

# The Power of Alternative Kolmogorov-Smirnov Tests Based on Transformations of the Data

Song-Hee Kim, Columbia University  
Ward Whitt, Columbia University

The Kolmogorov-Smirnov (KS) statistical test is commonly used to determine if data can be regarded as a sample from a sequence of i.i.d. random variables with specified continuous cdf  $F$ , but with small samples it can have insufficient power, i.e., its probability of rejecting natural alternatives can be too low. However, Durbin [1961] showed that the power of the KS test often can be increased, for given significance level, by a well-chosen transformation of the data. Simulation experiments reported here show that the power can often be more consistently and substantially increased by modifying the original Durbin transformation by first transforming the given sequence to a sequence of mean-1 exponential random variables, which is equivalent to a rate-1 Poisson process, and then applying the classical conditional-uniform transformation to convert the arrival times into the order statistics of i.i.d. uniform random variables. The new KS test often has much more power, because it focuses on the cumulative sums rather than the random variables themselves.

Categories and Subject Descriptors: I.6.5 [Simulation and Modeling]: Model Development

Additional Key Words and Phrases: Hypothesis tests, Kolmogorov-Smirnov statistical test, power, data transformations.

## ACM Reference Format:

Song-Hee Kim and Ward Whitt, 2013. The Power of Alternative Kolmogorov-Smirnov Tests Based on Transformations of the Data. *ACM Trans. Model. Comput. Simul.* V, N, Article A (January YYYY), 18 pages. DOI: <http://dx.doi.org/10.1145/0000000.0000000>

## 1. INTRODUCTION

The Kolmogorov-Smirnov (KS) statistical test is commonly used to determine if data can be regarded as a sample from a sequence of independent and identically distributed (i.i.d.) random variables  $\{X_n : n \geq 1\}$ , each distributed as a random variable  $X$  with a specified continuous cumulative distribution function (cdf)  $F(x) \equiv P(X \leq x)$ ,  $x \in \mathbb{R}$ . The test is based on the maximum difference between the empirical cdf (ecdf)

$$F_n(x) \equiv \frac{1}{n} \sum_{k=1}^n 1_{\{X_k \leq x\}}, \quad x \in \mathbb{R}, \quad (1)$$

and the underlying cdf  $F$ , where  $1_A$  is an indicator function, equal to 1 if the event  $A$  occurs, and equal to 0 otherwise, i.e.,

$$D_n \equiv \sup_x \{|F_n(x) - F(x)|\}, \quad (2)$$

This work is supported by the U.S. National Science Foundation grant CMMI 1066372 and the Samsung Foundation.

Author's addresses: Song-Hee Kim and Ward Whitt, Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or [permissions@acm.org](mailto:permissions@acm.org).

© YYYY ACM 1049-3301/YYYY/01-ARTA \$15.00

DOI: <http://dx.doi.org/10.1145/0000000.0000000>

which has a distribution that is independent of the cdf  $F$ , provided that the cdf is continuous. For any observed maximum  $y$  from a sample of size  $n$ , we compute the  $P$ -value  $P(D_n > y)$ , e.g., by using the Matlab program *ksstat*, and compare it to the significance level  $\alpha$ , i.e., for specified probability of rejecting the null hypothesis when it is in fact correct (type I error), which we take to be  $\alpha = 0.05$ . Sometimes it is preferable to use corresponding one-sided KS tests, but we will concentrate on the two-sided test. See [Simard and L'Ecuyer \[2011\]](#) and [Shorack and Wellner \[2009\]](#) for additional background and references on the KS test.

Alternative KS tests can be obtained by considering various transformations of the data, based on transformations of the hypothesized sequence of i.i.d. random variables  $\{X_n : n \geq 1\}$  with continuous cdf  $F$  into a new sequence of i.i.d. random variables  $\{Y_n : n \geq 1\}$  with continuous cdf  $G$ , while keeping the significance level  $\alpha$  unchanged. Since the KS test applies in both settings, we should prefer the new test based on the transformed data if it has substantially greater statistical power for contemplated alternatives, i.e., if it has higher probability of rejecting the null hypothesis when the null hypothesis is false. Specifically, for specified significance criterion  $\alpha$ , the power of a specified alternative is the probability  $1 - \beta$ , where  $\beta \equiv \beta(\alpha)$  is the probability of incorrectly accepting the null hypothesis (type II error) when it is false (which of course depends on the alternative as well as  $\alpha$ ).

[Durbin \[1961\]](#) suggested transforming the data to increase the power of the KS test and proposed a specific transformation for that purpose. In this paper we study the issue further. We conclude that a good data transformation can indeed significantly increase the power of the KS test, but that a modification of the [Durbin \[1961\]](#) transformation, proposed for testing a Poisson process by [Lewis \[1965\]](#), consistently has even more power.

### 1.1. Motivation: Arrival Processes in Service Systems

Our research was originally motivated by the desire to fit stochastic queueing models to data from large-scale service systems, such as telephone call centers and hospital emergency rooms, as discussed in [Brown et al. \[2005\]](#) and [Armony et al. \[2011\]](#). Since the arrival rate typically varies strongly by time of day in these service systems, the natural arrival process model is a nonhomogeneous Poisson process (NHPP) instead of a homogeneous Poisson process. Nevertheless, [Brown et al. \[2005\]](#) showed that the KS test can still be applied, provided that we transform the data.

Since the arrival rate in a service system typically changes relatively slowly compared to the overall arrival rate, it is often reasonable to assume that the arrival rate is piecewise-constant. A piecewise-constant NHPP can be regarded as a Poisson process over each subinterval. Given a Poisson process on any one subinterval, and conditional on the total number of arrivals in that interval, the arrival times divided by the length of that interval are distributed as the order statistics of i.i.d. random variables uniformly distributed on  $[0, 1]$ ; e.g., see §2.3 of [Ross \[1996\]](#). With that classical conditional-uniform (CU) approach, the data from all the subintervals can be combined to obtain a single sequence of i.i.d. random variables uniformly distributed on  $[0, 1]$ , to which the KS test can be applied directly. Moreover, the CU method eliminates the nuisance parameter; the method is independent of the rate of the PP. [Brown et al. \[2005\]](#) did not stop with the CU KS test, but instead proposed a (scaled) logarithmic transformation into a single sequence of i.i.d. exponential random variables for the KS test.

We wondered about the power of the passed-over CU KS test and the chosen logarithmic (Log) KS test of a NHPP. Thus, we conducted simulation experiments to study the power of these KS tests and various alternatives, and we reported the results in [Kim and Whitt \[2013c\]](#). Consistent with [Brown et al. \[2005\]](#), we found that the CU KS

test of a Poisson has remarkably little power, while the Log KS test has much greater power. We also found that there is a substantial history in the statistical literature. First, [Lewis \[1965\]](#) made a significant contribution for testing a Poisson process, recognizing that the [Durbin \[1961\]](#) transformation could be effectively applied after the CU transformation. Second, from [Lewis \[1965\]](#) we discovered that the direct CU KS test of a Poisson process was evidently first proposed by [Barnard \[1953\]](#); and [Lewis \[1965\]](#) showed that it had little power.

Upon discovering [Lewis \[1965\]](#), we first supposed that the Log KS test of [Brown et al. \[2005\]](#) would turn out to be equivalent to the [Lewis \[1965\]](#) transformation and that the KS test proposed by [Lewis \[1965\]](#), drawing upon [Durbin \[1961\]](#), would coincide with the KS test given in [Durbin \[1961\]](#), but neither is the case. Thus, this past work suggests several different KS tests. In [Kim and Whitt \[2013c\]](#) we concluded that the Lewis test of a Poisson process has the most power against stationary point processes having non-exponential interarrival distributions, providing a significant improvement over the Log KS test.

## 1.2. Standard KS Tests for i.i.d. Sequences with cdf $F$

Even though we were originally interested in tests of a Poisson process, because they yield tests of a piecewise-constant NHPP, the KS tests used to test a Poisson process can be also applied to test whether  $n$  observations can be regarded as a sample of size  $n$  from an i.i.d. sequence with arbitrary specified continuous cdf  $F$ . Such a KS test evidently has not been considered before.

Moreover, these new KS tests are also directly applicable to service systems, because the standard model for the service times is an i.i.d. sequence. The most convenient cdf for analysis is the exponential cdf, but data analysis often suggests a lognormal cdf instead, as in [Brown et al. \[2005\]](#). The new KS tests can be used to test these alternatives.

Just as it is common (and done by [Durbin \[1961\]](#)) to transform an initial sequence  $\{X_n : n \geq 1\}$  of i.i.d. random variables with cdf  $F$  into a sequence  $\{U_n : n \geq 1\}$  of i.i.d. random variables uniformly distributed on  $[0, 1]$  by letting  $U_n \equiv F(X_n)$ ,  $n \geq 1$ , so we can also transform the initial sequence into a sequence  $\{Y_n : n \geq 1\}$  of i.i.d. exponential random variables with mean 1 by letting  $Y_n \equiv -\log\{1 - F(X_n)\}$ ,  $n \geq 1$ . It is well known that the value of the KS statistic in (2) is unchanged by these transformations, provided of course that we use both the new ecdf and the new cdf in each case.

For applying the associated KS tests of a Poisson process, the key observation is that the sequence of partial sums  $\{T_n : n \geq 1\}$ , where  $T_n \equiv Y_1 + \dots + Y_n$ ,  $n \geq 1$ , constitute the arrival times of a rate-1 Poisson process. Moreover, for a fixed sample of size  $n$ , we can use a variant of the CU transformation, stating that the  $n-1$  random variables  $T_k/T_n$ ,  $1 \leq k \leq n-1$ , are distributed as the order statistics of  $n-1$  i.i.d. random variables uniformly distributed on  $[0, 1]$ . Thus, we can perform a new KS test based on the KS statistic in (2) with the new ecdf

$$F_n^{(CU)}(x) \equiv \frac{1}{n-1} \sum_{k=1}^{n-1} 1_{\{(T_k/T_n) \leq x\}}, \quad 0 \leq x \leq 1, \quad (3)$$

and the underlying uniform cdf  $F(x) \equiv x$ ,  $0 \leq x \leq 1$ ; that is the CU test in the new Poisson process context. Alternatively, instead of the CU test based on the new ecdf in (3), we can use the associated Log or Lewis tests considered in [Kim and Whitt \[2013c\]](#).

To understand all these transformations, it is good to start with  $n$  i.i.d. random variables  $U_k$ ,  $1 \leq k \leq n$ , each uniformly distributed on  $[0, 1]$ , obtained by letting  $U_k \equiv F(X_k)$ . The direct applications of the Log and Durbin transformations apply by *sorting* these uniform random variables, which is equivalent to sorting the original

$n$  observations. Accordingly, we call these the sort-Log KS test and the sort-Durbin KS test. Since the sort-Durbin test coincides with the original Durbin [1961] test, we simply call it the Durbin test, but the Log transformation used by Brown et al. [2005] was only proposed *after* the CU transformation.

