{ is absolutely continuous.} \tag{2.6}$$

From the results of Section 1 we see that, for an arbitrary fixed n and sufficiently large nm, the two-sample Kolmogorov–Smirnov test is biased (for alternative $F = F_a \neq G$ given in Section 1), because for $m \to \infty$ we obtain in the limit the Kolmogorov goodness-of-fit test.

In Section 3 we show that in the case where $m = n$ the Kolmogorov–Smirnov test is unbiased, at least for small values of α for any alternative (2.6). However, for the same values of α and fixed n, the test will no longer be unbiased if m is large enough. In other words, the power of the test for some alternatives will be smaller for a large $m \gg n$ than for $m = n$. This means, paradoxically, that using the Kolmogorov–Smirnov test one cannot benefit from the additional information contained in a much larger sample: vice versa, instead of gaining power, the test loses it. The situation here is in some sense similar to that in statistical estimation theory in the situation where non-convex loss functions are used (see, for example, [3]).

3. On the unbiasedness of two-sample Kolmogorov–Smirnov test for samples of the same size

Here we will show that in the case where $m = n$ the Kolmogorov–Smirnov test is unbiased, at least for small values of α, for any alternative satisfying (2.6).

Theorem 3.1. *For $m = n$ there exists $\alpha \in (0,1)$ such that the Kolmogorov–Smirnov test is unbiased for any alternative (2.6).*

Proof. Recall that the Kolmogorov-Smirnov statistic is of the form

$$D_n = \sup_x |F_n(x) - G_n(x)|,$$

where F_n and G_n are sample d.f.'s based on the samples X_j and Y_j ($j = 1, \dots, n$), respectively. Clearly, under the hypothesis H_0 the distribution of the Kolmogorov–Smirnov statistic is discrete and therefore for some $\alpha \in (0,1)$ the event $D_n > \delta_\alpha$ is equivalent to the event $D_n = 1$. The latter event takes place if and only if

$$\max(X_1, \dots, X_n) < \min(Y_1, \dots, Y_n) \text{ or } \max(Y_1, \dots, Y_n) < \min(X_1, \dots, X_n) \tag{3.7}$$

The probability of the event (3.7) equals

$$\int_0^1 \big(F^n(x)(1-x)^{n-1} + (1-F(x))^n x^{n-1}\big) \mathrm{d}x. \tag{3.8}$$

In (3.8) we suppose that Y_1 has d.f. (2.5) and X_1 has d.f. $F(x)$.

It is easy to see that the function $y^n(1-x)^{n-1} + (1-y)^n x^{n-1}$, for any x $(0 < x < 1)$ has a minimum in y $(0 < y < 1)$ at the point $y = x$. Therefore, the integral (3.8) attains its minimum in F for $F(x) \equiv x$. This minimum equals

$$\int_0^1 z^{n-1}(1-z)^{n-1}\,\mathrm{d}z = n\frac{\Gamma^2(n)}{\Gamma(2n)},$$

what can be easily seen from combinatorial considerations, too. The integral represents the probability of rejecting the alternative, and it is minimal when $F = G$, i. e., when the null hypothesis is true. □

Note that in the case $m = n = 2$ Theorem 3.1 establishes the unbiasedness of the Kolmogorov–Smirnov test for any alternative satisfying(2.6), because other values of δ_α lead to a trivial result. We believe that in the case $m = n$ the test is unbiased for any α and any continuous alternative.

4. Concluding remarks

It has been shown that for the two-sample Kolmogorov–Smirnov test a paradoxical situation takes place: one cannot use additional information contained in a very large sample if the second sample is relatively small.

This paradoxical situation takes place not only for the Kolmogorov–Smirnov test. A similar paradox takes place, e. g., for the Cramér–Von Mises two-sample test (see [4], where the biasedness of the Cramér–Von Mises goodness-of-fit test is proved). We believe that a new approach is needed for handling the case of substantially different sample sizes.

Acknowledgement

The second named author was supported by the Grant MSM 002160839 of the Ministry of Higher Education of Czech Republic.

References

[1] Massey, F.J., Jr (1950). A Note on the Power of a Non-Parametric Test. *Annals of Math. Statist.* **21** 440–443.

[2] Thompson, Roy O.R.Y (1979). Bias and Monotonicity of Goodness-of-Fit Tests. *Journal of Amer. Statist. Association* **74** 875–876.

[3] Klebanov, L., Rachev, S., Fabozzi, F. (2009). *Robust and Non-Robust models in Statistics*, Nova, New York.

[4] Thompson, Roy O.R.Y (1966) Bias of the One-Sample Cramér-Von Mises Test. *Journal of Amer. Statist. Association* **61** 246-247.