Statistica Sinica **14**(2004), 1085-1103

# ANALYSIS OF DISTRIBUTIONS IN FACTORIAL EXPERIMENTS

R. C. H. Cheng and O. D. Jones

*University of Southampton and University of Melbourne*

*Abstract:* The Cramer-von Mises statistic provides a useful goodness of fit test of whether a random sample has been drawn from some given null distribution. Its use in comparing several samples has also been studied, but not systematically. We show that the statistic is capable of significant generalization. In particular we consider the comparison of the distributions of observations arising from factorial experiments. Provided that observations are replicated, we show that our generalization yields a test statistic capable of decomposition like the sum of squares used in ANOVA. The statistic is calculated using ranked data rather than original observations. We give the asymptotic theory. Unlike ANOVA, the asymptotic distributional properties of the statistic can be obtained without the assumption of normality. Further, the statistic enables differences in distribution other than the mean to be detected. Because it is distribution free, Monte-Carlo sampling can be used to directly generate arbitrarily accurate critical test null values in online analysis irrespective of sample size. The statistic is thus easy to implement in practice. Its use is illustrated with an example based on a man-in-the-loop simulation trial where operators carried out self assessment of the workload that they experienced under different operating conditions.

*Key words and phrases:* Cramer-von Mises statistic, distribution free, factorial experiment, homoscedasticity, rank based, simulation experiment.

## 1. Introduction

A goodness of fit test, like that based on the Cramer-von Mises statistic, is a simple but useful basic tool of statistics. Such tests are often useful in initial, exploratory, analyses of data to establish their broad form, before more refined but possibly more elaborate, statistical procedures are utilized. Some limited extensions of the Cramer-von Mises statistic have been considered. Kiefer (1959) considered tests of several samples using statistics that are sums of independent Cramer-von Mises statistics, Anderson (1962) considered a generalisation to the two sample case, and a related result is given by Baumgartner, Weiss and Schindler (1998). The purpose of this paper is show that the Cramer-von Mises can be applied in a much wider context than those considered in the above references, and that, in consequence, more specific structural aspects of the data can be tested. Our approach is based on preliminary work, previously reported in Cheng and Jones (2000); however it develops the ideas outlined there more fully and in a much more general framework.

We suggest that our extension of the Cramer-von Mises test is especially appropriate for the exploratory analysis of data obtained from a factorial experiment. When a linear model can be used to model factorial data, then its use together with analysis of variance is one of the most common used statistical techniques. However a linear model is not adequate when differences in the behaviour of the response variable $Y$ at different factor level combinations is not explicable in terms of differences in the means alone, but requires a comparison of other features of a distribution, such as variance or shape.

We show that the Cramer-von Mises test can be extended, using a decomposition similar to an ANOVA decomposition, to test not only for differences in means, but more general differences. For convenience we call our extensions EDFIT statistics. The main points of note are the following.

(i) Our proposed EDFIT statistics are intended for exploratory work and, given that they are extensions of the Cramer-von Mises statistic, we focus on goodness of fit aspects. Our theoretical discussion is formulated in terms of factorial data.

(ii) Though our approach is similar to ANOVA, we can relax the usual assumptions of normality and homoscedasticity of errors. The proposed EDFIT statistics tests will detect general differences between distributions, not only differences between means, but also differences in variance or shape. We do not consider estimation problems here.

(iii) The proposed EDFIT statistics make use of ranks, rather than actual observations. The asymptotic theory is derived in this paper for the case when the distributions are continuous. In fact critical values are easily obtained to arbitrary accuracy by direct (computer based) resampling. Resampling can be done not only for continuous distributions but also when distributions are discrete, as is the case for the example in Section 7.
We discuss power issues briefly, and show in the two sample case that they are closely related to those of the Cramer von-Mises statistic.

(iv) It should be mentioned that the Anderson-Darling statistic (1952) is also calculated from the EDF, and is arguably more widely advocated than the Cramer-von Mises statistic. This seems to be based mainly on the former's sensitivity to tail behaviour, where differences often lie. We have not considered use of the Anderson-Darling statistic, mainly on grounds of technical difficulty, but also because use of the Cramer-von Mises statistic allows us to make the analogy of our decompositions to those found in analysis of variance more explicit.

We will denote the original Cramer-von Mises statistic by $W^2$. The next Section discusses the distributional behaviour of a statistic, $T^2$, that is a generalization of $W^2$ which covers the situation where there are several samples. A point of theoretical interest is that, asymptotically, $T^2$ is an integral involving a multivariate Brownian Bridge whose components are correlated. Section 3 makes explicit the similarity of $T^2$ to sums of squares encountered in linear models and contains the main result, Theorem 1, which characterises the asymptotic distribution of $T^2$ under the null. Section 4 gives the particular form that $T^2$ takes in a number of explicit cases. Section 5 extends the decomposition of $T^2$ using the elegant Fourier decomposition first proposed by Durbin and Knott (1972) for $W^2$. Asymptotic power is discussed in Section 6. A numerical example based on data obtained from a real-time man-in-the-loop simulation is given in Section 7.

## 2. Distributional Result

Our asymptotic results apply in general only when the distributions are continuous, and we assume this from now on as far as the theoretical derivations are concerned. The distributions of the example in Section 7 are in fact discrete but, as will become clear, our methodology still can be used provided we use resampling to calculate critical values.

Let there be $q$ separate design points (by a design point we mean a combination of factor levels at which observations are obtained). Suppose, at the $i$th design point, that $y_{ij}$, $j = 1, \ldots, n_i$, is a random sample from the distribution $G_i$. This will be our main requirement: that at each design point there is a random sample, rather than just one observation. From now on we consider only the situation where the sample sizes of the $q$ samples remain in the same proportion, that is, $n_i = p_i n$, $i = 1, \ldots, q$, with the $p_i > 0$ remaining fixed. Thus in what follows, $n \to \infty$ implies $n_i = p_i n \to \infty$ for all $i$. The most useful case is where the $p_i$ are all equal, but we derive the main results for the general case of unequal $p_i$ as this is not much more difficult to do.

Our starting point is tests of the null hypothesis $H_0 : G_1 = G_2 = \cdots = G_q$ ($= G_0$, say) using a generalisation of $W^2$, the Cramer-von Mises EDF goodness of fit test statistic. It is well known that $W^2$ has an asymptotic null distribution that is the same as that of the integral $\int_0^1 W^2(u)du$, where $W(u)$ is a Brownian bridge (see Shorack and Wellner (1986) for example). We consider extensions of $W^2$ to a framework analogous to that of testing in ANOVA.