In contrast, the alternative KS tests based on first transforming to  $n$  i.i.d. mean-1 exponential random variables  $Y_k$  by letting  $Y_k \equiv -\log \{1 - F(X_k)\}$  and then applying the CU transformation applies by first considering the *partial sums* of the random variables. We thus have three KS tests based on the CU transformation applied to the exponential variables: (i) Exp+CU, the CU transformation alone as in Barnard [1953], (ii) Exp+CU+Log, the CU transformation plus the Log transformation as in Brown et al. [2005], and (iii) Exp+CU+Durbin, the CU transformation plus the Durbin [1961] transformation, as in Lewis [1965]. Since the Exp+CU+Durbin test was not proposed by Durbin [1961], but coincides with the Lewis [1965] test (even though the setting is new), we call the Exp+CU+Durbin test the Lewis test. In this new setting, we again find that the Lewis [1965] test consistently has the highest power against alternatives with different marginal distributions. Thus, we conclude that the Lewis [1965] test has wider applicability than to just the Poisson process.

### 1.3. Organization

We now indicate how the rest of the paper is organized. We start in §2 by carefully defining the six different KS tests. Next in §3 we describe our first simulation experiment, which is a fixed-sample-size discrete-time stationary-sequence analog of the fixed-interval-length continuous-time stationary point process experiment, aimed at studying tests of a Poisson process, conducted in Kim and Whitt [2013c]. In addition to the natural null hypothesis of i.i.d. exponential random variables, we also consider i.i.d. non-exponential sequences with Erlang, hyperexponential and lognormal marginal cdf's. We report the results in §4, which surprisingly show that the original Durbin [1961] method performs poorly, but the new version of the Lewis [1965] test performs well, providing increased power. However, Durbin [1961] considered different examples. Motivated by the good results found for a standard normal null hypothesis by Durbin [1961], in §5 we consider a second experiment to test for a sequence of i.i.d. standard normal random variables. Consistent with Durbin [1961], we find that the original Durbin [1961] method performs much better for the standard normal null hypothesis, but again the new version of the Lewis [1965] test also performs well. We draw conclusions in §6. Additional information appears in appendices, Kim and Whitt [2013a; 2013b].

## 2. THE ALTERNATIVE KS TESTS

We consider the following six KS tests to determine whether  $n$  observations  $X_k$ ,  $1 \leq k \leq n$ , can be considered a sample from a sequence of i.i.d. random variables having a continuous cdf  $F$ . We start by forming the associated variables  $U_k \equiv F(X_k)$ , which are i.i.d. uniform variables on  $[0, 1]$  under the null hypothesis.

*Standard Test..* We use the standard KS test to test whether  $U_k \equiv F(X_k)$ ,  $1 \leq k \leq n$ , can be considered to be i.i.d. random variables uniformly distributed on  $[0, 1]$ .

*Sort-Log Test..* Starting with the  $n$  random variables  $U_k$ ,  $1 \leq k \leq n$ , in the standard test, let  $U_{(j)}$  be the  $j^{\text{th}}$  smallest of these, so that  $U_{(1)} < \dots < U_{(n)}$ . As in §3.1 of Brown et al. [2005], we use the fact that under the null hypothesis

$$Y_j^{(L)} \equiv -\log_e (U_j/U_{j+1}), \quad 1 \leq j \leq n-1,$$

are  $n-1$  i.i.d. rate-1 exponential random variables, to which we can apply the KS test with  $n$  replaced by  $n-1$ .

*Durbin ( $\equiv$  Sort-Durbin) Test..* This is the original test proposed by [Durbin \[1961\]](#), which also starts with  $U_k \equiv F(X_k)$  and  $U_{(k)}$  with  $U_{(1)} < \dots < U_{(n)}$  as above. In this context, look at the successive *intervals* between these ordered observations:

$$C_1 \equiv U_{(1)}, C_j \equiv U_{(j)} - U_{(j-1)}, \quad 2 \leq j \leq n, \quad \text{and} \quad C_{n+1} \equiv 1 - U_{(n)}.$$

Then let  $C_{(j)}$  be the  $j^{\text{th}}$  smallest of these intervals,  $1 \leq j \leq n$ , so that  $0 < C_{(1)} < \dots < C_{(n+1)} < 1$ . Now let  $Z_j$  be scaled versions of the intervals between these new ordered intervals, i.e., let

$$Z_j = (n+2-j)(C_{(j)} - C_{(j-1)}), \quad 1 \leq j \leq n+1, \quad (\text{with } C_{(0)} \equiv 0). \quad (4)$$

Remarkably, [Durbin \[1961\]](#) showed (by a simple direct argument giving explicit expressions for the joint density functions, exploiting the transformation of random vectors by a function) that, under the null hypothesis, the random vector  $(Z_1, \dots, Z_n)$  is distributed the same as the random vector  $(C_1, \dots, C_n)$ . Hence, again under the null hypothesis, the vector of associated partial sums  $(S_1, \dots, S_n)$ , where  $S_k \equiv Z_1 + \dots + Z_k$ ,  $1 \leq k \leq n$ , has the same distribution as the original random vector  $(U_{(1)}, \dots, U_{(n)})$  of ordered uniform random variables. Hence, we can apply the KS test with the ecdf

$$F_n(x) \equiv n^{-1} \sum_{k=1}^n 1_{\{S_k \leq x\}}, \quad 0 \leq x \leq 1,$$

for  $S_k$  above, comparing it to the uniform cdf  $F(x) \equiv x$ ,  $0 \leq x \leq 1$ .

*CU, (Conditional-Uniform  $\equiv$  Exp+CU) Test..* We start with  $Y_k \equiv -\log \{1 - F(X_k)\}$ ,  $1 \leq k \leq n$ , which are i.i.d. mean-1 exponential random variables under the null hypothesis. Thus, the cumulative sums  $T_k \equiv Y_1 + \dots + Y_k$ ,  $1 \leq k \leq n$ , are the arrival times of a rate-1 Poisson process. In this context, the conditional-uniform property states that  $T_k/T_n$ ,  $1 \leq k \leq n$ , are distributed as the order statistics of  $n-1$  i.i.d. random variables uniformly distributed on  $[0, 1]$ . Thus we can apply the KS statistic with the ecdf in (3).

*CU+Log (Exp+CU+Log) Test..* We start with the partial sums  $T_k$ ,  $1 \leq k \leq n$ , used in the CU test, which are the arrivals times of a rate-1 Poisson process under the null hypothesis. We again use the conditional-uniform property for fixed sample size to conclude that, under the null hypothesis,  $T_k/T_n$ ,  $1 \leq k \leq n-1$ , are distributed as  $U_{(k)}$ , the order statistics of  $n-1$  random variables uniformly distributed on  $[0, 1]$ , with  $U_{(1)} < \dots < U_{(n-1)}$ . Hence, just as in the Sort-Log test above,

$$Y_j^{(L)} \equiv -\log_e (T_j/T_{j+1}), \quad 1 \leq j \leq n-1,$$

should be  $n-1$  i.i.d. rate-1 exponential random variables, to which we can apply the KS test.

*Lewis (Exp+CU+Durbin) Test..* We again start with the partial sums  $T_k$ ,  $1 \leq k \leq n$ , used in the CU test, which are the arrivals times of a rate-1 Poisson process under the null hypothesis. We again use the conditional-uniform property for fixed sample size to conclude that, under the null hypothesis,  $T_k/T_n$ ,  $1 \leq k \leq n-1$ , are distributed as  $U_{(k)}$ , the order statistics of  $n-1$  random variables uniformly distributed on  $[0, 1]$ , with  $U_{(1)} < \dots < U_{(n-1)}$ . From this point, we apply the [Durbin \[1961\]](#) test above with  $n$  replaced by  $n-1$ , just as [Lewis \[1965\]](#) did in his test of a Poisson process.

### 3. THE FIRST EXPONENTIAL EXPERIMENT

Our first simulation experiment is for the discrete-time analog of the experiment for testing the continuous-time Poisson process in [Kim and Whitt \[2013c\]](#). Our base case



is a sample of size  $n = 200$  i.i.d. mean-1 exponential random variables, but to see the impact of the sample size, we also give results for the larger sample size of  $n = 2000$ .

### 3.1. The Cases Considered

We use the same alternative hypotheses to the continuous-time Poisson process used in [Kim and Whitt \[2013c\]](#), except that replace the time intervals of fixed length  $t$  by sample sizes of fixed size  $n$ . That is, we now consider stationary sequences of mean-1 random variables. There are 9 cases, each with from 1 to 5 subcases, yielding 29 cases in all. Using the same cases as before facilitates comparison. Before we considered the random number of points observed by a rate-1 stationary point process in the fixed interval  $[0, 200]$ . The fixed sample size here  $n \equiv 200$  coincides with the expected sample size before.

The first five cases involve i.i.d. mean-1 random variables; the last four cases involve dependent identically distributed mean-1 random variables. The first i.i.d. case is our null hypothesis with exponential random variables. The other i.i.d. cases have non-exponential random variables. Cases 2 and 3 contain Erlang and hyperexponential random variables, which are, respectively, stochastically less variable and stochastically more variable than the exponential distribution in convex stochastic order, as in §9.5 of [Ross \[1996\]](#). Thus, they have *squared coefficient of variation* (scv, variance divided by the square of the mean, denoted by  $c^2$ ), respectively,  $c^2 < 1$  and  $c^2 > 1$ . Cases 4 and 5 contain non-exponential cdf's with  $c_X^2 = 1$  as well as  $E[X] = 1$ , just like the exponential cdf.

**Case 1, Exponential..** The null hypothesis with i.i.d. mean-1 exponential random variables (Base Case).

**Case 2, Erlang,  $E_k$ ..** Erlang- $k$  ( $E_k$ ) random variables, a sum of  $k$  i.i.d. exponentials for  $k = 2, 4, 6$  with  $c_X^2 \equiv c_k^2 = 1/k$ .

**Case 3, Hyperexponential,  $H_2$ ..** Hyperexponential-2 ( $H_2$ ) random variables, a mixture of 2 exponential cdf's with  $c_X^2 = 1.25, 1.5, 2, 4$  and 10 (five cases). The cdf is  $P(X \leq x) \equiv 1 - p_1 e^{-\lambda_1 x} - p_2 e^{-\lambda_2 x}$ . We further assume balanced means ( $p_1 \lambda_1^{-1} = p_2 \lambda_2^{-1}$ ) as in (3.7) of [Whitt \[1982\]](#) so that given the value of  $c_X^2$ ,  $p_i = [1 \pm \sqrt{(c_X^2 - 1)/(c_X^2 + 1)}]/2$  and  $\lambda_i = 2p_i$ .

**Case 4, mixture with  $c_X^2 = 1$ ..** A mixture of a more variable cdf and a less variable cdf so that the  $c_X^2 = 1$ ;  $P(X = Y) = p = 1 - P(X = Z)$ , where  $Y$  is  $H_2$  with  $c_Y^2 = 4$ ,  $Z$  is  $E_2$  with  $c_Z^2 = 1/2$  and  $p = 1/7$ .

