Estimation and Testing of Elliptically Contoured Distribution

Jianqiang Wang

Department of Statistics

Iowa State University

August 29, 2007

Contents

1 Introduction

This paper deals with the estimaion and testing of elliptically contoured distributions. This topic has been heavily explored in literature. Anderson (2003) talks about estimating mean vector and covariance matrix assuming the distribution of $R^2 \overset{d}{=} (\boldsymbol{X}_i - \boldsymbol{\mu})'\Lambda^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu})$ is known and Liebscher (2005) talks about semiparametric density estimation which is very similar to our method here. In testing spherically or elliptially contoured distributions, Serfling (2006) gives a detailed review of existing papers with the concept of " symmetry".

Section 2 of this paper gives an overview of the method of estimating a multivariate density function, assume it is elliptically contoured distribution, using semiparametric approach. Then in Section 3, we talk about how to test spherical symmetry, with a ready generalization to testing elliptical symmetry. We present our simulation results in Section 4 with some concluding remarks given in Section 5.

2 Semiparemetric Estimation of Elliptically Contoured Densities

Let $\boldsymbol{X}_1, ..., \boldsymbol{X}_n \overset{iid}{\sim} EC_p(\boldsymbol{\mu}, \Lambda, \phi)$, where we assume Λ is a full rank matrix, and $rank(\Lambda) = p$. It is well known that $\boldsymbol{X}_i$ has a representation $\boldsymbol{X}_i \overset{d}{=} \boldsymbol{\mu} + RA'U^{(p)}$, where $\boldsymbol{\mu} = E\boldsymbol{X}_i$, $A'A = \Lambda$ and $U^{(p)}$ is uniformly distributed on unit hypersphere in $\Re^p$. The non-negative random variable R has the same distribution as $\sqrt{(\boldsymbol{X}_i - \boldsymbol{\mu})'\Lambda^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu})}$.

Assume that

$$\boldsymbol{X} \sim |\Sigma|^{-1/2} f\left((\boldsymbol{x} - \boldsymbol{\mu})'\Lambda^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right), \tag{1}$$

then we can establish that $R^2 \overset{d}{=} (\boldsymbol{X}_i - \boldsymbol{\mu})'\Lambda^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu})$ has density

$$g(r) = \frac{\pi^{p/2}}{\Gamma(p/2)} r^{\frac{p}{2}-1} f(r), \tag{2}$$

which gives

$$f(r) = \frac{\Gamma(p/2)}{\pi^{p/2}} r^{1-\frac{p}{2}} g(r). \tag{3}$$

In order to solve the identifiability problem of parameters R and A, we assume the parameters are such that $\Lambda = \Sigma$, which is the variance-covariance matrix of $\boldsymbol{X}_i$. So $\boldsymbol{\mu}$ and

Σ can be estimated by the usual method of moment estimators,

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{X}}, \hat{\Sigma} = S_{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'. \tag{4}$$

We define $\boldsymbol{Z}_i = \Sigma^{-1/2}(\boldsymbol{X}_i - \boldsymbol{\mu})$, which is estimated by

$$\hat{\boldsymbol{Z}}_i = S_{n-1}^{-1/2}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}) \tag{5}$$

. We can obtain that

$$\hat{\boldsymbol{Z}}_i = \boldsymbol{Z}_i + O_p\left(\frac{1}{\sqrt{n}}\right), \tag{6}$$

as $\bar{\boldsymbol{X}} = \boldsymbol{\mu} + O_p\left(\frac{1}{\sqrt{n}}\right)$ and $S_{n-1} = \Sigma + O_p\left(\frac{1}{\sqrt{n}}\right)$. Further, we define

$$\hat{R}_i^2 \equiv \hat{\boldsymbol{Z}}_i'\hat{\boldsymbol{Z}}_i = (\boldsymbol{X}_i - \bar{\boldsymbol{X}})'\hat{\Sigma}^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}) = R_i^2 + O_p\left(\frac{1}{\sqrt{n}}\right) \tag{7}$$

We can use a kernel estimator on $\hat{R}_i^2$ to estimate function $g(\cdot)$,

$$\hat{g}(r) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{r - \hat{R}_i^2}{h}\right), \tag{8}$$

where h is the bandwidth and $K(\cdot)$ is the kernel function we are going to use. The selection of bandwidth h has been heavily explored in statistical literature, and in our simulation, we use the "reference method" and use a normal distribution as reference.