Let $S_i(u)$ be the EDF of the $i$th sample and let $\mathbf{V}(u)$ be the multivariate process with (independent) components

$$V_i(u) = \sqrt{n_i}(S_i(u) - G_0(u)). \tag{1}$$

The EDF of the combined samples will be denoted $\bar{S}(u)$. The test statistics that we consider have the form

$$T^2 = \int_{-\infty}^{\infty} \mathbf{Y}^T(u)\mathbf{C}\mathbf{Y}(u)d\bar{S}(u), \tag{2}$$

where $\mathbf{C}$ is a positive semidefinite matrix, $\mathbf{Y} = \mathbf{H}\mathbf{V}$, $\mathbf{H} = \mathbf{I} - \mathbf{R}\mathbf{R}^T$ and $\mathbf{R} = (\sqrt{p_1}, \ldots, \sqrt{p_q})^T$. From (1), the components of $\mathbf{Y}$ are

$$Y_i(u) = \sqrt{n_i}(S_i(u) - \bar{S}(u)). \tag{3}$$

Suppose that all $q$ samples are combined. Let $r_{ij}$ be the rank of observation $y_{ij}$ in the combined sample and let $n = \sum_{i=1}^q n_i$. As $1 \le r_{ij} \le n$, it will be more convenient to work with scaled ranks: $s_{ij} = r_{ij}/n$. The EDF of the combined sample ranks is obviously the same as its CDF, and is given by $\bar{S}^*(u) = \lceil un \rceil / n$, $0 \le u \le 1$. This can be written as $\bar{S}^*(u) = \sum_{i=1}^q (n_i/n) S_i^*(u)$, where

$$S_i^*(u) = \frac{\sum_{j=1}^{n_i} I_{[0,u]}(s_{ij})}{n_i}, \quad 0 \le u \le 1$$

is the empirical distribution function (EDF) of the scaled ranks of the $i$ th sample and $I_{[0,u]}$ is the indicator function of the unit interval.

Now, if $y_{(k)}$ is the $k$th ordered point in the combined sample, then $S_i(y_{(k)}) = S_i^*(k/n)$ and $Y_i(y_{(k)}) = \sqrt{n_i}(S_i^*(k/n) - k/n)$. Thus from (2) we see that $T^2$ reduces to the simple sum

$$T^2 = \frac{1}{n}\sum_{k=1}^{n} \mathbf{Y}^{*T}(k/n)\mathbf{C}\mathbf{Y}^*(k/n),$$

where the components of $\mathbf{Y}^*$ are $Y_i^*(u) = \sqrt{n_i}(S_i^*(u) - \bar{S}^*(u))$. This shows that $T^2$ is *distribution free*. So, for the remainder of this section, in considering the asymptotic form of $T^2$, we can without loss of generality assume that $G_0$ is the uniform distribution on $[0,1]$. Then, it is well-known (see Shorack and Wellner (1986) for a full discussion) that $V_i(u) \to_w W_i(u)$ as $n_i \to \infty$, $i = 1, \ldots, q$, where $\to_w$ denotes weak convergence, and the $W_i(u)$ are independent, Brownian bridges.

Weak convergence then implies that, if $D$ is a continuous function, $D(V_i) \to_d D(W_i)$, where $\to_d$ denotes convergence in distribution. A precise statement of this result is given in Shorack and Wellner (1986, Theorem 2.3.5). We require a multivariate version of this result, and this is provided by the following Lemma (a related result is given as Lemma 2 of Kiefer (1959)).

**Lemma 1.** *Let $D : \mathbb{R}^q \to \mathbb{R}_+$ be continuous, and let $W_1, \ldots, W_q$ be independent Brownian bridges. Then for $G_1, \ldots, G_q$ uniform on $[0,1]$,*

$$\int_0^1 D(V_1(u), \ldots, V_q(u))dS_j(u) \to_d \int_0^1 D(W_1(u), \ldots, W_q(u))du \text{ as } n \to \infty.$$

**Proof.** We show first that we can assume $D$ is bounded and uniformly continuous. Define $D_r(x_1,\ldots,x_q)$ to be $D(x_1,\ldots,x_q)$ when each $|x_i|\le r$ and bounded and uniformly continuous elsewhere. From Dvoretzky, Kiefer and Wolfowitz (1956) (see Csörgö and Révész (1981), Theorem 4.1.3) we have the following large deviation rate: there exists a constant $C$ such that, for all $n$ and $r$,

$$P(\sup_u |V_i(u)|\ge r)\le Ce^{-2r^2}.$$

Thus for any $\epsilon>0$, we can choose $r$ large enough that

$$P(D(V_1(u),\ldots,V_q(u))=D_r(V_1(u),\ldots,V_q(u)) \text{ on } [0,1])>1-\epsilon,$$

whence $\int_0^1 D_r(V_1(u),\ldots,V_q(u))dS_j(u)\to\int_0^1 D(V_1(u),\ldots,V_q(u))dS_j(u)$ in probability as $r\to\infty$. Similarly $\int_0^1 D_r(W_1(u),\ldots,W_q(u))du\to\int_0^1 D(W_1(u),\ldots,W_q(u))$ $du$ in probability as $r\to\infty$.

Now, from Komlós, Major and Tusnády (1975) (see Csörgö and Révész (1981, Theorem 4.4.1)) we have the following strong approximation. For each $i$ we have a sequence of Brownian bridges $W_i^1,W_i^2,\ldots$ such that for all $\epsilon>0$, $P(\sup_u|V_i(u)-W_i^n(u)|>\epsilon)<\epsilon$ for all $n$ large enough. Thus, supposing $D$ is uniformly continuous, for all $\epsilon>0$ we have for all $n$ large enough. $P(\sup_u|D(V_1(u),\ldots,V_q(u))-D(W_1^n(u),\ldots,W_q^n(u))|>\epsilon)<\epsilon$, whence

$$P\left(\left|\int_0^1 D(V_1(u),\ldots,V_q(u))dS_j(u)-\int_0^1 D(W_1^n(u),\ldots,W_q^n(u))dS_j(u)\right|>\epsilon\right)<\epsilon. \tag{4}$$

It is well known that for any Brownian bridge $W$

$$\lim_{h\to 0}\sup_{0\le s\le 1-h}\sup_{0\le t\le h}\frac{|W(s+t)-W(s)|}{\sqrt{-2h\log h}}=1 \text{ almost surely.}$$

In particular, $W$ is uniformly continuous with probability 1. Thus, as $D$ is also uniformly continuous, for any $\epsilon>0$ we can find $h>0$ such that, for all $n$, $P(|D(W_1^n(u),\ldots,W_q^n(u))-D(W_1^n(v),\ldots,W_q^n(v))|<\epsilon \text{ for all } |u-v|<h)>1-\epsilon$. Finally, since $S_j(u)\to u$ uniformly on $[0,1]$ with probability 1, we can choose $n$ large enough that $|S_j(u)-u|<h$ on $[0,1]$, whence

$$P\left(\left|\int_0^1 D(W_1^n(u),\ldots,W_q^n(u))dS_j(u)-\int_0^1 D(W_1^n(u),\ldots,W_q^n(u))du\right|>\epsilon\right)<\epsilon. \tag{5}$$

Combining (4) and (5) gives the result.

Clearly $\mathbf{V}^T(u)\mathbf{H}^T\mathbf{CHV}(u)=\mathbf{Y}^T(u)\mathbf{CY}(u)$ satisfies the conditions of Lemma 1. Moreover, since $\bar{S}(u)=\sum_{i=1}^q(n_i/n)S_i(u)$, the expression (2) for $T^2$ is the sum of terms each of the form $\int_0^1 D(V_1(u),\ldots,V_q(u))dS_j(u)$ appearing in Lemma 1. Application of the lemma to $T^2$ thus yields the following.

**Lemma 2.** *Under $H_0$, for any common continuous distribution $G_0$, as all $n_j \to \infty$, $T^2 \to_d \int_0^1 \mathbf{W}^T(u)\mathbf{H}^T\mathbf{CHW}(u)d(u)$, where $\mathbf{W}(u) = (W_1(u),\ldots,W_k(u))^T$ is a vector of independent Brownian bridges.*

In the remainder of this section we derive an expansion for the limiting form of $T^2$. Write

$$\mathbf{Z}(u) = \mathbf{HW}(u). \tag{6}$$

Clearly $E(\mathbf{Z}(u)) = \mathbf{0}$. We have, for $u < v$, $E[\mathbf{Z}(u)\mathbf{Z}^T(v)] = \mathbf{H}E[\mathbf{W}(u)\mathbf{W}^T(v)]\mathbf{H}^T = u(1-v)\mathbf{H}^2 = u(1-v)\mathbf{H}$, on noting that $\mathbf{H}$ is symmetric and idempotent.