**Case 5, lognormal, LN..** Lognormal ( $LN(1, \sigma^2)$ ) random variables with mean 1 and variance  $\sigma^2$  for  $\sigma^2 = c_X^2 = 0.25, 1.0, 4.0, 10.0$  (four cases).

Cases 6 and 7 are dependent stationary sequences that deviate from the null hypothesis (Case 1) only through dependence among successive variables, each exponentially distributed with mean 1:

**Case 6, RRI, dependent exponential interarrival times..** Randomly Repeated Interarrival (RRI) times with exponential interarrival times, constructed by letting each successive interarrival time be a mixture of the previous interarrival time with probability  $p$  or a new independent interarrival time from an exponential distribution with mean 1, with probability  $1 - p$  (a special case of a first-order discrete autoregressive process, DAR(1), studied by [Jacobs and Lewis \[1978; 1983\]](#)). Its serial correlation is  $\text{Corr}(X_j, X_{j+k}) = p^k$ . We consider three values of  $p$ : 0.1, 0.5 and 0.9.

**Case 7, EARMA, dependent exponential interarrival times..** A stationary sequence of dependent exponential interarrival times with the correlation structure of an autoregressive-moving average process, called EARMA(1,1) in [Jacobs and](#)

[Lewis \[1977\]](#). Starting from three independent sequences of i.i.d. random variables  $\{X_n : n \geq 0\}$ ,  $\{U_n : n \geq 1\}$ , and  $\{V_n : n \geq 1\}$ , where  $X_n$  is exponentially distributed with mean  $m$ , while

$$P(U_n = 0) = 1 - P(U_n = 1) = \beta \quad \text{and} \quad P(V_n = 0) = 1 - P(V_n = 1) = \rho, \quad (5)$$

the EARMA sequence  $\{S_n : n \geq 1\}$  is defined recursively by

$$\begin{aligned} S_n &= \beta X_n + U_n Y_{n-1}, \\ Y_n &= \rho Y_{n-1} + V_n X_n, \quad n \geq 1. \end{aligned} \quad (6)$$

Its serial correlation is  $\text{Corr}(S_j, S_{j+k}) = \gamma \rho^{k-1}$  where  $\gamma = \beta(1 - \beta)(1 - \rho) + (1 - \beta)^2 \rho$ . We consider five cases of  $(\beta, \rho)$ :  $(0.75, 0.50)$ ,  $(0.5, 0.5)$ ,  $(0.5, 0.75)$ ,  $(0.00, 0.75)$ ,  $(0.25, 0.90)$  so that the cumulative correlations  $\sum_{k=1}^{\infty} \text{Corr}(S_j, S_{j+k})$  increase: 0.25, 0.50, 1.00, 3.00, and 5.25. For more details, see [Pang and Whitt \[2012\]](#). We specify these cases by these cumulative correlations.

The final two cases are stationary sequences that have *both* non-exponential marginal distributions and dependence among successive variables:

**Case 8,  $mH_2$ , superposition of  $m$  i.i.d.  $H_2$  renewal processes..** a stationary sequence of interarrival times from a superposition of  $m$  i.i.d. equilibrium renewal processes, where the times between renewals (interarrival times) in each renewal process has a hyperexponential ( $H_2$ ) distribution with  $c_a^2 = 4$  ( $m H_2$ ). As the number  $m$  of component renewal processes increases, the superposition process converges to a PP, and thus looks locally more like a PP, with the interarrival distribution approaching exponential and the lag- $k$  correlations approaching 0, but small correlations extending further across time, so that the superposition process retains an asymptotic variability parameter,  $c_A^2 = 4$ . We consider four values of  $m$ : 2, 5, 10 and 20.

**Case 9, RRI ( $H_2$ ), dependent  $H_2$  interarrival times with  $c^2 = 4$ .** Randomly Repeated Interarrival (RRI) times with  $H_2$  interarrival times, each having mean 1,  $c^2 = 4$  and balanced means (as specified in Case 3). The repetition is done just as in Case 6. We again consider three values of  $p$ : 0.1, 0.5 and 0.9.

Cases 6 and 7 above have short-range dependence, whereas Case 8 for large  $m$  tends to have nearly exponential interarrival times, but longer-range dependence. For small  $m$ , the  $mH_2$  superposition process should behave much like the  $H_2$  renewal process in Case 3 with the component  $c^2 = 4$ ; for large  $m$ , the  $mH_2$  superposition process should behave more like Cases 6 and 7 with dependence and exponential interarrival times.

Since the new KS tests apply to i.i.d. sequences with arbitrary continuous cdf's, we also consider alternative null hypotheses. In particular, here we report results for  $E_2$ ,  $H_2$  (with  $c^2 = 2$ ) and lognormal  $LN(1, 4)$  (with  $c^2 = 4$ ) marginal cdf's having mean 1 as well as the exponential base case.

### 3.2. Simulation Design

For each case, we simulated  $10^4$  replications of  $10^4$  interarrival times. We generate much more data than needed in order to get rid of any initial effects. We are supposing that we observe stationary sequence, which is achieved by having the system operate for some time before collecting data. The initial effect was observed to matter for the cases with dependent interarrival times and relatively small sample sizes.

We use this simulation output to generate sample sizes of a fixed size  $n$ . With fixed sample size  $n = 200$ , in each replication of the  $10^4$  simulated interarrival times we use interarrival times from the  $10^3$ th interarrival time to the  $10^3 + 200$ th interarrival time. To consider large sample sizes, we increased  $n$  from 200 to 2000. We then consider the

Table I. Summary of the performance of alternative KS tests of i.i.d. mean-1 exponential variables for the sample size  $n = 200$  with significance level  $\alpha = 0.05$ : the case of a renewal process with  $H_2$  interarrival times having  $c_X^2 = 2$ .

KS test	Lewis	Standard	CU	Durbin
Power	0.93	0.64	0.28	0.14
Average $p$ value	0.02	0.09	0.24	0.40

interarrival times from the  $10^3$ th interarrival time to the  $10^3 + 2000$ th interarrival time to observe the effect of larger sample size.

For each sample we checked our simulation results by estimating the mean and scv of each interarrival-time cdf both before and after transformations; tables of the results and plots of the average of the ecdf's appear in the appendix, [Kim and Whitt \[2013b\]](#).

#### 4. RESULTS OF THE FIRST EXPERIMENT

The long appendix [Kim and Whitt \[2013b\]](#) contains detailed results of the experiments; we present a summary here. First, we found that the sort-Log and Log tests were consistently dominated by the [Durbin \[1961\]](#) test or the [Lewis \[1965\]](#) test, so we do not present detailed results for those two Log cases here. For the CU, CU+Log and Lewis tests, we considered variants based on the exponential variables  $-\log\{F(X)\}$  and well as  $-\log\{1 - F(X)\}$ , but we did not find great differences, so we do not report those either. Thus, we present the results of four KS tests: (i) the standard test, using the variables  $U_k \equiv F(X_k)$ , (ii) the [Durbin \[1961\]](#) test, (iii) the CU test and (iv) the [Lewis \[1965\]](#) test, as specified in §2. Under the null hypotheses, the cdf in all four cases is uniform on  $[0, 1]$ .

##### 4.1. The Base Case: i.i.d. Mean-1 Exponential Variables

We report the number of KS tests passed out of 10,000 replications as well as the average  $p$ -value with associated 95% confidence intervals. Thus, the estimate of the power is  $1 - (\text{number passed}/10,000)$ . The  $p$ -value is the significance level below which the hypothesis would be rejected. Thus low  $p$ -values indicate greater power. Just as in Table 1 of [Kim and Whitt \[2013c\]](#), the differences in the tests is striking for the middle  $H_2$  alternative with  $c^2 = 2.0$ , as shown in Table I. The results for the Lewis, standard and CU tests are very similar to those for the corresponding KS tests of a Poisson process in Table 1 of [Kim and Whitt \[2013c\]](#), but the results for the [Durbin \[1961\]](#) test are new, and surprisingly bad.

The results for all 29 cases are given in Table II. The first “exponential” case is the i.i.d. exponential null hypothesis. The results show that all tests behave properly for the i.i.d. exponential null hypothesis. The results also show that the tests perform quite differently for the alternative hypotheses. Table II shows that the standard and Lewis tests all perform reasonably well for the i.i.d. cases with non-exponential interarrival-time cdf's, in marked contrast to the CU and Durbin tests. Table II also shows that the Lewis test is consistently most powerful for these cases. For  $H_2$  cdf's with lower scv, the power of all methods is less than in Table IV, but the ordering remains; for  $H_2$  cdf's with higher scv, the power of all methods is greater, but the ordering remains.

Just as in [Kim and Whitt \[2013c\]](#), the story is more complicated for the dependent sequences. The Durbin test performs remarkably well for the RRI cases, far better than all others. For the RRI( $H_2$ ) cases, all tests except CU perform very well. Hence the Lewis test is consistently superior against non-exponential marginals. As in [Kim](#)



Table II. Performance of alternative KS tests of i.i.d. mean-1 exponential variables for the sample size  $n = 200$ : Number of KS tests passed (denoted by  $\#P$ ) at significance level 0.05 out of 10,000 replications and the average  $p$ -values (denoted by  $E[p - \text{value}]$ ) with associated 95% confidence intervals.

