



University of Pennsylvania
ScholarlyCommons

Statistics Papers

Wharton Faculty Research

1990

Tests and Confidence Bands for Bivariate Cumulative Distribution Functions

Robert J. Adler

Lawrence D. Brown
University of Pennsylvania

Kun-Liang Lu

Follow this and additional works at: https://repository.upenn.edu/statistics_papers

 Part of the [Statistics and Probability Commons](#)

Recommended Citation

Adler, R. J., Brown, L. D., & Lu, K. (1990). Tests and Confidence Bands for Bivariate Cumulative Distribution Functions. *Communications in Statistics - Simulation and Computation*, 19 (1), 25-36. <http://dx.doi.org/10.1080/03610919008812843>

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/statistics_papers/517
For more information, please contact repository@pobox.upenn.edu.

Tests and Confidence Bands for Bivariate Cumulative Distribution Functions

Keywords

cumulative distribution functions, non-parametric tests, confidence bands, Kolmogorov-Smirnov statistic

Disciplines

Statistics and Probability

TESTS AND CONFIDENCE BANDS FOR BIVARIATE
CUMULATIVE DISTRIBUTION FUNCTIONS

Robert J. Adler ¹

(Technion, Haifa, Israel)

Lawrence D. Brown ²

(Cornell University, Ithaca, N.Y. U.S.A.)

Kun-Liang Lu ³

(Cornell University, Ithaca, N.Y. U.S.A.)

AMS Subject Classification: Primary: 62G15, 62G10, 62G30. Secondary: 60E99

Key Words and Phrases: Cumulative distribution functions; non-parametric tests; confidence bands; Kolmogorov-Smirnov statistic.

1. INTRODUCTION.

In the case of a univariate cumulative distribution function (C.D.F.) the Kolmogorov-Smirnov statistic yields tests of $H_0 : F = F_0$ and thus confidence bands for the C.D.F. These procedures are distribution free over the class of continuous C.D.F.'s. For a multivariate C.D.F. the deviation of the Kolmogorov-Smirnov statistic is no longer distribution free. (See, e.g., Kiefer and Wolfowitz (1958).) Furthermore, in general there is no tractable expression for the distribution of the Kolmogorov-Smirnov statistic. To the present these facts have generally prevented the construction of multivariate analogs of the univariate Kolmogorov-Smirnov tests and confidence bands.

Recent papers of Adler and Brown (1986) and Brown and Rinott (1988) make it possible to construct suitable tests and confidence bands. How this can be done is explained later.

Tables are provided giving the critical values needed for the construction of one-sided and two-sided tests and confidence bands.

The one-sided confidence sets constructed in this way are optimal in the sense of being the smallest sets of the given form having at least the desired probability of containing the true distribution. The two-sided confidence bands constructed in this way are not

¹ Supported by the Air Force Office of Scientific Research AFOSR 87-0298 and the U.S.-Israel Binational Science Foundation 86-00285/1.

² This author was supported in part by NSF DMS 8506847.

³ This author was supported by the Army Research Office through the Mathematical Sciences Institute of Cornell University.

exactly optimal, but are nearly so, in the sense that they are numerically close to being the narrowest constant width bands guaranteed to have at least the desired probability of containing the true distribution.

Tables are given which help describe how far these two sided bands are from being optimal and also give some information about other characteristics of the suggested tests and confidence procedures. These tables also include simulation results about the distribution of the one-sided and two-sided Komogorov-Smirnov statistics from the uniform distribution on the unit square. Some simulation results about this distribution were obtained earlier by Pyke and Wilbour (unpublished). Theoretical bounds (upper and lower) for this distribution have also been previously established. (See references in Adler and Brown (1986).)

Another method yielding tests and confidence sets for bivariate distributions has been proposed by Beran and Millar (1986). Their method, based on a bootstrap argument, is very different from ours, and produces quite different tests and confidence sets. Their procedure has the possible advantage of being rotation invariant and thus may perhaps produce tests and confidence sets which are more appealing in a number of potential applications. Our method is easier to apply and yields tests and confidence sets of an intuitive and prespecified form, which can be easily visualized and conveniently described in terms of the cumulative distribution function.