It is well-known (Mercer's Theorem, see Anderson and Darling (1952)) that the symmetric correlation function $k(u,v) = \min(u,v) - uv$ can be expressed as $k(u,v) = \sum_{j=1}^{\infty}(\lambda_j)^{-1}f_j(u)f_j(v)$, where $\lambda_j$ is an eigenvalue and $f_j(u)$ is the corresponding normalized eigenfunction of the integral equation

$$\lambda\int_0^1 k(u,v)f(u)du = f(v)$$

and one can write $W(u) = \sum_{j=1}^{\infty}(\lambda_j)^{-1/2}X_j f_j(u)$, where the $X_j$ are independent $N(0,1)$ variables. Applying this to $\mathbf{Z}(u) = \mathbf{HW}(u)$ we have

**Lemma 3.** *$\mathbf{Z}(u)$ has the same distribution as $\sum_{j=1}^{\infty}(\lambda_j)^{-1/2}\mathbf{X}_j f_j(u)$ where $\mathbf{X}_j \sim MVN(\mathbf{0},\mathbf{H})$ with $\mathbf{H} = \mathbf{I} - \mathbf{RR}^T$, $\mathbf{R} = (\sqrt{p_1},\sqrt{p_2},\ldots,\sqrt{p_q})^T$, and with $\mathbf{X}_i$ and $\mathbf{X}_j$ independent if $i \neq j$.*

## 3. A Linear Model Analogue for EDFs

Let $F_i(u)$ be the (discrete) cumulative distribution function (CDF) of observations in the $i$th sample of scaled ranks $\{s_{ij}\ j = 1,\ldots,n_i\}$. We write $\mathbf{F}(u) = (F_1(u),\ldots,F_q(u))^T$, and consider the model $\mathbf{F}(u) = \mathbf{Ab}(u)$, where

$$\mathbf{A} = [\mathbf{1}_q|\mathbf{A}_1] \tag{7}$$

is a given matrix and $\mathbf{b}(u) = \begin{pmatrix} b_0(u) \\ \mathbf{b}_1(u) \end{pmatrix}$ is a column vector of unknown functions. In analogy with the common assumption in the standard linear model, inclusion of the column unit vector $\mathbf{1}_q$ is to allow for explicit inclusion of an overall mean $b_0(u) = \lceil un \rceil/n$. We assume that $\mathbf{A}_1$ has $c$ columns and has column rank $r$, possibly less than $c$. (If $c = q$, then $r < q$, as $\text{rank}(\mathbf{A}_1) < q$.)

In this section, we obtain an estimator for $\mathbf{b}_1$, and a test for the null hypothesis

$$H_0 : \mathbf{b}_1(u) = \mathbf{0}, \text{ or equivalently } F_i(u) = b_0(u) \text{ for all } i. \tag{8}$$

Let

$$\mathbf{D} = \begin{pmatrix} \sqrt{p_1} & 0 & \cdot & 0 \\ 0 & \sqrt{p_2} & \cdot & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdot & \sqrt{p_q} \end{pmatrix}, \quad \mathbf{p} = \mathbf{DR} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_q \end{pmatrix}.$$

Elementary manipulations, using (7), yield $\mathbf{D}(\mathbf{F}(u) - \mathbf{1}_q b_0(u)) = \mathbf{D}(\mathbf{I} - \mathbf{1}_q \mathbf{p}^T)\mathbf{A}_1$ $\mathbf{b}_1(u)$. In this equation $\mathbf{F}(u)$ and $\mathbf{b}_1(u)$ are unknown. Now at each experimental design point $i$, our random sample of observations is represented by the EDF of scaled ranks $S_i^*(u)$. We therefore replace $\mathbf{F}(u)$ with its EDF $\mathbf{S}^*(u)$ and note that $b_0(u) = \bar{S}^*(u)$. By analogy with the conventional linear model, we write

$$\begin{aligned} \frac{1}{\sqrt{n}} \mathbf{Y}^*(u) &= \mathbf{D}(\mathbf{S}^*(u) - \mathbf{1}_q \bar{S}^*(u)) \\ &= \mathbf{B}\mathbf{b}_1(u) + \boldsymbol{\varepsilon}(u), \end{aligned}$$

where $\mathbf{B} = \mathbf{D}(\mathbf{I} - \mathbf{1}_q \mathbf{p}^T)\mathbf{A}_1$ and $\boldsymbol{\varepsilon}(u)$ is a vector of unknown error functions. The least squares estimate of $\mathbf{b}_1(u)$ is $\hat{\mathbf{b}}_1(u) = \mathbf{G}\mathbf{B}^T\mathbf{Y}^*(u)/\sqrt{n}$, where $\mathbf{G}$ is a generalised inverse (g-inverse) of $\mathbf{B}^T\mathbf{B}$. The following result gives the asymptotic distribution of a statistic $T^2$ which can be used to test $H_0$.

**Theorem 1.** *Let $\mathbf{C} = \mathbf{BGB}^T$. Then, under $H_0$, the statistic $T^2 = \int_{-\infty}^{\infty} \mathbf{Y}^T(u)$ $\mathbf{CY}(u)d\bar{S}(u) = n^{-1}\sum_{k=1}^{n} \mathbf{Y}^{*T}(k/n)\mathbf{CY}^*(k/n)$ has the asymptotic representation $T^2 = \sum_{j=1}^{\infty}(\lambda_j)^{-1}\chi_j^2$, where the $\lambda_j$ are as in Lemma 3, and the $\chi_j^2$ are independently distributed $\chi^2(m)$ variables, with $m = \mathrm{trace}(\mathbf{CH}) = \mathrm{rank}(\mathbf{B}^T\mathbf{B})$.*

**Proof.** Under $H_0$, from Lemma 2, as $n \to \infty$,

$$\int_{-\infty}^{\infty} \mathbf{Y}^T(u)\mathbf{CY}(u)d\bar{S}(u) \to_d \int_0^1 \mathbf{Z}^T(u)\mathbf{CZ}(u)d(u),$$

where $\mathbf{Z}(u)$ is as defined in (6). It follows from Lemma 3 that

$$\int_0^1 \mathbf{Z}^T(u)\mathbf{CZ}(u)d(u) = \sum_{j=1}^{\infty} \frac{1}{\lambda_j} \mathbf{X}_j^T \mathbf{C} \mathbf{X}_j.$$

It is known (See for example Searle (1971, Corollary 2.2)) that if $\mathbf{X}_j \sim MVN(\mathbf{0},$ $\mathbf{H})$ then $\mathbf{X}_j^T\mathbf{C}\mathbf{X}_j \sim \chi^2(\mathrm{trace}(\mathbf{CH}))$ if and only if $\mathbf{HCHCH} = \mathbf{HCH}$. Using the fact that a g-inverse $\mathbf{G}$ of $\mathbf{B}^T\mathbf{B}$ satisfies $\mathbf{BGB}^T\mathbf{B} = \mathbf{B}$, we have $\mathbf{C}^2 = \mathbf{BGB}^T\mathbf{BGB}^T = \mathbf{C}$ so that $\mathbf{C}$ is idempotent. Moreover, as $\mathbf{p} = \mathbf{DR}$ and $\mathbf{1}_q^T\mathbf{D} = \mathbf{R}^T$, we have $\mathbf{B}^T\mathbf{R} = \mathbf{A}_1^T(\mathbf{I} - \mathbf{p}\mathbf{1}_q^T)\mathbf{DR} = \mathbf{A}_1^T(\mathbf{DR} - \mathbf{DRR}^T\mathbf{R}) = \mathbf{0}_c$ as $\mathbf{R}^T\mathbf{R} = 1$. It follows that $\mathbf{CR} = \mathbf{0}_q$ and that $\mathbf{CHC} = \mathbf{C}(\mathbf{I} - \mathbf{RR}^T)\mathbf{C} = \mathbf{C}$ and hence $\mathbf{HCHCH} = \mathbf{HCH}$.

