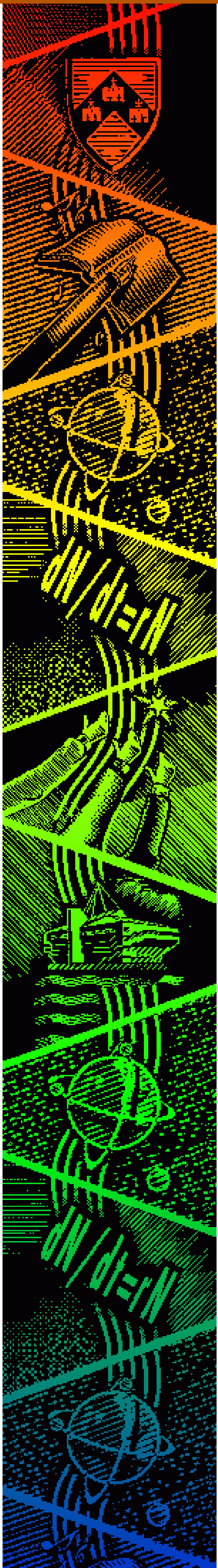

THE UNIVERSITY of York

Discussion Papers in Economics

No. 2005/25

A Two-Sample Non-Parametric Likelihood Ratio Test

by

Patrick Marsh

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

A Two-Sample Non-Parametric Likelihood Ratio Test[1]

Patrick Marsh

Department of Economics

University of York

Heslington, York

YO10 5DD

tel. +44 1904 433084

fax +44 1904 433759

e-mail: pwnm1@york.ac.uk

July 26, 2005

[1]Thanks are due to Francesco Bravo, Giovanni Forchini, Les Godfrey, Grant Hillier, Peter Phillips and participants at the York Econometrics Workshops, July 2005.

Abstract

This paper proposes a test for the hypothesis that two samples have the same distribution. The likelihood ratio test of Portnoy (1988) is applied in the context of the consistent series density estimator of Crain (1974) and Barron and Sheu (1991). It is proven that the test, when suitably standardised, is asymptotically standard normal and consistent against any complementary alternative. In comparison with the established Kolmogorov-Smirnov and Cramér-von Mises procedures the proposed test enjoys broadly comparable finite sample size properties, but vastly superior power properties.

1 Introduction

The problem of testing whether two independent samples are drawn from the same distribution is ubiquitous in applied statistics. Statistical tests for the two-sample hypothesis are commonly adaptations of tests that an identically and independently distributed sample has a particular, known, distribution. Examples are the Kolmogorov-Smirnov and Cramér-von Mises procedures, see Darling (1957) for detailed exposition in the single sample case and Kiefer (1959) for multi-sample extensions. A fuller account of these and other procedures can be found in Conover (1999).

This paper instead derives a test for the two-sample problem based upon the goodness of fit tests of Marsh (2005) and Claeskens and Hjort (2004). Those tests, although differing in terms of how the null hypothesis is imposed, are essentially the likelihood ratio test of Portnoy (1988), made non-parametric via the consistent exponential series density estimator of Crain (1974, 1976 & 1977) and Barron and Sheu (1991). Consequently it is relatively straightforward to establish the necessary asymptotic properties; first that the test, when appropriately standardised, is asymptotically standard normal and second that it is consistent against any complimentary alternative.

Of concern in applied research, for any suggested procedure, are three things; whether implementation is intuitive and straightforward, whether empirically relevant critical values (i.e. having actual size close to nominal) are readily available and whether the test offers power against theoretically relevant alternatives.

In terms of implementation, the Kolmogorov-Smirnov and Cramér-von Mises are based upon criteria utilising the sup and L_2 norms on the space of distributions. In particular the Cramér-von Mises has appeal, see Anderson (1962), in that the resultant test can be written entirely in terms of the respective ranks of observations from the two samples within a pooled sample. The proposed test is instead based upon the Kullback-Leibler (Entropy) distance, on the space of densities. Specifically, it is directly related to a likelihood ratio test for a simple hypothesis in the (albeit infinite) exponential family. In addition, as will be exposed below, the form of the statistic actually only depends upon the estimated parameters in that family and the raw moments of the samples themselves. In practice the proposed test involves only testing a simple hypothesis in the exponential family, with only the dimension of that family to be determined. Following Marsh (2005) the data driven selection criteria of both Akaike (1974) and Schwarz (1978) may be easily applied. Either criterion delivers test which is relatively straightforward to implement.

Regarding the availability of critical values, for the two-sample Cramér-von Mises procedure, Anderson (1962) provides a numerical approximation, while Kim (1969) provides an asymptotic distribution function for the Kolmogorov-Smirnov procedure. Although theoretically critical values may instead be found through simulation, having empirically relevant tabulated values is more convenient for applied problems. The proposed likelihood ratio statistic, when not standardised with respect to its degrees of freedom is asymptotically Chi-square. Critical values from standard Chi-square tables are shown, in this paper, to have finite sample numerical properties not dissimilar to those tabulated for the established tests. Further numerical evidence involving comparisons of these two established procedures while others can be found in Burr (1964) and Dufour and Farhat (2002).

Since all of these tests are distribution free, and hence critical values could be directly simulated, albeit at considerable numerical cost, we must consider numerical performance under the alternative. Power is examined by considering alternatives in which the samples have distributions differing in terms of their moments. Excepting the case of different means, where the performance is comparable, the proposed non-parametric likelihood ratio test has significantly more power. For certain alternatives involving distributions with different variances, skewness or kurtosis the proposed test may be as much as four or even five times more powerful than either of the established procedures.

The plan for the rest of the paper is as follows, the next section provides the main definitions and results for both the density estimator and the asymptotics for the resultant two-sample non-parametric likelihood ratio test. Section 3 details the numerical experiments of the paper and is followed by brief conclusions. An appendix contains the proof of the theorem containing the asymptotic results as well as tables containing the numerical results.

2 A Two-Sample Likelihood Ratio Test

Let $\{X_i\}_{i=1}^{n_X}$ and $\{Y_i\}_{i=1}^{n_Y}$ be i.i.d. samples taken from the random variables X and Y respectively, having common sample space $\mathbb{R}$. Let $F(\tau) = \Pr[X \leq \tau]$ and $G(\tau) = \Pr[Y \leq \tau]$, and suppose we wish to test

$$H_0 : F(\tau) = G(\tau) \quad \text{for all } \tau \in \mathbb{R},$$

against any complimentary alternative. In order to apply the likelihood ratio test of Portnoy (1988) in this context we will employ the exponential series density estimator first employed by Crain (1974, 1976 and 1977) and extended by Barron and Sheu (1991).

