

# A Semiparametric Approach to the One-Way Layout \*

Konstantinos Fokianos, Benjamin Kedem, Jing Qin, and David A. Short

June 1998  
Revised: April 2000

---

\*Konstantinos Fokianos is Lecturer, Department of Mathematics & Statistics, University of Cyprus, P.O. Box 20537, Nicosia 1678, Cyprus (email: fokianos@ucy.ac.cy). Benjamin Kedem is Professor, Department of Mathematics, University of Maryland, College Park, MD, 20742, U.S.A (email: bnk@math.umd.edu). Jing Qin is Assistant Attending Biostatistician, Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, U.S.A (email: qinj@biosta.mskcc.org). David Short is Senior Scientist, Applied Meteorology Unit, ENSCO Inc., 1980 N. Atlantic Avenue, Cocoa Beach, FL 32955, U.S.A. (email: shortd@fl.ensco.com).

### Abstract

We consider  $m$  distributions where the first  $m - 1$  are obtained by multiplicative exponential distortions of the  $m$ th distribution, it being a reference. The combined data from  $m$  samples, one from each distribution, are used in the semiparametric large sample problem of estimating each distortion and the reference distribution, and testing the hypothesis that the distributions are identical. The approach generalizes the classical normal based one way analysis of variance in the sense that it obviates the need for a completely specified parametric model. An advantage is that the probability density of the reference distribution is estimated from the combined data and not only from the  $m$ th sample. A power comparison with the t and F tests and with two nonparametric tests, obtained by means of a simulation, points to the merit of the present approach. The method is applied to rain rate data from meteorological instruments.

**Keywords:** exponential tilt, reference distribution, empirical likelihood, logistic regression, smooth goodness of fit, case-control data.

# 1 Introduction

Statistical techniques based on normal theory have been central to the development and teaching of statistics. All along, however, there have been numerous studies of the consequences of departure from the normal assumption, and of transformation methods that produce nearly normal data. As examples we mention in particular the work of Miller (1986) who discusses situations where the normal and other assumptions break down, and the well known normalizing Box-Cox transformation. The present work formulates an approach to analysis of variance that relaxes the normal assumption.

The present paper provides an alternative to the classical normal based one way analysis of variance by modeling the log ratio of the relevant probability densities with respect to a reference density. In the classical normal theory with equal variances, the log ratio takes on the form  $\alpha + \beta x$ . However as already was observed by Kay and Little (1987), there are cases where  $\alpha + \beta h(x)$ , for some  $h(x)$ , is more appropriate. For example, for certain lognormal and gamma populations,  $h(x) = \log(x)$  is precisely the right choice. In these cases, the classical F and t tests, where  $h(x) = x$  is used by default, have less power than the test provided here as simulation results show. Curiously, simulation results indicate that our test can be, approximately, as powerful as the classical t and F tests when the data are normal for moderate and large samples.

The approach presented here provides the mechanism for a general one way layout testing for any  $h(x)$ , including non-normal cases where we still have  $h(x) = x$ , without the knowledge of the reference distribution. This problem is better understood by taking a close look at the classical case first.

Consider the classical one-way analysis of variance with  $m = q + 1$  independent normal random samples,

$$\begin{aligned} x_{11}, \dots, x_{1n_1} &\sim g_1(x) \\ &\cdot \\ &\cdot \\ &\cdot \\ x_{q1}, \dots, x_{qn_q} &\sim g_q(x) \\ x_{m1}, \dots, x_{mn_m} &\sim g_m(x) \end{aligned}$$

where  $g_j(x)$  is the probability density of  $N(\mu_j, \sigma^2)$ ,  $j = 1, \dots, m$ . Then, holding  $g_m(x)$

as a reference,

$$\frac{g_j(x)}{g_m(x)} = \exp(\alpha_j + \beta_j x), \quad j = 1, \dots, q \quad (1)$$

where

$$\alpha_j = \frac{\mu_m^2 - \mu_j^2}{2\sigma^2}, \quad \beta_j = \frac{\mu_j - \mu_m}{\sigma^2}, \quad j = 1, \dots, q$$

It follows the test  $H_0 : \mu_1 = \dots = \mu_m$  is equivalent to  $H_0 : \beta_1 = \dots = \beta_q = 0$ . Clearly  $\beta_j = 0$  implies  $\alpha_j = 0$ ,  $j = 1, \dots, q$ .

An immediate generalization is obtained by eliminating the normal assumption, and regarding each  $g_j(x)$ ,  $j = 1, \dots, q$ , directly as an exponential distortion or tilt of a reference  $g_m(x)$ ,

$$\frac{g_j(x)}{g_m(x)} = \exp(\alpha_j + \beta_j h(x)), \quad j = 1, \dots, q \quad (2)$$

where  $h(x)$  is an arbitrary but known function of  $x$ . Again, since  $g_m(x)$  is a density,  $\beta_j = 0$  implies  $\alpha_j = 0$ ,  $j = 1, \dots, q$ , and the hypothesis  $H_0 : \beta_1 = \dots = \beta_q = 0$  implies that all the  $m$  populations are equidistributed, namely,  $g_j = g_m$ ,  $j = 1, \dots, q$ . The general unspecified form for  $g_m(x)$  constitutes the main departure from the classical one-way layout with normally distributed data.

An example of (2) is provided by multinomial logistic regression. Consider a categorical random variable  $y$  such that  $P(y = j) = \pi_j$  where  $f(x|y = j) = g_j(x)$ ,  $j = 1, \dots, m$ , and  $\sum_{j=1}^m \pi_j = 1$ . If

$$P(y = j|x) = \frac{\exp(\alpha_j^* + \beta_j h(x))}{1 + \sum_{k=1}^q \exp(\alpha_k^* + \beta_k h(x))}, \quad j = 1, \dots, m,$$

then an appeal to Bayes theorem shows that (2) holds with  $\alpha_j = \alpha_j^* + \log[\pi_m/\pi_j]$ ,  $j = 1, \dots, q$ .

Many authors have studied exponential distortions which resemble closely the form (2) in regard to goodness of fit, logistic regression, and classification models. An important example is the work of Neyman (1937) who introduced the notion of smooth goodness of fit test. Let us describe it briefly. Suppose we are interested in testing the null hypothesis that  $x_1, \dots, x_n$  is a random sample from a continuous distribution with probability density function  $f(x; \beta)$ , where  $\beta$  is a  $p \times 1$  vector. The first step is to embed the null probability density function in an order  $k$  alternative

$$g(x; \theta, \beta) = C(\theta, \beta) \exp\left\{\sum_{i=1}^k \theta_i h_i(x, \beta)\right\} f(x; \beta)$$

where  $\{h_i(x; \beta)\}$  are complete and orthonormal with respect to  $f(x; \beta)$  with  $h_0(x; \beta) \equiv 1$ , and  $C(\theta, \beta)$  is a normalized constant. Testing for  $f(x; \beta)$  is equivalent to testing  $H_0 : \boldsymbol{\theta} = (\theta_1, \dots, \theta_k)' = \mathbf{0}$ . Thus, we can see that in testing the equality of densities in  $m$ -sample problems, model (2) is a natural extension of Neyman's smooth goodness of fit test, where the form of the reference density  $g_m$  is left unspecified. For a concise description as well as a historical account of the development of smooth goodness of fit tests we refer the reader to the monograph by Rayner and Best (1989).

Other authors who have studied exponential distortions similar to (2) include Cox (1966), Anderson (1972, 1982), Prentice and Pyke (1979), Kay and Little (1987), Efron and Tibshirani (1996), and Qin and Zhang (1997).

For a 0-1 response  $Y$  and an explanatory variable  $X$ , Kay and Little (1987) observed that if the log ratio  $\log\{f(x|Y = 1)/f(x|Y = 0)\} = \alpha + \beta h(x)$ , the logistic regression model is the correct model for  $Y$  given  $X$ . Motivated by this fact they studied the improvement in the fit of logistic regression by considering models of the form (2) with transformation  $h(x)$ . They tabulate  $h(x)$  for some commonly used members of the exponential family. Efron and Tibshirani (1996) consider a model that has the form (2) but in the case of a single sample, where a probability density is a product of a *carrier* density and an exponential factor. Their idea is to estimate first the carrier density by a kernel density estimator and then estimate the parameters in the exponential factor by maximum likelihood, ignoring the fact that the carrier is data dependent. The resulting hybrid estimator, referred to as a specially designed exponential family, is a compromise between parametric and nonparametric density estimators. More recently, Qin and Zhang (1997) tested the validity of logistic regression under case-control sampling where (2) holds with  $m = 2$  and  $h(x) = x$ .