Finally trace($\mathbf{CH}$) = trace($\mathbf{BGB}^T - \mathbf{BGB}^T\mathbf{RR}^T$) = trace($\mathbf{BGB}^T$), using again the fact that $\mathbf{B}^T\mathbf{R} = \mathbf{0}_c$. Now for any two matrices $\mathbf{E}$ and $\mathbf{F}$, trace($\mathbf{EF}$) = trace($\mathbf{FE}$) provided $\mathbf{EF}$ and $\mathbf{FE}$ both exist. Therefore trace($\mathbf{BGB}^T$) = trace($\mathbf{GB}^T\mathbf{B}$), and by Searle (1971, Section 1.2, Lemma 1), trace($\mathbf{GB}^T\mathbf{B}$) = rank($\mathbf{B}^T\mathbf{B}$).

**Corollary 1.** *Under $H_0$, the asymptotic distribution of $T^2$ has the representation*

$$T^2 \to_d \sum_{i=1}^{m} \sum_{j=1}^{\infty} \frac{1}{\lambda_j} X_{ij}^2,$$

*where the $X_{ij}^2$ are independent $\chi^2(1)$ variables. Thus, asymptotically, $T^2$ behaves as the sum of $m$ independent and identically distributed $W^2$ statistics.*

The asymptotic distribution of a sum of independent $W^2$ statistics is given by Kiefer (1959). For completeness we restate it here.

**Theorem 2.** *Under $H_0$, $T^2$ has the asymptotic CDF $F_{T^2}(z) = \sum_{j=0}^{\infty} R_j(z)$, where*

$$R(j,z) = \frac{\Gamma(j+m/2)2^{1/2+m/2}z^{-m/4}}{\Gamma(j+1)\Gamma(m/2)\pi^{1/2}} \exp\left(-\frac{(4j+m)^2}{16z}\right) D_{m/2-1}\left(\frac{4j+m}{2z^{1/2}}\right).$$

*Here $D_a(x)$ is the Whittaker parabolic cylinder function, defined for example in Abramovitz and Stegun* (1992, *Chap.*19)*.*

Kiefer (1959) gives a method of calculating $D_a(x)$ using Hermite polynomials and Bessel functions. An alternative formula is the following.

**Corollary 2.** $D_{k/2-1}((4j+k)/2z^{1/2}) = U(1/2-k/2,(4j+k)/2z^{1/2})$, *where $U(\cdot,\cdot)$ is the cylinder function given in Abramovitz and Stegun* (1992, *Chap.*19)*. Then, using their formulas* 19.12.2, 13.133 *and* 13.2.5, *we find that the right hand side can be expressed as an integral*

$$U\left(\frac{1}{2}-\frac{k}{2},\frac{4j+k}{2z^{1/2}}\right) = \frac{2^{k/4-2}z^{-1/2}(4j+k)}{\Gamma(1-k/4)} \exp\left(-\frac{(4j+k)^2}{16z}\right)$$
$$\times \int_0^1 u^{-k/4}(1-u)^{-3/2} \exp\left(-\frac{(4j+k)^2}{8z}\frac{u}{1-u}\right) du.$$

*The only integer values of $k$ for which this formula holds are $k = 0,1,2,3$. However for $k \geq 4$ we can apply the recursion $U(1/2-k/2,x) = xU(3/2-k/2,x) + (2-k/2)U(5/2-k/2,x)$ repeatedly until all terms on the right have $U$'s with first arguments that are in the range $[-1,1/2]$. Thus for any $k \geq 4$, $U(1/2-k/2,x)$ can always be expressed as a linear combination of the pair $U(1/2,x)$ and $U(-1/2,x)$ when $k$ is odd, or else of the pair $U(0,x)$ and $U(-1,x)$ when $k$ is even.*

## 4. Examples

We give some examples which are likely to be of use in practice.

**One way classification.** Here we compare the distributions of $q$ separate samples by testing if $H_0$ as given in (8) is true.

**Lemma 4.** *Let* $\mathbf{A}_1 = \mathbf{I}_{q\times q}$. *Then the test statistic reduces to* $T^2 = \sum_{i=1}^{q} \int_{-\infty}^{\infty} Y_i^2(u)\, d\bar{S}(u) = \sum_{i=1}^{q} W_i^2$. *Under the null* $T^2$ *is asymptotically the sum of independent weighted* $\chi^2(q-1)$ *variables as given in Theorem* 1.

**Proof.** If $\mathbf{A}_1 = \mathbf{I}_{q\times q}$ then $\mathbf{B}^T\mathbf{B} = \mathbf{D}^2 - \mathbf{p}\mathbf{p}^T$. This is singular of rank $q-1$. It has g-inverse

$$\mathbf{G} = \begin{pmatrix} \mathbf{H} & \mathbf{0}_{q-1} \\ \mathbf{0}_{q-1}^T & \mathbf{0} \end{pmatrix},$$

where

$$\mathbf{H} = \begin{pmatrix} 1/p_1 + 1/p_q & 1/p_q & \cdots & 1/p_q \\ 1/p_q & 1/p_2 + 1/p_q & \cdots & 1/p_q \\ \vdots & \vdots & \ddots & \vdots \\ 1/p_q & 1/p_q & \cdots & 1/p_{q-1} + 1/p_q \end{pmatrix}.$$

$\mathbf{H}$ is of full rank $q-1$ so $\mathbf{G}$ and $\mathbf{B}^T\mathbf{B}$ are also of rank $q-1$. Some elementary algebra shows that $\mathbf{C} = \mathbf{B}\mathbf{G}\mathbf{B}^T = \mathbf{I} - \mathbf{R}\mathbf{R}^T$. Thus

$$\begin{aligned} \mathbf{Y}^T(u)\mathbf{C}\mathbf{Y}(u) &= \sum_{i=1}^{q} Y_i^2(u) - \left(\sum_{i=1}^{q} \sqrt{p_i}\, Y_i(u)\right)^2 \\ &= \sum_{i=1}^{q} Y_i^2(u), \end{aligned}$$

as $\sum_{i=1}^{q} \sqrt{p_i}\, Y_i(u) = 0$ from the definition (3) of $Y_i(u)$.

