Scientific
Research
Publishing

Pseudodistance Methods Using Simultaneously Sample Observations and Nearest Neighbour Distance Observations for Continuous Multivariate Models

Andrew Luong

École d'actuariat, Université Laval, Ste Foy, Québec, Canada
Email: Andrew.Luong@act.ulaval.ca

Abstract

Using the fact that a multivariate random sample of n observations also generates n nearest neighbour distance (NND) univariate observations and from these NND observations, a set of n auxiliary observations can be obtained and with these auxiliary observations when combined with the original multivariate observations of the random sample, a class of pseudodistance D_h is allowed to be used and inference methods can be developed using this class of pseudodistances. The D_h estimators obtained from this class can achieve high efficiencies and have robustness properties. Model testing also can be handled in a unified way by means of goodness-of-fit tests statistics derived from this class which have an asymptotic normal distribution. These properties make the developed inference methods relatively simple to implement and appear to be suitable for analyzing multivariate data which are often encountered in applications.

Keywords

Goodness-of-Fit Statistics, Robust Estimators, Multivariate Density Estimate, Information Matrix, Model Testing

1. Introduction

For statistical inferences methods for continuous multivariate models, we often assume to have a random sample of size n of multivariate observations $x_1, \cdots, x_n$ which are independent and identically distributed as the d-dimensional vector of random variable x with a d-dimensional density function $g(x)$.

For the parametric set-up $g(x) \in \{f_\theta\}, \theta = (\theta_1, \cdots, \theta_m)'$ and let the vector θ_0 denote the true vector of parameters, we would like to have statistical methods for estimating the vector θ_0 if the parametric model $\{f_\theta\}$ can be assumed and inference methods to validate the assumption of the model $\{f_\theta\}$ by means of various goodness-of-fit statistics. This leads to a composite null hypothesis and ideally we would like to use goodness-of-fit test statistics which follow a unique asymptotic distribution $\theta \in \Omega$, Ω assumed to be compact.

The multidimensional model testing often poses difficulties as often goodness-of-fit test statistics used either have very complicated distributions such as the case of statistics which make use of multivariate empirical characteristic functions, see Csörgö [1] or for the classical chi-square tests where the asymptotic distributions for simple and composite hypotheses are simple but observations must be grouped into cells and there is some arbitrariness on choosing cells, see Moore [2], Klugman *et al.* [3] (pages 208-209) on extending chi-square tests for continuous multidimensional models. Goodness of fit test statistics using multivariate sample distribution function often has a very complicated null distribution, see Babu and Rao [4] and extensive simulations are needed to obtain the p values of the tests. For applications in various fields, it appears that there is a need for developing goodness of fit tests statistics which are relatively simple to implement with the property of the tests based on such statistics which are consistent.

Multivariate modelling is used in many fields which include actuarial sciences and finance. For financial applications, Moore [2] used the chi-square tests for testing whether the joint weekly returns of two assets follow a bivariate normal distribution but as mentioned earlier, for chi-square tests we need to partition the sample space into cells and the tests are not consistent despite the asymptotic null distributions of the statistics are simple.

In this paper, we shall introduce a class of pseudodistance $D_h(g, f)$ constructed based on a class of convex functions $h(x)$ which measures the discrepancy between the two density functions g and f, see details in Section 2.3. Goodness-of-fit test statistics for model testing based on D_h will preserve the property of having a simple asymptotic null distribution comparable to chi-square tests but unlike chi-square tests, the tests based on D_h are consistent for model testing.

It is also interesting to note that within this class D_h, the statistic based on D_h with $h(x) = -\log(x)$ can also accommodate parameters being estimated using maximum likelihood (ML) method for composite hypothesis. On the estimation side, estimators based on D_h will have the potential of having good efficiencies and robustness properties. Furthermore, estimation and model testing can be handled in a unified way.

The inference methods proposed extends previous methods for the univariate models to multivariate continuous models. This paper can be considered as a follow up of previous papers by Luong [5], Luong [6]. The neighbour distance (NND) notion is used in this paper to replace a similar notion of distance which

f was used when considering spacings and order statistics, see Ranneby *et al.* [7]. Order statistics are used for defining spacings for univariate models.

The class of pseudo distances $D_h(g, f)$ is constructed using the following class of strictly convex functions $h(x)$ with its second derivative $h''(x) > 0$. For each chosen $h(x)$ we then have a corresponding pseudodistance $D_h(g, f)$ and $D_h(g, f)$ is a discrepancy measure between density g and density f.