To proceed define the monotone function $h(\tau) : \mathbb{R} \rightarrow (-a, a)$, $a < \infty$, so that

$$x_i = h(X_i) \quad \text{and} \quad y_i = h(Y_i),$$

or generically $x = h(X)$ and $y = h(Y)$ and denote their density functions as $p_x(h)$ and $p_y(h)$, respectively. The null hypothesis implies that $p_x(h) = p_y(h) = p(h)$. Let $\phi_j(h)$, $j = 1, ..., m$ be a set of linearly independent functions spanning $(-a, a)$ then according to Barron and Sheu (1991) the exponential series estimator for $p(h)$ is the maximum likelihood estimator (mle) in the family

$$p_h(\theta) = \exp\left\{\sum_{j=1}^{m} \theta_j \phi_j(h) - \psi_m(\theta)\right\}, \tag{1}$$

where $\theta = (\theta_1, ..., \theta_m)'$ and the cumulant function is defined by

$$\psi_m(\theta) = \ln \int_{-a}^{a} \exp\left\{\sum_{j=1}^{m} \theta_j \phi_j(h)\right\} dh. \tag{2}$$

Details on the implementation of the estimator may be found in Marsh (2005).

To proceed, given samples $\{x_i\}_{i=1}^{n_X}$ and $\{y_i\}_{i=1}^{n_Y}$ define the following vectors in $\mathbb{R}^m$

$$\left.\begin{array}{lll} \hat{\theta}_x & : & \int_{-a}^{a} \phi_j(h) p_h(\hat{\theta}_x) dh = \frac{1}{n_X} \sum_{i=1}^{n_X} \phi_j(x_i) \\ \\ \hat{\theta}_y & : & \int_{-a}^{a} \phi_j(h) p_h(\hat{\theta}_y) dh = \frac{1}{n_Y} \sum_{i=1}^{n_Y} \phi_j(y_i) \\ \\ \theta_{0x} & : & \int_{-a}^{a} \phi_j(h) p_h(\theta_{0x}) dh = \int_{-a}^{a} \phi_j(h) p_x(h) dh \\ \\ \theta_{0y} & : & \int_{-a}^{a} \phi_j(h) p_h(\theta_{0y}) dh = \int_{-a}^{a} \phi_j(h) p_y(h) dh \end{array}\right\} \quad j = 1, .., m, \tag{3}$$

and define the Sobolev space of functions, W_2^r, so that $f(x) \in W_2^r$ if the $(r-1)^{th}$ derivative of $f(.)$ is absolutely continuous and the r^{th} derivative is square integrable. The pertinent results of Crain (1973, 1974 and 1976) and Barron and Sheu (1991) can be summarized in the following lemma:

Lemma 1 *Let $\ln(p(h)) \in W_2^r$ with $r \geq 2$ and suppose that as n_X, n_Y and $m \to \infty$,*

$$\frac{n_X}{n_Y} = O(1) \text{ and } \frac{m^3}{n_X} = o(1),$$

then:

(i) For all n_X, n_Y and m, $\hat{\theta}_x$ and $\hat{\theta}_y$ exist and are unique and as $m \to \infty$, θ_{0x} and θ_{0y} exist and are unique.

(ii) Let $D(p_1|p_2)$ denote the Kullback-Leibler divergence then

$$\begin{aligned} D\left(p_h(\theta_{0x})|p_x(h)\right) &= O_r\left(m^{-2r}\right) \\ D\left(p_h(\hat{\theta}_x)|p_x(h)\right) &= O_p\left(\frac{m}{n_X} + m^{-2r}\right), \end{aligned}$$

and similarly for $D\left(p_h(\hat{\theta}_y)|p(h)\right)$.

(iii) Let $|.|$ *be Euclidean distance in* $\mathbb{R}^m$, *then* $|\hat{\theta}_x - \theta_{0x}|$, $|\hat{\theta}_y - \theta_{0y}|$ *are* $O_p\left(\sqrt{m/n_X}\right)$.

■

The importance of the set of results comprising Lemma 1 is that we can asymptotically approximate the densities $p_x(h)$ and $p_y(h)$ by $p_h(\theta_{0x})$ and $p_h(\theta_{0y})$. Moreover, since θ_{0x} and θ_{0y} are the unique solutions to lines 3 and 4 of (3) then the two-sample hypothesis can (asymptotically in m) be reformulated as

$$\lim_{m\to\infty} H_0^* : \theta_{0x} = \theta_{0y} \quad \text{vs.} \quad H_1 : \theta_{0x} \neq \theta_{0y}. \tag{4}$$

Consequently, the test statistic can be formulated in terms of likelihood ratio tests for the (asymptotically) simple hypothesis $\theta_{0x} = \theta_{0y}$ in the exponential family (1). That is, rather than being based upon either the sup or L_2 norms on distributions of the Kolmogorov-Smirnov and Cramér-von Mises procedures, here we exploit the Kullback-Leibler (relative entropy) distance on densities.

The crucial asymptotic results required for the two-sample case follow almost trivially from the single sample. First, notice that under the null hypothesis (4) and via the triangle inequality

$$|\hat{\theta}_x - \hat{\theta}_y| \leq |\hat{\theta}_x - \hat{\theta}_{x0}| + |\hat{\theta}_y - \hat{\theta}_{x0}| = O_p\left(\sqrt{\frac{m}{n_X}}\right), \tag{5}$$

while given the additivity of the Kullback-Leibler divergence

$$D\left(p_h(\hat{\theta}_x)|p_h(\hat{\theta}_y)\right) = O_p\left(\frac{m}{n_X} + m^{-2r}\right). \tag{6}$$

In order to test H_0^* in (4) define, for a sample $h_1, .., h_n$,

$$p_{\hat{\theta}}\left(\underline{h}\right) = p_{\hat{\theta}}\left(h_1, .., h_n\right) = \lim_{m,n\to\infty\,,\, m^3/n\to 0}\ \sup_{\theta\in\mathbb{R}^m} \exp\left\{\sum_{j=1}^{m}\theta_j \sum_{i=1}^{n}\phi_j(h_i) - n\psi_m\left(\theta\right)\right\}. \tag{7}$$