Motivated by the preceding discussion, we shall discuss a generalization of the classical one-way layout classification by considering the exponential tilt (2) with  $g \equiv g_m$ ,

$$g_j(x) = \exp(\alpha_j + \beta_j h(x))g(x), \quad j = 1, \dots, q \quad (3)$$

where  $\alpha_j$  depends on  $\beta_j$ . For  $h(x) = x$ ,  $\alpha_j$  is determined explicitly by  $\beta_j$  through the moment generating function  $M_g$  corresponding to  $g$ ,

$$\alpha_j = -\log\{M_g(\beta_j)\}, \quad j = 1, \dots, q$$

Denote the combined data from the  $m$  samples by  $\mathbf{t}$ ,

$$\mathbf{t} = (t_1, \dots, t_n)' = (\mathbf{x}'_1, \dots, \mathbf{x}'_q, \mathbf{x}'_m)'$$

where  $\mathbf{x}_j = (x_{j1}, \dots, x_{jn_j})'$ , and  $n = n_1 + \dots + n_q + n_m$ . In this paper we investigate the following semiparametric estimation/testing problems using the combined data  $\mathbf{t}$ .

1. Nonparametric estimation of  $G(x)$ , the cdf corresponding to  $g(x)$ .
2. Estimation of the parameters  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_q)'$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_q)'$ , and the study of the large sample properties of the estimators.
3. Testing of the hypothesis  $H_0 : \beta_1 = \dots = \beta_q = 0$ .

Evidently, the general construction does not require normality or even symmetry of the distributions, the variances need not be the same, and the model does not require knowledge of the reference distribution. The main assumption is the form of the distortion of the reference distribution, softened by the choice of the “distortion function”  $h(x)$ . Notice that the reference distribution may be any of the  $m$  distributions, leaving the exponential distortion intact but with shifted parameters.

## 1.1 An Application: Combination of Instruments

A possible application of (3) and the ensuing statistical analysis is in the combination of several instruments, a special case of which is discussed in Fokianos et al. (1997).

Suppose  $m$  instruments  $I_1, \dots, I_q, I_m$  measure the same quantity with the same resolution, where it is known that  $I_m$  is more reliable than the rest. The  $j$ 'th instrument  $I_j$  produces a set of measurements  $\mathbf{x}_j$ ,  $j = 1, \dots, q, m$ , and  $I_1, \dots, I_q$  are assumed a distortion of  $I_m$  as expressed by (3). The problem is to combine the information from all the instruments to increase the reliability of  $I_m$ , that is, to construct an improved estimate of  $g$  from  $\mathbf{t} = (\mathbf{x}'_1, \dots, \mathbf{x}'_q, \mathbf{x}'_m)'$ . Since the data from each instrument contain information about  $g$ , a more precise estimate of  $g$  can be obtained by using the combined data  $\mathbf{t}$ , and not just  $\mathbf{x}_m$  alone. The deviation of each instrument from the reference  $I_m$  can be quantified by the estimation of the  $\alpha_j$  and  $\beta_j$ , and used in calibration. An example where a radar and two radiometers are combined, all measuring rain rate, is discussed below.

## 2 Estimation and Large Sample Results

This section follows the construction in Qin and Lawless (1994), and Qin and Zhang (1997).

A maximum likelihood estimator of  $G(x)$  can be obtained by maximizing the likelihood over the class of step cdf's with jumps at the observed values  $t_1, \dots, t_n$ . Accordingly, if  $p_i = dG(t_i)$ ,  $i = 1, \dots, n$ , the likelihood becomes,

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, G) = \prod_{i=1}^n p_i \prod_{j=1}^{n_1} \exp(\alpha_1 + \beta_1 h(x_{1j})) \cdots \prod_{j=1}^{n_q} \exp(\alpha_q + \beta_q h(x_{qj})) \quad (4)$$

We follow a profiling procedure whereby first we express each  $p_i$  in terms of  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  and then we substitute the  $p_i$  back into the likelihood to produce a function of  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  only. When  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  are fixed, (4) is maximized by maximizing only the the product term  $\prod_{i=1}^n p_i$ , subject to the  $m$  constraints

$$\sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i [w_1(t_i) - 1] = 0, \dots, \sum_{i=1}^n p_i [w_q(t_i) - 1] = 0$$

where

$$w_j(t) = \exp(\alpha_j + \beta_j h(t)), \quad j = 1, \dots, q$$

The maximization employs the method of Lagrange multipliers the first of which becomes  $\lambda_0 = n$ , and the rest are expressed by construction as  $\lambda_j = \nu_j n$ ,  $j = 1, \dots, q$ , for some  $\nu_j$ . It follows that

$$p_i = \frac{1}{n} \frac{1}{1 + \nu_1 (w_1(t_i) - 1) + \cdots + \nu_q (w_q(t_i) - 1)} \quad (5)$$

which together with the constraints gives a set of equations satisfied by the  $\nu_j$ ,

$$\frac{1}{n} \sum_{i=1}^n \frac{w_j(t_i) - 1}{1 + \nu_1 (w_1(t_i) - 1) + \cdots + \nu_q (w_q(t_i) - 1)} = 0, \quad j = 1, \dots, q \quad (6)$$

Substituting  $p_i$  in  $\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, G)$ , the loglikelihood becomes up to a constant,

$$\begin{aligned} l \equiv \log \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, G) = & - \sum_{i=1}^n \log[1 + \nu_1 (w_1(t_i) - 1) + \cdots + \nu_q (w_q(t_i) - 1)] \\ & + \sum_{j=1}^{n_1} [\alpha_1 + \beta_1 h(x_{1j})] + \cdots + \sum_{j=1}^{n_q} [\alpha_q + \beta_q h(x_{qj})] \end{aligned}$$

To get expressions for the  $\nu_j$ , we set  $\partial l / \partial \alpha_j = 0$ ,  $j = 1, \dots, q$ , and using the equations (6), we obtain

$$\nu_j = \frac{n_j}{n}, \quad j = 1, \dots, q$$

Substituting these values of  $\nu_j$  in equation (5) we have

$$p_i = \frac{1}{n_m} \cdot \frac{1}{1 + \rho_1 w_1(t_i) + \cdots + \rho_q w_q(t_i)} \quad (7)$$

where  $\rho_j = n_j/n_m$ ,  $j = 1, \dots, q$ , and the value of the profile loglikelihood up to a constant as a function of  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  only is,

$$\begin{aligned} l = & - \sum_{i=1}^n \log[1 + \rho_1 w_1(t_i) + \cdots + \rho_q w_q(t_i)] \\ & + \sum_{j=1}^{n_1} [\alpha_1 + \beta_1 h(x_{1j})] + \cdots + \sum_{j=1}^{n_q} [\alpha_q + \beta_q h(x_{qj})] \end{aligned} \quad (8)$$

The score equations for  $j = 1, \dots, q$ , are therefore,

$$\begin{aligned} \frac{\partial l}{\partial \alpha_j} &= - \sum_{i=1}^n \frac{\rho_j w_j(t_i)}{1 + \rho_1 w_1(t_i) + \cdots + \rho_q w_q(t_i)} + n_j = 0 \\ \frac{\partial l}{\partial \beta_j} &= - \sum_{i=1}^n \frac{\rho_j h(t_i) w_j(t_i)}{1 + \rho_1 w_1(t_i) + \cdots + \rho_q w_q(t_i)} + \sum_{i=1}^{n_j} h(x_{ji}) = 0 \end{aligned} \quad (9)$$

The solution of the score equations gives the maximum likelihood estimators  $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}$ , and consequently by substitution also

$$\hat{p}_i = \frac{1}{n_m} \cdot \frac{1}{1 + \rho_1 \exp(\hat{\alpha}_1 + \hat{\beta}_1 h(t_i)) + \cdots + \rho_q \exp(\hat{\alpha}_q + \hat{\beta}_q h(t_i))} \quad (10)$$

and therefore,

$$\hat{G}(t) = \frac{1}{n_m} \cdot \sum_{i=1}^n \frac{I(t_i \leq t)}{1 + \rho_1 \exp(\hat{\alpha}_1 + \hat{\beta}_1 h(t_i)) + \cdots + \rho_q \exp(\hat{\alpha}_q + \hat{\beta}_q h(t_i))}. \quad (11)$$

Summarizing, by following a profiling procedure, we obtained a nonparametric estimator (11) for  $G(x)$ , and score estimating equations (9) for the parameters  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$ . It is argued in the appendix that the estimators  $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}$ , are asymptotically normal,

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} \Rightarrow N(\mathbf{0}, \boldsymbol{\Sigma}) \quad (12)$$

as  $n \rightarrow \infty$ . Here  $\boldsymbol{\alpha}_0$  and  $\boldsymbol{\beta}_0$  denote the true parameters and  $\boldsymbol{\Sigma} = \mathbf{S}^{-1} \mathbf{V} \mathbf{S}^{-1}$ , where the matrices  $\mathbf{S}$  and  $\mathbf{V}$  are defined in the appendix.