Case	Subcase	F(X)		Durbin		CU		Lewis	
		$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$
$Exp$	—	9487	$0.50 \pm 0.0057$	9515	$0.50 \pm 0.0056$	9511	$0.50 \pm 0.0056$	9493	$0.50 \pm 0.0057$
$E_k$	$k = 2$	28	$0.00 \pm 0.0001$	3320	$0.08 \pm 0.0029$	9985	$0.78 \pm 0.0045$	0	$0.00 \pm 0.0000$
	$k = 4$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	10000	$0.94 \pm 0.0021$	0	$0.00 \pm 0.0000$
	$k = 6$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	10000	$0.98 \pm 0.0011$	0	$0.00 \pm 0.0000$
$H_2$	$c^2 = 1.25$	8843	$0.42 \pm 0.0058$	9451	$0.49 \pm 0.0057$	8956	$0.41 \pm 0.0056$	7501	$0.30 \pm 0.0056$
	$c^2 = 1.5$	7204	$0.27 \pm 0.0053$	9331	$0.48 \pm 0.0058$	8418	$0.33 \pm 0.0053$	3966	$0.12 \pm 0.0039$
	$c^2 = 2$	3603	$0.09 \pm 0.0032$	8667	$0.40 \pm 0.0058$	7186	$0.24 \pm 0.0046$	695	$0.02 \pm 0.0013$
	$c^2 = 4$	90	$0.00 \pm 0.0003$	4569	$0.13 \pm 0.0039$	3648	$0.08 \pm 0.0027$	22	$0.00 \pm 0.0003$
	$c^2 = 10$	0	$0.00 \pm 0.0000$	878	$0.02 \pm 0.0012$	928	$0.02 \pm 0.0014$	67	$0.00 \pm 0.0006$
$Z$	—	1200	$0.02 \pm 0.0009$	7016	$0.26 \pm 0.0053$	9438	$0.57 \pm 0.0061$	187	$0.00 \pm 0.0004$
$LN$	(1, 0.25)	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	10000	$0.94 \pm 0.0022$	0	$0.00 \pm 0.0000$
	(1, 1)	98	$0.00 \pm 0.0002$	3482	$0.08 \pm 0.0025$	9517	$0.53 \pm 0.0058$	24	$0.00 \pm 0.0001$
	(1, 4)	176	$0.00 \pm 0.0005$	5542	$0.18 \pm 0.0047$	4742	$0.13 \pm 0.0036$	28	$0.00 \pm 0.0002$
	(1, 10)	0	$0.00 \pm 0.0000$	353	$0.01 \pm 0.0008$	2024	$0.04 \pm 0.0019$	0	$0.00 \pm 0.0000$
$RRI$	$p = 0.1$	9048	$0.41 \pm 0.0055$	1911	$0.03 \pm 0.0012$	9044	$0.42 \pm 0.0056$	9121	$0.41 \pm 0.0054$
	$p = 0.5$	4659	$0.11 \pm 0.0030$	0	$0.00 \pm 0.0000$	5587	$0.16 \pm 0.0039$	4624	$0.11 \pm 0.0030$
	$p = 0.9$	16	$0.00 \pm 0.0001$	0	$0.00 \pm 0.0000$	701	$0.01 \pm 0.0011$	13	$0.00 \pm 0.0001$
$EARMA$	0.25	9284	$0.47 \pm 0.0058$	9475	$0.50 \pm 0.0057$	8564	$0.36 \pm 0.0055$	9498	$0.50 \pm 0.0057$
	0.5	8865	$0.43 \pm 0.0059$	9516	$0.50 \pm 0.0057$	7519	$0.27 \pm 0.0050$	9393	$0.49 \pm 0.0058$
	1	8178	$0.37 \pm 0.0059$	9419	$0.50 \pm 0.0057$	6009	$0.19 \pm 0.0043$	8964	$0.44 \pm 0.0059$
	3	5209	$0.21 \pm 0.0055$	6356	$0.23 \pm 0.0050$	1896	$0.04 \pm 0.0018$	6796	$0.30 \pm 0.0061$
	5.25	4100	$0.14 \pm 0.0044$	8215	$0.38 \pm 0.0061$	1598	$0.03 \pm 0.0018$	5680	$0.21 \pm 0.0051$
$mH_2$	$m = 2$	4398	$0.14 \pm 0.0044$	8871	$0.42 \pm 0.0058$	4355	$0.11 \pm 0.0032$	1546	$0.04 \pm 0.0024$
	$m = 5$	7514	$0.32 \pm 0.0058$	9363	$0.48 \pm 0.0057$	5400	$0.17 \pm 0.0043$	7228	$0.29 \pm 0.0057$
	$m = 10$	7818	$0.35 \pm 0.0060$	9423	$0.49 \pm 0.0057$	6562	$0.24 \pm 0.0051$	9004	$0.44 \pm 0.0059$
	$m = 20$	7996	$0.37 \pm 0.0060$	9457	$0.50 \pm 0.0057$	7804	$0.33 \pm 0.0057$	9431	$0.49 \pm 0.0057$
$RRI(H_2)$	$p = 0.1$	104	$0.00 \pm 0.0003$	126	$0.00 \pm 0.0003$	2987	$0.07 \pm 0.0024$	37	$0.00 \pm 0.0003$
	$p = 0.5$	253	$0.00 \pm 0.0005$	0	$0.00 \pm 0.0000$	1105	$0.02 \pm 0.0013$	215	$0.00 \pm 0.0006$
	$p = 0.9$	4	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	229	$0.00 \pm 0.0005$	5	$0.00 \pm 0.0000$

and Whitt [2013c], none of the tests has much power against the EARMA alternatives, but the CU test has the most power.

#### 4.2. Plots of the Average Empirical Distributions

As in Kim and Whitt [2013c], we find that useful insight is provided by plots comparing the average of the ecdf's over all 10,000 replications to the cdf associated with the null hypothesis, which is uniform in each case here. Figures 1-4 illustrate for the i.i.d. variables having cdf's  $H_2$  with  $c^2 = 2$ ,  $E_2$  and  $LN(1, 4)$ , and for the dependent  $RRI(0.5)$  variables with  $n = 200$ . These figures show that the transformation in the Lewis KS test provides greater separation between the average ecdf and the cdf in the i.i.d. cases, while the Durbin KS test does so remarkably well for the RRI alternative. In each case, the Durbin and Lewis tests tend to produce stochastic order compared to the uniform cdf, whereas the ecdf crosses over for the standard KS test, which is especially evident for  $E_2$ .

#### 4.3. Erlang, Hyperexponential and Lognormal Alternatives

We now consider three different i.i.d. null hypotheses:  $E_2$ ,  $H_2$  with  $c^2 = 2$  and  $LN(1, 4)$ ; lognormal hypotheses are especially interesting for service systems, e.g., Brown et al. [2005]. The results are shown for the same 29 cases in Tables III-V below for the base case of  $n = 200$ . As before, all tests perform properly for the null hypotheses. The ordering of the tests by power when we consider the i.i.d. exponential alternative hypothesis is the same as before. Overall, these tables show that the previous conclusions for the

Fig. 1. Comparison of the average ecdf based on  $10^4$  replications for  $n = 200$  with the cdf of the null hypothesis;  $H_2$  ( $c^2 = 2$ ):  $F(X)$ , Durbin, CU, and Lewis Tests (from left to right).

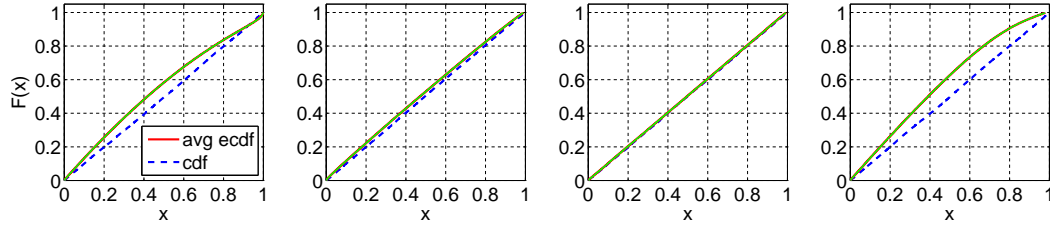


Fig. 2. Comparison of the average ecdf based on  $10^4$  replications for  $n = 200$  with the cdf of the null hypothesis;  $E_2$ :  $F(X)$ , Durbin, CU, and Lewis Tests (from left to right).

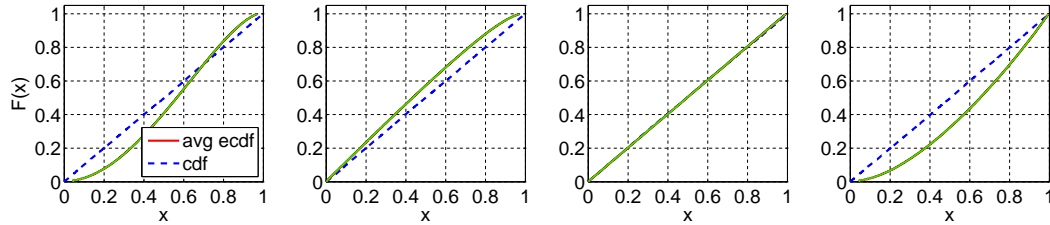


Fig. 3. Comparison of the average ecdf based on  $10^4$  replications for  $n = 200$  with the cdf of the null hypothesis;  $LN(1, 4)$ :  $F(X)$ , Durbin, CU, and Lewis Tests (from left to right).

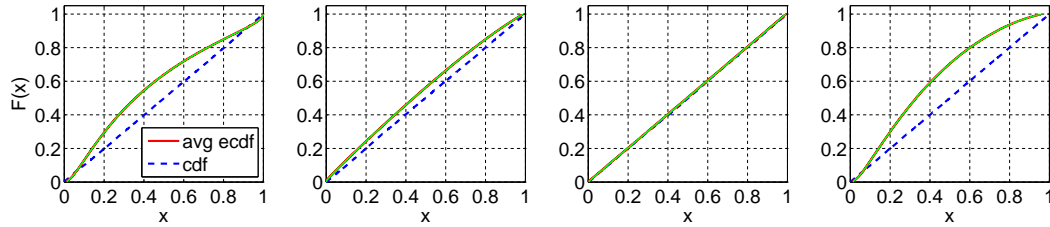
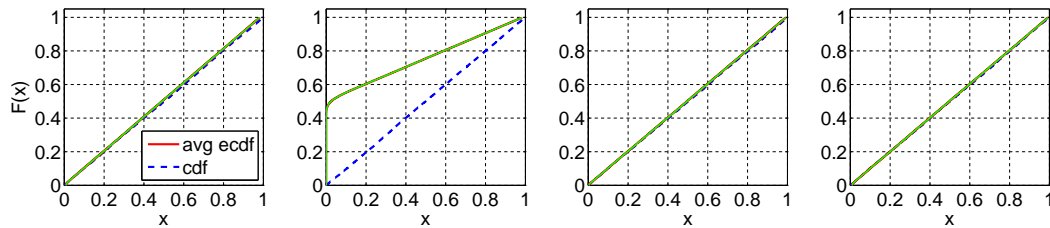


Fig. 4. Comparison of the average ecdf based on  $10^4$  replications for  $n = 200$  with the cdf of the null hypothesis;  $RR1(0.5)$ :  $F(X)$ , Durbin, CU, and Lewis Tests (from left to right).



i.i.d. exponential null hypothesis conclusions extend to i.i.d. null hypotheses with other marginal cdf's.

#### 4.4. Larger Sample Sizes

Table II-V clearly show how the power decreases as the alternative gets closer to the i.i.d. null hypothesis. For the i.i.d. exponential null hypothesis and the i.i.d. alternative hypotheses, we see this as the  $scv\ c_X^2$  approaches 1; for the dependent exponential sequences, we see this as the degree of dependence decreases. However, all of these are for the sample size  $n = 200$ . The power also increases as we increase the sample size, as we now illustrate by considering case  $n = 2000$  for the exponential null hypothesis

Table III. Performance of alternative KS tests of i.i.d.  $E_2$  variables for the sample size  $n = 200$ : Number of KS tests passed (denoted by  $\#P$ ) at significance level 0.05 out of 10,000 replications and the average  $p$ -values (denoted by  $E[p - \text{value}]$ ) with associated 95% confidence intervals.