Thus we can write a likelihood ratio

$$\Lambda^{\underline{h}}_{\theta_1,\theta_2} = \ln\left(\frac{p_{\theta_1}\left(\underline{h}\right)}{p_{\theta_2}\left(\underline{h}\right)}\right),$$

which is, implicitly, a test for the simple null hypothesis $\theta = \theta_1$ against $\theta = \theta_2$ using a sample of size n on h. From this can define our two-sample likelihood ratio test,

$$\Lambda_{\hat{\theta}_x,\hat{\theta}_y}^{\underline{x},\underline{y}} = \Lambda_{\hat{\theta}_x,\hat{\theta}_y}^{\underline{x}} + \Lambda_{\hat{\theta}_x,\hat{\theta}_y}^{\underline{y}} = \ln\left(\frac{p_{\hat{\theta}_x}(\underline{x})}{p_{\hat{\theta}_y}(\underline{x})}\right) + \ln\left(\frac{p_{\hat{\theta}_y}(\underline{y})}{p_{\hat{\theta}_x}(\underline{y})}\right). \tag{8}$$

Alternatively, in terms of Portnoy's (1988) test we can further decompose so that

$$\Lambda_{\hat{\theta}_x,\hat{\theta}_y}^{\underline{x},\underline{y}} = \ln\left(\frac{p_{\hat{\theta}_x}(\underline{x})}{p_{\theta_0}(\underline{x})}\right) + \ln\left(\frac{p_{\theta_0}(\underline{x})}{p_{\hat{\theta}_y}(\underline{x})}\right) + \ln\left(\frac{p_{\hat{\theta}_y}(\underline{y})}{p_{\theta_0}(\underline{y})}\right) + \ln\left(\frac{p_{\theta_0}(\underline{y})}{p_{\hat{\theta}_x}(\underline{y})}\right), \tag{9}$$

where under the null; $\theta_0 = \theta_{0x} = \theta_{0y}$. That is the decomposition of the likelihood ratio in (9) implies that the test can be interpreted as the sum of four likelihood ratios for testing the two hypotheses; $\theta_{0x} = \theta_0$ and $\theta_{0y} = \theta_0$ using both of the samples on X and Y.

Formally we will reject $H_0^* : \theta_{0x} = \theta_{0y}$ in favour of $H_1 : \theta_{0x} \neq \theta_{0y}$ if

$$\Lambda_{\hat{\theta}_x,\hat{\theta}_y}^{\underline{x},\underline{y}} > k, \tag{10}$$

where k is a suitably chosen critical value. As mentioned in the introduction, there is a relative dearth of distributional results for nonparametric two-sample tests, particularly in comparison with their one sample counterparts. Therefore the following Theorem demonstrates that a standardised version of the test is asymptotically standard normal under H_0^* and that under H_1 the test is consistent.

Theorem 1 *(i) Let*

$$\lambda_{n_X,n_Y} = \frac{\Lambda_{\hat{\theta}_x,\hat{\theta}_y}^{\underline{x},\underline{y}} - m}{\sqrt{m}},$$

then

$$\lambda_{n_X,n_Y} \to_d N(0,1),$$

(ii) define k_α *by* $\Pr[N[0,1] > k_\alpha | H_0] = \alpha > 0,$ *then*

$$\Pr[\lambda_{X,Y} > k_\alpha | H_1] \to 1,$$

as $n_X,\ n_Y\ \&\ m \to \infty$ *and* $m^3/n_X \to 0.$ ■

In summary, Theorem 1 establishes the necessary asymptotic theory for the two-sample test. Specifically, and unlike current tests, we have a standard asymptotic distribution under the null. Moreover consistency follows, almost trivially, from the properties of likelihood ratio tests in the exponential family.

Implementation of the test is particularly straightforward. From (7) we have

$$\Lambda^{\underline{x}}_{\hat{\theta}_x,\hat{\theta}_y} = n_X \left[\left(\hat{\theta}_x - \hat{\theta}_y\right)' \bar{x} - \left(\psi_m(\hat{\theta}_x) - \psi_m(\hat{\theta}_y)\right) \right],$$

and similarly for $\Lambda^{\underline{y}}_{\hat{\theta}_x,\hat{\theta}_y}$, so that from (8), we obtain

$$\Lambda^{\underline{x},\underline{y}}_{\hat{\theta}_x,\hat{\theta}_y} = n_X \left[\left(\hat{\theta}_x - \hat{\theta}_y\right)' (\bar{x} - \frac{n_Y}{n_X}\bar{y}) - (1 - \frac{n_Y}{n_X}) \left(\psi_m(\hat{\theta}_x) - \psi_m(\hat{\theta}_y)\right) \right].$$

Moreover, in the interesting special case of $n_X = n_Y$ the standardised test simplifies as

$$\lambda_n = \frac{n_X \left(\hat{\theta}_x - \hat{\theta}_y\right)' (\bar{x} - \bar{y}) - m}{\sqrt{m}} \to_d N(0,1),$$

as $m, n_X \to \infty$ and $m^3/n_X \to 0$.

3 Numerical Analysis

Theorem 1 demonstrates that asymptotically and under the null hypothesis λ_{n_X,n_Y} is standard normal. In practice, in order for the asymptotic normal distribution to serve as a reliable approximation m must be large. This implies a large number of estimating equations in (3) and hence the potential is for the whole process to become infeasible. Specifically, other experiments involving the single sample version of the Portnoy (1988) test, which will not be reported here, suggest that it is only for values of m above 15 with sample sizes above 500 that the standard normal provides an acceptable approximation. Instead, as in Portnoy (1988), we can instead utilise the asymptotic relation

$$\Lambda_m = 2\Lambda^{\underline{x},\underline{y}}_{\hat{\theta}_x,\hat{\theta}_y} \to_d \chi^2_{(2m)}, \tag{11}$$

as n_X, n_Y & $m \rightarrow \infty$ and $m^3/n_X \rightarrow 0$. That is the Chi-square may be used as a distributional approximation and, as will be demonstrated numerically, provides a reasonable approximation to the finite sample distribution of Λ_m.