After estimating $g(\cdot)$, we can directly plug it into (3), and obtain the following density estimator of $\boldsymbol{X}$,

$$\left|\hat{\Sigma}\right|^{-1/2}\frac{\Gamma(p/2)}{\pi^{p/2}}\left((\boldsymbol{x} - \bar{\boldsymbol{X}})'\hat{\Sigma}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{X}})\right)^{1-\frac{p}{2}}\hat{g}\left((\boldsymbol{x} - \bar{\boldsymbol{X}})'\hat{\Sigma}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{X}})\right). \tag{9}$$

Some problems with this "naive" estimator have been pointed out by Liebscher (2005) and he also suggested a couple of remedies in the same paper.

3 Testing Spherical Symmetry

Our approach of testing spherical symmetry would be based on characteristic functions (CF). We calculate a restriction-free estimate of characteristic function and estimate CF under spherical symmetry, then the discrepancy between these two estimates of characteristic function can be used to construct a test of spherical symmetry. To test elliptical symmetry, we could use pseudo data $\hat{\boldsymbol{Z}}_i$ as defined in (5) and test the spherical symmetry of $\hat{\boldsymbol{Z}}_i$.

Now, let us assume $\boldsymbol{X}_1, ..., \boldsymbol{X}_n \sim EC_p(\mathbf{0}, \mathrm{I}_p, \phi)$. We denote the characteristic function of $\boldsymbol{X}$ as $\phi_{\boldsymbol{X}}(\boldsymbol{t})$, and from Fang and Zhang (1990) we get

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \int_0^\infty \Omega_p(\|\boldsymbol{t}\|^2 r^2) g(r^2) 2r dr, \tag{10}$$

where $\Omega_p(\cdot)$ is defined in Theorem 2.5.1 of Fang and Zhang (1990) and $g(\cdot)$ is defined by (2).

Given an IID sample $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$, we can use an empirical estimator to estimate $\phi_{\boldsymbol{X}}(\boldsymbol{t})$. The empirical characteristic function (ECF) is defined as

$$\hat{\phi}_{\boldsymbol{X}}(\boldsymbol{t}) = \frac{1}{n} \sum_{i=1}^n e^{i\boldsymbol{t}'\boldsymbol{X}_i}, \tag{11}$$

and a "better" estimator of $\phi_{\boldsymbol{X}}(\boldsymbol{t})$ under spherical symmetry would be to start from (10) and replace the distribution function by an empirical distribution function,

$$\hat{\phi}_F(\boldsymbol{t}) = \frac{1}{n} \sum_{i=1}^n \Omega_p(\|\boldsymbol{t}\|^2 R_i^2), \tag{12}$$

where R_i^2 is defined by $R_i^2 = \boldsymbol{X}_i'\boldsymbol{X}_i$.

In (12), function $\Omega_p(\cdot)$ is a known real function, but the ECF defined in (11) may have a non-zero imaginary part.

The statistic we use in testing spherical symmetry would be a discrepancy measure between ECF and ECF under H_0,

$$D\left(\hat{\phi}_{\boldsymbol{X}}, \hat{\phi}_F\right) = \int \left\{\hat{\phi}_{\boldsymbol{X}}(\boldsymbol{t}) - \hat{\phi}_F(\boldsymbol{t})\right\}^2 w(\boldsymbol{t}) d\boldsymbol{t}, \tag{13}$$

where $w(\boldsymbol{t})$ is a weight function, and the integral is over $\Re^p$.

The distribution of $D\left(\hat{\phi}_{\boldsymbol{X}}, \hat{\phi}_F\right)$ under H_0 is need to construct a test of spherical symmetry. We need some derivation to find out the theoretical properties of $D\left(\hat{\phi}_{\boldsymbol{X}}, \hat{\phi}_F\right)$, and an alternative way to obtain this distribution is to use the Bootstrap method.

