THE MULTIVARIATE FERMI-DIRAC DISTRIBUTION AND ITS STATISTICAL APPLICATIONS

by

Mauro Gasparini and Peiming Ma
Purdue University and University of Wisconsin-Stout

Technical Report #96-3

Department of Statistics
Purdue University
West Lafayette, IN

February 1996

o

The multivariate Fermi-Dirac distribution and its statistical applications

Mauro Gasparini[1] and Peiming Ma
Purdue University and *University of Wisconsin-Stout*

Abstract

We investigate the multivariate elliptically contoured generalization of a parametric family of univariate distributions proposed by Ferreri in 1964. Such a $p-$variate Fermi-Dirac distribution has density

$$f(\mathbf{x}) = \frac{\Gamma(p/2)\lambda^p(\alpha)}{\pi^{p/2}F(p/2-1,\alpha)|\mathbf{\Sigma}|^{1/2}} \frac{1}{1+\exp\{\alpha+\lambda^2(\alpha)(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\}}$$

where $\mathbf{x}, \boldsymbol{\mu} \in R^p$, $\alpha \in R$, $\mathbf{\Sigma}$ is a $p \times p$ positive definite matrix of rank p and

$$F(p,\alpha) := \int_0^\infty \frac{u^p}{1+\exp\{\alpha+u\}} du$$

is the Fermi-Dirac integral used in statistical physics.

The Fermi-Dirac family provides a one-dimensional continuous parametrization that joins the multivariate uniform distribution on an ellipsoid to the multivariate normal distribution.

A discussion of maximum likelihood estimation of its parameters illustrates some interesting nonstandard phenomena. For example, as a by-product, a possible solution to the problem of circumscribing the smallest ellipsoid to a set of points in R^p is obtained. The method is illustrated with a multivariate quality control example.

[1]Research partially supported by the National Science Foundation, Grant DMS-9303556.

[0]*Key words and phrases.* Fermi-Dirac distribution, elliptically symmetric distributions, platikurtosis, smallest enclosing ellipsoid, profile likelihood.

1. Introduction

In recent years, there has been a lot of interest in studying multivariate distributions alternative to the normal model. Motivations may be found in the development of modern computing and in the advances in theoretical statistics, which allow for more flexibility in the choice of a parametric model for observable data. In particular, we currently have a deeper understanding of the implications of certain symmetry assumptions on the multivariate structure and of the extent to which some properties of the multivariate normal distributions are, in reality, properties of a larger class of distributions displaying certain regularities. These theoretical achievements are the effect of a spur of activities in multivariate analysis summarized in, among others, the recent books by Fang, Kotz and Ng (1990) and Fang and Zhang (1990).

On the applied side, more models available simply mean more freedom for the practitioner, enabled to choose the parametric model that best fits a certain situation, whether for empirical or theoretical reasons. Admittedly, this happens in spite of the enormous growth of modern nonparametric techniques.

In the field of quality control, for example, it is recognized that manufacturing processes give rise to distributions of continuous quality characteristics generally different from the normal model, because of the possibility of a distinct flatness in the center of the distribution. Such a characteristic is one of the aspects of platikurtosis; see for example Balanda and MacGillivray (1988) for a recent discussion of the many ramifications. An example by Taguchi is part of the contemporary quality control folklore; we were not able to find references other than Taguchi (1979), reported by several textbooks in industrial statistics. Analyzing a univariate color-intensity quality characteristic, he noticed that television sets manufactured in the United States tend to display a distribution which has approximately the same mean and standard deviation as television sets manufactured in Japan, but is in general flatter (more platikurtic) in the center, close to target specification. Lower variability in a neighborhood of the target helped Taguchi to explain the greater market success of Japanese sets, despite the approximate equality of proportion of nonconforming items (i.e., the probability of the quality characteristic lying outside a certain specification interval centered at target) in the two distributions. It should be noticed that, notwithstanding this famous example, the use of platikurtic distributions in quality control is quite rare, and usually limited to finite mixtures of normal distributions.

That platikurtic distributions are the norm, rather than the exception, in the manufacturing world was noticed before by Ferreri (1964), who defined a system of distributions described in the following section. An important half of that system are the Fermi-Dirac distributions, which are extended in this paper to the multivariate case.

In particular, after reviewing in Section 2 Ferreri's definition of the Fermi-Dirac distribution, we present its elliptically contoured multivariate generalization in Section 3, prove that it is not a scale mixture of normals in Section 4 and present some interesting limiting cases in Section 5. Maximum likelihood issues are analyzed in Sections 6 and 7 and an interesting geometrical connection is illustrated in Section 8.