Explicitly, the function $h(x)$ takes the form

$$h(x) = -x^\alpha, \tag{1}$$

α is a known constant with $0 < \alpha < 1$ and in practice we choose α near 0, and $h(x)$ can also have the form

$$h(x) = -\log(x). \tag{2}$$

Note that $-\dfrac{x^\alpha - 1}{\alpha} \to -\log(x)$ as α decreases to 0 and $h(x)$ only needs to be defined up to an additive and a positive multiplicative constant and provided that these constants are known inference procedures based on D_h with $h(x)$ using α decreasing to 0 have efficiencies very close to inference procedures using D_h with $h(x) = -\log(x)$.

Furthermore, if $h(x) = -\log(x)$ is used to construct the pseudo distance D_h, estimation using this pseudo distance will give the maximum likelihood estimators. This D_h as a pseudodistance is up to a few terms which does not depend on θ the Kullback-Leibler (KL) distance used to generate ML estimators. These few terms without involving θ do not affect estimation but they are very significant for construction of goodness-of-fit test statistics as goodness of fit test statistics constructed using D_h will have an asymptotic normal distribution for model testing meanwhile goodness-of-fit test statistics using the KL pseudodistance do not have a simple asymptotic distribution especially for the composite null hypothesis case where parameters must be estimated using the ML estimators. We shall give more discussions in Section 2.2.

The paper is organized as follows. In Section 2, we introduce the auxiliary observations obtained from the NND observations. The class of pseudodistances D_h is also introduced in this section. Asymptotic properties of estimators based on D_h are considered in Section 3. Estimators obtained using D_h with $h(x) = -\log(x)$ are identical to ML estimators which are fully efficient. If other $h(x)$ is used for D_h, the corresponding estimators have the potential of good efficiencies and some robustness properties. These properties allow flexibility for balancing efficiency and robustness. In Section 4, goodness-of-fit statistics based on the class D_h are shown to have an asymptotic normal distribution and in Section 5 an example is provided for illustration of the proposed techniques.

2. Pseudo Distances

2.1. Nearest Neighbour Distance (NND) Observations

For each vector of observation $x_i, i = 1, \cdots, n$ in the random sample, we define

r_i the nearest neighbour distance (NND) to x_i with

$$r_i = \min_{x_j \neq x_i} \left\| x_j - x_i \right\|,$$

$\left\| \ \right\|$ is the commonly used Euclidean distance
and r_i clearly can be obtained using the sample of d-variate observations.

In the literature, these r_i's have been used to construct goodness of fit statistics, see Bickel and Breiman [8], Zhou and Jammalamadaka [9] but often these statistics for multivariate models do not have a simple asymptotic distribution which might create difficulties for applications. Now, we can define y_i as given by proposition 2 by Ranneby *et al.* [7] (page 433) with $y_i = n c_d r_i$ or equivalent-

ly $\dfrac{y_i}{n} = c_d r_i, d \geq 2$, $c_d = \dfrac{\pi^{\frac{d}{2}}}{\Gamma\left(\dfrac{d}{2}+1\right)}$, π is the usual constant pi used in

formulas to find volume or area and $\Gamma(.)$ is the commonly used gamma function.

Note that we have $y_1, \cdots, y_n$ which are n univariate auxiliary observations obtained from NND observations. Therefore, from the original observations of the sample $x_1, \cdots, x_n$ and using the n auxiliary observations, we can form the following $d+1$ multivariate observations

$$\left(x_1, y_1 \right), \cdots, \left(x_n, y_n \right).$$

These n observations for $n \to \infty$ are asymptotically independent and have a common density function given by the density of $\left(X, Y \right)$ below,

$$p_0 \left(x, y \right) = g^2 \left(x \right) \mathrm{e}^{-g(x)}, \tag{3}$$

see the end of Section 2 given by Kuljus and Ranneby [10], (p1094). In fact, the situation is similar to the univariate case where spacings were used, see Luong [5] (pages 619-620).

Now we can consider the random criterion function

$$Q_n^h \left(\boldsymbol{\theta} \right) = \frac{1}{n} \sum_{i=1}^{n} h \left(y_i f_{\boldsymbol{\theta}} \left(x_i \right) \right) \tag{4}$$

for the class of function h defined by expressions (1) and (2), we shall see subsequently that inference methods based on the objective functions (4) are pseudodistance methods based on a class of pseudodistance $D_h \left(g, f \right)$ where g and f are density functions.