2. SETTING.

For $x = (x_1, x_2) \in \mathbb{R}^2$ and $y = (y_1, y_2) \in \mathbb{R}^2$ write $x \leq y$ to mean $x_1 \leq y_1, x_2 \leq y_2$. The cumulative distribution function of a random variable X , taking values in \mathbb{R}^2 is defined by $F(x) = P(X \leq x)$. The positive (resp., negative) Kolmogorov-Smirnov distance between two C.D.F.'s, F and G , on \mathbb{R}^2 is defined by

$$|F - G|^+ = \sup_{x \in \mathbb{R}^2} (F(x) - G(x)) \quad (\text{resp.}, |F - G|^- = \sup_{x \in \mathbb{R}^2} (G(x) - F(x))).$$

The ordinary Kolmogorov-Smirnov distance is

$$|F - G| = \max (|F - G|^+, |F - G|^-) = \sup_{x \in \mathbb{R}^2} |F(x) - G(x)|.$$

Let X_1, \dots, X_n be a random sample from some bivariate population. The empirical C.D.F. is defined as

$$\hat{F}_n(t) = n^{-1} \sum_{i=1}^n \chi_{\{t \geq X_i\}}(t) = n^{-1} \cdot \#\{i : X_i \leq t\}$$

where $\#$ denotes "cardinality of" and χ_S denotes the indicator function of S .

Tests will be constructed of the null hypothesis $H_0 : F = F_0$ where F_0 is a specified C.D.F. on \mathbb{R}^2 . The alternative may be either one sided, of the form $H_{1+} : F \geq F_0 (F \neq$

F_0) or of the form $H_{1-} : F \leq F_0 (F \neq F_0)$, or two sided as defined by $H_1 : F \neq F_0$. The two sided alternative is of course much more commonly of interest than the one sided one. The one sided alternative has been introduced here in part to facilitate presentation of the theory in Section 4.

Confidence bands are constructed by inverting the tests of H_0 versus H_1 . Also constructed are one sided confidence sets based on inversion of the tests of H_0 versus H_{1+} or of H_0 versus H_{1-} . These are of theoretical interest and may occasionally also be of interest in practice.

3. THE PROCEDURES.

(3.1) For testing $H_0 : F = F_0$ versus the one sided alternative H_{1+} (resp., H_{1-}) at level α : Reject H_0 when $\sqrt{n} |\hat{F}_n - F_0|^+ \geq C_n^+(\alpha)$ (resp., $\sqrt{n} |\hat{F}_n - F_0|^- \geq C_n^-(\alpha)$) where Table 1 gives values of $C_n^\pm(\alpha)$ for selected values of α, n . The values in this table are evaluated numerically (via simulation) based on the theory described in the next section. Various considerations indicate that this numerical determination is generally accurate up to a possible relative error of $\pm 1\%$.

(3.2) For testing $H_0 : F = F_0$ versus the two sided alternative $H_1 : F \neq F_0$ at level α : Reject H_0 when $\sqrt{n} |\hat{F}_n - F_0| \geq C_n(\alpha)$. Values of $C_n(\alpha)$ are given in Table 2. These values also appear to be generally accurate to within $\pm 1\%$.

Confidence sets and bands are determined by inverting these tests:

(3.3) A lower (resp., upper) $100(1 - \alpha)\%$ confidence set for F is a set of the form

$$\{F : F(t) \leq \hat{F}_n(t) + C_n^+(\alpha)/\sqrt{n} \forall t\}$$

$$\text{(resp., } \{F : F(t) \geq \hat{F}_n(t) - C_n^-(\alpha)/\sqrt{n} \forall t\} \text{)}.$$

(3.4) A $100(1 - \alpha)\%$ confidence band for F is a set of the form

$$\{F : \hat{F}_n(t) - C_n(\alpha)/\sqrt{n} \leq F(t) \leq \hat{F}_n(t) + C_n(\alpha)/\sqrt{n} \forall t\}.$$