## 2.1 Some Simulation Results

To illustrate empirically the asymptotic normality result (12) we performed a small simulation study with 500 runs in each of four cases with  $h(x) = x$  throughout the simulation.

Consider first the case of three uniform populations on  $(0, 1)$ , that is  $g(x) = 1$  for  $0 \leq x \leq 1$  and  $g(x) = 0$ , otherwise, with  $g_1(x) = g_2(x) = g(x)$  and  $q = 2$ . It follows that  $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 0$ . We generated  $n_1 = n_2 = n_3 = 200$  observations from each population. The combined sample therefore consists of 600 observations. The theoretical variance–covariance matrix can be calculated easily (see Section 3). It turns out that

$$\frac{1}{600}\Sigma = \begin{pmatrix} 0.030 & 0.015 & -0.060 & -0.030 \\ 0.015 & 0.030 & -0.030 & -0.060 \\ -0.060 & -0.030 & 0.120 & 0.060 \\ -0.030 & -0.060 & 0.060 & 0.120 \end{pmatrix}.$$

It follows that the corresponding theoretical standard errors are 0.1732, 0.1732, 0.3464 and 0.3464 for  $\hat{\alpha}_1$ ,  $\hat{\alpha}_2$ ,  $\hat{\beta}_1$  and  $\hat{\beta}_2$ , respectively. The first row of Table 1 summarizes our simulation results in this case. The last column gives the estimated parameters together with their corresponding standard errors. There is a good agreement between the theoretical and estimated standard errors. The second row of Table 1 reports results from the same scenario as before but now  $n_1 = 200$ ,  $n_2 = 300$  and  $n_3 = 100$ . Consequently  $\rho_1 = 2$  and  $\rho_2 = 3$ . In this case, we have that

$$\frac{1}{600}\Sigma = \begin{pmatrix} 0.045 & 0.030 & -0.090 & -0.060 \\ 0.030 & 0.040 & -0.060 & -0.080 \\ -0.090 & -0.060 & 0.180 & 0.120 \\ -0.060 & -0.080 & 0.120 & 0.160 \end{pmatrix}.$$

Here, the corresponding theoretical standard errors are 0.2121, 0.2, 0.4242 and 0.4 for  $\hat{\alpha}_1$ ,  $\hat{\alpha}_2$ ,  $\hat{\beta}_1$  and  $\hat{\beta}_2$ , respectively. The estimated standard errors are given in the second row of Table 1. Again, there is a good agreement between the theoretical and estimated standard errors. The last two rows of Table 1 refer to the situation where the reference distribution  $g(x)$  is  $N(0, 1)$ ,  $g_1(x)$  is  $N(3, 1)$  and  $g_2(x)$  is  $N(2, 1)$ . In this case,  $\alpha_1 = -4.5$ ,  $\beta_1 = 3$ ,  $\alpha_2 = -2$  and  $\beta_2 = 2$ . The third row summarizes the results when  $n_1 = n_2 = n_3 = 200$  and the last row for  $n_1 = 200$ ,  $n_2 = 300$  and  $n_3 = 100$ . In all the cases, we see that the estimators are close to the true values.

## 2.2 The Case $m=2$

The case  $m = 2, q = 1$  requires a slight change in notation. For  $k = 0, 1, 2$ , define,

$$A_k = \int \frac{h^k(t) \exp(\alpha + \beta h(t))}{1 + \rho \exp(\alpha + \beta h(t))} dG(t)$$

and,

$$\mathbf{A} = \begin{pmatrix} A_0 & A_1 \\ A_1 & A_2 \end{pmatrix}$$

With  $\rho \equiv \rho_1$ , Qin and Zhang (1997) showed

$$\boldsymbol{\Sigma} = \mathbf{S}^{-1} \mathbf{V} \mathbf{S}^{-1} = \frac{1 + \rho}{\rho} \left[ \mathbf{A}^{-1} - \begin{pmatrix} 1 + \rho & 0 \\ 0 & 0 \end{pmatrix} \right]$$

so that under some regularity conditions and regardless of  $h(x)$ ,

$$\sqrt{n} \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} \Rightarrow N(\mathbf{0}, \boldsymbol{\Sigma}) \quad (13)$$

where  $\alpha_0, \beta_0$  are the true parameters.

As an illustration consider the case where  $\mathbf{x}_2$  is uniformly distributed in  $[0, 1]$ , so that  $g(x) = 1, 0 \leq x \leq 1$ , and  $g(x) = 0$  otherwise. Assume  $\rho = 1, \alpha = \beta = 0$ , and  $h(x) = x$ , and observe that when  $\alpha = \beta = 0$  the two populations are identical. As  $n \rightarrow \infty$ , the asymptotic covariance matrix  $\boldsymbol{\Sigma}$  can be obtained exactly from

$$A_k = \int_0^1 \frac{t^k \exp(\alpha + \beta t)}{1 + \rho \exp(\alpha + \beta t)} dG(t) = \frac{1}{2(k+1)}, \quad k = 0, 1, 2$$

so that

$$\mathbf{A} = \begin{pmatrix} 1/2 & 1/4 \\ 1/4 & 1/6 \end{pmatrix}$$

and

$$\boldsymbol{\Sigma} = \frac{1 + \rho}{\rho} \left[ \mathbf{A}^{-1} - \begin{pmatrix} 1 + \rho & 0 \\ 0 & 0 \end{pmatrix} \right] = \begin{pmatrix} 12 & -24 \\ -24 & 48 \end{pmatrix}$$

### 2.3 Mean Estimation

Consider  $h(\cdot)$  appearing in (3). The first two moments of  $h(t)$  with respect to  $g$  are needed for hypothesis testing in the next section. The mean of  $h(t)$ ,

$$\int h(t)dG(t)$$

can be estimated from the combined data using the estimator  $\sum_{i=1}^n h(t_i)\hat{p}_i$ , or by taking the average of  $h(x_{m1}), \dots, h(x_{mnm})$ . Interestingly, the two estimates are identical. To see this, notice that from (10) we can get an expression for  $h(t_i)[1 - n_m\hat{p}_i]$ . Summing this over  $i$  and invoking (9) for  $\beta_j, j = 1, \dots, q$ , we have

$$\sum_{i=1}^n h(t_i)[1 - n_m\hat{p}_i] = \sum_{i=1}^{n_1} h(x_{1i}) + \dots + \sum_{i=1}^{n_q} h(x_{qi}) = \sum_{i=1}^n h(t_i) - \sum_{i=1}^{n_m} h(x_{mi})$$

since  $(t_1, \dots, t_n) = (\mathbf{x}'_1, \dots, \mathbf{x}'_q, \mathbf{x}'_m)$ . Therefore,

$$\sum_{i=1}^n h(t_i)\hat{p}_i = \frac{1}{n_m} \sum_{i=1}^{n_m} h(x_{mi}) \tag{14}$$

This however is not the case for higher order moments of  $h(t)$ , and the combined estimate is not the same as the corresponding estimate from the  $m$ th sample (see also the discussion in Efron and Tibshirani (1996)).