Minimizing $Q_n^h \left(\boldsymbol{\theta} \right)$ with $h \left(x \right) = -\log \left(x \right)$, we obtain the pseudodistance D_h estimators which are identical to maximum likelihood (ML) estimators, ML estimators can be viewed as pseudodistance estimators based on the Kullback-Leibler (KL) distance but we shall see goodness-of-fit tests statistics are complicated with the use of the KL distance unlike the ones which are based on $Q_n^h \left(\boldsymbol{\theta} \right)$ and consequently based on $D_h \left(g, f_{\boldsymbol{\theta}} \right)$. The KL pseudodistance used to derive ML estimators will be discussed in Section 2.2 and the class of pseudodistances will be introduced in Section 2.2.

Furthermore, if we use $h(x) = -x^\alpha$ to construct D_h then α should be set near 0 but within the range $0 < \alpha < 1$ for robust estimation without relying on *a*, explicit multivariate density estimate which is needed for the minimum Hellinger method as proposed by Tamura and Boos [11]. Therefore, it appears that the class of pseudodistance methods being considered are very useful for applications and they are relatively simple to implement so that practitioners might want to use them for applied works.

2.2. Kullback-Leibler (KL) Pseudo-Distance

The negative of the log likelihood function can be expressed as

$$Q_n^{ML}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} -\log f_{\boldsymbol{\theta}}(x_i)$$

and ML estimators can be viewed as the values obtained by minimizing the observed version which can also be called sample version of the Kullback-Leibler (KL) pseudo-distance (D_{KL}), *i.e.*, $D_{KL}^o(g, f_{\boldsymbol{\theta}})$ defined as

$$D_{KL}^o(g, f_{\boldsymbol{\theta}}) = Q_n^{ML}(\boldsymbol{\theta}) + \frac{1}{n}\sum_{i=1}^{n} \log(g(x_i)) \xrightarrow{p} D_{KL}(g, f_{\boldsymbol{\theta}}), \qquad (5)$$

$\xrightarrow{p}$ denotes convergence in probability,

and minimizing $D_{KL}^o(g, f_{\boldsymbol{\theta}})$ is equivalent to maximize the log of the likelihood function.

The KL pseudo-distance is defined as $D_{KL}(g, f_{\boldsymbol{\theta}}) = -E_g\left(\log\left(\dfrac{f_{\boldsymbol{\theta}}}{g}\right)\right)$, is the KL pseudo-distance.

Howewer, for testing the validity of the model with the null composite hypothesis given by

$$H_0 : g(\boldsymbol{x}) \in \{f_{\boldsymbol{\theta}}\}, \ \boldsymbol{\theta} \in \Omega$$

and since $g(\boldsymbol{x})$ appears in the LHS of expression (5), it must be estimated and replacing $g(\boldsymbol{x})$ say by a multivariate density estimate $\hat{g}(\boldsymbol{x})$ will make the distribution of the LHS of expression (5) complicated despite that we can replace $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}_{ML}$. This might explain the limited use of the KL pseudo-distance for construction of statistics for model testing with the use of $\hat{\boldsymbol{\theta}}_{ML}$.

2.3. The Class of Pseudo-Distances D_h

We shall focus on pseudo-distance methods based on $D_h(g, f_{\boldsymbol{\theta}})$ for parametric model with emphasis on continuous multivariate models but some of the previous univariate results which are scattered can also be unified by viewing them as pseudo distance methods.

In general for pseudodistances we require the following property:

$$D_h(g, f) > 0 \ \text{ if } \ g \neq f, \ D_h(g, f) = 0 \ \text{ if and only if } \ g = f, \qquad (6)$$

g and *f* are density functions. The property given by (6) are needed for estab-

lishing consistency of estimation and for consitency goodness of fit tests, see Broniatowski *et al.* [12] for more notions and properties of pseudodistances.

Since $D_h(g,f)$ in general is not observable if g is unknown, we shall see at the end of this section that can define an observed version $D_h^0(g,f)$ with the property $D_h^0(g,f) \xrightarrow{p} D_h(g,f)$. $D_h^0(g,f)$ will satisfy the property given by expression (6) in probability which similar to $D_{KL}^o(g,f_\theta)$ with the property $D_{KL}^o(g,f_\theta) \xrightarrow{p} D_{KL}(g,f_\theta)$ for the KL distance.