2. A UNIVARIATE SYSTEM OF DISTRIBUTIONS BY FERRERI

Ferreri (1964) observed that distributions with exponential tails similar to the normal but with different degrees of kurtosis are important for applied work. In particular, platikurtic distributions are needed in manufacturing applications, whereas quantitative characteristics in the biological world have, quite often, leptokurtic distributions. Drawing from these empirical considerations and from an interesting derivation *á la* Pearson, based on differential equations for the frequency function, Ferreri proposes the system of univariate distributions

$$f(x) = \frac{\lambda_c(\alpha)}{\sigma G(c,-1/2,\alpha)} \frac{1}{c+\exp\{\alpha+\lambda_c^2(\alpha)(x-\mu)^2/\sigma^2\}} \tag{1}$$

where $-\infty < \mu < \infty$ is the mean, $\sigma > 0$ is the standard deviation, c can be either 1 or -1; for $c = 1$ and $p > -1$, we define the Fermi-Dirac integral

$$G(1,p,\alpha) := F(p,\alpha) := \int_0^\infty \frac{u^p}{1+\exp\{\alpha+u\}} du$$

and, correspondingly, the function

$$\lambda_1^2(\alpha) := \frac{F(1/2,\alpha)}{F(-1/2,\alpha)}.$$

Similarly, for $c = 1$ and $p > -1$, we define the Bose-Einstein integral

$$G(-1,p,\alpha) := \int_0^\infty \frac{u^p}{-1+\exp\{\alpha+u\}} du$$

and, correspondingly, the function $\lambda_{-1}^2(\alpha) := G(-1,1/2,\alpha)/G(-1,-1/2,\alpha)$. The two cases $c = 1$ and $c = -1$ represent two subfamilies of distributions, the first platikurtic and the second leptokurtic. The normal is a limiting case for both. The Fermi-Dirac and Bose-Einstein integrals are special functions extensively studied in the physics literature.

Ferreri's system of distributions, although mentioned in Johnson and Kotz (1970), is not well known. Part of the reason is that the computational requirements for the special functions above are very demanding, and investigations on their efficient computation

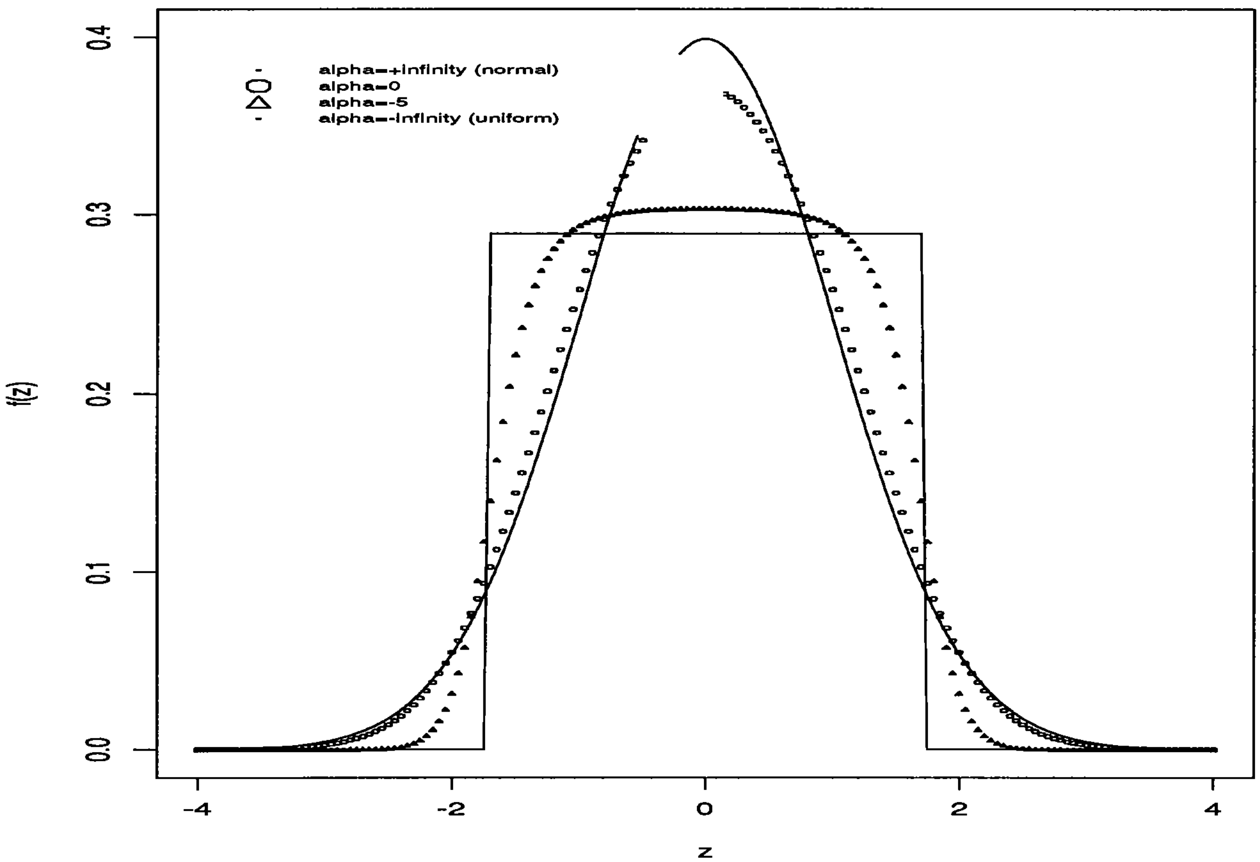

FIGURE 1. Univariate Fermi-Dirac densities with $\mu = 0$ and $\sigma^2 = 1$.

have continued until the present day. See for example the review paper by Blakemore (1982).