More important than the choice of asymptotic benchmark will be how to choose the dimension of the model, m. Since the likelihood ratios into which the statistic $\Lambda^{\underline{x},\underline{y}}_{\hat{\theta}_x,\hat{\theta}_y}$ may be partitioned are all likelihood ratio statistics for testing simple hypotheses in the exponential well established data driven selection criteria may be employed. Specifically, here we will use both the Akaike (1974) information criterion (AIC) as well as the Bayesian information criterion (BIC) of Schwarz (1978). To implement these, define the set of integers $\mathbb{M} = \{1, 2, ..., \bar{m}\}$, and let the estimated dimensions based upon these criteria be $\hat{m}_A$ and $\hat{m}_B$, respectively, then they satisfy

$$\begin{aligned} \hat{m}_A &= \arg\max_{m\in\mathbb{M}} \left[L_X(\hat{\theta}_x) + L_Y(\hat{\theta}_y) - 2m \right] \\ \hat{m}_B &= \arg\max_{m\in\mathbb{M}} \left[L_X(\hat{\theta}_x) + L_Y(\hat{\theta}_y) - m\ln(n_X n_Y) \right], \end{aligned} \tag{12}$$

where

$$\begin{aligned} L_X(\hat{\theta}_x) &= \sum_{i=1}^{n_X}\sum_{k=1}^{m}(\hat{\theta}_x)_k\phi_k(x_i) - n_X\psi_m\left(\hat{\theta}_x\right) \\ L_Y(\hat{\theta}_y) &= \sum_{i=1}^{n_Y}\sum_{k=1}^{m}(\hat{\theta}_y)_k\phi_k(y_i) - n_Y\psi_m\left(\hat{\theta}_y\right). \end{aligned}$$

Notice that the criteria in (12) impose a common dimension on the estimators for both samples. It would be possible to optimise separately, however this would merely introduce an unnecessary complication in terms of using the chi-square distribution as an approximation. In any case, since both $L_X(\theta) = O(n_X)$ and $L_Y(\theta) = O(n_Y)$ for all θ, then

$$\text{as } n_X, \bar{m} \rightarrow \infty \ , \ \bar{m}^3/n_X \rightarrow 0 \quad \begin{array}{c} \hat{m}_A \rightarrow \infty \\ \hat{m}_B \rightarrow \infty \end{array}.$$

That is if we allow $\bar{m}$ to grow slowly relative to n_X then both criteria will deliver a consistent density estimator for each sample. We shall denote the two resulting test statistics as $\Lambda_{\hat{m}_A}$ for that based upon the AIC and $\Lambda_{\hat{m}_B}$ for that based upon the BIC.

The numerical properties of $\Lambda_{\hat{m}_A}$ and $\Lambda_{\hat{m}_B}$ will be compared with those of the two-sample version of the Kolmogorov-Smirnov and Cramér–von Mises tests, defined in our notation by

$$\begin{aligned} KS &= \sqrt{\frac{n_X n_Y}{n_X + n_Y}} \sup_i |F_{n_X}(x_i) - G_{n_Y}(y_i)|, \\ CM &= \frac{n_X n_Y}{(n_X + n_Y)^2} \left[\sum_{i=1}^{n_X} (F_{n_X}(x_i) - F_{n_X}(y_i))^2 + \sum_{i=1}^{n_Y} (G_{n_Y}(y_i) - G_{n_Y}(x_i))^2 \right], \end{aligned} \tag{13}$$

where $F_{n_X}(.)$ and $G_{n_Y}(.)$ are the empirical distribution functions of the samples obtained from X and Y respectively. Although these are not the only two-sample tests currently available, they do enjoy the advantage of having relatively well understood asymptotic distributions. For the KS test from Kim (1969) as $n_X,\ n_Y\ \rightarrow \infty$ then

$$\Pr[KS \leq s] \rightarrow G(s) = 1 - 2\sum_{r=1}^{\infty} (-1)^{r-1} e^{-2r^2 s^2} \quad ; \quad s > 0.$$

Asymptotic critical values of nominal size α can be obtained via solution for c_α^{KS} of

$$\alpha = 2\sum_{r=1}^{k} (-1)^{r-1} e^{-2r^2 \left(c_\alpha^{KS}\right)^2},$$

and to 4 decimal places we find

$$c_{0.05}^{KS} = 1.358 \quad ; \quad cv_{0.1}^{KS} = 1.220.$$

Also from Anderson (1962) CM has the same asymptotic distribution as the single sample Cramér–von Mises test, as tabulated in Anderson and Darling (1952). Thus for the CM test we have asymptotic critical values

$$c_{0.05}^{CM} = 0.4614 \quad ; \quad cv_{0.1}^{CM} = 0.3473.$$

Asymptotic critical values for any Λ_m test say cv_α^m, are obtained from $\Pr[\chi_m^2 \leq cv_\alpha^m] = 1 - \alpha$.

The first set of three experiments concern the finite sample performance of the asymptotic critical values as approximations to the finite sample distribution. For the purposes all of the experiments to follow the set $M = \{3, 4, 5\}$ was optimised over to construct both the $\Lambda_{\hat{m}_A}$ and $\Lambda_{\hat{m}_B}$ tests, for sample sizes of $n_X = n_Y = n = 50, 100, 200, 400$. All of the experiments are based upon 5000 Monte Carlo replications. All computations were performed on a Pentium IV 3.0GHz P.C. running Mathematica 4.0. A single replication of all 4 statistics took between 1 and 2 seconds, depending upon the sample size.

The first experiment consisted of generating samples $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ from $X \sim Y \sim N(0, 1)$, and then samples $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$ lying in $(0, 1)$ via

$$x_i = \sqrt[3]{\Phi(X_i)} \quad ; \quad y_i = \sqrt[3]{\Phi(Y_i)} \tag{14}$$

where $\Phi(.)$ is the standard normal CDF. The second experiment used the standard exponential distribution, i.e. $X \sim Y \sim Exp[1]$, with then the Exponential CDF replacing the standard normal in (14). Lastly, the third experiment varies the set up slightly, in that we assume that $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ are generated from

$$X \sim Y \sim \begin{cases} N(-1, 1) & \text{with prob. } 0.5 \\ N(1, 1) & \text{with prob. } 0.5 \end{cases},$$

and now $x_i = F(X_i)$ where $F(.)$ is the CDF of a $N(0, 2)$ random variable, the y_i are defined similarly.

Deliberately neither x_i nor y_i are constructed to have the uniform distribution on $(0, 1)$. The reason, as Marsh (2005) highlights, is that for the selection criteria for the likelihood ratio tests cannot be consistent, since the uniform is embedded within the exponential with $m = 0$ and $\mathbb{M}$ cannot contain 0. For the tests based upon the empirical distribution functions this issue is irrelevant.