Now we work with following pair of observations to develop D_h methods,

$$(X_1, Y_1)', \cdots, (X_n, Y_n)'.$$

For this sample, the observations are asymptotically independent using Propostion 3 by Ranneby *et al.* [7] (p413) and as $n \to \infty$, the distribution of (X_i, Y_i) tends to a common distribution, *i.e.*, the common distribution is the distribution of the random vector (X, Y) with joint density function given by

$$p(x, y) = g^2(x) e^{-yg(x)},$$

see Kuljus and Ranneby [10] (p1101).

Therefore, the results are very similar to the univariate case with the interpretation $g(x)$ being a multivariate density here instead of a univariate density, the results given by Luong [5] (page 624) continue to hold and we also have:

1) $Z = Yg(X)$ follows a standard exponential distribution with density $f(z) = e^{-z}, z > 0$.

2) Z and X are independent.

If we use Jensen's inequality it follows that

$$E\big(h(yf(x))\big) = E_Z\left(E\left(h(z)\frac{f(x)}{g(x)}\right)\middle| Z\right) > E_Z\left(h(z)E\left(\frac{f(x)}{g(x)}\middle| Z\right)\right) = E_z\big(h(z)\big)$$

for $f \neq g$, since $E\left(\frac{f(x)}{g(x)}\middle| Z\right) = E\left(\frac{f(x)}{g(x)}\right) = 1$

and $E\big(h(yf(x))\big) = E_z\big(h(z)\big)$ for $f = g$.

Now, we can define

$$D_h(g, f) = E\big(h(yf(x))\big) - E_z\big(h(z)\big)$$

and $E_z\big(h(z)\big)$ is a known constant which does not depend on the parameters given by the vector θ.

Under the parametric model, if we consider to minimize $D_h(g, f_\theta)$ but g is unknown, it leads to consider the observed objective function $D_h^o(g, f_\theta)$ defined below which is based on $D_h(g, f_\theta)$ with

$$D_h^o(g, f_\theta) = \frac{1}{n}\sum_{i=1}^{n}\big(h(y_i f_\theta(x_i))\big) - E_z\big(h(z)\big),$$

note that we have

$$D_h^o(g, f_\theta) \xrightarrow{p} D_h(g, f_\theta).$$

The pseudodistance D_h estimators given by the vector $\hat{\boldsymbol{\theta}}$ based on $D_h(g, f_{\boldsymbol{\theta}})$ is obtained by minimizing $D_h^o(g, f_{\boldsymbol{\theta}})$. Equivalently, it is obtained by minimizing

$$Q_n^h(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} h\left(y_i f_{\boldsymbol{\theta}}(x_i)\right)$$

which is expression (4).

3. Asymptotic Properties of D_h estimators

3.1. Consistency

It is not difficult to see that the D_h estimators given by the vector $\hat{\boldsymbol{\theta}}$ which minimizes expression (4) is consistent by defining $h(z_i(\boldsymbol{\theta}, n)) = h(y_i f_{\boldsymbol{\theta}}(x_i))$ and by using assumptions and results of Section 3.1 as given by Luong [5] (pages 622-624).

The limit laws like uniform weak law of large numbers UWLLN and Central limit Theorem (CLT) are applicable by using the property of $\{z_i(\boldsymbol{\theta}, n)\}$ being a mixing sequence which is due to $(X_i, Y_i), i = 1, \cdots, n$ are asymptotically independent with a common distribution as $n \to \infty$. Therefore, $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$.

3.2. Asymptotic Normality

Using CLT and results given by Section 2 in Luong [5] (pages 626-631), we can conclude that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{L} N\left(\sigma_h^2 \left[\boldsymbol{I}(\boldsymbol{\theta}_0)\right]^{-1}\right), \quad \xrightarrow{L} \text{ denotes convergence in law,}$$

$\boldsymbol{I}(\boldsymbol{\theta}_0) = -E_{\boldsymbol{\theta}_0}\left(\dfrac{\partial^2 \ln f(x; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}' \partial \boldsymbol{\theta}}\right)$ is the commonly used information matrix with

$\sigma_h^2 = 1$, if the function $h(x) = -\log(x)$ is used to define D_h and D_h^0 and

$\sigma_h^2 = \dfrac{E_Z\left[(h'(Z))^2 Z^2\right]}{\left[E_Z\left(Z^2 h''(Z)\right)\right]^2}$ if the function $h(x) = -x^\alpha, 0 < \alpha < 1$ is used to define

D_h and D_h^0 with the first and second derivatives of h denoted respectively by h' and h''. The random variable Z follows a standard exponential distribution as given by expression 25 in Luong [5] (page 631) and from the standard exponential distribution, we also have $E_Z(Z^k) = \Gamma(1+k), k > -1$ and note that $\sigma_h^2 \to 1$ as $\alpha \to 0$.