Only the Fermi-Dirac case ($c = 1$), more suitable for our statistical applications, is considered in this paper. To simplify the notation, we use $F(p, \alpha)$ instead of $G(1, p, \alpha)$ and drop the subscript from $\lambda_1(\alpha)$. A plot of some Fermi-Dirac univariate densities, i.e. formula (1) with $c = 1$, is shown in Figure 1 for $\alpha = 0, -5$ and the limiting cases $\alpha = -\infty$ and $\alpha = +\infty$, which represent, respectively, the uniform over the interval $[-\sqrt{3}, \sqrt{3}]$ and the standard normal.

3. The multivariate Fermi-Dirac distribution

The p-variate Fermi-Dirac distribution is the elliptically symmetric multivariate version of (1), with density

$$f(\mathbf{x}) = \frac{\Gamma(p/2)\lambda^p(\alpha)}{\pi^{p/2}F(p/2-1,\alpha)|\mathbf{\Sigma}|^{1/2}} \frac{1}{1+\exp\{\alpha+\lambda^2(\alpha)(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\}} \tag{2}$$

where $\mathbf{x}, \boldsymbol{\mu} \in R^p$, $\alpha \in R$, $\mathbf{\Sigma}$ is a $p \times p$ positive definite matrix of rank p, F is the Fermi-Dirac integral (2), discussed in the Appendix, and $\lambda^2(\alpha) := F(1/2,\alpha)/F(-1/2,\alpha)$. The corresponding spherically symmetric random vector $\mathbf{Z}$, obtained when $\boldsymbol{\mu} = 0$ and $\mathbf{\Sigma}$ is the unit matrix, has density

$$f(\mathbf{z}) = \frac{\Gamma(p/2)\lambda^p(\alpha)}{\pi^{p/2}F(p/2-1,\alpha)} \frac{1}{1+\exp\{\alpha+\lambda^2(\alpha)\mathbf{z}'\mathbf{z}\}}. \tag{3}$$

By standard results in the theory of elliptically symmetric distributions (see for example Fang, Kotz and Ng (1990)), if $\mathbf{X}$ has density (2), then it has the stochastic representation

$$\mathbf{X} \stackrel{d}{=} R\mathbf{\Sigma}^{1/2}\mathbf{U} + \boldsymbol{\mu} \tag{4}$$

where $\stackrel{d}{=}$ means equality in distribution, R is a positive random variable such that

$$R^2 \stackrel{d}{=} (\mathbf{X}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})$$

and $\mathbf{U}$ is a p-variate random vector uniformly distributed on the unit hypersphere independent of R. The density of R^2, calculated in t, is

$$h(t) = \frac{\lambda^p(\alpha)}{F(p/2-1,\alpha)} \frac{t^{p/2-1}}{1+\exp\{\alpha+\lambda^2(\alpha)t\}}$$

and plays a role similar to the $\chi^2(p)$ density in the normal theory. For example, R is the random variable one should generate in order to simulate from the multivariate Fermi-Dirac distribution [see Johnson (1987), chapter 6].

The mixed moments of the spherically symmetric version are (see Fang, Kotz and Ng (1990), page 34)

$$\begin{aligned} E(\prod \mathbf{Z}_i^{2s_i}) &= E(R^{2s})\pi^{-p/2}\frac{\Gamma(p/2)}{\Gamma(p/2+s)}\prod_{i=1}^{p}\Gamma(1/2+s_i) \\ &= \frac{F(s+p/2-1,\alpha)\Gamma(p/2)}{F(p/2-1,\alpha)\lambda^{2s}\pi^{p/2}\Gamma(p/2+s)}\prod_{i=1}^{p}\Gamma(1/2+s_i) \end{aligned} \tag{5}$$

where $s = s_1 + s_2 + \dots s_p$. The moments of (2) can be calculated from (5) using standard methods.

4. The symmetric Fermi-Dirac distribution is not a scale mixture of normals

The distribution of the random vector $(X_1, \ldots, X_p)$ is said to be a scale mixture of central normal distributions if there exists a random vector $(N_1, \ldots, N_p)$ of independent standard normal variables and an independent variable V such that $(X_1, \ldots, X_p)$ has the same distribution as $(N_1/V, \ldots, N_p/V)$.

Scale mixtures of normal distributions are very important as models for heterogeneous populations and as marginals of exchangeable random variables in Bayesian models with normal likelihoods. They represent an important subfamily of spherically symmetric distributions. In particular, they are the only possible choice for the finite dimensional distributions of an infinite sequence of random variables with the property of spherical symmetry [see for example Kingman (1972)]. It is therefore quite interesting to investigate whether a given spherically symmetric distribution is a scale mixture of normals.

A viable necessary and sufficient condition for a given univariate symmetric density to be a scale mixture of central normal distributions was given by Andrews and Mallows (1974). It is possible to extend their results to the p-dimensional case, but that is not needed here, since we will actually prove that the Fermi-Dirac density is not a scale mixture of normals: if a vector $(X_1, \ldots, X_p)$ is a scale mixture of normals, the same is true for X_1, so it is enough to prove the negative result in the univariate case. The condition given by Andrews and Mallows is recalled in the next theorem.