Let $\boldsymbol{X}^{(k)} = (\boldsymbol{X}_1^{(k)}, \boldsymbol{X}_2^{(k)}, ..., \boldsymbol{X}_n^{(k)})$ denote the k^{th} Bootstrap resample, then the ECF based on $\boldsymbol{X}^{(k)}$ is defined as,

$$\hat{\phi}_{\boldsymbol{X}}^{(k)}(\boldsymbol{t}) = \frac{1}{n} \sum_{i=1}^n e^{i\boldsymbol{t}'\boldsymbol{X}_i^{(k)}},$$

and the ECF under H_0 is defined as,

$$\hat{\phi}_F^{(k)}(\boldsymbol{t}) = \frac{1}{n} \sum_{i=1}^n \Omega_p(\|\boldsymbol{t}\|^2 {R_i^{(k)}}^2).$$

Then the k^{th} replicate of discrepancy measure is defined as,

$$D^{(k)}\left(\hat{\phi}_{\boldsymbol{X}}^{(k)}, \hat{\phi}_F^{(k)}\right) = \int \left\{\hat{\phi}_{\boldsymbol{X}}^{(k)}(\boldsymbol{t}) - \hat{\phi}_F^{(k)}(\boldsymbol{t})\right\}^2 w(\boldsymbol{t})d\boldsymbol{t},$$

so we can use the distribution of $D^{(k)}\left(\hat{\phi}_{\boldsymbol{X}}^{(k)}, \hat{\phi}_F^{(k)}\right)$ to approximate the distribution of $D\left(\hat{\phi}_{\boldsymbol{X}}, \hat{\phi}_F\right)$ and test the spherical symmetry of an underlying distribution.

4 Simulation Results

4.1 Density estimation

In density estimation, we compare the semiparametric estimation (SE) method with a nonparametric kernel estimation (NP) in terms of bias, variance, MSE and Mean Integrated Squared Error (MISE). In the fully nonparametric estimation, we use product of epanechnikov kernels.

We draw samples of size $n = 400$ from each of the two bivariate normal distributions,

- $$\boldsymbol{X} \sim \begin{bmatrix} -1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix} N_2\left(\mathbf{0}, \mathrm{I}_2\right), \tag{14}$$

- $$\boldsymbol{X} \sim \begin{bmatrix} -1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} N_2\left(\mathbf{0}, \mathrm{I}_2\right), \tag{15}$$

calculate both estimators on a wide range of grid points and use contour plots and summary measures to compare the two.

Figure 1 shows the contour plot of normal densities 14 and 15 with scatterred points, the plot on the left has more skiny contours and the one on the right has more spherical contours.

Figure 2 shows the kernel estimated density of $\hat{R}^2$ and the true density ignoring the estimation of mean vector and covriance matrix under 4 different bandwidths $h = 0.2, 0.5, 0.8, 1.2$. We can see that a bandwidth between 0.8 and 1.2 works well for distribution (14). In our simulation, we dynamicly choose bandwidths for both semiparametric estimation and bivariate density estimation. A histogram of resulting bandwidths is shown in Figure 3. From

	SE	NP
Model 1	0.1727	0.5590
Model 2	0.1356	0.6538

Table 1: MISE for semiparametric density estimation and nonparametric density estimation under two different models

Figure 3, we can see that the bandwidths chosen by reference method are very stable and we do not have extreme bandwidths.

Then we compare semiparametric estimation with nonparametric estimation in terms of absolute bias, variance, MSE and MISE. Figures 4 and 5 show the difference of bias, variance and MSE (SE-NP) under two distributions. And Table 4.1 shows the comparison of MISE (actually, it is the summation of all MSE's on grid points, which is proportional to MISE). We can see that semiparametric approach is outperforming the usual bivariate denstiy estimation in having much smaller MISEs.

4.2 Testing spherical symmetry

4.2.1 Distribution of discrepancy measure

In Section 3, we defined a discrepancy measure between an empirical characteristic function (ECF) $\hat{\phi}_{\boldsymbol{X}}(\boldsymbol{t})$ and an empirical characteristic function under H_0 $\hat{\phi}_F(\boldsymbol{t})$ in (13), and it can be used to test the asymmetry of underlying distribution.

As has been pointed out by Epps and Pulley (1983), we assign high weight where $|\phi_1(\boldsymbol{t}) - \phi_0(\boldsymbol{t})|$ is large and where the mean squared error of $\hat{\phi}_{\boldsymbol{X}}(\boldsymbol{t})$ is small. In the simulation study, we will use $w(\boldsymbol{t}) \equiv 1$ for simplicity.