The Monte Carlo rejection probabilities are presented in, respectively, Tables 1,2 and 3 in the Appendix. The asymptotic critical values for KS are generally undersized while those for CM are slightly oversized. The asymptotic Chi-square critical values used for the $\Lambda_{\hat{m}_A}$ and $\Lambda_{\hat{m}_B}$ tests are oversized, having Monte Carlo size slightly closer to nominal than is the case for those for the KS test, slightly further than those for the CM test. Since the Chi-square approximation seems to work better for smaller m, for a given n, the BIC, which favours parsimony, has a very slender advantage. However, the only tangible conclusion that may be reached is that the results in Tables 1 through 3 indicate a consistency in performance in a variety of circumstances.

On the sole basis of the finite sample performance of asymptotic critical values there is very little basis for preferring one procedure over another. However, five further experiments examine the comparative power of the proposed tests. For each of these experiments, again assuming $n_X = n_Y = n$, the i.i.d. sample $\{X_i\}_{i=1}^n$ was generated according to

$$X \sim N(0,1),$$

while alternatives were considered by generating i.i.d. samples $\{Y_i\}_{i=1}^n$ according to

$$\begin{array}{llll}
H_1^A: & Y \sim N(\mu,1) & ; & \mu = .1, .2, .3, .4, .5 \\
H_1^B: & Y \sim N(0,(1+\mu)^2) & ; & \mu = .1, .2, .3, .4, .5 \\
H_1^C: & Y \sim \frac{\chi^2(v)-v}{\sqrt{2v}} & ; & v = 35,\ 30, 25, 20, 15, 10, 5 \\
H_1^D: & Y \sim \sqrt{\frac{v-2}{v}}t_v & ; & v = 12, 10, 8, 6, , 4 \\
H_1^E: & Y \sim \begin{cases} N(-\mu,1) & \text{with prob. } 0.5 \\ N(\mu,1) & \text{with prob. } 0.5 \end{cases} & ; & \mu = .1, .3, .5, .7, .9,
\end{array},$$

where $\chi^2(v)$ and t_v represent Chi-Square and Student-t random variables on v degrees of freedom.

The first four alternatives attempt to classify alternatives in terms of departures in successive moments, the mean, variance skewness and kurtosis, while holding other moments constant. Notice though that the excess kurtosis of the standardised Chi-Square is in fact $12/v$. The final experiment considers alternatives which are bi-modal. For alternatives A, B and D, the null hypothesis is satisfied for $\mu = 0$, while for alternatives C and D is obtained in the limit as $v \to \infty$.

The simulated powers, based upon Monte Carlo critical values at the 5% significance level, of the likelihood ratio tests $\Lambda_{\hat{m}_A}$ and $\Lambda_{\hat{m}_B}$ and of the KS and CM statistics are presented in Tables 3 through 8 in the Appendix. Collectively the powers of the two likelihood ratio information criteria based tests are similar. Likewise the KS and CM have similar power properties to each other. For alternative A, differing means, in fact the established tests have a slight power advantage. However for all other moment departures the $\Lambda_{\hat{m}_A}$ and $\Lambda_{\hat{m}_B}$ tests enjoy a significant power advantage. In fact, in many cases, the power is several orders of magnitude higher. The same is true for the bi-modal alternatives.

As with the one-sample versions of the Kolmogorov-Smirnov and Cramér-von Mises tests, it would be possible, in principle, to utilise weighting functions, other than the unit, such as the particular case in Anderson and Darling (1952). For example, we might expect that tests with weight specifically in one tail or the other should fair better against skewed alternatives.

However, before concluding that it must therefore be possible to find versions of the established tests with powers comparable with those proposed here several limitations must be considered. Such tests are not yet really feasible, although Canner (1975) has some limited numerical evidence for a particularly weighted version of the Kolmogorov-Smirnov. Their asymptotic distributions will be nonstandard, and moreover as in Anderson and Darling (1952) will have to be developed on a case-by-case basis. Even then, as Marsh (2005) finds in the single sample goodness of fit case, a particular weighting function might deliver high power against certain alternatives, but only at the expense of power against other alternatives. Consequently, in the absence of explicit knowledge of the direction, at least, of the alternative it would be difficult to justify any particular weighted version of either the Kolmogorov-Smirnov and Cramér-von Mises tests.

Notice also that under the assumptions of the paper the two-sample hypothesis can be parameterised in terms of an infinite exponential family. Save for the use of the density estimator the test is essentially the Portnoy (1988) test applied to the problem of testing the simple hypothesis $H_0^* : \theta_{0x} = \theta_{0y}$. Although no nonparametric test, in this setting, can claim optimality, at least in the asymptotic regime $m, n \to \infty$, $m^3/n \to 0$ the proposed test will be coincident with a point optimal test for H_0^*.

4 Conclusions

This paper has proposed nonparametric two-sample tests based upon the criteria of Portnoy (1988), exploiting the series density estimator of rain (1974) and Barron and Sheu (1991). The asymptotic distribution of the test is standard under the null and diverges under the alternative, ensuring consistency. Numerical evidence suggests the finite sample performance of asymptotic critical values for the test is at least equivalent to those for the Kolmogorov-Smirnov test, slightly worse than those for the Cramér–von Mises tests. On the other hand evidence is presented which indicates a clear power superiority for the tests proposed in this paper.

References

Akaike, H. 1974. A new look at the statistical model identification. System identification and time-series analysis. IEEE Trans. Automatic Control AC-19, 716–723

Anderson, T.W. 1962. On the distribution of the two-sample Cramér-von Mises criterion. Annals of Mathematical Statistics, 33, 1148-1159.

Anderson, T.W. and D.A. Darling 1952. Asymptotic theory of certain 'goodness-of-fit' criteria based on stochastic processes. Annals of Mathematical Statistics, 23, 193-212.

Barron, A.R. and C-H. Sheu 1991. Approximation of density functions by sequences of exponential families. Annals of Statistics, 19, 1347-1369.

Burr, E.J. 1964. Small-sample distributions of the two-sample Cramér-von Mises' W^2 and Watson's U^2. Annals of Mathematical Statistics, 35, 1091–1098.

Canner, P.L. 1975. A simulation study of one- and two-sample Kolmogorov-Smirnov Statistics with a particular weight function. Journal of the American Statistical Association, 70, 209-211.

Chow, Y.S. and H. Teicher 1988. Probability theory, 2^{nd} ed., Springer-Verlag, New York.

Claeskens, G. and N.L. Hjort 2004. Goodness of fit via nonparametric likelihood ratios. Skandanavian Journal of Statistics, 31, 487-513.

Crain, B.R. 1974. Estimation of distributions using orthogonal expansions. Annals of Statistics, 2, 454–463.