Table 1: Critical Values, $C_n^\pm(\alpha)$, for One Sided Tests

100α	$C_n^\pm(\alpha)$	$n = 10$	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 500$
10.0%	$C_n^+(\alpha)$	1.453	1.474	1.485	1.494	1.501	1.507
	$C_n^-(\alpha)$	1.365	1.409	1.430	1.455	1.472	1.500
5.0%	$C_n^+(\alpha)$	1.576	1.603	1.615	1.625	1.634	1.640
	$C_n^-(\alpha)$	1.488	1.538	1.558	1.586	1.604	1.635
2.5%	$C_n^+(\alpha)$	1.685	1.720	1.731	1.744	1.755	1.752
	$C_n^-(\alpha)$	1.599	1.656	1.677	1.707	1.723	1.753
1.0%	$C_n^+(\alpha)$	1.815	1.855	1.875	1.884	1.898	1.904
	$C_n^-(\alpha)$	1.721	1.791	1.813	1.846	1.865	1.911
0.5%	$C_n^+(\alpha)$	1.903	1.950	1.976	1.978	2.001	2.004
	$C_n^-(\alpha)$	1.809	1.885	1.907	1.943	1.970	2.023

Table 2: Critical Values, $C_n(\alpha)$, for Two Sided Tests

100α	$n = 10$	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 500$
10.0%	1.536	1.574	1.588	1.607	1.620	1.638
5.0%	1.646	1.690	1.705	1.727	1.739	1.752
2.0%	1.776	1.826	1.845	1.865	1.882	1.906
1.0%	1.864	1.919	1.942	1.961	1.984	2.012

4. THEORY

Let G denote the bivariate C.D.F. corresponding to the distribution which is uniform over the negative diagonal of the unit square. Formally,

$$(4.1) \quad G(x_1, x_2) = (x_1 + x_2 - 1)^+, \quad 0 \leq x_1, x_2 \leq 1.$$

Let F be any other bivariate C.D.F. It is shown in Brown and Rinott (1988, Section 5) that for any $0 \leq c \leq 1$

$$(4.2) \quad P_G(\sqrt{n} |\hat{G}_n - G|^\pm > c) \geq P_F(\sqrt{n} |\hat{F}_n - F|^\pm > c).$$

(The validity of this statement in the limit as $n \rightarrow \infty$ was earlier established in Adler and Brown (1986, Section 3).) Consequently

$$(4.3) \quad \begin{aligned} & P_F(\sqrt{n} |\hat{F}_n - F| > c) \\ & \leq P_F(\sqrt{n} |\hat{F}_n - F|^+ > c) + P_F(\sqrt{n} |F_n - F|^- > c) \\ & \leq P_G(\sqrt{n} |\hat{G}_n - G|^+ > c) + P_G(\sqrt{n} |G_n - G|^- > c) \end{aligned}$$

In order to construct Table 1 values of $P_G(\sqrt{n} |\hat{G}_n - G|^\pm > c)$ were found numerically by repeated simulation. The entries are the numerically obtained values $C_n^\pm(\alpha)$ for which

$$(4.4) \quad P_G(\sqrt{n} |\hat{G}_n - G|^\pm > C_n^\pm(\alpha)) = \alpha.$$

The entries for $n = 5, \dots, 100$ (resp., $n = 500$) in the table are the result of 100,000 (resp., 20,000) simulations of samples of size n . The accuracy of these simulations was checked by comparing for some entries a first group of simulations against a second group. Nearly all numbers agreed to within 1%.

Furthermore, one of us had earlier performed a similar simulation of 100,000 repetitions to calculate $C_n^+(\alpha)$ for all n as above and $\alpha \geq .01$. Those results agreed to within 1% (and generally closer) with all numbers in Table 1. (The probabilities on which these numbers were based may have a somewhat larger percentagewise simulation error. See the end of Remark (5.4).)

The entries in Table 2 are values $C_n(\alpha)$ from the preceding simulation experiment for which

$$(4.5) \quad P_G(\sqrt{n} |\hat{G}_n - G|^+ > C_n(\alpha)) + P_G(\sqrt{n} |\hat{G}_n - G|^- > C_n(\alpha)) = \alpha.$$

According to (4.1) and (4.3)

$$(4.6) \quad P_F(\sqrt{n} |\hat{F}_n - F|^\pm > C_n^\pm(\alpha)) \leq \alpha$$

with equality if $F = G$. (If the marginals of F are uniform on $(0,1)$ then equality holds only if $F = G$.) Hence the one-sided tests proposed in (3.1) are conservative, as are the confidence sets proposed in (3.3). Because there is equality in (4.6) when $F = G$, those sets are the smallest confidence sets of the form $\{F : F(\cdot) \leq \hat{F}_n(\cdot) + c\}$ and $\{F : F(\cdot) \geq \hat{F}_n(\cdot) - c\}$, respectively, having the prescribed confidence coefficient.