We are interested in the "true" distribution of $D\left(\hat{\phi}_{\boldsymbol{X}}, \hat{\phi}_F\right)$ under different spherically contoured distributions, or in other words, the distribution of $D\left(\hat{\phi}_{\boldsymbol{X}}, \hat{\phi}_F\right)$ under H_0. We also want to get an idea of the MSE of $\hat{\phi}_{\boldsymbol{X}}(\boldsymbol{t})$ and $\hat{\phi}_F(\boldsymbol{t})$, respectively. The definition of the MSE's are both respect to the true characteristic function,

$$MSE(\hat{\phi}_{\boldsymbol{X}}(\boldsymbol{t})) = E\left(\hat{\phi}_{\boldsymbol{X}}(\boldsymbol{t}) - \phi_{\boldsymbol{X}}(\boldsymbol{t})\right)^2,$$

and

$$MSE(\hat{\phi}_F(\boldsymbol{t})) = E\left(\hat{\phi}_F(\boldsymbol{t}) - \phi_{\boldsymbol{X}}(\boldsymbol{t})\right)^2.$$

We generate IID samples of size 100 from the following SC distributions,

1. Bivirate normal distribution with mean vector $\mathbf{0}$ and variance matrix I_2, or $g(r) = \frac{1}{2}e^{(-r/2)}$.

2. Uniform distribution on a unit disk in $\Re^2$, or $g(r) = \mathrm{I}_{(r\leq 1)}$.

3. Bivariate distribution where R^2 has density $g(r) = \frac{r}{2}\mathrm{I}_{(r\leq 1)}$.

4. Bivariate distribution where the density of R^2 is a mixture of two normals left censored at 0.

Then we calculate $\hat{\phi}_{\boldsymbol{X}}(\boldsymbol{t}), \hat{\phi}_F(\boldsymbol{t})$ and $D\left(\hat{\phi}_{\boldsymbol{X}}, \hat{\phi}_F\right)$ for each sample, and approximate MSE as well as the null distribution of $D\left(\hat{\phi}_{\boldsymbol{X}}, \hat{\phi}_F\right)$ from Monte Carlo simulations.

Figure 6 shows the scatterplot of samples of size 400 drawn from each of the four distributions described above. The density of the "triangular" distribution is diminishing as $\boldsymbol{X}$ approaches $\mathbf{0}$, but increase as $\|\boldsymbol{X}\|$ increases until it hits the boundary of support. The distribution of R^2 has two modes and the similar is observed on the scatterplot of " bimodal" distribution.

Figure 7 shows the contour plot of $MSE(\hat{\phi}_{\boldsymbol{X}}(\boldsymbol{t}))$ and $MSE(\hat{\phi}_F(\boldsymbol{t}))$. We can see that the contour of $MSE(\hat{\phi}_F(\boldsymbol{t}))$ has the shape of concentrical circles, but the contour of $MSE(\hat{\phi}_{\boldsymbol{X}}(\boldsymbol{t}))$ is concentrical circle around the origin, but does not follow this pattern on the peripheral. The MSE of ECF is generally greater than that of ECF under H_0, indicating that we can improve our estimated characteristic function by using the information on spherical symmetry. Another thing to point out is that for both contour plots, the MSE is small for $\boldsymbol{t}$ close to the origin but increases when $\|\boldsymbol{t}\|$ increases.

We want to get an idea of the distribution of $D\left(\hat{\phi}_{\boldsymbol{X}}, \hat{\phi}_F\right)$, so we repeatedly draw samples, calculate $D\left(\hat{\phi}_{\boldsymbol{X}}, \hat{\phi}_F\right)$ under each realized sample and plotted them on Figure 9. We can see from the histograms that the distribution of $D\left(\hat{\phi}_{\boldsymbol{X}}, \hat{\phi}_F\right)$ are unimodal with a heavier right tail in all four different configurations (the right skewness in Bimodal case is not that significant, though). So in deriving the theory of test, we can think about using normal, Gamma or χ^2 distribution to approximate the distribution of the discrepancy measure.

We are also interested in how well the distribution of Bootstrapped discrepancy measures can approximate the true distribution. Figure 10 illustrates the distribution of $D\left(\hat{\phi}_{\boldsymbol{X}}, \hat{\phi}_F\right)$ through repeated sampling and Bootstrapping from a finite sample. The two histograms resembles a lot in terms of shape and magnitude and similar pattern is seen for other finite samples as well. The resemblance validates the use of Bootstrap in our situation.