Crain, B.R. 1976. More on estimation of distributions using orthogonal expansions. Journal of the American Statistical Association, 71, 741–745.

Crain, B.R. 1977. An information theoretic approach to approximating a probability distribution. SIAM Journal of Applied Mathematics, 32, 339–346.

Conover, W.J. 1999. Practical Nonparametric Statistics, John Wiley and Sons, New York.

Darling, D.A. 1957. The Kolmogorov-Smirnov, Cramér-von Mises tests. Annals of Mathematical Statistics, 28, 223-238.

Dufour, J-M. and A. Farhat 2002. Exact nonparametric two-sample homogeneity tests. Goodness-of-fit tests and model validity (Paris, 2000), Statistics for Industry and Technology, Birkhäuser Boston, 435–448.

Kiefer, J. 1959. K-sample analogues of the Kolmogorov-Smirnov and Cramér-v. Mises tests. Annals of Mathematical Statistics, 30, 420-447.

Kim, P.J. 1969. On the exact and approximate sampling distribution of the two-sample Kolmogorov-Smirnov criterion D_{mn}, $m \leq n$. Journal of the American Statistical Association, 64, 1625–1637.

Marsh, P.W.N. 2005. Goodness of fit tests via exponential series density estimation, 05/24, University of York.

Portnoy, S. 1988. Asymptotic behaviour of likelihood methods for exponential families when the number of parameters tends to infinity. Annals of Statistics, 16, 356-366.

Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics, 6, 461-464.

Appendix

Proof of Theorem 1.

Proof. For part (i) we shall consider the likelihood ratios based on the two samples separately. Define $\bar{x} = \frac{1}{n_X} \sum_{i=1}^{n_X} \phi_i^x$, where $\phi_i^x = (\phi_1(x_i), .., \phi_m(x_i))'$, and so for the sample on X we have the series density estimator

$$p_{\hat{\theta}_x}(\underline{x}) = \exp\left\{ n_X \left(\hat{\theta}_x' \bar{x} - \psi_m\left(\hat{\theta}_x\right) \right) \right\},$$

where $\hat{\theta}_x$ is the solution to the first line in (3). On basis of this estimated density we can define the (log-)likelihood ratio by,

$$\Lambda^{\underline{x}}_{\hat{\theta}_x, \hat{\theta}_y} = \ln\left(\frac{p_{\hat{\theta}_x}(\underline{x})}{p_{\hat{\theta}_y}(\underline{x})} \right) = n_X \left\{ (\hat{\theta}_x - \hat{\theta}_y)'\bar{x} - \left(\psi_m(\hat{\theta}_x) - \psi_m(\hat{\theta}_y) \right) \right\},$$

where $\hat{\theta}_y$ now solves the second line in (3). The expansions given in equations (2.1)-(2.3) of Portnoy (1988) hold for any two values in $\mathbb{R}^m$, and so we can write

$$\psi_m'\left(\hat{\theta}_x\right) = \psi_m'\left(\hat{\theta}_y\right) + (\hat{\theta}_x - \hat{\theta}_y)'\psi_m''\left(\hat{\theta}_y\right) + \frac{1}{2} E_{\hat{\theta}_y}\left[\left((\hat{\theta}_x - \hat{\theta}_y)' U_x \right)^2 U_x \right] + \Lambda(\tilde{\theta}), \quad (15)$$

where $\Lambda(.)$ is a remainder and $\tilde{\theta}$ lies on a line segment joining $\hat{\theta}_x$ and $\hat{\theta}_y$ and $U_x = V_x - E_{\hat{\theta}_y}[V_x]$, $V_x \sim p_h\left(\hat{\theta}_y\right)$ and V_x is distributed independently of X.

Since $\Lambda^{\underline{x}}_{\hat{\theta}_x, \hat{\theta}_y}$ is a likelihood ratio it is invariant to reparametrisations of the form $\theta \to \alpha + \beta\theta$, which will be exploited here. Moreover, $p_{\hat{\theta}_x}(x)$ is in the exponential family and hence so is $p_{\hat{\theta}_x}(a+bx)$, consequently and without loss of generality we can assume that $\bar{x}$ is standardised in that

$$\psi'_m\left(\hat{\theta}_y\right)=0 \quad \text{and} \quad \psi''_m\left(\hat{\theta}_y\right)=I_m. \tag{16}$$

Therefore, and since by definition $\psi'_m\left(\hat{\theta}_x\right)=\bar{x}$ we can rewrite (15) as

$$(\hat{\theta}_x-\hat{\theta}_y)=\bar{x}-\frac{1}{2}E_{\hat{\theta}_y}\left[\left((\hat{\theta}_x-\hat{\theta}_y)'U_x\right)^2U_x\right]+\Lambda(\tilde{\theta}). \tag{17}$$

Multiplying (17) first by $\bar{x}'$ we have

$$(\hat{\theta}_x-\hat{\theta}_y)'(\hat{\theta}_x-\hat{\theta}_y)=(\hat{\theta}_x-\hat{\theta}_y)'\bar{x}-\frac{1}{2}E_{\hat{\theta}_y}\left[\left((\hat{\theta}_x-\hat{\theta}_y)'U_x\right)^3\right]+(\hat{\theta}_x-\hat{\theta}_y)'\Lambda(\tilde{\theta}) \tag{18}$$

while instead multiplying (17) by $(\hat{\theta}_x-\hat{\theta}_y)'$ we get

$$(\hat{\theta}_x-\hat{\theta}_y)'\bar{x}=\bar{x}'\bar{x}-\frac{1}{2}E_{\hat{\theta}_y}\left[\left((\hat{\theta}_x-\hat{\theta}_y)'U_x\right)^2(\bar{x}'U_x)\right]+\bar{x}'\Lambda(\tilde{\theta}). \tag{19}$$

Thus subtracting (19) from (18) yields

$$\begin{aligned}|(\hat{\theta}_x-\hat{\theta}_y)'-\bar{x}| \;&=\; \frac{1}{2}E_{\hat{\theta}_y}\left[\left((\hat{\theta}_x-\hat{\theta}_y)'U_x\right)^2\left(\bar{x}-(\hat{\theta}_x-\hat{\theta}_y)\right)'U_x\right]\\ &\quad+\left((\hat{\theta}_x-\hat{\theta}_y)-\bar{x}\right)'\Lambda(\tilde{\theta}).\end{aligned}$$