According to (4.2) and (4.4)

$$(4.7) \quad P_F(\sqrt{n} |\hat{F}_n - F| > C_n(\alpha)) \leq \alpha.$$

(Actually, strict inequality always holds in (4.7) because the second inequality in (4.3) is strict whenever F has uniform marginals (w.l.o.g.) and $F \neq G$; and when $F = G$ the first inequality is strict because $P_G(\sqrt{n} |\hat{G}_n - G|^+ > c$ and $\sqrt{n} |\hat{G}_n - G|^- > c) > 0$ for $c < \sqrt{n}$.) Hence the two sided tests of (3.2) and the confidence bands of (3.4) are conservative. But it is not the case here that the two sided bands are the smallest possible of the given form since actually

$$(4.8) \quad \sup_F P_F(\sqrt{n} |\hat{F}_n - F| > C_n(\alpha)) < \alpha.$$

as indicated above. (It can be seen by reference to (4.3) that as $\alpha \rightarrow 0$ for fixed n the ratio of the two sides of (4.6) converges to 1, so the extra physical size of the proposed bands becomes relatively negligible.)

5. FURTHER NUMERICAL PROPERTIES OF THE PROCEDURES.

Table 3 gives values for $D_n(\alpha)$ determined by simulation to satisfy

$$(5.1) \quad P_G(\sqrt{n} |\hat{G}_n - G| > D_n(\alpha)) = \alpha.$$

Note that $D_n(\alpha) < C_n(\alpha)$, as predicted by (4.8). These values could be used to test $H_0 : F = G$ versus $H_1 : F \neq G$. They could also be used to construct confidence bands with nominal confidence coefficient $100(1 - \alpha)\%$. However, there is no guarantee that these bands would be conservative in the sense that

(I) $P_F(\sqrt{n} |\hat{F}_n - F| > D_n(\alpha)) \leq \alpha$ for all F and for all α . In fact, we conjecture they are not conservative, but we suspect that the desired inequality (i) fails by at most a numerically very small amount, and only for special choices of F .

The actual size at $F_0 = G$ of the two-sided test proposed in (3.2) is $P_G(\sqrt{n} |\hat{G}_n - G| > C_n(\alpha))$. Selected values of this quantity appear in Table 4. This also gives another indication of the extent of inequality in (4.4).

Let U denote the uniform distribution on the unit square. Let F by any other continuous bivariate distribution under which the coordinates X_1 and X_2 are independent. Then the distribution under U of $|\hat{U}_n - U|$ (and of $|\hat{U}_n - U|^\pm$, resp.) is the same as that under F of $|\hat{F}_n - F|$ ($|\hat{F}_n - F|^\pm$). Consequently it is of special interest to examine the performance of the tests suggested in Section 3 as tests of $H_0 : F = U$.

In order to do so we simulated the required distributions. The results for $n = 10, \dots, 100$ summarize 20,000 replications; those for $n = 500$ summarize only $n = 10,000$. (This smaller value was chosen because of constraints on computer time. In consequence the probabilities reported for $n = 500$ and $\alpha = .01$ have a standard deviation approximately 10% of their value.)

Table 5 gives values of $E_n^+(\alpha)$ and $E_n^-(\alpha)$ numerically satisfying

$$(5.2) \quad P_U(|\hat{U}_n - U|^\pm > E_n^\pm(\alpha)) = \alpha.$$

These could be used to construct one-sided exactly level α tests of $H_0 : F = U$. Pyke and Wilbour (unpublished) earlier obtained extensive tables of $E_n^+(\alpha)$ based on 10,000 repetitions for $n = 5, 10, 20, 50, 100$, which they have kindly shown us. With two exceptions their values and ours are within 1% of each other. (They have $E_{20}^+(\alpha) = 1.825$, and $E_{50}^+(\alpha) = 1.800$.)