**Comparison of two samples.** Another obvious application is the comparison of distribution functions of the responses at different experimental design points. In general this is not straightforward when the proportions $p_i$ are unequal. However there are simple special cases. An elementary case is when there are just two unequal samples. We can then set $A_1 = (-1/p_1, 1/p_2)$, in which case

$$\begin{aligned} T^2 &= \int_{-\infty}^{\infty} p_1 p_2 \left(Y_1(u)/\sqrt{p_1} - Y_2(u)/\sqrt{p_2}\right)^2 d\bar{S}(u) \\ &= \int_{-\infty}^{\infty} p_1 p_2 n \left(S_1(u) - S_2(u)\right)^2 d\bar{S}(u) \qquad (9) \\ &\to_d \sum_{j=1}^{\infty} \frac{1}{\lambda_j} X_j^2 \text{ as } n \to \infty, \end{aligned}$$

where $X_j^2$ are independent $\chi^2$ variates each with one degree of freedom. Thus in this case $T^2$ provides a direct measure of $G_1(u) - G_2(u)$ through the difference $S_1(u) - S_2(u)$ between the sample EDFs. We note that the form of $T^2$ in this example reduces to exactly that of the statistic $T$ discussed by Anderson (1962). Anderson tabulates critical values for this statistic showing they remain very stable even for small sample sizes.

**Two-way classification.** Limitations of space prevents our discussing this in full generality. (The authors know of no larger published matrix than the full design matrix for a two-way cross-classification as given by Stuart and Ord (1991, equation 29.35)) However, the two factor case each at two levels, clearly illustrates the EDFIT version. Let the number of observations in each of the four cells (1,1), (1,2), (2,1), (2,2) corresponding to the different (Factor1, Factor2) level combinations be respectively: $rsn$, $r(1-s)n$, $(1-r)sn$, $(1-r)(1-s)n$. In this case let

$$\mathbf{A}_1 = \begin{pmatrix} 1-r & 1-s & (1-r)(1-s) \\ 1-r & -s & -(1-r)s \\ -r & 1-s & -r(1-s) \\ -r & -s & rs \end{pmatrix}.$$

Then $\mathbf{B}^T\mathbf{B}$ is diagonal, and

$$\begin{aligned} T^2 &= r(1-r)\sum_{k=1}^{n}(S^*_{1\cdot}(k/n) - S^*_{2\cdot}(k/n))^2 + s(1-s)\sum_{k=1}^{n}(S^*_{\cdot 1}(k/n) - S^*_{\cdot 2}(k/n))^2 \\ &\quad + rs(1-r)(1-s)\sum_{k=1}^{n}(S^*_{11}(k/n) - S^*_{12}(k/n) - S^*_{21}(k/n) + S^*_{22}(k/n))^2 \\ &= T_1^2 + T_2^2 + T_3^2, \text{ say,} \end{aligned}$$

where $S^*_{ij}(k/n)$, $i,j = 1,2$ are the EDFs of the ranked observations at level $i$ of Factor1 and level $j$ of Factor2; $S^*_{i\cdot}(k/n), i = 1,2$, are the EDFs of the ranked observations at level $i$ of Factor1 with both levels of Factor2 combined; and $S^*_{\cdot j}(k/n)$ $j = 1,2$ is similarly defined for Factor2. Under the null, the three $T_j^2$ are asymptotically independently distributed as $W^2$ statistics. $T_1^2$ tests whether there is a difference between the two levels of Factor1 whilst $T_2^2$ tests whether there is a difference between the two levels of Factor2. $T_3^2$ tests the interaction.

**Orthogonal contrasts.** A more general set of contrasts can be simultaneously tested using the following.

**Lemma 5.** *If all the* $p_i = 1/q$ *and* $\mathbf{A}_1 = (\mathbf{a}_1, \ldots, \mathbf{a}_c) = (a_{ij})$ *is a set of* $c < q$ *orthonormal contrast vectors, so that* $\mathbf{a}_i^T\mathbf{a}_j = \delta_{ij}$ *and* $\sum_{i=1}^{q} a_{ij} = 0$, *then* $\mathbf{B}^T\mathbf{B} =$

$q^{-1}\mathbf{I}_c$, *whence* $\mathbf{C} = \mathbf{A}_1\mathbf{A}_1^T$ *and*

$$\begin{aligned} T^2 &= \sum_{j=1}^{c} \frac{1}{q} \sum_{k=1}^{n} \Big( \sum_{i=1}^{q} a_{ij} S_i^*(k/n) \Big)^2 \\ &= \sum_{j=1}^{c} T_j^2 \qquad \textit{say.} \end{aligned}$$

*Under the null the* $T_j^2 \to_d \sum_{i=1}^{\infty} (\lambda_i)^{-1} X_{ij}^2$ *as* $n \to \infty$ *where the* $X_{ij}^2$ *are independent* $\chi^2(1)$ *variables.*

Typical application of Lemma 5 is in polynomial regression where the contrasts can be formed to assess linear, quadratic and higher order effects. In ANOVA, regression coefficients can be checked sequentially with the power of one test not depending on the power of others. Likewise in our situation different contrasts can be checked sequentially for significance using the above decomposition of the $T^2$ statistic. Note however that significance of one contrast, through $T_i^2$ say, can affect the power of another statistic, $T_j^2$, to detect a difference.

For example, consider the case of three samples (with $p_i = 1/3$ for $i = 1, 2, 3$). Then for

$$\mathbf{A}_1 = \begin{pmatrix} -1 & 1 \\ 0 & -2 \\ 1 & 1 \end{pmatrix}$$

we have $T_1^2 = (1/3) \sum_{k=1}^{n} (S_3^*(k/n) - S_1^*(k/n))^2$ and $T_2^2 = (1/3) \sum_{k=1}^{n} (S_3^*(k/n) + S_1^*(k/n) - 2S_2^*(k/n))^2$. Thus $T_2^2$ tests if $F_2(u) = (1/2)(F_1(u) + F_3(u))$ so that $T_2^2$ gives information about the relative position of the ranks of the second sample compared with those of the other two samples. Irrespective of whether $F_2(u) = (1/2)(F_1(u) + F_3(u))$ or not, $T_1^2$ will provide a test whether $F_1(u) = F_3(u)$ or not.

## 5. Finite Sample Analysis

The above asymptotic analysis can be replaced by a more exact finite sample analysis which gives a precise decomposition of individual components. We shall not attempt full generality but focus on the decomposition of the individual components $Y_h$. We adopt the approach of Durbin and Knott (1972).

**Theorem 3.**

$$\begin{aligned} W_h^2 &= \int_{-\infty}^{\infty} n_h (S_h(u) - \bar{S}(u))^2 d\bar{S}(u) \\ &= \frac{n_h}{n} \sum_{k=1}^{n} (S_h^*(k/n) - k/n)^2 \;=\; \sum_{j=1}^{n-1} \frac{z_{nj}^2}{(j\pi)^2}, \end{aligned} \tag{10}$$

where

$$z_{nj} = \sum_{l=1}^{n_h} \frac{j\pi[\sin\{j\pi r_{hl}/n\} - \sin\{j\pi(r_{hl}-1)/n\}]}{n\sqrt{2n_h}[1-\cos(j\pi/n)]} \text{ for } j = 1, \dots, n-1 \tag{11}$$

and $r_{hl}$, $l = 1, \dots, n_h$, are the ranks of the $h$th sample in the total combined sample. Moreover $E(z_{nj}) = 0$ and $E(z_{nj}z_{nk}) = 0$ if $j \neq k$.