Noting that from (5) under H_0^* we have $|\hat{\theta}_x-\hat{\theta}_y|=O_p\left(\sqrt{m/n_X}\right)$ and moreover since the elements of U_x are bounded the moment condition required for Theorem 3.1 of Portnoy (1988) are automatically satisfied, then as there it is true that

$$\left((\hat{\theta}_x-\hat{\theta}_y)-\bar{x}\right)'\Lambda(\tilde{\theta})=O_p\left(\left(\frac{m}{n_X}\right)^2\right). \tag{20}$$

Using the inequality

$$\begin{aligned}&\left|E_{\hat{\theta}_y}\left[\left(\bar{x}-(\hat{\theta}_x-\hat{\theta}_y)\right)'U_x\left((\hat{\theta}_x-\hat{\theta}_y)'U_x\right)^2\right]\right|\\ \leq\;& E_{\hat{\theta}_y}\left[\left(\left(\bar{x}-(\hat{\theta}_x-\hat{\theta}_y)\right)'U_x\right)^2\right]^{1/2}E_{\hat{\theta}_y}\left[\left((\hat{\theta}_x-\hat{\theta}_y)'U_x\right)^4\right]^{1/2},\end{aligned}$$

we then find,

$$\left|\frac{1}{2}E_{\hat{\theta}_y}\left[\left((\hat{\theta}_x-\hat{\theta}_y)'U_x\right)^2\left(\bar{x}-(\hat{\theta}_x-\hat{\theta}_y)\right)'U_x\right]\right|=O_p\left(\left(\frac{m}{n_X}\right)^2\right),$$

similar to equation (3.7) of Portnoy (1988), although it should be noted that this applies for our standardising coordinates implying (16).

If instead we substitute (19) into (18) rather than subtracting then on account of (20) we also obtain

$$|(\hat{\theta}_x - \hat{\theta}_y)'(\hat{\theta}_x - \hat{\theta}_y) - \bar{x}'\bar{x}| = O_p\left(\left(\frac{m}{n_X}\right)^2\right).$$

Again noting (16) the likelihood ratio permits a Taylor expansion with remainder of the form,

$$\begin{aligned} 2\Lambda^{\underline{x}}_{\hat{\theta}_x,\hat{\theta}_y} &= 2\left\{\ln\left(p_{\hat{\theta}_x}\left(\underline{x}\right)\right) - \ln\left(p_{\hat{\theta}_y}\left(\underline{x}\right)\right)\right\} \\ &= n_X\left\{(\hat{\theta}_x - \hat{\theta}_y)'(\hat{\theta}_x - \hat{\theta}_y) + \frac{1}{3}E_{\theta^*}\left[\left(\left(\hat{\theta}_x - \hat{\theta}_y\right)' U_{x^*}\right)^3\right]\right\}, \end{aligned}$$

where $\theta^* \in \left(\hat{\theta}_x, \hat{\theta}_y\right)$ and $U_{x^*} = V_{x^*} - E_{\theta^*}\left[V_{x^*}\right]$, while also noting that again since U_{x^*} is a bounded random variable, then

$$E_{\theta^*}\left[\left(\left(\hat{\theta}_x - \hat{\theta}_y\right)' U_{x^*}\right)^3\right] = O_p\left(|\hat{\theta}_x - \hat{\theta}_y|^3\right) = O_p\left(\left(\frac{m}{n_X}\right)^{3/2}\right),$$

and so

$$2\Lambda^{\underline{x}}_{\hat{\theta}_x,\hat{\theta}_y} = n_X\bar{x}'\bar{x} + O_p\left(\sqrt{\frac{m^3}{n_X}}\right).$$

Since

$$\frac{2\Lambda^{\underline{x}}_{\hat{\theta}_x,\hat{\theta}_y} - m}{\sqrt{2m}} = \frac{n_X\bar{x}'\bar{x} - m}{\sqrt{2m}} + O_p\left(\sqrt{\frac{m^2}{n_X}}\right),$$

and given the martingale central limit theorem, for example Theorem 9.3.1 of Chow and Teicher (1988), which implies that

$$\frac{n_X\bar{x}'\bar{x} - m}{\sqrt{2m}} \to_d N(0,1),$$

so that noting $m^2/n_X \to 0$ then

$$\frac{2\Lambda^{\underline{x}}_{\hat{\theta}_x,\hat{\theta}_y} - m}{\sqrt{2m}} \sim N(0,1) + o_p(1).$$

To complete part (i) consider the sample $\{y_i\}_{i=1}^n$ derived form Y, then proceeding exactly as above, for

$$\Lambda^{y}_{\hat{\theta}_x,\hat{\theta}_y} = n_Y \left\{ (\hat{\theta}_y - \hat{\theta}_x)'\bar{y} - \left(\psi_m(\hat{\theta}_y) - \psi_m(\hat{\theta}_x) \right) \right\},$$

we can reparameterise according to $\theta \to \gamma = a + b\theta$, so that now

$$\psi'_m(\hat{\gamma}_x) = 0 \quad \text{and} \quad \psi''_m(\hat{\gamma}_x) = I_m.$$

Proceeding exactly as above and noting the invariance of the likelihood ratio we will have,

$$\frac{2\Lambda^{y}_{\hat{\theta}_x,\hat{\theta}_y} - m}{\sqrt{2m}} = \frac{n_X \bar{x}'\bar{x} - m}{\sqrt{2m}} + O_p\left(\sqrt{\frac{m^2}{n_Y}}\right) \sim N(0,1) + o_p(1).$$

Then since X and Y are independent, adding the likelihood ratios gives

$$\lambda = \frac{2\left(\Lambda^{x}_{\hat{\theta}_x,\hat{\theta}_y} + \Lambda^{y}_{\hat{\theta}_x,\hat{\theta}_y}\right) - 2m}{\sqrt{2m}} \to_d N(0,2) + o_p(1),$$

which proves part (i). ■

For part (ii) put $\theta_{0x} = \theta_0 \neq \theta_{0y}$ under H_1. Though $\hat{\theta}_x$ and $\hat{\theta}_y$ still exist and are unique, now $|\hat{\theta}_x - \theta_0|$ and $|\hat{\theta}_y - \theta_0|$ are $O_p(m)$. The uniqueness of θ_{0x} and θ_{0y} and convexity of the exponential density imply that

$$\theta_{0y} \neq \theta_{0x} \Rightarrow \psi_m(\theta_{0x}) \neq \psi_m(\theta_{0y}),$$

and hence

$$n_X \left(\psi_m(\theta_{0x}) - \psi_m(\theta_{0y})\right) = O(n_X).$$

Further, since the $x_i's$ and hence the $\phi_k(x_i)'s$ are i.i.d. with mean zero, the individual elements of $\sqrt{n_X}\bar{x}$ satisfy