## 3 Hypothesis Testing

We are now in a position to test the hypothesis  $H_0 : \boldsymbol{\beta} = \mathbf{0}$  that all the  $m$  populations are equidistributed. Several possibilities exist; perhaps the simplest is to use the score test using the score equations (9) for the  $\beta_j, j = 1, \dots, q$ , thus eliminating the need to evaluate  $\hat{\boldsymbol{\beta}}$ . Accordingly, under  $H_0 : \boldsymbol{\beta} = \mathbf{0}$ , the score equations reduce to

$$\frac{\partial l}{\partial \beta_j} \Big|_{\boldsymbol{\beta}=\mathbf{0}} = n_j \{ \overline{h(x_j)} - \overline{h(t)} \}, \quad j = 1, \dots, q$$

The basis for the test is the fact  $E[\partial l / \partial \boldsymbol{\beta}] = \mathbf{0}$ , which implies that the score equations should themselves be close to zero as well. However, having gone through the estimation and large sample study, we opt for the more direct and more intuitive alternative which relies on the asymptotic properties of  $\hat{\boldsymbol{\beta}}$ .

We shall use the following notation for the moments of  $h(t)$  with respect to the reference distribution:

$$E(t^k) \equiv \int h^k(t) dG(t)$$

$$Var(t) \equiv E(t^2) - E^2(t)$$

Under  $H_0 : \beta = \mathbf{0}$ —so that all the moments of  $h(t)$  are taken with respect to  $g$ —consider the  $q \times q$  matrix  $\mathbf{A}_{11}$  whose  $j$ th diagonal element is

$$\frac{\rho_j [1 + \sum_{k \neq j}^q \rho_k]}{[1 + \sum_{k=1}^q \rho_k]^2}$$

and otherwise for  $j \neq j'$ , the  $jj'$  element is

$$\frac{-\rho_j \rho_{j'}}{[1 + \sum_{k=1}^q \rho_k]^2}$$

The elements are bounded by 1 and the matrix is nonsingular,

$$|\mathbf{A}_{11}| = \frac{\prod_{k=1}^q \rho_k}{[1 + \sum_{k=1}^q \rho_k]^m} > 0$$

and can be used to represent  $\mathbf{S}$ ,

$$\mathbf{S} = \begin{pmatrix} 1 & E(t) \\ E(t) & E(t^2) \end{pmatrix} \otimes \mathbf{A}_{11}$$

with  $\otimes$  denoting the Kronecker product. It follows that  $\mathbf{S}$  is nonsingular,

$$|\mathbf{S}| = \{Var(t)\}^q |\mathbf{A}_{11}|^2$$

and,

$$\mathbf{S}^{-1} = \frac{1}{Var(t)} \begin{pmatrix} E(t^2) & -E(t) \\ -E(t) & 1 \end{pmatrix} \otimes \mathbf{A}_{11}^{-1}$$

On the other hand,  $\mathbf{V}$  is singular,

$$\mathbf{V} = Var(t) \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{11} \end{pmatrix}$$

as is

$$\boldsymbol{\Sigma} = \mathbf{S}^{-1} \mathbf{V} \mathbf{S}^{-1} = \frac{1}{Var(t)} \begin{pmatrix} E^2(t) & -E(t) \\ -E(t) & 1 \end{pmatrix} \otimes \mathbf{A}_{11}^{-1}$$

Luckily the right component is nonsingular and we finally have from (12),

$$\sqrt{n}\hat{\boldsymbol{\beta}} \Rightarrow N\left(\mathbf{0}, \frac{1}{\text{Var}(t)} \mathbf{A}_{11}^{-1}\right) \quad (15)$$

It follows under  $H_0 : \boldsymbol{\beta} = \mathbf{0}$

$$\mathcal{X}_1 = n\text{Var}(t)\hat{\boldsymbol{\beta}}' \mathbf{A}_{11}\hat{\boldsymbol{\beta}} \quad (16)$$

is approximately distributed as  $\chi^2(q)$ , and  $H_0$  can be rejected for large values of  $n\text{Var}(t)\hat{\boldsymbol{\beta}}' \mathbf{A}_{11}\hat{\boldsymbol{\beta}}$ .

### 3.1 Some Comments About $\mathbf{A}_{11}$

Due to its importance in testing  $H_0 : \boldsymbol{\beta} = \mathbf{0}$ , it is instructive to consider some special cases of  $\mathbf{A}_{11}$ . For  $m = 2, q = 1$ ,  $\mathbf{A}_{11}$  reduces to a scalar  $\rho_1/(1 + \rho_1)^2$ . For  $m = 3, q = 2$ ,

$$\mathbf{A}_{11} = \frac{1}{(1 + \rho_1 + \rho_2)^2} \cdot \mathbf{M}_2$$

where

$$\mathbf{M}_2 = \begin{pmatrix} \rho_1(1 + \rho_2) & -\rho_1\rho_2 \\ -\rho_1\rho_2 & \rho_2(1 + \rho_1) \end{pmatrix} = \begin{pmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{pmatrix} \begin{pmatrix} 1 + \rho_2 & -\rho_2 \\ -\rho_1 & 1 + \rho_1 \end{pmatrix}$$

and the eigenvalues of the matrix on the right are  $1, 1 + \rho_1 + \rho_2$ . For  $m = 4, q = 3$ ,

$$\mathbf{A}_{11} = \frac{1}{(1 + \rho_1 + \rho_2 + \rho_3)^2} \cdot \mathbf{M}_3$$

where

$$\mathbf{M}_3 = \begin{pmatrix} \rho_1(1 + \rho_2 + \rho_3) & -\rho_1\rho_2 & -\rho_1\rho_3 \\ -\rho_1\rho_2 & \rho_2(1 + \rho_1 + \rho_3) & -\rho_2\rho_3 \\ -\rho_1\rho_3 & -\rho_2\rho_3 & \rho_3(1 + \rho_1 + \rho_2) \end{pmatrix}$$

Note that  $\mathbf{M}_3$  can be decomposed as,

$$\mathbf{M}_3 = \begin{pmatrix} \rho_1 & 0 & 0 \\ 0 & \rho_2 & 0 \\ 0 & 0 & \rho_3 \end{pmatrix} \begin{pmatrix} 1 + \rho_2 + \rho_3 & -\rho_2 & -\rho_3 \\ -\rho_1 & 1 + \rho_1 + \rho_3 & -\rho_3 \\ -\rho_1 & -\rho_2 & 1 + \rho_1 + \rho_2 \end{pmatrix}$$

where the eigenvalues of the matrix on the right are

$$1, 1 + \rho_1 + \rho_2 + \rho_3, 1 + \rho_1 + \rho_2 + \rho_3$$

In general

$$\mathbf{A}_{11} = \frac{1}{(1 + \sum_{k=1}^q \rho_k)^2} \cdot \mathbf{M}_q$$

where  $\mathbf{M}_q$  can be decomposed into a diagonal matrix  $Diag(\rho_1, \dots, \rho_q)$  times a  $q \times q$  “matrix on the right” that has eigenvalues 1 and  $1 + \sum_{k=1}^q \rho_k$  with multiplicity  $q - 1$ .

### 3.2 Testing the Linear Hypothesis

We can further test the general linear hypothesis  $\mathbf{H}\boldsymbol{\theta} = \mathbf{c}$  where  $\mathbf{H}$  is a  $p \times 2q$  predetermined matrix of rank  $p$  ( $p < 2q$ ),  $\boldsymbol{\theta} = (\alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_q)'$ , and  $\mathbf{c}$  is a vector in  $R^p$ . Then, using (12), we have under the hypothesis

$$\sqrt{n}(\mathbf{H}\hat{\boldsymbol{\theta}} - \mathbf{c}) \Rightarrow N(\mathbf{0}, \mathbf{H}\boldsymbol{\Sigma}\mathbf{H}')$$

Thus, the random variable

$$\mathcal{X}_2 = n(\mathbf{H}\hat{\boldsymbol{\theta}} - \mathbf{c})'(\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}')^{-1}(\mathbf{H}\hat{\boldsymbol{\theta}} - \mathbf{c}) \quad (17)$$

has an asymptotic chi-square distribution with  $p$  degrees of freedom provided the inverse exists (Sen and Singer 1993 p. 239). A consistent estimator of  $\boldsymbol{\Sigma}$  can be obtained by replacing all the parameters by their maximum likelihood estimates.

Note that in general the results obtained from (16) and (17) are different, since in (16) we substitute the exact value  $\boldsymbol{\beta} = \mathbf{0}$  in  $\boldsymbol{\Sigma}$ , while (17) requires the maximum likelihood estimate of  $\boldsymbol{\theta}$  instead.