Case	Subcase	F(X)		Durbin		CU		Lewis	
		$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$
$Exp$	—	129	$0.00 \pm 0.0003$	2596	$0.06 \pm 0.0027$	7421	$0.24 \pm 0.0046$	0	$0.00 \pm 0.0000$
$E_k$	$k = 2$	9492	$0.50 \pm 0.0056$	9500	$0.49 \pm 0.0057$	9497	$0.50 \pm 0.0057$	9506	$0.50 \pm 0.0057$
	$k = 4$	155	$0.00 \pm 0.0003$	4100	$0.11 \pm 0.0034$	9977	$0.77 \pm 0.0046$	0	$0.00 \pm 0.0000$
	$k = 6$	0	$0.00 \pm 0.0000$	7	$0.00 \pm 0.0001$	9999	$0.88 \pm 0.0033$	0	$0.00 \pm 0.0000$
$H_2$	$c^2 = 1.25$	17	$0.00 \pm 0.0001$	1181	$0.03 \pm 0.0016$	6106	$0.17 \pm 0.0040$	0	$0.00 \pm 0.0000$
	$c^2 = 1.5$	0	$0.00 \pm 0.0000$	539	$0.01 \pm 0.0008$	4905	$0.12 \pm 0.0033$	0	$0.00 \pm 0.0000$
	$c^2 = 2$	0	$0.00 \pm 0.0000$	129	$0.00 \pm 0.0004$	3336	$0.07 \pm 0.0024$	0	$0.00 \pm 0.0000$
	$c^2 = 4$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	752	$0.01 \pm 0.0009$	0	$0.00 \pm 0.0000$
	$c^2 = 10$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	67	$0.00 \pm 0.0004$	0	$0.00 \pm 0.0000$
$Z$	—	8069	$0.32 \pm 0.0054$	9286	$0.46 \pm 0.0058$	7152	$0.28 \pm 0.0054$	4466	$0.15 \pm 0.0046$
$LN$	(1, 0.25)	0	$0.00 \pm 0.0000$	425	$0.01 \pm 0.0006$	9973	$0.75 \pm 0.0048$	0	$0.00 \pm 0.0000$
	(1, 1)	3086	$0.07 \pm 0.0027$	8424	$0.37 \pm 0.0058$	6809	$0.22 \pm 0.0045$	331	$0.01 \pm 0.0009$
	(1, 4)	0	$0.00 \pm 0.0000$	3	$0.00 \pm 0.0000$	1507	$0.03 \pm 0.0014$	0	$0.00 \pm 0.0000$
	(1, 10)	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	408	$0.01 \pm 0.0006$	0	$0.00 \pm 0.0000$
$RRI$	$p = 0.1$	135	$0.00 \pm 0.0003$	24	$0.00 \pm 0.0001$	6455	$0.19 \pm 0.0042$	5	$0.00 \pm 0.0000$
	$p = 0.5$	164	$0.00 \pm 0.0004$	0	$0.00 \pm 0.0000$	2429	$0.05 \pm 0.0020$	45	$0.00 \pm 0.0002$
	$p = 0.9$	3	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	142	$0.00 \pm 0.0004$	3	$0.00 \pm 0.0000$
$EARMA$	0.25	108	$0.00 \pm 0.0002$	2552	$0.06 \pm 0.0027$	5494	$0.15 \pm 0.0037$	1	$0.00 \pm 0.0000$
	0.5	114	$0.00 \pm 0.0003$	2614	$0.07 \pm 0.0027$	4064	$0.10 \pm 0.0029$	0	$0.00 \pm 0.0000$
	1	135	$0.00 \pm 0.0003$	2597	$0.07 \pm 0.0028$	2670	$0.06 \pm 0.0022$	6	$0.00 \pm 0.0001$
	3	918	$0.02 \pm 0.0015$	3573	$0.12 \pm 0.0043$	508	$0.01 \pm 0.0008$	585	$0.02 \pm 0.0018$
	5.25	432	$0.01 \pm 0.0007$	2347	$0.07 \pm 0.0032$	374	$0.01 \pm 0.0006$	339	$0.01 \pm 0.0007$
$mH_2$	$m = 2$	0	$0.00 \pm 0.0000$	289	$0.01 \pm 0.0007$	1248	$0.02 \pm 0.0013$	0	$0.00 \pm 0.0000$
	$m = 5$	23	$0.00 \pm 0.0001$	1179	$0.03 \pm 0.0015$	2356	$0.05 \pm 0.0022$	0	$0.00 \pm 0.0000$
	$m = 10$	63	$0.00 \pm 0.0002$	1684	$0.04 \pm 0.0020$	3581	$0.09 \pm 0.0031$	0	$0.00 \pm 0.0000$
	$m = 20$	96	$0.00 \pm 0.0002$	2070	$0.05 \pm 0.0024$	4884	$0.14 \pm 0.0038$	0	$0.00 \pm 0.0000$
$RRI(H_2)$	$p = 0.1$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	557	$0.01 \pm 0.0007$	0	$0.00 \pm 0.0000$
	$p = 0.5$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	151	$0.00 \pm 0.0003$	0	$0.00 \pm 0.0000$
	$p = 0.9$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	23	$0.00 \pm 0.0002$	1	$0.00 \pm 0.0000$

in Table VI. Corresponding results for Erlang, hyperexponential and lognormal null hypotheses appear in Kim and Whitt [2013b]. When the sample size is increased to  $n = 2000$ , all the tests except the CU test reject the alternative hypotheses in all  $10^4$  replications for most of the alternatives. Nevertheless, the superiority of the Lewis test for non-exponential marginals is evident from the  $H_2$  case with  $c^2 = 1.25$ , the superiority of the Durbin test for the RRI cases is evident, and the superiority of the CU test for the EARMA cases is evident, consistent with the previous results for  $n = 200$ .

## 5. THE SECOND NORMAL EXPERIMENT

The poor results for the Durbin [1961] test for the i.i.d. cases in §4 seem inconsistent with the results in Durbin [1961] and the enthusiastic endorsement by Lewis [1965], so we decided to repeat some of the experiments actually performed by Durbin [1961]. We now consider the same four KS tests applied to the i.i.d. standard normal ( $N(0, 1)$ ) null hypothesis. To keep the same mean equal to 0 for all alternatives, we consider all the previous 29 cases after subtracting 1 to make them all have mean 0. Indeed, the first alternative considered by Durbin [1961] was an i.i.d. sequence of random variables distributed as  $Y - 1$ , where  $Y$  is a mean 1 exponential variable; it has the same mean and variance as  $N(0, 1)$ . We summarize the results for this alternative with the sample size  $n = 50$  used by Durbin [1961] in Table VII. Table VII shows that now the Durbin [1961] and Lewis [1965] have essentially the same power, which is far greater than for the standard and CU tests.

Table IV. Performance of alternative KS tests of i.i.d.  $H_2$  with  $c^2 = 2$  variables for the sample size  $n = 200$ : Number of KS tests passed (denoted by  $\#P$ ) at significance level 0.05 out of 10,000 replications and the average  $p$ -values (denoted by  $E[p - \text{value}]$ ) with associated 95% confidence intervals.

Case	Subcase	F(X)		Durbin		CU		Lewis	
		$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$
$Exp$	—	3661	$0.10 \pm 0.0034$	8951	$0.43 \pm 0.0058$	9935	$0.69 \pm 0.0051$	1613	$0.03 \pm 0.0014$
$E_k$	$k = 2$	0	$0.00 \pm 0.0000$	92	$0.00 \pm 0.0003$	10000	$0.89 \pm 0.0032$	0	$0.00 \pm 0.0000$
	$k = 4$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	10000	$0.98 \pm 0.0012$	0	$0.00 \pm 0.0000$
	$k = 6$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	10000	$0.99 \pm 0.0005$	0	$0.00 \pm 0.0000$
$H_2$	$c^2 = 1.25$	6574	$0.23 \pm 0.0052$	9433	$0.49 \pm 0.0057$	9850	$0.63 \pm 0.0055$	5543	$0.15 \pm 0.0038$
	$c^2 = 1.5$	8530	$0.39 \pm 0.0059$	9497	$0.50 \pm 0.0057$	9750	$0.58 \pm 0.0056$	8307	$0.34 \pm 0.0055$
	$c^2 = 2$	9511	$0.50 \pm 0.0056$	9482	$0.50 \pm 0.0057$	9482	$0.50 \pm 0.0057$	9507	$0.50 \pm 0.0056$
	$c^2 = 4$	4983	$0.14 \pm 0.0040$	9107	$0.44 \pm 0.0058$	8143	$0.31 \pm 0.0052$	3888	$0.11 \pm 0.0038$
	$c^2 = 10$	269	$0.01 \pm 0.0005$	6142	$0.19 \pm 0.0046$	5098	$0.15 \pm 0.0039$	1221	$0.04 \pm 0.0024$
$Z$	—	0	$0.00 \pm 0.0000$	1932	$0.04 \pm 0.0021$	9989	$0.80 \pm 0.0043$	0	$0.00 \pm 0.0000$
$LN$	(1, 0.25)	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	10000	$0.98 \pm 0.0011$	0	$0.00 \pm 0.0000$
	(1, 1)	0	$0.00 \pm 0.0000$	585	$0.01 \pm 0.0006$	9982	$0.77 \pm 0.0046$	0	$0.00 \pm 0.0000$
	(1, 4)	5685	$0.18 \pm 0.0045$	9051	$0.44 \pm 0.0059$	8493	$0.36 \pm 0.0055$	5281	$0.16 \pm 0.0043$
	(1, 10)	13	$0.00 \pm 0.0001$	4888	$0.15 \pm 0.0043$	5824	$0.17 \pm 0.0042$	11	$0.00 \pm 0.0001$
$RRI$	$p = 0.1$	3400	$0.09 \pm 0.0032$	1352	$0.02 \pm 0.0010$	9804	$0.61 \pm 0.0056$	1410	$0.03 \pm 0.0013$
	$p = 0.5$	2058	$0.05 \pm 0.0020$	0	$0.00 \pm 0.0000$	7608	$0.28 \pm 0.0050$	883	$0.02 \pm 0.0012$
	$p = 0.9$	9	$0.00 \pm 0.0001$	0	$0.00 \pm 0.0000$	1282	$0.03 \pm 0.0017$	6	$0.00 \pm 0.0000$
$EARMA$	0.25	3697	$0.10 \pm 0.0035$	8922	$0.43 \pm 0.0058$	9684	$0.56 \pm 0.0056$	1577	$0.03 \pm 0.0014$
	0.5	3839	$0.11 \pm 0.0037$	8872	$0.42 \pm 0.0059$	9216	$0.45 \pm 0.0057$	1630	$0.03 \pm 0.0015$
	1	3755	$0.11 \pm 0.0037$	8629	$0.40 \pm 0.0059$	8364	$0.34 \pm 0.0055$	1607	$0.03 \pm 0.0017$
	3	3607	$0.13 \pm 0.0044$	5683	$0.19 \pm 0.0047$	3333	$0.08 \pm 0.0028$	2577	$0.07 \pm 0.0032$
	5.25	2770	$0.08 \pm 0.0032$	6642	$0.27 \pm 0.0056$	3118	$0.08 \pm 0.0029$	1690	$0.05 \pm 0.0025$
$mH_2$	$m = 2$	8771	$0.42 \pm 0.0058$	9466	$0.49 \pm 0.0057$	7788	$0.29 \pm 0.0052$	9091	$0.43 \pm 0.0057$
	$m = 5$	6227	$0.24 \pm 0.0053$	9290	$0.47 \pm 0.0058$	7974	$0.33 \pm 0.0056$	5465	$0.16 \pm 0.0041$
	$m = 10$	5052	$0.18 \pm 0.0047$	9032	$0.44 \pm 0.0058$	8543	$0.40 \pm 0.0061$	3210	$0.07 \pm 0.0025$
	$m = 20$	4598	$0.15 \pm 0.0044$	9013	$0.43 \pm 0.0058$	9265	$0.50 \pm 0.0061$	2263	$0.05 \pm 0.0018$
$RRI(H_2)$	$p = 0.1$	4641	$0.14 \pm 0.0040$	1227	$0.02 \pm 0.0010$	7377	$0.26 \pm 0.0048$	3720	$0.11 \pm 0.0037$
	$p = 0.5$	2542	$0.05 \pm 0.0022$	0	$0.00 \pm 0.0000$	3586	$0.09 \pm 0.0029$	2467	$0.05 \pm 0.0022$
	$p = 0.9$	13	$0.00 \pm 0.0001$	0	$0.00 \pm 0.0000$	440	$0.01 \pm 0.0008$	9	$0.00 \pm 0.0001$