**Proof.** In this proof alone the letter $i$ is used to denote $\sqrt{-1}$. Consider the Fourier expansion $S_h^*(k/n) - k/n = \sum_{j=1}^n \hat{\beta}_j \sin(j\pi k/n)$. If we multiply this equation by $\sin(l\pi k/n)$ and sum over all $k$, we find, on using the usual orthogonality property of the sine basis functions, that $\hat{\beta}_l = (2/n)\sum_{k=1}^n (S_h^*(k/n) - k/n)\sin(l\pi k/n)$, $l = 1, \dots, n$. Now $\sum_{k=1}^n [S_h^*(k/n) - k/n]\sin(l\pi k/n)$ is simply the imaginary part of $\sum_{k=1}^n [S_h^*(k/n) - k/n]\exp(i\,l\pi k/n)$. Moreover from the form of $S_h^*(k/n)$ we have that $\sum_{k=1}^n S_h^*(k/n)\exp(i\,l\pi k/n) = (n_h)^{-1}\sum_{j=1}^{n_h}\sum_{k=r_{hj}}^n \exp(i\,l\pi k/n)$. Also we have $\sum_{k=1}^n (k/n)\exp(i\,l\pi k/n) = n^{-1}\sum_{j=1}^n \sum_{k=j}^n \exp(i\,l\pi k/n)$. Using these two expression we find, after some not inconsiderable manipulation, that

$$\begin{aligned} \hat{\beta}_l &= \mathrm{Im}\frac{2}{n}\sum_{k=1}^n (S_h^*(k/n) - k/n)\exp(i\,l\pi k/n) \\ &= \frac{1}{nn_h[1-\cos(j\pi/n)]}\sum_{j=1}^{n_h}[\sin(l\pi r_{hj}/n) - \sin(l\pi(r_{hj}-1)/n)]. \end{aligned} \tag{12}$$

From Parseval's Theorem we have that $W_h^2 = (n_h/2)\sum_{l=1}^n \hat{\beta}_l^2$. Using the expression (12) for $\hat{\beta}_l$ (where $\hat{\beta}_n = 0$) and rearranging then yields (10).

$E(z_{nj})$ is evaluated by summing the expression for $z_{nj}$ over all possible combinations of (ranked) positions that the $n_h$ observations of the $h$th sample can take in the combined sample of size $n$. By symmetry each possible value, viz $1, \dots, n$, that $r_{hl}$ can take, occurs the same number of times in the total summation. Thus the expectation takes the form

$$E(z_{nj}) = A(j, n_h, n)\sum_{l=1}^n [\sin\{j\pi l/n\} - \sin\{j\pi(l-1)/n\}] = 0.$$

Similarly we have

$$\begin{aligned} E(z_{nj}z_{nk}) = A'(j, k, n_h, n) \sum_{l,m=1}^n &[\sin\{j\pi l/n\} - \sin\{j\pi(l-1)/n\}] \times \\ &[\sin\{j\pi m/n\} - \sin\{j\pi(m-1)/n\}] = 0. \end{aligned}$$

**Remark.** Here the $z_{nj}$,to within a scale factor, are the discrete analogues of the regression coefficients defined in Durbin and Knott (1972, Equation 2.11). They therefore are essentially the principal components of the $n$ dimensional vector with components $\sqrt{n_h/n}(S_h^*(k/n) - k/n)$, $k = 1, \ldots, n$. (There are only $n-1$ non-trivial components, the $n$th being identically zero.)

The properties of the $z_{nj}$ are not as simple as in the Cramer-von Mises $W^2$ case, though there are obvious similarities. Thus the $z_{nj}$ are uncorrelated with mean zero, and they are similarly, but not identically, distributed; in particular their variances are all close to unity. Each $z_{nj}$ is the sum of $n_h$ identically distributed random variables, but they are not independently distributed as in the case of $W^2$. This makes difficult analytic calculation of the precise distribution of any given $z_{nj}$, and we do not attempt to do so here. However, from simulations not reported here, it is clear that the distributions, like those of the Durbin-Knott $z_j$, are very nearly normal.

However it is quite easy to calculate, by Monte-Carlo simulation, the distribution of any given $z_{nj}$, and hence its percentage points. This can be done to arbitrary accuracy as follows: $z_{nj}$ variates can be sampled from their null distribution simply by sampling $n_h$ values uniformly from the set $\{1, \ldots, n\}$ without replacement, and treating these as the ranks $r_{hl}$, $l = 1, \ldots, n_h$, at (11). We can therefore generate a random sample of $z_{nj}$'s, of size $M$ say, and use the EDF of this random sample to estimate the CDF. This can be done to arbitrary accuracy by choosing $M$ sufficiently large. This method of calculating the percentage points of the distribution of $z_{nj}$ is sufficiently easy to do to incorporate into online analyses.

When the null hypothesis is not satisfied it is clear from the form of the $z_{nj}$ that $z_{n1}$ will be correlated with, and hence will measure deviation of the $h$th sample mean from, the mean of other samples. Similarly $z_{n2}$ measures deviation of the variance of the $h$th sample from that of other samples.

## 6. Power

We briefly consider the limiting power of EDFIT statistics as $n \to \infty$. The analysis used by Durbin and Knott for the one-sample case can be easily extended to our case. We first outline their method.

Durbin and Knott suppose that the distribution is of the form $G(y, \theta)$ where $\theta$ is a vector of parameters that specifies the distribution completely, and that $G$ is a differentiable function of $\boldsymbol{\theta}$. They consider the null hypothesis $H_0$: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus the alternative $H_1 \boldsymbol{\theta} = \boldsymbol{\theta}_0 + n^{-1/2}\boldsymbol{\gamma}$, where $\gamma$ is a constant vector. We put $x = G(y, \boldsymbol{\theta}_0)$, $Y(x) = \sqrt{n}\{G(x) - x\}$ and $\mathbf{g}(x) = \partial G(y, \boldsymbol{\theta})/\partial \boldsymbol{\theta}|_{\theta_0}$, expressing this derivative as a function of $x$ by means of the transform $x = G(y, \boldsymbol{\theta}_0)$. Then under $H_1$ as $n \to \infty$, $E\{Y(x)\} \to \boldsymbol{\gamma}^T \mathbf{g}(x)$. The asymptotic covariance function of $Y(x)$ on $H_1$ is the same as that on $H_0$. It then follows that the limiting distribution of

$W^2$ on $H_1$ is that of $W^2 = \sum_{j=1}^{\infty} z_j^2/(j\pi)^2$, where the $z_j$ are independent normal $N(\boldsymbol{\gamma}^T\boldsymbol{\delta}_j, 1)$ variables, with $\boldsymbol{\delta}_j = \sqrt{2}j\pi\int_0^1 \mathbf{g}(x)\sin(j\pi x)\mathrm{dx}$.

For the EDFIT case we take as our hypotheses $H_0 : F_i(u) = b_0(u)$ all $i$, and $H_1 : \mathbf{F}(u) = \mathbf{A}\mathbf{b}(u)$, where

$$\mathbf{b}(u) = \begin{pmatrix} b_0(u) \\ n^{-1/2}\tilde{\mathbf{b}}_1(u) \end{pmatrix}. \tag{13}$$

$\mathbf{A}$ and $\tilde{\mathbf{b}}_1(u)$ are independent of $n$ but will depend on $\mathbf{p}$. From Theorem 1 we have that $T^2 = n^{-1}\sum_{k=1}^{n} \mathbf{Y}^{*T}(k/n)\mathbf{C}\mathbf{Y}^*(k/n) = \sum_{i=1}^{m} W_i^2$, where the $W_i^2$ are independent $W^2$ statistics under $H_0$. We can apply Durbin and Knott's argument to each $W_i^2$. The precise detail depends on the precise decomposition used. We only consider the most useful case in practice which is where each takes the simple form

$$W_i^2 = n^{-1}\sum_{k=1}^{n}\{\mathbf{c}_i^T\mathbf{Y}^*(k/n)\}^2. \tag{14}$$

All the examples of Section 4 are of this form.