Theorem 1 (Andrews and Mallows). *The distribution of the random variable X is a scale mixture of central normal distributions if and only if its density $f(x)$ satisfies the following condition:*

$$(-\frac{d}{dx})^k f(\sqrt{x}) \geq 0 \tag{6}$$

for all $k \geq 1$ and all $x > 0$.

For the univariate spherically symmetric Fermi-Dirac distribution, we can apply the test to the function

$$\frac{\lambda(\alpha)}{F(-1/2, \alpha)} \frac{1}{1 + \exp\{\alpha + \lambda^2(\alpha) u\}}, \qquad u > 0.$$

or equivalently to the function

$$h(x) := \frac{1}{1 + e^{\alpha + u}}, \qquad u > 0.$$

Now, it is easy to see that the test applied to the second derivative $h''(u) = e^{\alpha+u}(-1 + e^{\alpha+u})/(1 + e^{\alpha+u})^3$ rules out negative values of α. Taking one more derivative, $-h'''(u) =$

$e^{\alpha+u}(1-4e^{\alpha+u}+e^{2(\alpha+u)})/(1+e^{\alpha+u})^4$ we can see that the test is not satisfied for $\alpha < 2+\sqrt{3}$. Taking further derivatives has the practical effect of bounding more and more the set of α satisfying the test; such a set is actually empty, as it is shown in the following theorem.

Theorem 2. *The Fermi-Dirac distribution is not a scale mixture of central normal distributions for any* α.

Proof. By the above remarks, we only need to consider $\alpha > 3$. In particular, letting $y := \alpha + u$, we have to show that there is no $\alpha > 3$ such that

$$(-\frac{d}{dy})^k \frac{1}{1+e^y} \geq 0 \quad \forall k \;\; \forall y \geq \alpha.$$

Writing $(1+e^y)^{-1}$ as a geometric series, it is clear that the condition can be rewritten

$$\sum_{n=0}^{\infty}(-1)^n(n+1)^k e^{-(n+1)y} \geq 0 \quad \forall k \;\; \forall y \geq \alpha.$$

Now, for $y \geq 3$, take $k = 2y$ and write

$$\sum_{0}^{\infty}(-1)^n(n+1)^{2y}e^{-(n+1)y} = e^{-y} - 2^{2y}e^{-2y} + 3^{2y}e^{-3y} - \sum_{j=4}^{\infty}(j^{2y}e^{-jy} - (j+1)^{2y}e^{-(j+1)y}).$$

The sum of the first three terms is negative and each term of the remainder series is positive, proving the result. ∎

5. A BRIDGE BETWEEN THE UNIFORM AND THE NORMAL

It was shown in the previous section that the Fermi-Dirac model cannot be reduced to a mixture of normals. One consequence is that its platikurtosis cannot be explained in terms of heterogeneity, but it is an inherent property of the distribution.

The model is one of many possible responses to the observation that "not only there is a paucity of multivariate nonnormal distributional models, but also most of the proposed alternative distributions (e.g., multivariate lognormal, exponential) are defined so as to have properties that are similar to those of the multivariate normal (e.g., that all marginal distributions belong to the same class)" contained in Gnanadesikan (1977, page 161). The Fermi-Dirac distribution has exponential tails, like the normal, but it differs from it in the area close to the center of the distribution which, from the applied point of view, is of primary importance. The marginal distributions of a multivariate Fermi-Dirac are not of the same kind, and are quite complicated indeed.

The most desirable feature of the Fermi-Dirac model is that it provides a continuous parametrization bridging the normal and the uniform, two models a practitioner often has in mind as competitive alternatives. This is a consequence of the following

Theorem 3. *If* $\mathbf{X}$ *has a Fermi-Dirac distribution, its density (2) converges to a normal density with parameters* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$ *as* $\alpha \to \infty$*, and to a uniform density inside the p-dimensional ellipsoid*

$$\{\mathbf{x} \in R^p : (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) < 3\} \tag{7}$$

as $\alpha \to -\infty$.

Proof. Due to the approximations illustrated in the appendix, we have the asymptotic equivalences $\lambda^2(\alpha) \to .5$ as $\alpha \to \infty$ and $\lambda^2(\alpha) \sim -\alpha/3$ as $\alpha \to -\infty$. Thus

$$\lim_{\alpha\to\infty} f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}$$

$$\lim_{\alpha\to-\infty} f(\mathbf{x}) = \lim_{\alpha\to-\infty} \frac{p/2\Gamma(p/2)(-\alpha/3)^{p/2}}{\pi^{p/2}|\Sigma|^{1/2}(-\alpha)^{p/2}} \frac{1}{1 + \exp\{\alpha - \alpha(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})/3\}}$$

$$= \frac{\Gamma(p/2+1)}{(3\pi)^{p/2}} \frac{((\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) < 3)}{|\boldsymbol{\Sigma}|^{1/2}} \tag{8}$$

where (A) denotes the indicator function of the event A. ∎

The univariate exponential power family of distributions, considered, among others, by Box and Tiao (1973, chapter 3.1) also contains the normal and the uniform as a special and a limiting case. The tail behavior varies widely within that family, though, and it is not clear how to generalize it to higher dimensions.