### 3.3 Power Comparison With the t and F Tests

We report here simulation results in which the power of  $\chi_1$  defined in (16) is compared with the power of the two sample t and Wilcoxon rank sum (W) tests for  $m = 2$ , and with the F and Kruskal–Wallis (K–W) tests for  $m = 3$ ; see Randles and Wolfe (1979). As is the case in practice,  $Var(t)$  needed for  $\chi_1$ , and defined above as the variance of  $h(t)$  (not of  $t$  unless  $h(t) = t$ ), is estimated from

$$\sum_{i=1}^n h^2(t_i)\hat{p}_i - \left( \sum_{i=1}^n h(t_i)\hat{p}_i \right)^2$$

Three cases are considered, normal, lognormal, and gamma, where the respective reference distributions are  $N(0,1)$ ,  $LN(0,1)$ , and  $\text{Gamma}(3,1)$ .

Consider first the two sample case,  $m = 2, q = 1$ . In the normal case,  $\mathbf{x}_1 \sim N(\beta_1, 1)$ ,  $\mathbf{x}_2 \sim N(0, 1)$ , in the lognormal case  $\mathbf{x}_1 \sim LN(\beta_1, 1)$ ,  $\mathbf{x}_2 \sim LN(0, 1)$ , and in the gamma case,  $\mathbf{x}_1 \sim \text{Gamma}(3+\beta_1, 1)$ ,  $\mathbf{x}_2 \sim \text{Gamma}(3, 1)$ . Under our formulation we test  $H_0 : \beta_1 = 0$ , while under the t-test we test the equivalent hypothesis  $H_0 : \mu_1 = \mu_2$ . Both hypotheses imply that the respective distributions are identical.

The power results as a function of  $\beta_1$  are given in Table 2 for a nominal level of 0.05. Each power entry in the table was obtained from 150 independent runs. The fact that the  $\chi_1$ -test displays more power than the t-test in the lognormal and gamma cases shows that a departure from the classical normal and variance equality assumptions can weaken the t-test considerably. Apparently, our test dominates the Wilcoxon rank sum test in all the cases considered. Interestingly, the  $\chi_1$ -test is not dominated by the t-test in the present normal example with equal variances. More precisely, if  $p$  is a power entry in Table 2 corresponding to the t-test under normality, then the standard error  $\sqrt{p(1-p)/150}$  ranges from 0.023 ( $p = 0.087$ ) to 0.009 ( $p = 0.987$ ). Whence differences in power between the  $\chi_1$ -test and the t-test in the normal case are largely insignificant, but those for lognormal and gamma data are.

Very similar power results were obtained in the normal case even after making sure that the observed size was identical for both the  $\chi_1$ - and t-tests. Thus for an observed size of 0.05333333 in both cases, the power corresponding to the  $\beta_1$  values in Table 2 was 0.08(0.1), 0.167(0.1), 0.273(0.207), 0.38(0.267), 0.487(0.56), 0.847(0.76), 0.833(0.867), 0.98(0.98), where the power for the t-test is given in parentheses. Again, the difference in power is largely not significant.

Next we consider the power results in the three sample case  $m = 3, q = 2$ , comparing the  $\chi_1$  with the F-test (F) and the Kruskal–Wallis (K–W) tests. In the normal case,  $\mathbf{x}_1 \sim N(\beta_1, 1)$ ,  $\mathbf{x}_2 \sim N(\beta_2, 1)$ ,  $\mathbf{x}_3 \sim N(0, 1)$ , in the lognormal case  $\mathbf{x}_1 \sim LN(\beta_1, 1)$ ,  $\mathbf{x}_2 \sim LN(\beta_2, 1)$ ,  $\mathbf{x}_3 \sim LN(0, 1)$ , and in the gamma case,  $\mathbf{x}_1 \sim \text{Gamma}(3+\beta_1, 1)$ ,  $\mathbf{x}_2 \sim \text{Gamma}(3+\beta_2, 1)$ ,  $\mathbf{x}_3 \sim \text{Gamma}(3, 1)$ . The hypothesis is now  $H_0 : \mu_1 = \mu_2 = \mu_3$  or  $H_0 : \beta_1 = \beta_2 = 0$ . The power results as a function of  $\beta_1, \beta_2$  are given in Table 3, where again each power entry in the table was obtained from 150 independent runs. Again, the  $\chi_1$ -test displays more power than the F-test in the lognormal and gamma cases, while it is not dominated by the F-test in the normal example. Evidently, the nonparametric test has less power than the  $\chi_1$  test in all the simulated cases.

## 4 Testing in Radar/Radiometer Data

This section uses space-time colocated independent radar and radiometer data of rain rate spatially averaged to a 12.5 km resolution, described in detail in Fokianos et al. (1998). The data consist of 500 radar, and 700 radiometer observations. For illustration purposes we consider both large and moderate sample sizes. In the first two cases  $h(x) = x$ , in the third  $h(x) = \log(x)$ . Although the testing conclusions using either  $h(x) = x$  or  $h(x) = \log(x)$  are the same, there is an indication, as expressed by more pronounced  $\chi_1$  and  $\chi_2$  values, that  $h(x) = \log(x)$  is more suitable, reflecting the fact the data are highly skewed; see Fokianos et al. (1998).

### 4.1 Large Sample Results

In this subsection we use  $h(x) = x$  throughout.

Let the radar data with  $n_1 = 500$  be the first sample. The radiometer data are now divided into two samples so that all together there are 3 samples,  $m = 3$ , where the two radiometer samples are from the same population by construction. The first sample from the radiometer data has  $n_2 = 400$  observations and the remaining  $n_3 = 300$  radiometer observations serve as the “reference sample” from the reference distribution. In essence, we can think of the data as coming from three different instruments, all measuring rain rate, that may or may not perform similarly. By construction, the two radiometers perform equally giving rise to data from the same distribution.

The resulting estimates are  $\hat{\alpha}_1 = 0.508$ ,  $\hat{\beta}_1 = -0.463$ ,  $\hat{\alpha}_2 = -0.113$ ,  $\hat{\beta}_2 = 0.066$ . The relatively small value of  $\hat{\beta}_2$  indicates that the reference sample and the second sample most likely come from the same distribution, which is the case by construction. The matrix  $\Sigma$  is estimated by

$$\Sigma = \begin{pmatrix} 5.580 & 2.914 & -4.852 & -1.791 \\ 2.914 & 5.325 & -1.836 & -3.130 \\ -4.852 & -1.836 & 4.665 & 1.112 \\ -1.791 & -3.130 & 1.112 & 1.849 \end{pmatrix}.$$

In the following, we discuss some hypotheses of interest. We first consider  $H_0 : \beta_1 = \beta_2 = 0$ . This hypothesis implies that all the instruments are alike. By using

$$\mathbf{H} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and  $\mathbf{c} = (0, 0)'$ ,  $\chi_2$  in (17) is equal to 78.94 with p-value 0 at 2 degrees of freedom. By using (16),  $\chi_1 = 31.72$  giving a very small p-value at 2 degrees of freedom. Both tests correctly reject, rather strongly, the hypothesis that all the populations are identical, that is, the instruments do not perform in the same manner.

Next we test  $H_0 : \beta_2 = 0$ , meaning that the second and third samples were drawn from the same distribution. Now we use  $\mathbf{H} = (0 \ 0 \ 0 \ 1)$  and get from (17) that the test statistic  $\chi_2$  is equal to 2.860. The p-value, using the chi-square distribution with one degree of freedom, is 0.091. Thus we do not reject the hypothesis, as should be the case since the second and third populations are the same by construction. The value of the test statistic  $\chi_1$  in (16), properly modified to account for a single parameter, is 0.139 with p-value equal to 0.709, consistent with the result from (17).

To test the hypothesis that the first and third populations are alike, consider  $H_0 : \beta_1 = 0$ . We use now  $\mathbf{H} = (0 \ 0 \ 1 \ 0)$ . The value of the test statistic (17) is  $\chi_2 = 67.88$  which rejects the hypothesis as it should giving a p-value close to 0, while from (16)  $\chi_1 = 8.953$  with p-value 0.002, consistent with the previous test.