Table VIII below shows all the results for our original 29 cases with  $n = 50$ . Since those alternatives have quite a different shape from the symmetric  $N(0, 1)$  distributions, we also considered i.i.d. sequences of random variables distributed as  $Z_k - 1 + \sqrt{1 - (1/k)}N(0, 1)$ , where  $Z_k$  has an  $E_k$  cdf, for  $k = 2, 4, 6$ . These have the same first two moments and approximately the same shape.

The new base case is the i.i.d. standard normal null hypothesis; it appears just below the previous alternatives in Table VIII. Just as in the previous tables, the results show that all tests behave properly for the standard normal null hypothesis. Overall, Table VIII shows that the Durbin [1961] test performs much better now, just as originally reported. In this case both the Durbin [1961] and Lewis [1965] KS tests perform much better than the standard and CU alternatives. An exception is the set of three modified Erlang cases, with the same shape and first two moments as  $N(0, 1)$ . The Lewis test has most power, but all four tests have low power for these cases.

As in §4, the power increases as the sample size increases. Table IX shows the results for  $n = 200$ . With the larger sample size  $n = 200$ , all tests except CU have estimated perfect power except in the last three modified Erlang cases, where the Lewis test stands out with power 0.375 for the modified  $E_2$  case compared to 0.13 for standard and CU, and only 0.055 for Durbin. Figures 5 and 6 show that the reason can be seen in the average of the ecdf's of the transformed data.

Table V. Performance of alternative KS tests of i.i.d.  $LN(1, 4)$  variables for the sample size  $n = 200$ : Number of KS tests passed (denoted by  $\#P$ ) at significance level 0.05 out of 10,000 replications and the average  $p$ -values (denoted by  $E[p - \text{value}]$ ) with associated 95% confidence intervals.

Case	Subcase	F(X)		Durbin		CU		Lewis	
		$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$
$Exp$	—	181	$0.00 \pm 0.0005$	5509	$0.18 \pm 0.0046$	9972	$0.75 \pm 0.0047$	38	$0.00 \pm 0.0002$
$E_k$	$k = 2$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	10000	$0.93 \pm 0.0024$	0	$0.00 \pm 0.0000$
	$k = 4$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	10000	$0.99 \pm 0.0007$	0	$0.00 \pm 0.0000$
	$k = 6$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	10000	$1.00 \pm 0.0003$	0	$0.00 \pm 0.0000$
$H_2$	$c^2 = 1.25$	811	$0.02 \pm 0.0012$	7382	$0.29 \pm 0.0056$	9939	$0.70 \pm 0.0051$	513	$0.01 \pm 0.0007$
	$c^2 = 1.5$	2340	$0.05 \pm 0.0023$	8354	$0.37 \pm 0.0058$	9895	$0.66 \pm 0.0053$	2255	$0.05 \pm 0.0020$
	$c^2 = 2$	5665	$0.17 \pm 0.0043$	9006	$0.43 \pm 0.0058$	9788	$0.59 \pm 0.0055$	6140	$0.19 \pm 0.0043$
	$c^2 = 4$	9164	$0.36 \pm 0.0048$	8864	$0.41 \pm 0.0058$	9294	$0.46 \pm 0.0056$	8783	$0.31 \pm 0.0046$
	$c^2 = 10$	3774	$0.08 \pm 0.0023$	6700	$0.23 \pm 0.0050$	8538	$0.35 \pm 0.0054$	5450	$0.13 \pm 0.0032$
$Z$	—	0	$0.00 \pm 0.0000$	196	$0.00 \pm 0.0005$	10000	$0.87 \pm 0.0034$	0	$0.00 \pm 0.0000$
$LN$	(1, 0.25)	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	10000	$0.99 \pm 0.0005$	0	$0.00 \pm 0.0000$
	(1, 1)	0	$0.00 \pm 0.0000$	90	$0.00 \pm 0.0003$	9999	$0.85 \pm 0.0037$	0	$0.00 \pm 0.0000$
	(1, 4)	9508	$0.50 \pm 0.0056$	9508	$0.50 \pm 0.0056$	9508	$0.50 \pm 0.0057$	9490	$0.50 \pm 0.0057$
	(1, 10)	232	$0.01 \pm 0.0005$	6261	$0.22 \pm 0.0051$	8094	$0.30 \pm 0.0051$	185	$0.00 \pm 0.0004$
$RRI$	$p = 0.1$	193	$0.00 \pm 0.0005$	346	$0.01 \pm 0.0004$	9921	$0.68 \pm 0.0053$	47	$0.00 \pm 0.0001$
	$p = 0.5$	408	$0.01 \pm 0.0007$	0	$0.00 \pm 0.0000$	8255	$0.34 \pm 0.0054$	120	$0.00 \pm 0.0003$
	$p = 0.9$	13	$0.00 \pm 0.0001$	0	$0.00 \pm 0.0000$	1738	$0.04 \pm 0.0021$	3	$0.00 \pm 0.0001$
$EARMA$	0.25	206	$0.00 \pm 0.0006$	5443	$0.18 \pm 0.0046$	9866	$0.64 \pm 0.0054$	34	$0.00 \pm 0.0001$
	0.5	312	$0.01 \pm 0.0007$	5388	$0.17 \pm 0.0045$	9571	$0.53 \pm 0.0058$	44	$0.00 \pm 0.0002$
	1	436	$0.01 \pm 0.0009$	5032	$0.16 \pm 0.0045$	9023	$0.43 \pm 0.0058$	72	$0.00 \pm 0.0003$
	3	1594	$0.04 \pm 0.0024$	4073	$0.13 \pm 0.0041$	4018	$0.10 \pm 0.0033$	647	$0.01 \pm 0.0012$
	5.25	1220	$0.03 \pm 0.0019$	3612	$0.12 \pm 0.0042$	4027	$0.11 \pm 0.0036$	469	$0.01 \pm 0.0013$
$mH_2$	$m = 2$	4930	$0.15 \pm 0.0040$	8640	$0.39 \pm 0.0058$	8786	$0.39 \pm 0.0057$	4425	$0.12 \pm 0.0035$
	$m = 5$	1706	$0.04 \pm 0.0022$	7193	$0.27 \pm 0.0055$	8677	$0.40 \pm 0.0059$	606	$0.01 \pm 0.0008$
	$m = 10$	1083	$0.03 \pm 0.0017$	6179	$0.22 \pm 0.0051$	9085	$0.48 \pm 0.0062$	178	$0.00 \pm 0.0004$
	$m = 20$	808	$0.02 \pm 0.0013$	5752	$0.19 \pm 0.0049$	9572	$0.57 \pm 0.0060$	79	$0.00 \pm 0.0002$
$RRI(H_2)$	$p = 0.1$	8581	$0.29 \pm 0.0046$	834	$0.02 \pm 0.0008$	8830	$0.39 \pm 0.0055$	8117	$0.26 \pm 0.0044$
	$p = 0.5$	3857	$0.08 \pm 0.0024$	0	$0.00 \pm 0.0000$	5080	$0.14 \pm 0.0036$	3547	$0.07 \pm 0.0024$
	$p = 0.9$	17	$0.00 \pm 0.0001$	0	$0.00 \pm 0.0000$	658	$0.01 \pm 0.0010$	5	$0.00 \pm 0.0001$

## 6. CONCLUSIONS

We have conducted simulation experiments to study the power of alternative Kolmogorov-Smirnov (KS) statistical tests of a sequence of i.i.d. random variables with continuous cdf  $F$ , focusing on the exponential and standard normal null hypotheses. Our analysis strongly supports the data-transformation approach proposed by Durbin [1961], but we find that another related KS test proposed by Lewis [1965], which uses the original Durbin [1961] transformation after converting to exponential random variables and applying the conditional-uniform (CU) property, is usually superior, often markedly so. Thus, we recommend the Lewis test, implemented as described in §2. Since there is some variation in the results, we recommend applying simulation as we have done in this paper, if there is the opportunity, in order to assess what KS test has the most power and what that power should be in a new setting of interest. The tables and plots based on  $10^4$  replications give a very clear picture.

Both in Kim and Whitt [2013c] and here we have focused on the two-sided KS test, but we also conducted one-sided KS tests. We found that the one-sided test can further increase power when it is justified. As usual with statistical tests, the power increases with the sample size, so that some sample sizes may be too small to have any power, whereas other sample sizes may be too large to accept even the slightest deviation from a null hypothesis. Thus, as many have discovered before, judgment is required in the use of statistical tests.



Table VI. Performance of alternative KS tests of i.i.d. mean-1 exponential variables for the sample size  $n = 2000$ : Number of KS tests passed (denoted by  $\#P$ ) at significance level 0.05 out of 10,000 replications and the average  $p$ -values (denoted by  $E[p - \text{value}]$ ) with associated 95% confidence intervals.