Table 6 gives values of $E_n(\alpha)$ numerically satisfying

$$(5.3) \quad P_U(|\hat{U}_n - U| > E_n(\alpha)) = \alpha.$$

These could be used to construct tests of $H_0 : F = U$ versus $H_0 : F \neq U$; or, as noted, tests of $H_0 : F = F_0$ versus $H_0 : F \neq F_0$ for any F_0 having independent marginals.

Tables 7 and 8 provide information about the true size when $F_0 = U$ of the omnibus tests proposed in Section 3. Table 7 gives the size (= power) of the one-sided tests (3.1) when used for testing $H_0 : F = U$. The entries of this table consequently describe the

probability when the true C.D.F. is U that it is in the confidence sets of (3.3). Table 8 provides the analogous information for the two sided test of (3.2) and the confidence bands of (3.4).

It is, of course, of interest to compare these true sizes to the nominal size (= level of the test). It is also of interest to compare them to the true sizes when the true C.D.F. is G , as given in Tables 3 and 4. For example, Tables 4 and 8 can be roughly summarized as follows: When $F_O = G$ the two-sided test has true power which is usually from 80 to 90% of its nominal power (α). The true power is further reduced when $F_O = U$. This reduction is as much as another 40% for small α and n but is a much smaller percent for larger n or larger α .

(5.4) (Remark) Kac, Kiefer, and Wolfowitz (1955, equation (4.6)) give the following expression:

$$(5.5) \quad \lim_{n \rightarrow \infty} P_G(\sqrt{n}|\hat{G}_n - G| > c) = \sum_{n=1}^{\infty} (8n^2c^2 - 2)e^{-2n^2c^2}.$$

The values $c = C_{500}(\alpha)$ were substituted in (5.5) and the result was entered in the column headed $n = \infty$ of Table 4. The discrepancy between the entries for $n = 500$ and $n = \infty$ is much larger than one would normally expect to be due to simulation error from a simulation involving 20,000 replications of events having probability α . However, perhaps this discrepancy is not surprising in view of the fact that $\sqrt{n}(\hat{G}_n - G_n)$ has jumps of size $1/\sqrt{n}$ whereas the limiting process on which (5.5) is based is a continuous process. (Note that $1/\sqrt{500} = .045$. One might therefore expect $P_G(\sqrt{500}|\hat{G}_n - G| > c) \approx \lim_{n \rightarrow \infty} P_G(\sqrt{n}|\hat{G}_n - G| > c + (.045/2))$. The corresponding values of this for $c = C_{500}(\alpha)$ and $\alpha = .10, .05, .02, .01$ are $100P = 8.085, 4.297, 1.645, .849$. Two of these four numbers are now noticeably less than the corresponding entries for $n = 500$! This change in the relative size of the figures also demonstrates how sensitive the probabilities are to small changes in the constant c .)

More extensive tables of probabilities from which the above were extracted may be found in Adler, Brown, and Lu (1988).

Table 3: Values of $D_n(\alpha)$ Defined in (5.1)

100α	$n = 10$	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 500$
10.0%	1.507	1.541	1.556	1.571	1.584	1.604
5.0%	1.625	1.663	1.682	1.699	1.711	1.727
2.0%	1.761	1.809	1.824	1.843	1.860	1.884
1.0%	1.854	1.906	1.929	1.944	1.967	1.991

Table 4: Actual Size (in %) of the Two Sided Test (3.2)
When $F_o = G$. (See (5.5) for $n = \infty$)

100α	$n = 10$	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 500$	$n = \infty$
10.0%	8.480	8.397	8.411	8.296	8.250	8.410	9.094
5.0%	4.323	4.282	4.287	4.252	4.242	4.295	4.894
2.0%	1.788	1.747	1.762	1.739	1.743	1.730	1.904
1.0%	0.910	0.892	0.900	0.874	0.868	0.855	0.993

Table 5: Critical Values, $E_n^\pm(\alpha)$, for One Sided Tests
for $H_0 : F = U$