Next we use the radar data for reference, that is  $n_3 = 500$ , partitioning the radiometer data into two independent samples where  $n_1 = 400$  and  $n_2 = 300$ . The maximum likelihood estimates are  $\hat{\alpha}_1 = -0.555$ ,  $\hat{\beta}_1 = 0.492$ ,  $\hat{\alpha}_2 = -0.595$ ,  $\hat{\beta}_2 = 0.514$ . Note that  $\hat{\beta}_1$  is close to  $\hat{\beta}_2$  since they correspond to identical populations. The matrix  $\Sigma$  is now estimated by

$$\Sigma = \begin{pmatrix} 4.817 & 2.677 & -4.289 & -3.038 \\ 2.677 & 5.688 & -3.034 & -4.721 \\ -4.289 & -3.034 & 4.280 & 3.550 \\ -3.038 & -4.721 & 3.550 & 4.497 \end{pmatrix}.$$

We again tested the same hypotheses as before obtaining the expected results in all cases. For example, the hypothesis  $H_0 : \beta_2 = 0$ , which implies that the second and third populations are identical, was correctly rejected. Another hypothesis of interest, in this setting, is  $H_0 : \beta_1 = \beta_2$ . By choosing  $\mathbf{H} = (0 \ 0 \ 1 \ -1)$ , the test statistic (17) is equal to 0.363 (p-value=0.546), with one degree of freedom. Therefore we do not reject the hypothesis. This should be the case since the first and second populations are identical.

## 4.2 Moderate Sample Results I

The exact same analysis was repeated with  $h(x) = x$  and  $m = 3$  using much smaller samples.

We first sampled  $n_1 = 25$  observations from the radar data. Next we obtained two samples of size  $n_2 = n_3 = 25$  from the radiometer data. Again, we first keep the second radiometer sample as the “reference sample” from the reference distribution.

The resulting estimates are  $\hat{\alpha}_1 = 1.227$ ,  $\hat{\beta}_1 = -1.254$ ,  $\hat{\alpha}_2 = 0.031$ ,  $\hat{\beta}_2 = -0.017$ .

$$\Sigma = \begin{pmatrix} 10.238 & 3.607 & -11.067 & -2.056 \\ 3.607 & 7.284 & -1.950 & -4.143 \\ -11.067 & -1.950 & 14.293 & 1.137 \\ -2.056 & -4.143 & 1.137 & 2.356 \end{pmatrix}.$$

Again, the relatively small value of  $\hat{\beta}_2$  indicates that the reference sample and the second sample most likely come from the same distribution, which is the case by construction.

Testing  $H_0 : \beta_1 = \beta_2 = 0$ , and with  $\mathbf{H}$  as above,  $\chi_2$  in (17), with 2 degrees of freedom, is equal to 8.479, giving a p-value of 0.014. Also with 2 degrees of freedom,  $\chi_1$  from equation (16) is equal to 66.794, giving a very small p-value. Again both tests correctly reject the hypothesis.

In testing  $H_0 : \beta_2 = 0$ ,  $\mathbf{H} = (0 \ 0 \ 0 \ 1)$ ,  $\chi_2 = 0.010$ . The p-value, using the chi-square distribution with one degree of freedom, is 0.919. Thus we do not reject the hypothesis, as should be the case since the second and third populations are the same by construction. The value of the test statistic  $\chi_1$  in (16), properly modified to account for a single parameter, is 0.281 with p-value equal to 0.596, consistent with the result from (17).

To test the hypothesis that the first and third populations are alike,  $H_0 : \beta_1 = 0$ ,  $\mathbf{H} = (0 \ 0 \ 1 \ 0)$ ,  $\chi_2 = 8.257$  with p-value 0.004, and  $\chi_1 = 13.719$  with p-value 0.0002, both rejecting correctly under a single degree of freedom.

With the radar data as the “reference” 3rd distribution, the estimates are  $\hat{\alpha}_1 = -1.408$ ,  $\hat{\beta}_1 = 1.659$ ,  $\hat{\alpha}_2 = -1.488$ ,  $\hat{\beta}_2 = 1.706$ ,

$$\Sigma = \begin{pmatrix} 10.502 & 6.799 & -12.977 & -10.833 \\ 6.799 & 10.674 & -10.639 & -12.867 \\ -12.97 & -10.639 & 20.025 & 18.729 \\ -10.83 & -12.767 & 18.729 & 10.957 \end{pmatrix}.$$

Again  $\hat{\beta}_1$  is not far from  $\hat{\beta}_2$  since they correspond to identical populations. We again tested the same hypotheses as before obtaining the expected results in all cases. For example, consider  $H_0 : \beta_1 = \beta_2$ . By choosing  $\mathbf{H} = (0 \ 0 \ 1 \ -1)$ , the test statistic (17) is equal to 0.064 (p-value=0.799), with one degree of freedom. Therefore we do not

reject the hypothesis. This should be the case since the first and second populations are identical.

### 4.3 Moderate Sample Results II

We repeat the analysis in the previous subsection with three samples of size  $n_1 = n_2 = n_3 = 25$ , but this time with  $h(x) = \log(x)$ .

With the first sample from the radar, and the other two from the radiometer, we let the second radiometer sample represent the reference distribution. The resulting estimates are  $\hat{\alpha}_1 = -0.709$ ,  $\hat{\beta}_1 = -1.336$ ,  $\hat{\alpha}_2 = 0.001$ ,  $\hat{\beta}_2 = -0.006$ ,

$$\Sigma = \begin{pmatrix} 12.067 & 3.663 & -9.374 & -1.906 \\ 3.660 & 7.328 & -1.888 & -3.816 \\ -9.374 & -1.888 & 10.192 & 0.986 \\ -1.906 & -3.816 & 0.986 & 1.988 \end{pmatrix}.$$

Again  $\hat{\beta}_2$  is relatively small as it should, coming from a radiometer sample.

In testing  $H_0 : \beta_1 = \beta_2 = 0$ ,  $\chi_2 = 13.749$  with p-value 0.001 at 2 degrees of freedom, and  $\chi_1 = 20.2$  with p-value 0.00004 also at 2 degrees of freedom, both rejecting correctly.

In testing  $H_0 : \beta_2 = 0$ ,  $\chi_2 = 0.001$ , giving a p-value of 0.968 at one degree of freedom. Also,  $\chi_1 = 0.0092$  with a p-value of 0.924, consistent with the result from (17). Thus both tests correctly do not reject the hypothesis.

For  $H_0 : \beta_1 = 0$ ,  $\chi_2 = 13.15$  giving a p-value of 0.0002, and  $\chi_1 = 15.22$  with p-value 0.0001, consistent with the previous test. Thus both tests reject correctly.

With the radar data as reference, the estimates are  $\hat{\alpha}_1 = 1.382$ ,  $\hat{\beta}_1 = 2.212$ ,  $\hat{\alpha}_2 = 1.415$ ,  $\hat{\beta}_2 = 2.108$ , and

$$\Sigma = \begin{pmatrix} 12.067 & 3.663 & -9.374 & -1.906 \\ 3.660 & 7.328 & -1.888 & -3.816 \\ -9.374 & -1.888 & 10.192 & 0.986 \\ -1.906 & -3.816 & 0.986 & 1.988 \end{pmatrix}.$$

Note that  $\hat{\beta}_1$  is close to  $\hat{\beta}_2$  since they correspond to identical populations. We again tested the same hypotheses as before obtaining the expected results in all cases. In particular, in testing  $H_0 : \beta_1 = \beta_2$ ,  $\chi_2 = 0.326$  with a p-value of 0.567 at one degree of freedom, echoing the fact that the first and second populations are identical.

## 5 Summary

We have outlined a method for testing the similarity of  $m$  populations given  $m$  independent random samples, where the first  $q = m - 1$  populations deviate from the  $m$ th one by an exponential distortion as in (3). The analysis was illustrated using radar/radiometer rain rate data. The same development goes through for deviations of a more general form including for example,

$$g_j(x) = \exp(\alpha_j + h(\beta_j, x))g(x), \quad j = 1, \dots, q$$

with  $h(0, x) = 0$ , or even ,

$$g_j(x) = \varphi(\alpha_j + h(\beta_j, x))g(x), \quad j = 1, \dots, q$$

These and related problems are investigated presently.

Finally, we would like to touch upon the following points and in doing so provide some additional useful references.