6. Maximum likelihood estimation

The loglikelihood function of a random sample of n p-dimensional vectors $\mathbf{X}_1, \ldots \mathbf{X}_n$ from density (2) is defined as usual:

$$\ell(\alpha, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = n \log(\frac{\Gamma(p/2)\lambda^p(\alpha)}{\pi^{p/2}F(p/2-1,\alpha)}) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \sum_{i=1}^{n} \log(1 + \exp\{\alpha + \lambda^2(\alpha)(\mathbf{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\}). \tag{9}$$

Due to the limiting results illustrated in the previous Section, we can extend the parameter space by stipulating that $\alpha = -\infty$ means $\mathbf{X}$ is multivariate uniform inside ellipsoid (7) and $\alpha = \infty$ means $\mathbf{X}$ is multivariate normal. It is possible to observe a likelihood increasing as $\alpha \to \infty$, which would be a strong indication for switching to the normal model, instead of the more complex Fermi-Dirac model. Similarly, if we observe a likelihood growing as $\alpha \to -\infty$, we could opt for the uniform model. In other words, inspection of the *profile likelihood*

$$\wp\ell(\alpha) = \sup_{\boldsymbol{\mu},\boldsymbol{\Sigma}} \ell(\alpha, \boldsymbol{\mu}, \Sigma) \tag{10}$$

serves both the purposes of estimation and model selection.

For these reasons and because of considerable computational problems created by differentiating expression (9) with respect to α, it is advisable to separate the maximum likelihood analysis into two steps: firstly, maximizing $l(\alpha, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ for every fixed α; secondly, inspecting the profile likelihood (10) to gain information on the appropriate model and to approximate the maximum likelihood estimate of α. Both steps require intensive numerical computations.

Using some results illustrated, for example, by Mardia, Kent and Bibby (1979, page 103-104), the likelihood equations in $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, for a fixed α, can be written

$$\mathbf{0} = \frac{\partial l}{\partial \boldsymbol{\mu}} = -2\lambda^2(\alpha)\boldsymbol{\Sigma}^{-1} \sum_{i=1}^{n} \frac{\exp\{\alpha + \lambda^2(\alpha)(\mathbf{x}_i - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\}}{1 + \exp\{\alpha + \lambda^2(\alpha)(\mathbf{x}_i - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\}}(\mathbf{x}_i - \boldsymbol{\mu})$$

(11)

$$\mathbf{0} = \frac{\partial l}{\partial \boldsymbol{\Sigma}^{-1}} = n\boldsymbol{\Sigma} - \frac{n}{2}\operatorname{diag} \boldsymbol{\Sigma} - \lambda^2(\alpha) \sum_{i=1}^{n} \frac{\exp\{\alpha + \lambda^2(\alpha)(\mathbf{x}_i - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\}}{1 + \exp\{\alpha + \lambda^2(\alpha)(\mathbf{x}_i - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\}} \times \\ (2(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})' - \operatorname{diag}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})')$$

These equations do not have an explicit solution, and are not very transparent indeed. Considering the univariate case $p = 1$ may help. For the sake of simplicity, let

$$w_i := w_i(\alpha, \mu, \sigma^2) := \frac{\exp\{\alpha + \lambda^2(\alpha)(x_i - \mu)^2/\sigma^2\}}{1 + \exp\{\alpha + \lambda^2(\alpha)(x_i - \mu)^2/\sigma^2\}}, \quad i = 1, \dots, n.$$

Then the likelihood equations can be rewritten

$$\hat{\mu} = \sum_{i=1}^{n} x_i w_i / \sum_{i=1}^{n} w_i$$

$$\hat{\sigma}^2 = 2\frac{\lambda^2(\alpha)}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2 w_i, \tag{12}$$

from which we can clearly see that the maximum likelihood estimators of μ and σ^2 are weighted versions of the sample mean and the sample variance, according to weights w_i that depend on α but also on μ and σ^2 themselves. Notice that equations (12) only define an implicit relation the maximum likelihood estimators must satisfy and are not useful in actually computing the MLEs.

It is interesting to notice that the sample mean and the sample variance are actual solutions of equations (12) if we let $\alpha \to \infty$, since $\lim_{\alpha\to\infty} \lambda^2(\alpha) = .5$ and $\lim_{\alpha\to\infty} w_i = 1$, $i = 1, \ldots, n$ uniformly with respect to μ and σ. This is expected, since for $\alpha = \infty$ the model is the normal distribution. At the other extreme, numerical computations suggest that the solutions to equations (12) converge to the MLE estimators of μ and σ in the uniform case as $\alpha \to -\infty$, which are

$$\hat{\mu} = \frac{\min_i x_i + \max_i x_i}{2} \qquad (13)$$

$$\hat{\sigma^2} = \frac{(\max_i x_i - \min_i x_i)^2}{12}.$$

For $p = 1$ but also for moderate dimension $p > 1$, the Newton-Raphson method gives good results in solving likelihood equations (11), but some caution is needed.