The D_h estimators using $h(x) = -x^\alpha, 0 < \alpha < 1$, might have some robustness property using M-estimation theory, see Luong [5] (page 632) and might be preferred over the ML estimators.

From the fact that the proposed D_h methods are density based but without requiring an explicit density estimate to implement hence they appear to be simpler for practitioners and can be used as alternative to other robust methods such as the Hellinger methods as proposed by Tamura and Boos [11]. Besides, the observed pseudodistances D_h^o based on D_h can also be used for construc-

tion of goodness-of-fit statistics and lead to statistics which are relatively simple to implemement.

4. Goodness-of-Fit Tests Statistics Using D_h^0

For model selection and model testing we are primary interested on testing the null composite hypothesis

$$H_0 : g(x) \in \{f_{\boldsymbol{\theta}}\}, \ \boldsymbol{\theta} \in \Omega.$$

Howewer it might be easier to follow the procedure to construct test statistics by first consider the test based on D_h^0 which is also implicitly based on D_h for the simple hypothesis first where there is no unkown parameter.

4.1. Simple Null Hypothesis

For simple $H_0 : g = f$,

A natural statistics to use can be based on
$D_h^0(g, f) = \frac{1}{n} \sum_{i=1}^{n} \left(h(y_i f(x_i)) - E_z(h(z)) \right)$ and since $\left\{ h(y_i f(x_i)), i = 1, 2, \cdots \right\}$ forms a mixing sequence, CLT can be applied with the distribution of each $h(y_i f(x_i))$ tends to a standard exponential random variable and Slutsky's Theorem can also be used if needed. Therefore, the following test statistic $\sqrt{n} D_h^0(g, f)$ can be used and $\sqrt{n} D_h^0(g, f) \xrightarrow{L} N\left(0, v_Z^h\right)$ where v_Z^h is the variance of $h(Z)$ where Z follows a standard exponential distribution.

For an α level test, we can reject H_0 if

$$\frac{\sqrt{n} D_h^0(g, f)}{\sqrt{v_Z^h}} > z_{1-\alpha}$$

where $z_{1-\alpha}$ is the $(1-\alpha)$ the percentile of the standard normal distribution.

Equivalently, we can reject H_0 if

$$-\frac{\sqrt{n} D_h^0(g, f)}{\sqrt{v_Z^h}} < z_\alpha. \tag{7}$$

Note that $-D_h^0(g, f) = \frac{1}{n} \sum_{i=1}^{n} \left(-h(y_i f(x_i)) - E_z(-h(z)) \right)$ and v_Z^h is also the variance of $-h(Z)$.

Now if we use D_h^0 with $h(x) = -\log(x)$, $E_z(-h(z)) = E_z(\log(Z))$ and using the moment generating function of $\log Z$ which is given by $M_{\log Z}(t) = E_Z\left(e^{t \log z}\right) = \Gamma(1+t)$ with $\Gamma(.)$ being the gamma function so that the cumulant generating function $\log M_{\log Z}(t) = \log \Gamma(1+t)$ and by differentiating it, we can obtain the first two cumulants which are given by $E_z(\log(Z)) = \psi(1)$, $v_Z^h = \psi'(1)$, $\psi(.)$ and $\psi'(.)$ are respectively the digamma function and the trigamma function and they are available in most of the statistical packages.

The test statistic given by expression (7) can be expressed explicilty as

$$\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}y_{i}+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\log f\left(x_{i}\right)-\sqrt{n}\psi\left(1\right)}{\sqrt{\psi'\left(1\right)}} \tag{8}$$

and reject the simple H_{0}

if $\quad\dfrac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}y_{i}+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\log f\left(x_{i}\right)-\sqrt{n}\psi\left(1\right)}{\sqrt{\psi'\left(1\right)}}<z_{\alpha}\quad$ for an α level test,

$0<\alpha<1$.

Note that the test is consistent as $\sqrt{n}D_{h}^{0}\left(g,f\right)\to\infty$ as $n\to\infty$ if $g\neq f$, so we will reject H_{0} with probability 1 should $g\neq f$ but this property is not shared by chi-square tests. Also, there is also the difficulty of arbitrariness of grouping observations into cells for chi-square tests, see Bickel and Breiman [8] for more discussions.