1. Under (3) with  $m = 2$  and  $h(x) = x$ , it can be shown the score based test reduces to a test in terms of the statistic  $\bar{x}_1 - \bar{x}_2$ . But this is precisely the test statistic obtained from the score based test derived under the normal assumption. This shows that tests based on  $\bar{x}_1 - \bar{x}_2$  are appropriate also for some non-normal distributions that give rise to the exponential tilt (3).
2. Again under (3) with  $h(x) \neq x$ , tests based on  $\bar{x}_1 - \bar{x}_2$ , such as the t-test, may not necessarily perform well relative to tests derived from the profile likelihood, such as the  $\chi_1$ -test, as our simulation results indicate.
3. Apparently, (3) provides a semiparametric alternative to Cox proportional hazards model and location shift models with the added advantage that the reference distribution is estimated.
4. Another advantage of the present approach is that it provides different tests of similarity or “sameness” of  $k$  distributions by using different distortion functions  $h(x)$ , an example of which we saw in the radar/radiometer example. This is a generalization of Neyman’s smooth goodness of fit test advocated by Rayner and Best (1989).
5. The choice/estimation of  $h(x)$  can be approached in several ways. One possibility is to approximate  $h(x)$  by a polynomial or by B-splines adopting some of the methods of Stone (1990). Another possibility is to employ some type of kernel estimation applied to the log-ratio of probability densities. In this regard, Tibshirani and Hastie (1987) and Fan and Gijbels (1996) used a local likelihood procedure

while Silverman (1978) considered log-ratio estimation using nonparametric penalized maximum likelihood estimation. For skewed geophysical data  $h(x) = \log(x)$  may be helpful.

6. Similarly to Neyman's goodness of fit applied in testing for distribution equality, a wrong choice of  $h(x)$  will most likely lead to a power loss in testing and to a bias in estimation problems. Yet, as demonstrated in the real data example, our procedure can still lead to sensible conclusions in testing.

7. In principle, it does not matter which distribution is taken as the reference distribution  $g(x)$  as the difference in the beta-values remains constant; the alphas depend on the betas and  $g(x)$ . However, choosing the distribution that gives the most reliable data as reference is sensible.

8. The fact that the profile likelihood for the finite dimensional parameter in a semi-parametric setting behaves as a parametric likelihood has been studied theoretically in Murphy and van der Vaart (1997).

## Appendix

We prove the asserted asymptotic normality (12). First define

$$\nabla \equiv \left( \frac{\partial}{\partial \alpha_1}, \dots, \frac{\partial}{\partial \alpha_q}, \frac{\partial}{\partial \beta_1}, \dots, \frac{\partial}{\partial \beta_q} \right)'$$

Then it is easy to see  $E[\nabla l(\boldsymbol{\alpha}, \boldsymbol{\beta})] = \mathbf{0}$ . To obtain the score second moments it is convenient to define  $\rho_m \equiv 1$ ,  $w_m(t) \equiv 1$ ,

$$E_j(t) \equiv \int h(t)w_j(t)dG(t), \quad \text{Var}_j(t) \equiv \int h^2(t)w_j(t)dG(t) - E_j^2(t)$$

and,

$$\begin{aligned} A_0(j, j') &\equiv \int \frac{w_j(t)w_{j'}(t)dG(t)}{1 + \sum_{k=1}^q \rho_k w_k(t)} \\ A_1(j, j') &\equiv \int \frac{h(t)w_j(t)w_{j'}(t)dG(t)}{1 + \sum_{k=1}^q \rho_k w_k(t)} \\ A_2(j, j') &\equiv \int \frac{h^2(t)w_j(t)w_{j'}(t)dG(t)}{1 + \sum_{k=1}^q \rho_k w_k(t)} \end{aligned}$$

for  $j, j' = 1, \dots, q$ . Then, the entries in

$$\mathbf{V} \equiv Var \left[ \frac{1}{\sqrt{n}} \nabla l(\boldsymbol{\alpha}, \boldsymbol{\beta}) \right]$$

are,

$$\begin{aligned} \frac{1}{n} Var \left( \frac{\partial l}{\partial \alpha_j} \right) &= \frac{\rho_j^2}{1 + \sum_{k=1}^q \rho_k} [A_0(j, j) - \sum_{r=1}^m \rho_r A_0^2(j, r)] \\ \frac{1}{n} Cov \left( \frac{\partial l}{\partial \alpha_j}, \frac{\partial l}{\partial \alpha_{j'}} \right) &= \frac{\rho_j \rho_{j'}}{1 + \sum_{k=1}^q \rho_k} [A_0(j, j') - \sum_{r=1}^m \rho_r A_0(j, r) A_0(j', r)] \\ \frac{1}{n} Cov \left( \frac{\partial l}{\partial \alpha_j}, \frac{\partial l}{\partial \beta_j} \right) &= \frac{\rho_j^2}{1 + \sum_{k=1}^q \rho_k} [A_0(j, j) E_j(t) - \sum_{r=1}^m \rho_r A_0(j, r) A_1(j, r)] \\ \frac{1}{n} Cov \left( \frac{\partial l}{\partial \alpha_j}, \frac{\partial l}{\partial \beta_{j'}} \right) &= \frac{\rho_j \rho_{j'}}{1 + \sum_{k=1}^q \rho_k} [A_0(j, j') E_{j'}(t) - \sum_{r=1}^m \rho_r A_0(j, r) A_1(j', r)] \\ \frac{1}{n} Var \left( \frac{\partial l}{\partial \beta_j} \right) &= \frac{\rho_j^2}{1 + \sum_{k=1}^q \rho_k} [-A_2(j, j) + 2A_1(j, j) E_j(t) - \sum_{r=1}^m \rho_r A_1^2(j, r)] \\ &\quad + \frac{\rho_j}{1 + \sum_{k=1}^q \rho_k} Var_j(t) \\ \frac{1}{n} Cov \left( \frac{\partial l}{\partial \beta_j}, \frac{\partial l}{\partial \beta_{j'}} \right) &= \frac{\rho_j \rho_{j'}}{1 + \sum_{k=1}^q \rho_k} [-A_2(j, j') + A_1(j, j') (E_j(t) + E_{j'}(t)) \\ &\quad - \sum_{r=1}^m \rho_r A_1(j, r) A_1(j', r)] \end{aligned}$$

Next, as  $n \rightarrow \infty$ ,

$$-\frac{1}{n} \nabla \nabla' l(\boldsymbol{\alpha}, \boldsymbol{\beta}) \rightarrow \mathbf{S}$$

where  $\mathbf{S}$  is a  $2q \times 2q$  matrix with entries corresponding to  $j, j' = 1, \dots, q$ ,

$$\begin{aligned} -\frac{1}{n} \frac{\partial^2 l}{\partial \alpha_j^2} &\rightarrow \frac{\rho_j}{1 + \sum_{k=1}^q \rho_k} \int \frac{[1 + \sum_{k \neq j}^q \rho_k w_k(t)] w_j(t)}{1 + \sum_{k=1}^q \rho_k w_k(t)} dG(t) \\ -\frac{1}{n} \frac{\partial^2 l}{\partial \alpha_j \partial \alpha_{j'}} &\rightarrow \frac{-\rho_j \rho_{j'}}{1 + \sum_{k=1}^q \rho_k} \int \frac{w_j(t) w_{j'}(t)}{1 + \sum_{k=1}^q \rho_k w_k(t)} dG(t) \\ -\frac{1}{n} \frac{\partial^2 l}{\partial \alpha_j \partial \beta_j} &\rightarrow \frac{\rho_j}{1 + \sum_{k=1}^q \rho_k} \int \frac{[1 + \sum_{k \neq j}^q \rho_k w_k(t)] h(t) w_j(t)}{1 + \sum_{k=1}^q \rho_k w_k(t)} dG(t) \end{aligned}$$

$$\begin{aligned}
-\frac{1}{n} \frac{\partial^2 l}{\partial \alpha_j \partial \beta_{j'}} &\rightarrow \frac{-\rho_j \rho_{j'}}{1 + \sum_{k=1}^q \rho_k} \int \frac{h(t) w_j(t) w_{j'}(t)}{1 + \sum_{k=1}^q \rho_k w_k(t)} dG(t) \\
-\frac{1}{n} \frac{\partial^2 l}{\partial \beta_j^2} &\rightarrow \frac{\rho_j}{1 + \sum_{k=1}^q \rho_k} \int \frac{[1 + \sum_{k \neq j}^q \rho_k w_k(t)] h^2(t) w_j(t)}{1 + \sum_{k=1}^q \rho_k w_k(t)} dG(t) \\
-\frac{1}{n} \frac{\partial^2 l}{\partial \beta_j \partial \beta_{j'}} &\rightarrow \frac{-\rho_j \rho_{j'}}{1 + \sum_{k=1}^q \rho_k} \int \frac{h^2(t) w_j(t) w_{j'}(t)}{1 + \sum_{k=1}^q \rho_k w_k(t)} dG(t)
\end{aligned}$$

The entries are obtained by a repeated application of the facts  $\int dG(t) = 1$  and  $\int w_j(t) dG(t) = 1$ ,  $j = 1, \dots, q$ . It should be noted that due to profiling, the matrix  $\mathbf{S}$  is not the usual information matrix although it plays a similar role.