In this case $W_i^2 = \sum_{j=1}^{n-1} z_j^{(i)2}/(j\pi)^2$ and we find, following Durbin and Knott's argument, that as $n \to \infty$, the $z_j^{(i)}$ are asymptotically normal and independent with mean

$$E(z_j^{(i)}) = \mathbf{c}_i^T\mathbf{B}\{\sqrt{2}j\pi\int_0^1 \tilde{\mathbf{b}}_1(x)\sin(j\pi u)\mathrm{du}\}.$$

This is the equivalent of Durbin and Knott's expression $E(z_{nj}) = \gamma^T\boldsymbol{\delta}_j$.

Power considerations follow easily from this result. We consider only the two sample case of Section 4 in any detail, where two distributions are compared. Here $T^2 = W_1^2$ where $W_1^2$ is of the form (14) with $\mathbf{c}_1 = (-\sqrt{p_2}, \sqrt{p_1})^T$. Moreover from (13), we have $F_1(u) - F_2(u) = n^{-1/2}(p_1p_2)^{-1}\tilde{b}_1(u)$. We therefore take $\tilde{b}_1(u) = \tilde{\gamma}p_1p_2g(u)$, so that we are considering a difference $n^{-1/2}\tilde{\gamma}g(u)$ which does not depend on $p_1$ and $p_2$. Then $E(z_j^{(1)}) = \tilde{\gamma}(p_1p_2)^{1/2}\delta_j$, where $\delta_j = \sqrt{2}j\pi\int_0^1 g(u)\sin(j\pi u)\mathrm{du}$ is the direct analogue of $\delta_j$ given by Durbin and Knott. The interesting power results given by Durbin and Knott concerning the use of different components to investigate differences between means and variances therefore apply here. The only difference is that $\gamma$ as given by Durbin and Knott is replaced by $\tilde{\gamma}(p_1p_2)^{1/2}$. This shows there is some loss of power especially in comparing two distributions using unbalanced samples.

## 7. Application

As an example of the above discussion we consider data from a real-time man-in-the-loop simulation trial comparing two different methods of operating a certain piece of equipment processing a certain product. The particular data set to be considered here, being essentially ordinal in form, might of course be analysed by statistical methods developed for this specific form of data (see for example McCullagh and Nelder (1983, Chap.5 and Chap.6)). We should point out, therefore, that the data is only a small part of a much bigger data set measuring a very large number of different characteristics of process (and operator) behaviour under the two methods of equipment operation. The full data set comprised a large number of samples of all sorts, including continuous as well as ordinal data, and it was not clear at the outset which samples and which characteristics would be important to look at. In this situation, the use of EDFIT statistics allowed a rapid initial exploration of the full data set, with both continuous and discrete data handled in the same way. Thus it was possible to rapidly determine those characteristics which appeared to behave differently under the two methods of equipment operation, whatever the form these differences took. This then allowed special follow-up analyses to be more effectively employed, focusing on just those characteristics revealed by the EDFIT analysis to behave differently under the two modes of operation.

A very brief preliminary analysis of the specific data set considered here has previously appeared in Cheng and Jones (2000). Note that the analysis used there does not include a Fourier decomposition, as described below.

The output comprised 782 independent observations of (real-time simulated) operator activity for each method, where each observation was on a simple integral scale ranging from 0 through 5 (0 indicating low activity, 5 indicating high activity). The observations were assumed to be independent. The observed activity level was also expected to be dependent on the production level which could be set at one of four levels. The two methods of operation were each simulated 782 times, with production levels 1, 2, 3 and 4 occurring a total of $50, 140, 346$ and 246 times respectively.

The number of observed activity levels for the different combinations of method ($i = 1, 2$) and production ($j = 1, 2, 3, 4$) levels are given in Table 1.

Use of ANOVA is clearly not appropriate in this case. A possible, though not ideal, method of analysis is to use the well known Friedman non-parametric test for a two-way layout (see for example Hollander and Wolfe (1973)). This test allows a number of matched samples to be compared. If we take the two operating methods as being the two 'treatments' and pair off each of the 50 observations at Production Level 1 using Operating Method 1 with a randomly selected observation using Operating Method 2, then this gives us 50 'matched' pairs. Doing the same thing with the observations at other Production Levels gives us a total of 782 'matched' pairs. The test statistic is calculated from this set of matched samples. (Details of the calculation are given in Hollander and Wolfe (1973).)

Table 1. Structure of data in the example.

| Operating Method 1 | | | | | Operating Method 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Production Level | 1 | 2 | 3 | 4 | Production Level | 1 | 2 | 3 | 4 |
| Activity Level | | | | | Activity Level | | | | |
| 0 | 0 | 1 | 6 | 4 | 0 | 1 | 3 | 6 | 7 |
| 1 | 7 | 9 | 8 | 1 | 1 | 0 | 7 | 25 | 7 |
| 2 | 20 | 83 | 200 | 153 | 2 | 24 | 86 | 188 | 150 |
| 3 | 23 | 43 | 132 | 79 | 3 | 25 | 36 | 110 | 72 |
| 4 | 0 | 4 | 0 | 9 | 4 | 0 | 8 | 17 | 9 |
| 5 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 1 |
| Total | 50 | 140 | 346 | 246 | Total | 50 | 140 | 346 | 246 |

There will be a large number of ties in the observations, and these are handled using the correction method given in Hollander and Wolfe (1973). The test value obtained was 0.465. To determine the level of significance, we used bootstrap resampling to form $1,000$ bootstrap test values under the null. This was done by obtaining two samples each of size 782, using bootstrap resampling, only with *both* bootstrap samples obtained from the one original sample corresponding to Operating Method 1. (To make sure that corresponding observations in each bootstrap sample were matched, each such pair was sampled from observations made at the same production level. This ensured that there was proper matching of production levels as in the original samples.)

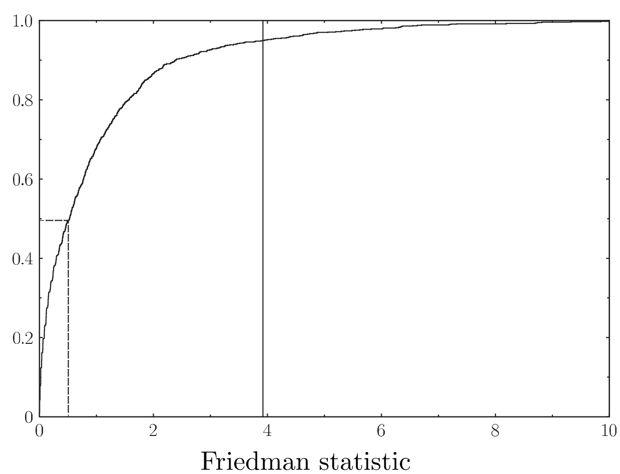


Figure 1. EDF for Friedman statistic and observed value.