Case	Subcase	F(X)		Durbin		CU		Lewis	
		$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$
$Exp$	—	9515	$0.50 \pm 0.0056$	9495	$0.50 \pm 0.0057$	9481	$0.50 \pm 0.0057$	9495	$0.50 \pm 0.0057$
$E_k$	$k = 2$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	9985	$0.79 \pm 0.0044$	0	$0.00 \pm 0.0000$
	$k = 4$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	10000	$0.95 \pm 0.0019$	0	$0.00 \pm 0.0000$
	$k = 6$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	10000	$0.98 \pm 0.0009$	0	$0.00 \pm 0.0000$
$H_2$	$c^2 = 1.25$	3380	$0.08 \pm 0.0029$	9360	$0.48 \pm 0.0057$	8957	$0.40 \pm 0.0055$	281	$0.01 \pm 0.0006$
	$c^2 = 1.5$	68	$0.00 \pm 0.0002$	8320	$0.36 \pm 0.0059$	8313	$0.32 \pm 0.0051$	0	$0.00 \pm 0.0000$
	$c^2 = 2$	0	$0.00 \pm 0.0000$	3425	$0.08 \pm 0.0030$	6893	$0.21 \pm 0.0043$	0	$0.00 \pm 0.0000$
	$c^2 = 4$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	2788	$0.05 \pm 0.0019$	0	$0.00 \pm 0.0000$
	$c^2 = 10$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	34	$0.00 \pm 0.0002$	0	$0.00 \pm 0.0000$
$Z$	—	0	$0.00 \pm 0.0000$	4	$0.00 \pm 0.0001$	9450	$0.52 \pm 0.0058$	0	$0.00 \pm 0.0000$
$LN$	(1, 0.25)	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	10000	$0.95 \pm 0.0019$	0	$0.00 \pm 0.0000$
	(1, 1)	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	9501	$0.51 \pm 0.0057$	0	$0.00 \pm 0.0000$
	(1, 4)	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	2610	$0.06 \pm 0.0023$	0	$0.00 \pm 0.0000$
	(1, 10)	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	242	$0.00 \pm 0.0005$	0	$0.00 \pm 0.0000$
$RRI$	$p = 0.1$	9010	$0.41 \pm 0.0055$	0	$0.00 \pm 0.0000$	9129	$0.41 \pm 0.0055$	9014	$0.40 \pm 0.0055$
	$p = 0.5$	4410	$0.10 \pm 0.0028$	0	$0.00 \pm 0.0000$	4666	$0.11 \pm 0.0030$	4531	$0.10 \pm 0.0028$
	$p = 0.9$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	25	$0.00 \pm 0.0001$	0	$0.00 \pm 0.0000$
$EARMA$	0.25	9336	$0.47 \pm 0.0057$	9483	$0.50 \pm 0.0057$	8326	$0.33 \pm 0.0052$	9429	$0.49 \pm 0.0057$
	0.5	8806	$0.42 \pm 0.0059$	9505	$0.50 \pm 0.0057$	7063	$0.22 \pm 0.0044$	9408	$0.49 \pm 0.0057$
	1	8210	$0.37 \pm 0.0059$	9488	$0.50 \pm 0.0057$	4722	$0.12 \pm 0.0031$	8901	$0.43 \pm 0.0058$
	3	5247	$0.21 \pm 0.0054$	6406	$0.22 \pm 0.0049$	822	$0.01 \pm 0.0008$	6715	$0.29 \pm 0.0061$
	5.25	4111	$0.14 \pm 0.0045$	9290	$0.47 \pm 0.0058$	193	$0.00 \pm 0.0003$	5769	$0.21 \pm 0.0051$
$mH_2$	$m = 2$	0	$0.00 \pm 0.0000$	5272	$0.16 \pm 0.0042$	3029	$0.06 \pm 0.0022$	0	$0.00 \pm 0.0000$
	$m = 5$	3135	$0.09 \pm 0.0032$	9281	$0.46 \pm 0.0058$	3434	$0.07 \pm 0.0024$	182	$0.00 \pm 0.0004$
	$m = 10$	6428	$0.25 \pm 0.0054$	9471	$0.49 \pm 0.0057$	3732	$0.09 \pm 0.0027$	4432	$0.13 \pm 0.0040$
	$m = 20$	7364	$0.31 \pm 0.0058$	9470	$0.50 \pm 0.0057$	4365	$0.11 \pm 0.0033$	8127	$0.35 \pm 0.0058$
$RRI(H_2)$	$p = 0.1$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	1897	$0.03 \pm 0.0015$	0	$0.00 \pm 0.0000$
	$p = 0.5$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	177	$0.00 \pm 0.0003$	0	$0.00 \pm 0.0000$
	$p = 0.9$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$

Table VII. Summary of the performance of alternative KS tests of i.i.d. standard normal  $N(0, 1)$  variables for the sample size  $n = 50$  with significance level  $\alpha = 0.05$ : the case of the random variable  $Y - 1$ , where  $Y$  is a mean-1 exponential random variable.

KS test	Lewis	Standard	CU	Durbin
Power	0.885	0.443	0.328	0.894
Average $p$ value	0.02	0.07	0.23	0.02

Table VIII. Performance of alternative KS tests of i.i.d.  $N(0, 1)$  variables for the sample size  $n = 50$ : Number of KS tests passed (denoted by  $\#P$ ) at significance level 0.05 out of 10,000 replications and the average  $p$ -values (denoted by  $E[p - \text{value}]$ ) with associated 95% confidence intervals.

<i>Case</i>	<i>Subcase</i>	F(X)		Durbin		CU		Lewis	
		$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$
<i>Exp</i>	—	5576	$0.07 \pm 0.0010$	1871	$0.04 \pm 0.0016$	6716	$0.23 \pm 0.0049$	1154	$0.02 \pm 0.0006$
$E_k$	$k = 2$	3813	$0.04 \pm 0.0006$	2953	$0.05 \pm 0.0018$	9364	$0.52 \pm 0.0059$	376	$0.01 \pm 0.0004$
	$k = 4$	20	$0.01 \pm 0.0002$	336	$0.01 \pm 0.0004$	9977	$0.81 \pm 0.0043$	0	$0.00 \pm 0.0000$
	$k = 6$	0	$0.00 \pm 0.0000$	5	$0.00 \pm 0.0000$	10000	$0.92 \pm 0.0026$	0	$0.00 \pm 0.0000$
$H_2$	$c^2 = 1.25$	4188	$0.05 \pm 0.0010$	1004	$0.02 \pm 0.0011$	5051	$0.16 \pm 0.0043$	417	$0.01 \pm 0.0004$
	$c^2 = 1.5$	3100	$0.04 \pm 0.0009$	629	$0.01 \pm 0.0009$	4022	$0.12 \pm 0.0039$	174	$0.00 \pm 0.0003$
	$c^2 = 2$	1747	$0.02 \pm 0.0008$	221	$0.00 \pm 0.0004$	2639	$0.07 \pm 0.0031$	36	$0.00 \pm 0.0001$
	$c^2 = 4$	222	$0.00 \pm 0.0003$	17	$0.00 \pm 0.0001$	1237	$0.04 \pm 0.0027$	1	$0.00 \pm 0.0000$
	$c^2 = 10$	7	$0.00 \pm 0.0001$	1	$0.00 \pm 0.0000$	1870	$0.09 \pm 0.0046$	0	$0.00 \pm 0.0000$
$Z$	—	4836	$0.05 \pm 0.0008$	2671	$0.05 \pm 0.0018$	7273	$0.37 \pm 0.0065$	533	$0.01 \pm 0.0004$
$LN$	(1, 0.25)	0	$0.00 \pm 0.0001$	41	$0.00 \pm 0.0001$	9915	$0.76 \pm 0.0050$	0	$0.00 \pm 0.0000$
	(1, 1)	1722	$0.03 \pm 0.0005$	700	$0.01 \pm 0.0007$	5971	$0.24 \pm 0.0055$	89	$0.00 \pm 0.0002$
	(1, 4)	460	$0.01 \pm 0.0004$	31	$0.00 \pm 0.0002$	2027	$0.06 \pm 0.0028$	5	$0.00 \pm 0.0000$
	(1, 10)	24	$0.00 \pm 0.0001$	0	$0.00 \pm 0.0000$	1168	$0.03 \pm 0.0021$	1	$0.00 \pm 0.0000$
$RRI$	$p = 0.1$	5219	$0.06 \pm 0.0010$	763	$0.01 \pm 0.0009$	6239	$0.21 \pm 0.0049$	1152	$0.02 \pm 0.0007$
	$p = 0.5$	2791	$0.03 \pm 0.0008$	0	$0.00 \pm 0.0000$	4283	$0.13 \pm 0.0039$	788	$0.01 \pm 0.0007$
	$p = 0.9$	62	$0.00 \pm 0.0001$	0	$0.00 \pm 0.0000$	3696	$0.17 \pm 0.0057$	15	$0.00 \pm 0.0001$
$EARMA$	0.25	5395	$0.07 \pm 0.0010$	1813	$0.04 \pm 0.0016$	5820	$0.20 \pm 0.0048$	1120	$0.02 \pm 0.0007$
	0.5	5296	$0.06 \pm 0.0011$	1872	$0.04 \pm 0.0016$	5140	$0.17 \pm 0.0045$	1192	$0.02 \pm 0.0008$
	1	5028	$0.06 \pm 0.0011$	1884	$0.04 \pm 0.0017$	4883	$0.17 \pm 0.0047$	1370	$0.02 \pm 0.0010$
	3	3034	$0.03 \pm 0.0008$	2492	$0.05 \pm 0.0019$	2970	$0.09 \pm 0.0034$	1474	$0.03 \pm 0.0014$
	5.25	3446	$0.04 \pm 0.0010$	2049	$0.05 \pm 0.0020$	4275	$0.17 \pm 0.0053$	2115	$0.05 \pm 0.0023$
$mH_2$	$m = 2$	2363	$0.03 \pm 0.0009$	460	$0.01 \pm 0.0007$	2777	$0.09 \pm 0.0038$	76	$0.00 \pm 0.0002$
	$m = 5$	4045	$0.05 \pm 0.0010$	1109	$0.02 \pm 0.0012$	4591	$0.16 \pm 0.0046$	421	$0.01 \pm 0.0005$
	$m = 10$	4667	$0.06 \pm 0.0010$	1477	$0.03 \pm 0.0015$	5682	$0.20 \pm 0.0049$	706	$0.01 \pm 0.0006$
	$m = 20$	4932	$0.06 \pm 0.0010$	1636	$0.03 \pm 0.0015$	6361	$0.23 \pm 0.0050$	891	$0.02 \pm 0.0007$
$RRI(H_2)$	$p = 0.1$	302	$0.01 \pm 0.0003$	3	$0.00 \pm 0.0001$	1306	$0.04 \pm 0.0028$	3	$0.00 \pm 0.0000$
	$p = 0.5$	454	$0.01 \pm 0.0004$	0	$0.00 \pm 0.0000$	2009	$0.07 \pm 0.0037$	12	$0.00 \pm 0.0001$
	$p = 0.9$	17	$0.00 \pm 0.0001$	0	$0.00 \pm 0.0000$	4063	$0.21 \pm 0.0065$	1	$0.00 \pm 0.0000$
$N(0, 1)$	—	9447	$0.50 \pm 0.0057$	9460	$0.50 \pm 0.0057$	9501	$0.50 \pm 0.0056$	9492	$0.50 \pm 0.0057$
$E_k - 1$	$k = 2$	9336	$0.47 \pm 0.0057$	9472	$0.49 \pm 0.0057$	8782	$0.40 \pm 0.0057$	8393	$0.38 \pm 0.0058$
$+\sqrt{1 - 1/k}$	$k = 4$	9526	$0.51 \pm 0.0056$	9493	$0.50 \pm 0.0057$	9330	$0.47 \pm 0.0057$	9410	$0.48 \pm 0.0057$
$\times N(0, 1)$	$k = 6$	9503	$0.50 \pm 0.0057$	9476	$0.50 \pm 0.0057$	9427	$0.49 \pm 0.0057$	9445	$0.49 \pm 0.0057$

Table IX. Performance of alternative KS tests of i.i.d.  $N(0, 1)$  variables for the sample size  $n = 200$ : Number of KS tests passed (denoted by  $\#P$ ) at significance level 0.05 out of 10,000 replications and the average  $p$ -values (denoted by  $E[p - \text{value}]$ ) with associated 95% confidence intervals.