Furthermore, if we use D_{h}^{0} with $h\left(x\right)=-x^{\alpha},0<\alpha<1$, $E_{Z}\left(Z^{\alpha}\right)=\Gamma\left(1+\alpha\right)$, $E_{Z}\left(Z^{2\alpha}\right)=\Gamma\left(1+2\alpha\right)$ and $v_{Z}^{h}=\Gamma\left(1+2\alpha\right)-\left[\Gamma\left(1+\alpha\right)\right]^{2}$.

The corresponding test statistic given by expression (7) can be expressed explicitly as

$$\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(y_{i}f\left(x_{i}\right)\right)^{\alpha}-\sqrt{n}\Gamma\left(1+\alpha\right)}{\sqrt{\Gamma\left(1+2\alpha\right)-\left[\Gamma\left(1+\alpha\right)\right]^{2}}} \tag{9}$$

and reject the simple H_{0} if

$$\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(y_{i}f\left(x_{i}\right)\right)^{\alpha}-\sqrt{n}\Gamma\left(1+\alpha\right)}{\sqrt{\Gamma\left(1+2\alpha\right)-\left[\Gamma\left(1+\alpha\right)\right]^{2}}}<z_{\alpha}.$$

4.2. Composite Null Hypothesis

For model testing, we consider the composite $H_{0}:g\left(x\right)\in\left\{f_{\theta}\right\},\theta\in\Omega$, since θ_{0} is unknown, first we estimate θ_{0} by $\hat{\theta}$ which minimizes $D_{h}^{0}\left(g,f_{\theta}\right)$, then we can form the following statistic, $\sqrt{n}D_{h}^{0}\left(g,f_{\hat{\theta}}\right)$ and we shall show that $\sqrt{n}D_{h}^{0}\left(g,f_{\hat{\theta}}\right)\overset{L}{\longrightarrow}N\left(0,v_{Z}^{h}\right)$ which is similar to the statistic for the simple H_{0} and unlike other statistics when parameters are estimated lead to complicated null distribution, we shall show that the statistics behave like the one used for simple H_{0} in Section 4.1 and the equivalent rejection rules are similar to the ones given by expression (8) and expression (9) depending on the choice of $h\left(x\right)$ used for D_{h}^{0},D_{h}.

In fact, these expressions remain valid for the composite H_{0} provided that we replace $f\left(x_{i}\right)$ by $f_{\hat{\theta}}\left(x_{i}\right),i=1,\cdots,n$ when they appear in these expressions.

As we have seen by using a version of CLT for mixing sequences if needed, $\sqrt{n}D_{h}^{0}\left(g,f_{\theta_{0}}\right)\overset{L}{\longrightarrow}N\left(0,v_{Z}^{h}\right)$, now if we can establish

$$\sqrt{n}D_h^0\left(g,f_{\hat{\boldsymbol{\theta}}}\right)=\sqrt{n}D_h^0\left(g,f_{\boldsymbol{\theta}_0}\right)+o_p\left(1\right) \tag{10}$$

with $o_p\left(1\right)$ being a term which converges to 0 in probability and by Slutzky theorem,

$$\sqrt{n}D_h^0\left(g,f_{\hat{\boldsymbol{\theta}}}\right)\xrightarrow{L}N\left(0,v_Z^h\right).$$

Now, we will proceed to establish the property by expression (10). Using the Mean Value Theorem and the following expansion around $\hat{\boldsymbol{\theta}}$, we have

$$\sqrt{n}D_h^0\left(g,f_{\boldsymbol{\theta}_0}\right)=\sqrt{n}D_h^0\left(g,f_{\hat{\boldsymbol{\theta}}}\right)+\frac{1}{2}\sqrt{n}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'\frac{\partial^2 D_h^0\left(g,f_{\bar{\boldsymbol{\theta}}}\right)}{\partial\boldsymbol{\theta}'\partial\boldsymbol{\theta}}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)$$

with $\bar{\boldsymbol{\theta}}$ lies on the line segment joining $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$.

Since $\dfrac{\partial D_h^0\left(g,f_{\hat{\boldsymbol{\theta}}}\right)}{\partial\boldsymbol{\theta}}=0$ and using $\dfrac{\partial^2 D_h^0\left(g,f_{\bar{\boldsymbol{\theta}}}\right)}{\partial\boldsymbol{\theta}'\partial\boldsymbol{\theta}}\xrightarrow{p}E_Z\left(Z^2h''\left(Z\right)\right)I\left(\theta_0\right)$,

see Luong [5] (page 630). Also, since $I\left(\theta_0\right)$ is positive definite, $E_Z\left(Z^2h''\left(Z\right)\right)>0$ and $\hat{\boldsymbol{\theta}}\xrightarrow{p}\boldsymbol{\theta}_0$, we then have the relation as given by expression (10).