Figure 1 shows the bootstrap EDF of $1,000$ values of the Friedman test statistic calculated from paired samples formed in this way. This EDF yields an estimate of the critical value at the 90% level of significance as 2.806, and a value of 3.704 at the 95% level of significance. This latter value is depicted in the figure. The actual test value of 0.465 is therefore nowhere near significant at either of these levels.

We also applied an EDFIT test. This was done by comparing, for each production level, the distribution of observations obtained under Operating Method 1 with that obtained under Operating Method 2, using $T^2$ as defined in (9). Thus there were four such statistics $T_i^2$ $i = 1, 2, 3, 4$ corresponding to the four production levels.

We then decomposed each into Fourier components using Theorem 3, $T_i^2 = \sum_{j=1}^{n_i} t_{ij}^2$. For simplicity, as the example is for illustration only, we did not consider the $T_i^2$ separately but considered the combined Fourier coefficients $C_j = \sum_{i=1}^{4} t_{ij}^2$, $j = 1, 2, 3$, $R = T^2 - C_1 - C_2 - C_3$, where $T^2 = \sum_{i=1}^{4} T_i^2$. Thus $C_1$, $C_2$ and $C_3$ provide overall measures of the difference in means, variances and shapes, respectively, of the distributions of activity level between the two operating methods across all four production levels. $R$ is a more complicated measure of any remaining overall difference between the distributions under the two operating methods. Table 2 gives the observed values of these quantities together with their p-values. The table also gives 90% and 95% critical values. The p-values and critical values were all obtained by bootstrap resampling. Figure 2 shows the null EDFs of $T^2$, $C_1$, $C_2$, $C_3$ and $R$, together with the observed values.

Table 2. EDFIT analysis of real-time simulation data.

| Component | Observed Value | p-value | 90% point | 95% point |
|---|---|---|---|---|
| $T^2$ | 0.756 | 0.164 | 0.90 | 1.08 |
| $C_1$ | 0.280 | 0.441 | 0.68 | 0.86 |
| $C_2$ | 0.347 | 0.002 | 0.16 | 0.19 |
| $C_3$ | 0.007 | 0.930 | 0.07 | 0.09 |
| $R$ | 0.123 | 0.092 | 0.12 | 0.13 |

It is seen that $T^2$, the overall measure of difference between distributions under the two operating methods, has a value of 0.756 which is not significant at the 90% level. The measure of difference between means, $C_1$ is not significant either. However $C_2$, the difference in variability, is significant at 95%. The value of $R$ is just significant at 90% but not 95%.

Closer examination of the spread of observations in the data set corroborates these findings. Though there is little to choose between the means under the two different operating methods, the variability about the mean does appear to be greater for the second operating method than for the first, and this feature of the data has been clearly exposed by the EDFIT analysis.

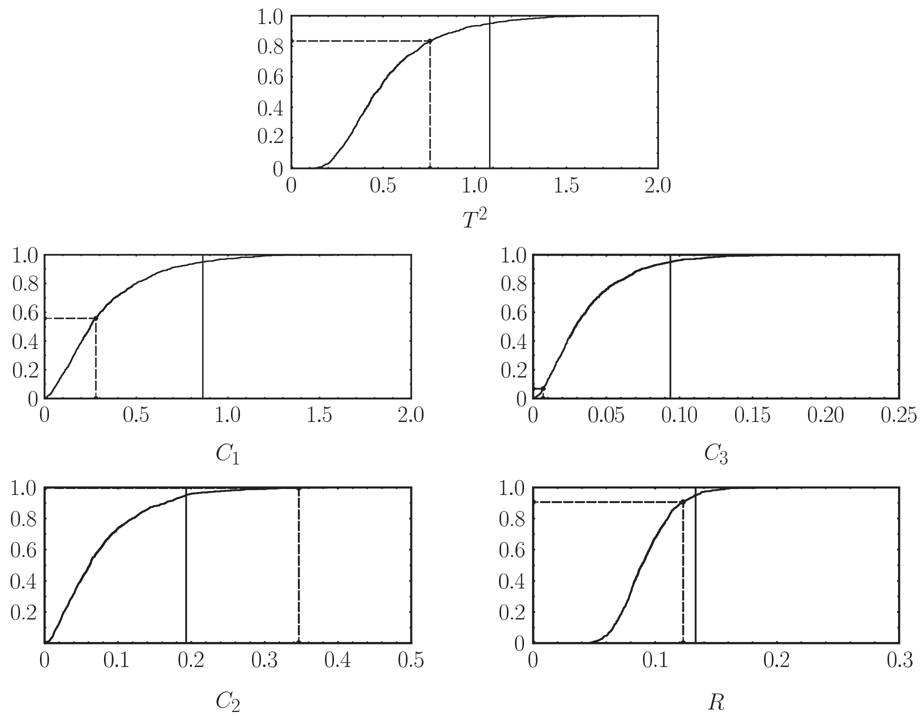


Figure 2. EDFs of $T^2$ and its components.

## 8. Conclusions

In factorial experiments where observations are replicated, the discussion suggests that the $EDFIT$ statistic can provide a much more sensitive test for distinguishing differences between responses at different factor combinations. This is backed up by the example, where a standard, normally quite powerful nonparametric test did not reveal differences between two treatments. In contrast the $EDFIT$ test indicated a significant difference. The reason why it gives a much more statistically significant result in this case is because the difference is not all that great between the means of the two samples. However there is a significant difference in the variability of the two samples. The sensitivity of the $EDFIT$ statistic to any difference between samples has therefore enabled this difference to be detected.

## References


Abramowitz, M. and Stegun, I. A. Eds. (1992). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Reprint of the 1972 edition. Dover, New York.

Anderson, T. W. (1962). On the distribution of the two-sample Cramer-von Mises criterion. *Ann. Math. Statist.* **33**, 1148-1159.

Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *Ann. Math. Statist.* **23**, 193-212.

Baumgartner, W., Weiss, P. and Schindler, H. (1998). A nonparametric test for the general two-sample problem. *Biometrics* **54**, 1129-1135.

Cheng, R. C. H. and Jones, O. D. (2000). Analysis of simulation factorial experiments by EDF resample statistics. *Proceedings of the* 2000 *Winter Simulation Conference* (Edited by J. A. Joines, R. Barton, P. Fishwick and K. Kang), 697-703.

Csörgö, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York.

Durbin, J. and Knott, M. (1972). Components of Cramer-von Mises statistics I. *J. Roy. Statist. Soc.* **34**, 290-307.

Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the multinomial estimator. *Ann. Math. Statist.* **27**, 642-669.

Hollander, M. and Wolfe, D. A. (1973). *Nonparametric Statistical Methods.* Wiley, New York.

Kiefer, J. (1959). $K$-sample analogues of the Kolmogorov-Smirnov and Cramér-von Mises tests. *Ann. Math. Statist.* **30**, 420-447.

Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent R.V.'s and the sample DF. I. *Z. Wahr. Verw. Gebiete* **32**, 111-131.

McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models.* Chapman and Hall, London.

Searle, S. R. (1971). *Linear Models.* Wiley, New York.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

Stuart, A. and Ord, J. K. (1991). *Kendall's Advanced Theory of Statistics*, Vol. 2. Edward Arnold, London.

Faculty of Mathematical Studies, University of Southampton, Southampton SO17 1BJ, U.K.

E-mail: rchc@maths.soton.ac.uk

Department of Mathematics and Statistics, The University of Melbourne, Parkville, VIC, 3010, Australia.

E-mail: O.D.Jones@ms.unimelb.edu.au

(Received November 2002; accepted October 2003)