As it is usual for Newton-Raphson, the choice of the starting values of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a crucial one. The naive choice, i.e. using the vector of sample means, variances and covariances as initial value, causes the Newton Raphson algorithm to diverge for $\alpha < 0$. A more careful strategy is contained in the following.

Algorithm 1. *To compute a numerical approximation of the profile likelihood of* α*:*

1. *start the Newton-Raphson computation of* $\hat{\boldsymbol{\mu}}$ *and* $\hat{\boldsymbol{\Sigma}}$ *at a large value of* α*, using the vector of sample means, variances and covariances as starting values;*
2. *decrease* α*; use as starting values* $\hat{\boldsymbol{\mu}}$ *and* $\hat{\boldsymbol{\Sigma}}$ *obtained from the previous value of* α*;*
3. *repeat step 2 until convergence is apparent.*

The algorithm usually terminates at a a negative value of α with a large absolute value. An example is illustrated in the next Section.

7. An application in multivariate quality control

Tracy and Young (1992) present multivariate data taken from a real chemical process. $n = 13$ observations collected on the variables X_1 = % impurities, X_2 = temperature and X_3 = concentration are shown in Table 1. The authors write "preliminary tests provided no reason to doubt that the data follow a multivariate normal distribution", but for the sake of comparison we want to fit a multivariate Fermi-Dirac distribution with $p = 3$ using Algorithm 1 to maximize the likelihood.

Table 2 contains selected values of α, its profile likelihood and the corresponding MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Figure 2 is a plot of the approximate profile likelihood of α over a much finer

% impurities	temperature	concentration
16.90	83.77	43.44
17.38	84.46	42.74
16.90	86.27	43.60
16.92	85.23	43.18
16.71	83.81	43.72
17.07	86.08	43.33
16.93	85.85	43.41
16.71	85.73	43.28
16.88	86.27	42.59
16.73	83.46	44.00
17.07	85.81	42.78
17.60	85.92	43.11
16.90	84.23	43.48

TABLE 1. Tracy data.

grid. It is very interesting to notice that the plot shows a decreasing profile likelihood over the relevant range $\alpha \in [-100, 100]$. This is to say that the data show more evidence for the uniform rather than the normal model, the MLE of α being $-\infty$. In other words, for such a small sample size, the data cannot discriminate between the usually assumed normal model and a uniform over an ellipsoid. The usual tests of multivariate normality are powerful against alternative distributions with a different tail behavior, and that is probably the reason why preliminary tests of normality were not significant. On the other hand, the Taguchi example illustrated in the Introduction provides some motivations for concentrating on a more careful modelling of the center of the distribution. For practical purposes, given only the evidence provided by the small sample, the uniform model over the ellipsoid approximated by the line $\alpha = -100$ of Table 2 should be preferred over the normal model of $\alpha \approx 100$.

8. Smallest enclosing ellipsoids

One interesting byproduct of the algorithm illustrated in the previous section is the approximation of the maximum likelihood estimator of $\boldsymbol{\mu}$ and Σ when sampling from a uniform distribution on a p-dimensional ellipsoid, with density (8). In that case, for practical purposes, the MLEs of $\boldsymbol{\mu}$ and Σ are equivalent to the MLE of $\boldsymbol{\mu}$ and Σ for a value of α negative and large in absolute value. For the Tracy data, as remarked

alpha	pl(alpha)	mu1	mu2	mu3
-100	-16.86805	17.08297	85.22751	43.21661
-50	-17.42532	17.08309	85.22804	43.21657
-20	-18.51286	17.08325	85.22949	43.21642
-10	-19.60991	17.07892	85.22378	43.21764
-5	-20.36843	17.06036	85.19454	43.22815
-2	-20.74008	17.02197	85.16261	43.25253
-1	-20.83072	17.00510	85.15617	43.26298
0	-20.90925	16.99158	85.15175	43.27163
1	-20.96650	16.98342	85.14860	43.27707
2	-20.99812	16.97952	85.14677	43.27974
5	-21.01991	16.97706	85.14546	43.28144
10	-21.02115	16.97692	85.14539	43.28154
20	-21.02116	16.97692	85.14539	43.28154
50	-21.02116	16.97692	85.14539	43.28154
100	-21.02116	16.97692	85.14539	43.28154

alpha	Sigma11	Sigma22	Sigma33	Sigma12	Sigma13	Sigma23
-100	0.117626	1.394761	0.306755	0.103767	-0.086140	-0.366880
-50	0.119621	1.418129	0.311870	0.105213	-0.087635	-0.372263
-20	0.121703	1.440630	0.316404	0.106009	-0.089195	-0.375023
-10	0.121597	1.423967	0.306059	0.104954	-0.088047	-0.358483
-5	0.113344	1.319191	0.262673	0.098902	-0.080698	-0.307158
-2	0.091551	1.169762	0.205005	0.086612	-0.067744	-0.248875
-1	0.081019	1.114526	0.184860	0.080954	-0.061644	-0.230361
0	0.072123	1.067848	0.168858	0.076146	-0.056274	-0.215841
1	0.066502	1.036536	0.158833	0.073027	-0.052744	-0.206631
2	0.063734	1.020229	0.153862	0.071449	-0.050963	-0.201997
5	0.061951	1.009335	0.150640	0.070413	-0.049799	-0.198963
10	0.061853	1.008721	0.150460	0.070355	-0.049734	-0.198794
20	0.061852	1.008717	0.150459	0.070355	-0.049734	-0.198793
50	0.061852	1.008717	0.150459	0.070355	-0.049734	-0.198793
100	0.061852	1.008717	0.150459	0.070355	-0.049734	-0.198793

TABLE 2. Profile likelihood and maximum likelihood estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ for selected values of α.