$$\frac{1}{\sqrt{n_X}} \left(\sum_{i=1}^{n_X} \phi_k(x_i) \right) = O_p(1),$$

which follows from a standard central limit theorem. Differentiability of $\psi_m(.)$ over $\mathbb{R}^m$ ensures that under H_1

$$2\Lambda^{x}_{\hat{\theta}_x,\hat{\theta}_y} \quad = \quad n_X \left\{ (\hat{\theta}_x - \hat{\theta}_y)'\bar{x} - \left(\psi_m(\hat{\theta}_x) - \psi_m(\hat{\theta}_y) \right) \right\}$$

$$
\begin{aligned}
\rightarrow\ & {}_p n_X \left\{(\theta_{0x}-\theta_{0y})'\bar{x} - (\psi_m(\theta_{0x})-\psi_m(\theta_{0y}))\right\} \\
= \ & \sqrt{n_X}\sum_{k=1}^{m}\left(\theta_{0x}^{(k)}-\theta_{0y}^{(k)}\right)\frac{1}{\sqrt{n_X}}\sum_{i=1}^{n_X}\phi_k(x_i)+O(n_X) \\
= \ & O_p(m\sqrt{n_X}+n_X).
\end{aligned}
$$

Finally, since by assumption $m^3/n_X \rightarrow 0$, then

$$
2\Lambda^{x}_{\hat{\theta}_x,\hat{\theta}_y} = O_p\left(n_X\right),
$$

and by an identical argument also $2\Lambda^{y}_{\hat{\theta}_y,\hat{\theta}_x} = O_p\left(n_Y\right)$, which is sufficient for consistency under H_1. ■

Tables 1-3: Monte Carlo rejection probabilities of the asymptotic critical values.

Table 1: $H_0 : X \sim Y \sim N(0,1)$

		sample size			
Test	sig. level	50	100	200	400
$\Lambda_{\hat{m}_A}$	0.10	0.126	0.124	0.118	0.108
	0.05	0.087	0.078	0.070	0.060
$\Lambda_{\hat{m}_B}$	0.10	0.128	0.123	0.117	0.107
	0.05	0.081	0.077	0.072	0.058
KS	0.10	0.068	0.073	0.077	0.087
	0.05	0.033	0.036	0.039	0.042
CM	0.10	0.110	0.108	0.105	0.104
	0.05	0.056	0.055	0.055	0.053

Table 2: $H_0 : X \sim Y \sim Exp(1)$

		sample size			
Test	sig. level	50	100	200	400
$\Lambda_{\hat{m}_A}$	0.10	0.133	0.125	0.118	0.110
	0.05	0.088	0.075	0.068	0.057
$\Lambda_{\hat{m}_B}$	0.10	0.131	0.123	0.116	0.105
	0.05	0.083	0.072	0.060	0.055
KS	0.10	0.065	0.069	0.077	0.087
	0.05	0.033	0.035	0.038	0.041
CM	0.10	0.115	0.110	0.107	0.105
	0.05	0.058	0.055	0.054	0.054

Table 3: $H_0 : X \sim Y \sim \begin{cases} N(-1,1) & \text{with probability } 0.5 \\ N(1,1) & \text{with probability } 0.5 \end{cases}$

		sample size			
Test	sig. level	50	100	200	400
$\Lambda_{\hat{m}_A}$	0.10	0.123	0.120	0.114	0.111
	.05	0.084	0.075	0.064	0.059
$\Lambda_{\hat{m}_B}$	0.10	0.123	0.113	0.111	0.108
	0.05	0.081	0.066	0.063	0.053
KS	0.10	0.069	0.082	0.086	0.083
	0.05	0.036	0.038	0.040	0.043
CM	0.10	0.114	0.110	0.108	0.106
	0.05	0.058	0.055	0.054	0.054

Tables 4-8: Monte Carlo rejection probabilities under the alternative hypotheses.

Table 4: $H_1^A : Y \sim N(\mu, 1)$

	μ				
Test	0.1	0.2	0.3	0.4	0.5
$\Lambda_{\hat{m}_A}$	0.125	0.357	0.670	0.919	0.986
$\Lambda_{\hat{m}_B}$	0.119	0.329	0.643	0.897	0.981
KS	0.137	0.410	0.718	0.922	0.991
CM	0.149	0.442	0.771	0.945	0.996

Table 5: $H_1^B : Y \sim N(0, (1+\mu)^2)$

	μ				
Test	0.1	0.2	0.3	0.4	0.5
$\Lambda_{\hat{m}_A}$	0.160	0.482	0.806	0.938	0.992
$\Lambda_{\hat{m}_B}$	0.163	0.477	0.799	0.932	0.988
KS	0.066	0.139	0.274	0.413	0.624
CM	0.053	0.096	0.231	0.465	0.708

Table 6: $H_1^C : Y \sim \frac{\chi^2(v)-v}{\sqrt{2v}}$

	v						
Test	35	30	25	20	15	10	5
$\Lambda_{\hat{m}_A}$	0.192	0.209	0.251	0.295	0.393	0.619	0.928
$\Lambda_{\hat{m}_B}$	0.171	0.198	0.234	0.274	0.394	0.620	0.922
KS	0.077	0.098	0.103	0.117	0.151	0.204	0.373
CM	0.075	0.083	0.094	0.113	0.133	0.175	0.349

Table 7: $H_1^D : Y \sim \sqrt{\frac{v-2}{v}} t_v$

	v				
Test	12	10	8	6	4
$\Lambda_{\hat{m}_A}$	0.141	0.174	0.299	0.552	0.964
$\Lambda_{\hat{m}_B}$	0.139	0.172	0.307	0.565	0.969
KS	0.050	0.056	0.073	0.077	0.213
CM	0.050	0.053	0.057	0.064	0.141

Table 8: $H_1^E : Y \sim \begin{cases} N(-\mu, 1) & \text{with probability } 0.5 \\ N(\mu, 1) & \text{with probability } 0.5 \end{cases}$

	μ				
Test	.1	.3	.5	.7	.9
$\Lambda_{\hat{m}_A}$	0.055	0.124	0.426	0.846	0.994
$\Lambda_{\hat{m}_B}$	0.062	0.130	0.419	0.853	0.996
KS	0.050	0.068	0.117	0.321	0.740
CM	0.050	0.055	0.076	0.182	0.542