Furthermore, if we use $h\left(x\right)=-\log\left(x\right)$ for D_h, $\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}_{ML}$,

$$\sqrt{n}D_h^0\left(g,f_{\hat{\boldsymbol{\theta}}_{ML}}\right)=\sqrt{n}D_h^0\left(g,f_{\hat{\boldsymbol{\theta}}}\right)\xrightarrow{L}N\left(0,v_Z^h\right).$$

The use of the ML estimators $\hat{\boldsymbol{\theta}}_{ML}$ for chi-square distance type statistics often create complications when comes to derive the asymptotic distributions of these statistics, see Chernoff and Lehmann [13] (p580), Luong and Thompson [14] (p249-251).

For applications, it has been recognized that the maximum value attained by the log of the likelihood function can provide information on goodness-of-fit for the model being used, the test as given by expression (8) with $f\left(x_i\right)$ replaced by $f_{\hat{\boldsymbol{\theta}}}\left(x_i\right)$ formalizes the informal procedures on the use of the maximum value of the log likelihood function for assessing goodness-of-fit of the model, see Klugman *et al.* [15] but note that the condition of no tied observation is needed for the use of test based on the log of likelihood function as given by expression (8) with $f\left(x_i\right)$ replaced by $f_{\hat{\boldsymbol{\theta}}}\left(x_i\right)$, otherwise there are some values of $y_i=0$ and the log of these values are undefined meanwhile test based on D_h^o with $h\left(x\right)=-x^\alpha,0<\alpha<1$ is well defined even with the presence of tied obsevations and in general we should fix a value for α near 0 for balancing efficiency and robustness for the estimation procedures.

5. Illustration

For illustration of the proposed methods, we use the multivariate normal model with d dimension; its density function is often parameterized using the mean $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ and it is given by

$$f\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}\right)=\left(2\pi\right)^{-\frac{1}{2}d}\left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})},\quad \boldsymbol{x}\in R^d. \tag{11}$$

$\left|\boldsymbol{\Sigma}\right|$ is the determinant of the matrix $\boldsymbol{\Sigma}$, see Anderson [16] (page 20).

There is redundancy when using elements of the matrix $\boldsymbol{\Sigma}$ as parameters as $\boldsymbol{\Sigma}$ being a covariance matrix; it is symmetric.

We can eliminate the redundancy by defining the vector of parameters as $\boldsymbol{\theta}$ with

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\mu} \\ Vech\boldsymbol{\Sigma} \end{pmatrix}.$$

The Vech operator when applied to $\boldsymbol{\Sigma}$ extracts the lower triangular elements of $\boldsymbol{\Sigma}$ and stacks them in a vector. Equivalently, we can use the vector of parameters $\boldsymbol{\theta}$ instead of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and express the multivariate normal density as $f(\boldsymbol{x};\boldsymbol{\theta})$ to avoid redundancy of the previous parameterization. We assume that we have a random sample of size n which allows us to obtain the auxiliary univariate observations $y_1,\cdots,y_n$ from NND observations and there is no tied observation so that

$$y_i > 0, i = 1,\cdots,n$$

For illustration say we use D_h^0 with $h(x) = -\log x$, the vector of estimators in this case coincides with maximum likelihood (ML) estimators, *i.e.*, $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{ML}$ but for multivariate normal model, it is well known $\hat{\boldsymbol{\theta}}_{ML}$ can be obtained explicitly, see Anderson [16] (page 112).

Explicitly,

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{ML}, \quad \hat{\boldsymbol{\theta}}_{ML} = \begin{pmatrix} \overline{x} \\ Vech\boldsymbol{S} \end{pmatrix}, \quad \overline{x} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i ,$$

$\overline{x}$ is the sample mean and $\boldsymbol{S}$ is the sample covariance matrix which can also be expressed as

$$\boldsymbol{S} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{x})(\boldsymbol{x}_i - \overline{x})' .$$

For model testing then we can use the test statistic

$$\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} y_i + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \log f_{\hat{\boldsymbol{\theta}}_{ML}}(\boldsymbol{x}_i) - \sqrt{n}\psi(1)}{\sqrt{\psi'(1)}}$$

and reject the model if the statistics gives a value smaller than z_α for an α level test.