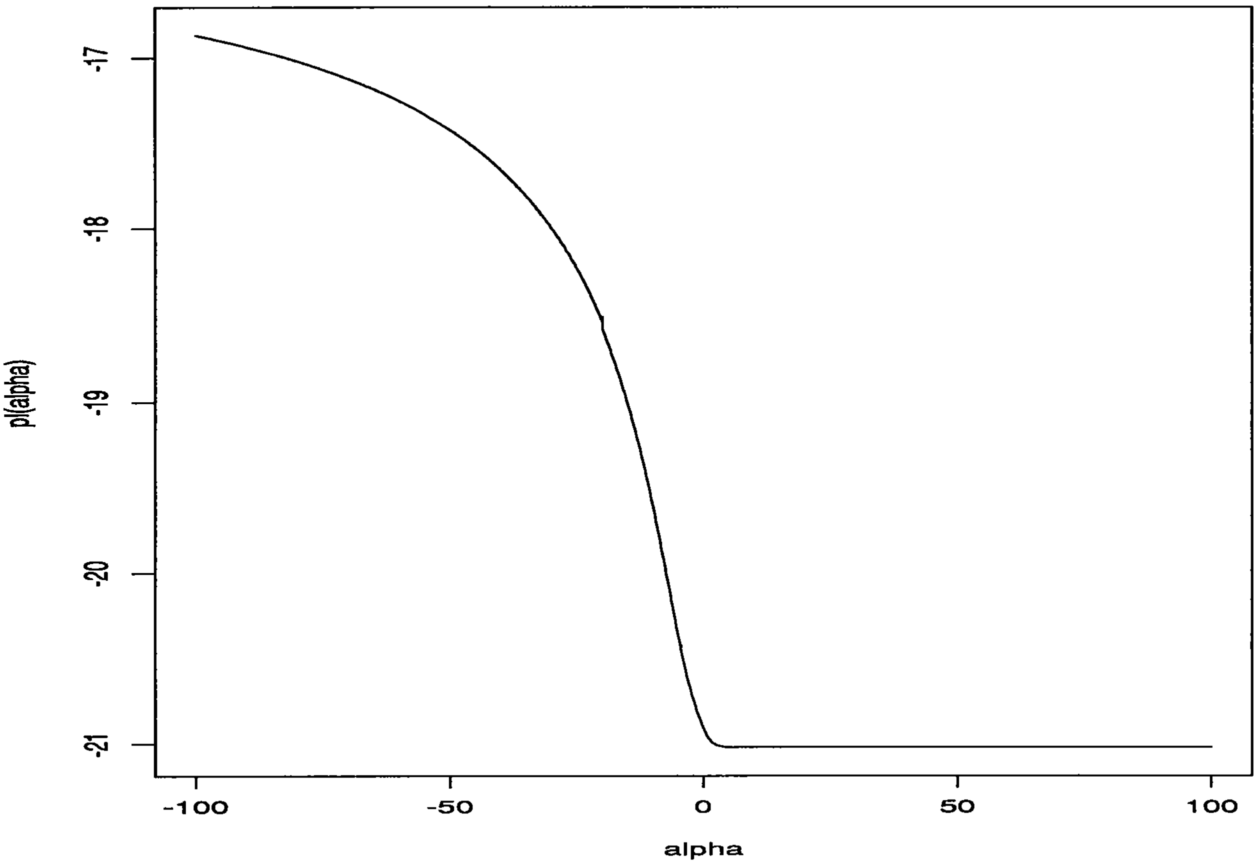

FIGURE 2. Profile likelihood of α for the Tracy data.

above, a Newton Raphson approximation of $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ from a uniform distribution on a 3-dimensional ellipsoid can be read off the lines of Table 2 corresponding to $\alpha = -100$.

Until recently, the computation of maximum likelihood estimators for a uniform distribution on a p-dimensional ellipsoid was a challenging problem, except for the case $p = 1$, when they reduce, trivially, to expressions (13). For a discussion of the case $p = 2$ see instead Silverman and Titterington (1981). From the point of view of Computational Geometry, the problem is tantamount to the calculation of the smallest ellipsoid enclosing n points in R^p. A constructive solution has been given recently by Welzl (1991). See also references thereof. For the Tracy data, an implementation [2] of the Welzl algorithm for $p = 3$ provides results equivalent to the lines of Table 2 corresponding to $\alpha = -100$.

---

[2]gently provided by the author.