<i>Case</i>	<i>Subcase</i>	<b>F(X)</b>		<b>Durbin</b>		<b>CU</b>		<b>Lewis</b>	
		$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$	$\#P$	$E[p - \text{value}]$
<i>Exp</i>	—	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	5524	$0.16 \pm 0.0039$	0	$0.00 \pm 0.0000$
<i>E<sub>k</sub></i>	$k = 2$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	9262	$0.48 \pm 0.0059$	0	$0.00 \pm 0.0000$
	$k = 4$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	9990	$0.82 \pm 0.0041$	0	$0.00 \pm 0.0000$
	$k = 6$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	10000	$0.93 \pm 0.0024$	0	$0.00 \pm 0.0000$
<i>H<sub>2</sub></i>	$c^2 = 1.25$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	3240	$0.08 \pm 0.0028$	0	$0.00 \pm 0.0000$
	$c^2 = 1.5$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	1850	$0.04 \pm 0.0019$	0	$0.00 \pm 0.0000$
	$c^2 = 2$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	702	$0.01 \pm 0.0010$	0	$0.00 \pm 0.0000$
	$c^2 = 4$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	33	$0.00 \pm 0.0003$	0	$0.00 \pm 0.0000$
	$c^2 = 10$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	27	$0.00 \pm 0.0004$	0	$0.00 \pm 0.0000$
<i>Z</i>	—	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	4471	$0.18 \pm 0.0053$	0	$0.00 \pm 0.0000$
<i>LN</i>	(1, 0.25)	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	9926	$0.76 \pm 0.0050$	0	$0.00 \pm 0.0000$
	(1, 1)	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	3684	$0.10 \pm 0.0035$	0	$0.00 \pm 0.0000$
	(1, 4)	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	248	$0.00 \pm 0.0006$	0	$0.00 \pm 0.0000$
	(1, 10)	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	48	$0.00 \pm 0.0002$	0	$0.00 \pm 0.0000$
<i>RRI</i>	$p = 0.1$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	4784	$0.13 \pm 0.0035$	0	$0.00 \pm 0.0000$
	$p = 0.5$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	2118	$0.05 \pm 0.0021$	0	$0.00 \pm 0.0000$
	$p = 0.9$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	416	$0.01 \pm 0.0009$	0	$0.00 \pm 0.0000$
<i>EARMA</i>	0.25	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	4026	$0.10 \pm 0.0032$	0	$0.00 \pm 0.0000$
	0.5	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	2997	$0.07 \pm 0.0026$	0	$0.00 \pm 0.0000$
	1	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	2143	$0.05 \pm 0.0022$	0	$0.00 \pm 0.0000$
	3	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	625	$0.01 \pm 0.0009$	0	$0.00 \pm 0.0000$
	5.25	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	632	$0.01 \pm 0.0011$	5	$0.00 \pm 0.0000$
<i>mH<sub>2</sub></i>	$m = 2$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	315	$0.01 \pm 0.0008$	0	$0.00 \pm 0.0000$
	$m = 5$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	1395	$0.03 \pm 0.0018$	0	$0.00 \pm 0.0000$
	$m = 10$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	2470	$0.06 \pm 0.0026$	0	$0.00 \pm 0.0000$
	$m = 20$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	3603	$0.10 \pm 0.0033$	0	$0.00 \pm 0.0000$
<i>RRI(H<sub>2</sub>)</i>	$p = 0.1$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	42	$0.00 \pm 0.0002$	0	$0.00 \pm 0.0000$
	$p = 0.5$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	93	$0.00 \pm 0.0005$	0	$0.00 \pm 0.0000$
	$p = 0.9$	0	$0.00 \pm 0.0000$	0	$0.00 \pm 0.0000$	472	$0.01 \pm 0.0012$	0	$0.00 \pm 0.0000$
<i>N(0, 1)</i>	—	9491	$0.51 \pm 0.0056$	9476	$0.50 \pm 0.0057$	9522	$0.50 \pm 0.0056$	9498	$0.50 \pm 0.0057$
<i>E<sub>k</sub> - 1</i>	$k = 2$	8702	$0.39 \pm 0.0056$	9366	$0.48 \pm 0.0057$	8631	$0.37 \pm 0.0055$	6247	$0.21 \pm 0.0049$
$+\sqrt{1 - 1/k}$	$k = 4$	9474	$0.49 \pm 0.0057$	9480	$0.50 \pm 0.0056$	9370	$0.47 \pm 0.0056$	9233	$0.46 \pm 0.0057$
$\times N(0, 1)$	$k = 6$	9514	$0.50 \pm 0.0056$	9515	$0.50 \pm 0.0057$	9416	$0.49 \pm 0.0057$	9436	$0.49 \pm 0.0057$

Fig. 5. Comparison of the average ecdf based on  $10^4$  replications with  $n = 200$  for the  $Exp - 1$  alternative and the  $N(0, 1)$  null hypothesis:  $F(X)$ , Durbin, CU, and Lewis Tests (from left to right).

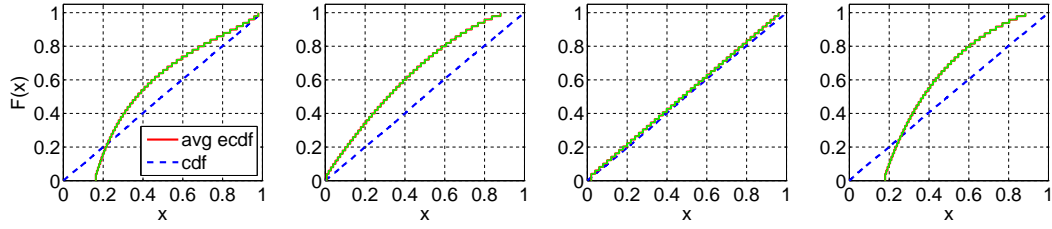
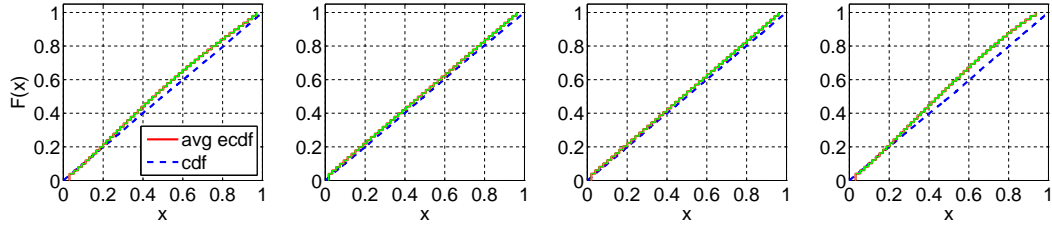


Fig. 6. Comparison of the average ecdf based on  $10^4$  replications with  $n = 200$  for the  $Exp - 1$  alternative and the  $N(0, 1)$  null hypothesis:  $E_2 - 1 + \sqrt{1 - 1/k}N(0, 1)$ :  $F(X)$ , Durbin, CU, and Lewis Tests (from left to right).



## REFERENCES

- M. Armony, S. Israelit, A. Mandelbaum, Y. Marmor, Y. Tseytlin, and G. Yom-Tov. 2011. Patient Flow in Hospitals: a Data-based Queueing-science Perspective. (2011). New York University, <http://www.stern.nyu.edu/om/faculty/armony/>.
- G. A. Barnard. 1953. Time Intervals Between Accidents—A Note on Maguire, Pearson & Wynn's Paper. *Biometrika* 40 (1953), 212–213.
- L. Brown, N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. Zhao. 2005. Statistical analysis of a telephone call center: a queueing-science perspective. *J. Amer. Stat. Assoc.* 100 (2005), 36–50.
- J. Durbin. 1961. Some methods for constructing exact tests. *Biometrika* 48, 1 (1961), 41–55.
- P. A. Jacobs and P. A. W. Lewis. 1977. A Mixed Autoregressive-moving Average Exponential Sequence and Point Process (EARMA 1,1). *Advances in Applied Probability* 9, 9 (1977), 87–104.
- P. A. Jacobs and P. A. W. Lewis. 1978. Discrete Time Series Generated by Mixtures. I: Correlational and Runs Properties. *Journal of the Royal Statistical Society, Series B* 40, 1 (1978), 94–105.
- P. A. Jacobs and P. A. W. Lewis. 1983. Stationary Autoregressive Discrete Moving Average Time Series Generated by Mixtures. *Journal of the Royal Statistical Society, Series B* 4, 1 (1983), 19–36.
- S.-H. Kim and W. Whitt. 2013a. Appendix to Choosing Arrival Process Models for Service Systems: Tests of a Nonhomogeneous Poisson Process. (2013). Columbia University, <http://www.columbia.edu/~ww2040/allpapers.html>.
- S.-H. Kim and W. Whitt. 2013b. Appendix to the Power of Alternative Kolmogorov-Smirnov Tests Based on Transformations of the Data. (2013). Columbia University, <http://www.columbia.edu/~ww2040/allpapers.html>.
- S.-H. Kim and W. Whitt. 2013c. Choosing Arrival Process Models for Service Systems: Tests of a Nonhomogeneous Poisson Process. (2013). Columbia University, <http://www.columbia.edu/~ww2040/allpapers.html>.
- S.-H. Kim and W. Whitt. 2013d. The Power of Alternative Kolmogorov-Smirnov Tests Based on Transformations of the Data. (2013). Columbia University, <http://www.columbia.edu/~ww2040/allpapers.html>.
- P. A. W. Lewis. 1965. Some results on tests for Poisson processes. *Biometrika* 52, 1 (1965), 67–77.
- G. Pang and W. Whitt. 2012. The impact of dependent service times on large-scale service systems. *Manufacturing and Service Oper. Management* 14, 2 (2012), 262–278.
- S. M. Ross. 1996. *Stochastic Processes* (second ed.). Wiley, New York.
- G. R. Shorack and J. A. Wellner. 2009. *Empirical Processes with Applications in Statistics* (siam classics in applied mathematics ed.). SIAM, Philadelphia.
- R. Simard and P. L'Ecuyer. 2011. Computing the Two-Sided Kolmogorov-Smirnov Distribution. *Journal of Statistical Software* 39, 11 (2011), 1–18.
- W. Whitt. 1982. Approximating a point process by a renewal process: two basic methods. *Oper. Res.* 30 (1982), 125–147.