As Tamura and Boos [11] have pointed out that, $\hat{\boldsymbol{\theta}}_{ML}$ might not be robust and hence proposed multivariate Hellinger density estimators but a multivariate density estimate is needed for their procedures. For robust estimation or in case of having tied observation we might want to use D_h^0 with $h(x) = -x^\alpha$ with α being a positive number but near 0.

In this paper, we focus on presentations of methodologies of D_h, leaving simulation studies for assessing power of the tests, the use of other distributions than the normal distribution for the null distribution of goodness-of-fit tests statistics and assessing efficiency when sample sizes are small or in finite samples for subsequent works. Practitioners might be encouraged to use these D_h me-

thods.

Acknowledgements

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Csörgö, S.L. (1986) Testing for Normality in Arbitrary Dimension. *The Annals of Statistics*, **14**, 708-723. https://doi.org/10.1214/aos/1176349948

[2] Moore, D.S. and Stubblebine, J.B. (1981) Chi-Square Tests for Multivariate Normality with Applications to Common Stocks Prices. *Communications in Statistics, Theory and Methods*, **10**, 713-738. https://doi.org/10.1080/03610928108828070

[3] Klugman, S.A., Panjer, H.H. and Willmott, G.E. (2013) Loss Models: Further Topics. Wiley, New York. https://doi.org/10.1002/9781118787106

[4] Babu, G.J. and Rao, C.R. (2004) Goodness-of-Fit Tests When Parameters Are Estimated. *Sankhya*, **66**, 63-74.

[5] Luong, A. (2018) Unified Asymptotic Results for Maximum Spacing and Generalized Spacing Methods for Continuous Models. *Open Journal of Statistics*, **8**, 614-639. https://doi.org/10.4236/ojs.2018.83040

[6] Luong, A. (2018) Asymptotic Results for Goodness-of-Fit Tests Using a Class of Generalized Spacing Methods. *Open Journal of Statistics*, **8**, 731-746. https://doi.org/10.4236/ojs.2018.84048

[7] Ranneby, B., Jammalamadaka, S.R. and Tetterukovskiy, A (2005) The Maximum Spacing Estimation for Multivariate Observations. *Journal of Statistical Planning Inference*, **129**, 427-446. https://doi.org/10.1016/j.jspi.2004.06.059

[8] Bickel, P.J. and Breiman, L. (1983) Sum of Functions of Nearest Neighbor Distances, Moment Bounds, Limit Theorem and a Goodness of Fit Test. *The Annals of Probability*, **11**, 185-114. https://doi.org/10.1214/aop/1176993668

[9] Zhou, S. and Jammalamadakala, S.R. (1993) Goodness of Fit in Multi-Dimensions Based on Nearest Neighbour Distances. *Nonparametric Statistics*, **2**, 271-284. https://doi.org/10.1080/10485259308832558

[10] Kuljus, K. and Ranneby, B. (2015) Generalized Maximum Spacing Estimation for Multivariate Observations. *Scandinavian Journal of Statistics*, **42**, 1092-1108. https://doi.org/10.1111/sjos.12153

[11] Tamura, R.N. and Boos, D.D. (1986) Minimum Hellinger Distance Estimation for Multivariate Location and Covariance. *Journal of the American Statistical Association*, **81**, 223-229. https://doi.org/10.1080/01621459.1986.10478264

[12] Broniatowski, M., Toma, A. and Vajda, I. (2012) Decomposable Pseudodistances and Applications in Statistical Estimation. *Journal of Statistical Planning and Inference*, **142**, 2574-2585. https://doi.org/10.1016/j.jspi.2012.03.019

[13] Chernoff, H. and Lehmann, E.L. (1954) The Use of Maximum Estimates in

Chi-Square Tests for Goodness of Fit. *Annals of Mathematical Statistics*, **25**, 579-586. https://doi.org/10.1214/aoms/1177728726

[14] Luong, A. and Thompson, M.E. (1987) Minimum Distance Methods Based on Quadratic Distances for Transforms. *Canadian Journal of Statistics*, **15**, 239-251. https://doi.org/10.2307/3314914

[15] Klugman, S.A., Panjer, H.H. and Willmott, G.E. (2012) Loss Models: From Data to Decision. Wiley, New York. https://doi.org/10.1002/9781118787106

[16] Anderson, T.W. (2003) An Introduction to the Multivariate Statistical Analysis. 3rd Edition, Wiley, New York.