4.2.2 α-level error rate and power

In calculating the α-level error rate, we use a bivariate normal distribution with mean $\mathbf{0}$ and identity variance-covariance matrix. The actual type I error rate is 0.045 from 400 simulations with a nominal level of 0.05.

We have many different alternative non-spherically symmetric distributions to choose from, like shifting or rotating a spherically contoured distribution, or a distribution that is not elliptically contoured at all. The distribution we use is $RU^{(p)}$, where R is the square root of a χ_2^2 random variable, $U^{(p)} = (cos\theta, sin\theta)$ with $f(\theta) = \frac{1}{4\pi} + \frac{1}{2\pi^2}\theta, \theta \in [0, 2\pi)$, and Figure 11 shows the scatterplot of two samples of sizes 100 and 2000 from this distribution. At a sample size of 100, our test have a power of 0.27 approximated from 200 simulations with $\alpha = 0.05$.

5 Concluding Remarks

In this paper, we derived methods of estimating elliptically contoured densities and compared its performance with usual multivariate kernel density estimator. The simulation results favor the semiparametric estimation approach as it has a much smaller Mean Integrated Square Error. But its performance near the center still has room to be improved, compared to the multivariate kernel density estimator.

Concerning testing spherical symmetry, we examined the distribution of the discrepancy measure $D^{(k)}\left(\hat{\phi}^{(k)}_{\boldsymbol{X}}, \hat{\phi}^{(k)}_{F}\right)$ under several different null distributions, and illustrated the use of Bootstrap approach to approximate this distribution. A couple of simulations have been conducted to find out the type I error rate and power under specific examples, but we still need more simulations as well as theoretical justifications for this method.

References

Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis.* John Wiley & Sons.

Figure 1: Contour plot of true density of two normal distributions in subsection 4.1 with scatterred points.

Epps, T. W. and L. B. Pulley (1983). A test for normality based on the empirical characteristic function. *Biometrika 70*, 723–726.

Fang, K. and Y. Zhang (1990). *Generalized Multivariate Analysis.* Science Press and Springer-Verlag.

Liebscher, E. (2005). A semiparametric density estimator based on elliptical distributions. *Journal of Multivariate Analysis 92*(1), 205–225.

Serfling, R. J. (2006). Multivariate symmetry and asymmetry. *In Encyclopedia of Statistical Sciences, Second Edition*, 5338–5345.

Figure 2: Kernel density estimates for $\hat{R}^2$ as defined in (7), where the distribution of $\boldsymbol{X}$ is given in (14).

Figure 3: Histogram of selected bandwidth using reference method; the data are simulated from 14.

Figure 4: Surface plot of the difference in absolute bias and variance between semiparametric estimation and nonparametric estimation for distributions (14) and (15); the left column are from distribution drawn from (14), and the right from (15).

Figure 5: Surface plot of the difference in MSE between semiparametric estimation and non-parametric estimation for distributions (14) and (15); the left column are from distribution drawn from (14), and the right from (15).

Figure 6: Scatterplot of different null distributions used in subsection 4.2.1 of testing spherical symmetry.

Figure 7: Contour plot of $MSE(\hat{\phi}_{\boldsymbol{X}}(\boldsymbol{t}))$ and $MSE(\hat{\phi}_F(\boldsymbol{t}))$, where $\boldsymbol{X} \sim N(\mathbf{0}, \mathrm{I}_2)$

Figure 8: Contour plot of $MSE(\hat{\phi}_{\boldsymbol{X}}(\boldsymbol{t}))$ and $MSE(\hat{\phi}_F(\boldsymbol{t}))$, where $\boldsymbol{X}$ is uniformly distributed on a unit disk.

Figure 9: Histogram of $D\left(\hat{\phi}_{\boldsymbol{X}}, \hat{\phi}_F\right)$; the plot on the upper left has underlying distribution being normal, the one on the upper right is from uniform distribution, the one on the lower left is from the triangular distribution and last one from the bimodal distribution.

Figure 10: Histogram of $D\left(\hat{\phi}_{\boldsymbol{X}}, \hat{\phi}_F\right)$; the plot on the left is the simulated distribution of discrepancy measure and the plot on the right is Bootstrapped from a finite sample of size 100.

Figure 11: Scatterplot of an alternative distribution in subsection 4.2.2 under sample sizes 100 and 2000.