Thus, when (3) holds with true parameters  $\alpha_0, \beta_0$ , it follows under regularity conditions  $\hat{\alpha}, \hat{\beta}$  are both consistent and asymptotically normal (See Sen and Singer 1993, Ch. 5),

$$\sqrt{n} \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} \Rightarrow N(\mathbf{0}, \Sigma) \tag{18}$$

where  $\Sigma = \mathbf{S}^{-1} \mathbf{V} \mathbf{S}^{-1}$ .

## Acknowledgments

We would like to thank the Editor, Associate Editor, and two reviewers for their thoughtful and constructive remarks. We are grateful to Jeff Haferman for his effort in putting together the rain rate data. The work of the first two authors was supported by NASA contract NAG52783.

## References

1. Anderson, J. A. (1972), "Separate sample logistic discrimination", *Biometrika*, **59**, 19-35.
2. Anderson, J. A. (1982), "Logistic Discrimination." In P. R. Krishnaiah and L. N. Kanal, eds., *Handbook of Statistics*, Vol 2, pp. 169-191, Amsterdam: North-Holland.
3. Cox, D.R., (1966), "Some procedures associated with the logistic qualitative response curve". In: F.N. David, ed., *Research Papers in Statistics: Festschrift for J. Neyman*, pp. 55-71, New York: John Wiley.

4. Efron, B. and Tibshirani, R. (1996) "Using specially designed exponential families for density estimation", *The Annals of Statistics*, **24**, 2431-2461.
5. Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and its Applications*, London: Chapman & Hall.
6. Fokianos, K., B. Kedem, J. Qin, J. Haferman, and D. A. Short (1998), "On Combining Instruments." *J. Appl. Meteor.*, **37**, 220-226.
7. Kay, R., and S. Little (1987), " Transformations of the explanatory variables in the logistic regression model for binary data. *Biometrika*, **74**, 495-501.
8. Miller, R. G. (1986), *Beyond ANOVA, Basics Of Applied Statistics*, New York: John Wiley.
9. Murphy, S. A. and van der Vaart, A. W. (1997), "Semiparametric likelihood ratio inference," *The Annals of Statistics*, **25**, 1471-1509.
10. Neyman, J. (1937), " 'Smooth' tests for goodness of fit," *Skandinavisk aktuarietidskrift*, **20**, 149-199.
11. Prentice, R. L. and R. Pyke (1979), "Logistic disease incidence models and case-control studies," *Biometrika*, **66**, 403-411.
12. Qin, J., and J.F. Lawless (1994), "Empirical likelihood and general estimating equations," *The Annals of Statistics*, **22**, 300-325.
13. Qin, J., and B. Zhang (1997), "A goodness of fit test for logistic regression models based on case-control data." in *Biometrika*, **84**, 609-618.
14. Randles, R. H. and Wolfe, D. A. (1979), *Introduction to the Theory of Non-parametrics Statistics*, New York: John Wiley.
15. Rayner, J. C. W. and Best, D. J. (1989), *Smooth Tests Of Goodness Of Fit*, Oxford: Oxford University Press.
16. Sen, P. K. and J. M. Singer (1993), *Large Sample Methods in Statistics, An Introduction With Applications*, London: Chapman & Hall.
17. Silverman, B. W. (1978), "Density ratios, empirical likelihood and cot death," *Applied Statistics*, **27**, 26-33.

18. Stone, C. J. (1990), "Large-sample inference for log-spline models," *The Annals of Statistics*, **18**, 717-741.
19. Tibshirani, R. and Hastie, T. (1987), "Local likelihood estimation," *Journal of the American Statistical Association*, **82**, 559-567.

Table 1: Parameter estimation where the respective reference distributions are  $U(0, 1)$  and  $N(0, 1)$ . No distortion occurs for  $\beta_1 = \beta_2 = 0$ .

Population	Sample Sizes			Parameters				Estimated Parameters			
	$n_1$	$n_2$	$n_3$	$\alpha_1$	$\alpha_2$	$\beta_1$	$\beta_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
$g(x) \sim U(0, 1)$ $g_1(x) \sim U(0, 1)$ $g_2(x) \sim U(0, 1)$	200	200	200	0	0	0	0	0.0063 (0.1824)	-0.0006 (0.1812)	-0.0127 (0.3569)	0.0017 (0.3627)
$g(x) \sim U(0, 1)$ $g_1(x) \sim U(0, 1)$ $g_2(x) \sim U(0, 1)$	200	300	100	0	0	0	0	-0.0026 (0.2068)	-0.0110 (0.1995)	0.0099 (0.4161)	0.0280 (0.4006)
$g(x) \sim N(0, 1)$ $g_1(x) \sim N(2, 1)$ $g_2(x) \sim N(3, 1)$	200	200	200	-4.5	-2	3	2	-4.5527 (0.3811)	-2.0327 (0.2169)	3.0454 (0.2739)	2.0370 (0.1994)
$g(x) \sim N(0, 1)$ $g_1(x) \sim N(2, 1)$ $g_2(x) \sim N(3, 1)$	200	300	100	-4.5	-2	3	2	-4.5541 (0.3851)	-2.0356 (0.2162)	3.0427 (0.2584)	2.0360 (0.2173)

Table 2: Power comparison with the t-test as a function of  $\beta_1$ .  $m = 2$ , nominal level=0.05,  $n_1 = n_2 = 30$ . The reference distributions are N(0,1), LN(0,1), Gamma(3,1), respectively.

$\beta_1$	Normal			Lognormal			Gamma		
	$\chi_1$	t-Test	W-test	$\chi_1$	t-Test	W-test	$\chi_1$	t-Test	W-test
0.1	0.093	0.087	0.086	0.147	0.027	0.066	0.100	0.047	0.046
0.2	0.140	0.133	0.133	0.173	0.093	0.100	0.127	0.067	0.060
0.3	0.247	0.240	0.200	0.247	0.140	0.226	0.140	0.087	0.120
0.4	0.387	0.367	0.326	0.360	0.287	0.346	0.240	0.153	0.160
0.5	0.473	0.447	0.466	0.506	0.347	0.493	0.253	0.160	0.213
0.7	0.793	0.787	0.700	0.760	0.587	0.706	0.493	0.380	0.393
0.8	0.840	0.813	0.806	1.000	0.687	0.813	0.513	0.340	0.480
1.0	0.987	0.987	0.966	1.000	0.860	0.953	0.693	0.527	0.520

Table 3: Power comparison with the F-test as a function of  $\beta_1, \beta_2$ .  $m = 3$ . The reference distributions are N(0,1), LN(0,1), Gamma(3,1), respectively.

$\beta_1$	$\beta_2$	Sample Size	Level	Normal			Lognormal			Gamma		
				$\chi_1$	F	K-W	$\chi_1$	F	K-W	$\chi_1$	F	K-W
0.2	0.2	n1=n2=n3=15	0.05	0.087	0.060	0.026	0.067	0.047	0.046	0.093	0.060	0.053
0.1	0.4			0.253	0.193	0.106	0.180	0.127	0.173	0.167	0.087	0.060
0.2	0.5			0.313	0.260	0.140	0.247	0.113	0.200	0.233	0.093	0.100
0.5	0.5			0.300	0.260	0.240	0.293	0.107	0.220	0.246	0.093	0.146
0.7	0.5			0.420	0.360	0.346	0.393	0.200	0.333	0.287	0.140	0.113
0.2	0.2	n1=n2=n3=30	0.01	0.053	0.047	0.046	0.073	0.027	0.066	0.047	0.000	0.020
0.1	0.4			0.207	0.147	0.080	0.153	0.073	0.106	0.067	0.020	0.060
0.2	0.5			0.207	0.180	0.160	0.240	0.093	0.140	0.127	0.033	0.026
0.5	0.5			0.307	0.273	0.300	0.287	0.067	0.246	0.113	0.060	0.053
0.7	0.5			0.447	0.380	0.373	0.413	0.160	0.400	0.227	0.120	0.106