100α	$E_n^\pm(\alpha)$	$n = 10$	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 500$
10.0%	$E_n^+(\alpha)$	1.354	1.381	1.389	1.405	1.424	1.434
	$E_n^-(\alpha)$	1.301	1.344	1.377	1.388	1.395	1.422
5.0%	$E_n^+(\alpha)$	1.488	1.524	1.530	1.546	1.564	1.586
	$E_n^-(\alpha)$	1.443	1.476	1.520	1.536	1.538	1.566
2.5%	$E_n^+(\alpha)$	1.602	1.643	1.653	1.689	1.688	1.708
	$E_n^-(\alpha)$	1.555	1.597	1.640	1.659	1.673	1.698
2.0%	$E_n^+(\alpha)$	1.634	1.676	1.695	1.730	1.724	1.747
	$E_n^-(\alpha)$	1.598	1.634	1.681	1.701	1.710	1.734
1.0%	$E_n^+(\alpha)$	1.725	1.778	1.805	1.834	1.834	1.855
	$E_n^-(\alpha)$	1.695	1.734	1.781	1.803	1.822	1.828

Table 6: Critical Values E_n^\pm for Two Sided Test
for $H_0 : F = U$

100α	$n = 10$	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 500$
10.0%	1.467	1.499	1.524	1.541	1.552	1.573
5.0%	1.582	1.624	1.647	1.674	1.679	1.703
2.0%	1.710	1.757	1.795	1.817	1.827	1.848
1.0%	1.798	1.865	1.891	1.919	1.924	1.943

Table 7: Actual Size (in %) of the One Sided Test (3.1)
When $F_0 = U$

100α	$H_n^\pm(\alpha)$	$n = 10$	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 500$
10.0%	$H_n^+(\alpha)$	6.000	6.345	6.280	6.635	6.820	7.020
	$H_n^-(\alpha)$	7.395	7.235	7.680	7.370	7.025	6.690
5.0%	$H_n^+(\alpha)$	2.935	3.145	3.185	3.360	3.350	3.710
	$H_n^-(\alpha)$	3.900	3.550	3.960	3.745	3.600	3.360
2.0%	$H_n^+(\alpha)$	1.030	1.135	1.265	1.515	1.310	1.630
	$H_n^-(\alpha)$	1.680	1.310	1.575	1.535	1.505	1.260
1.0%	$H_n^+(\alpha)$	0.500	0.660	0.640	0.715	0.655	0.660
	$H_n^-(\alpha)$	0.795	0.670	0.810	0.685	0.770	0.570

Table 8: Actual Size (in %) of the Two Sided Test (3.2)
When $F_0 = U$

100α	$n = 10$	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 500$
10.0%	6.615	6.505	6.955	7.035	6.975	7.110
5.0%	3.235	3.135	3.530	3.675	3.475	3.600
2.0%	1.175	1.275	1.390	1.440	1.385	1.240
1.0%	0.590	0.675	0.655	0.740	0.635	0.700

ACKNOWLEDGEMENT

The computations reported above were performed using the Cornell National Supercomputer Facility which receives major funding from the National Science Foundation and the IBM Corporation.

6. REFERENCES

- Adler, R.J. and Brown, L.D. (1986). Tail behavior for suprema of empirical processes. *Ann. Prob.* 14, 1-30.
- Adler, R.J., Brown, L.D., and Lu, K-L. (1988). Tables to accompany "Tests and Confidence Bands for Bivariate Cumulative Distribution Functions". Preprint (available from Cornell Mathematical Sciences Institute or from Cornell Statistics Center).
- Beran, R. and Millar, P.W. (1986). Confidence sets for a multivariate distribution. *Ann. Statist.* 14, 431-443.
- Brown, L.D. and Rinott, Y. (1988). Inequalities for multivariate infinitely divisible processes, to appear in *Ann. Prob.*
- Kac, M., Kiefer, J.C. and Wolfowitz, J. (1955). On tests of normality and other tests of goodness of fit based on distance methods. *Ann. Math. Statist.* 26, 189-211.
- Kiefer, J.C. and Wolfowitz, J. (1958). On the deviations of the empiric distribution function of vector chance variables. *Trans. Amer. Math. Soc.* 87, 173-186.