9. Appendix: the Fermi-Dirac integral

The Fermi-Dirac integral

$$F(p,\alpha) := \int_0^\infty \frac{u^p}{1+\exp\{\alpha+u\}}du$$

is real-valued for any α and $p > -1$. Let $\Gamma(p) := \int_0^\infty u^{p-1}e^{-u}du$ be the gamma function. Asymptotic expressions for the Fermi-Dirac integral are given in the following proposition.

Proposition 1. *The following equalities hold*

$$\lim_{\alpha\to\infty} \frac{F(p,\alpha)e^\alpha}{\Gamma(p+1)} = 1 \quad \textit{and} \quad \lim_{\alpha\to-\infty} \frac{F(p,\alpha)(p+1)}{-\alpha^{(p+1)}} = 1$$

Proof. The first limit is obtained from the following series expansion, for $\alpha > 0$:

$$F(p,\alpha) := \Gamma(p+1)\sum_{n=1}^\infty (-1)^{n-1}\frac{e^{-na}}{n^{p+1}}.$$

The expansion can be derived by writing the integrand as a geometric series.

For the second limit, write instead, for $\alpha < 0$,

$$F(p,\alpha) := \int_0^{-2\alpha} \frac{u^p}{1+\exp\{\alpha+u\}}du + \int_{-2\alpha}^\infty \frac{u^p}{1+\exp\{\alpha+u\}}du =: I_\alpha + I'_\alpha, \text{ say.} \tag{14}$$

We then have, by the change of variable $x = u + 2\alpha$,

$$I'_\alpha = \int_0^\infty \frac{(-2\alpha+x)^p}{1+\exp\{-\alpha+x\}}du \to 0 \text{ as } \alpha \to -\infty$$

since, if $p \le 0$, then $(-2\alpha+x)^p/(1+\exp\{-\alpha+x\}) \le (-2\alpha)^p e^{-x}/e^\alpha$. On the other hand, if $p > 0$, then $(-2\alpha+x)^p/(1+\exp\{-\alpha+x\}) \le p!2^p e^{-x/2}$ which is integrable, and the Dominated Convergence theorem applies.

For the first term on the right hand side of equation (14), we have instead, by the change of variable $x = -u/\alpha$,

$$\frac{I'_\alpha}{(-\alpha)^{p+1}} = \int_0^2 \frac{x^p}{(1+\exp\{-\alpha(x-1)\})}dx \to \int_0^1 x^p dx = \frac{1}{p+1}$$

since $x^p/(1+\exp\{-\alpha(x-1)\}) \le x^p$ and the Dominated Convergence theorem applies. ∎

Proposition (1) gives an approximation $F(p,\alpha) \sim \Gamma(p+1)e^{-\alpha}$ which is already very good for $\alpha > 4$, and another approximation $F(p,\alpha) \sim (-\alpha)^{p+1}/(p+1)$, good for $\alpha < -20$. In the intermediate range $-20 < \alpha < 4$, which is also the range of greater statistical interest, the Fermi-Dirac integral must be evaluated numerically. See Van Cong and Doan-Khanh (1992).

REFERENCES

Andrews, D.F. and Mallows, C.L. (1974). Scale mixtures of normal distributions. *J. Roy. Statist. Soc.* B, **36**, 99-102.

Balanda, P.K. and MacGillivray, H. L. (1988). Kurtosis: A critical review. *Amer. Statistician*, 111-119.

Blakemore, J.S. (1982). Approximations for Fermi-Dirac integrals, especially the function $F_{1/2}(\eta)$ used to describe elctron density in a semiconductor. *Solid State Elelctronics*, 111-119.

Box and Tiao (1973). *Bayesian inference in statistical analysis.* Addison-Wesley, Reading.

Feller, W. (1966). *An introduction to probability theory and its applications.* Wiley, New York.

Ferreri, C. (1964).

Fang, K.-T., Kotz, S. and Ng, K.W. (1990). *Symmetric Multivariate and Related Distributions.* Chapman and Hall, London.

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

Gnanadesikan, R. (1977). *Methods for statistical data analysis of multivariate observations.* Wiley, New York.

Johnson, Mark E. (1987). *Multivariate statistical simulation.* Wiley, New York.

Johnson, N.L. and Kotz, S. (1970). *Continuous univariate distributions-1.* Wiley, New York.

Kingman, J.F.C. (1972). On random sequences with spherical symmetry. *Biometrika*, **59**, 2, 492-494.

Taguchi, G. (1979). *The Asahi.* Japanese newspaper, April 15, 1979.

Silverman, B.W. and Titterington, D.M. (1980). Minimum covering ellipses. *SIAM J. Sci. Stat. Comput.* , **1**, 4, 401-409.

Tracy, Nola D. and Young, John C. (1992). Multivariate control charts for individual observations. *J. Qual. Tech.*, **24**, 2, 88-95.

Van Cong, H. and Doan-Khanh, B. (1992). Simple accurate general expression of the Fermi-Dirac Integral $F_j(a)$ for arbitrary a and $j > -1$. *Solid-State Electronics*, **35**, 7, 949-951.

Welzl, E.(1991). Smallest enclosing disks (balls and ellipsoids). In *New results and new trends in Computer Science*, Maurer, H. *ed.*, 359-370, Springer Verlag, Berlin.