

INFERENCE FOR LINEAR MODELS WITH RADIALLY DECOMPOSABLE ERROR

BY K. W. NG AND D. A. S. FRASER

University of Hong Kong and York University

This paper considers a linear model with an error term following a so-called radially decomposable multivariate distribution which can be represented as an independent product of a random scalar component, called the radial component, and a vector component, called the base component. The radially decomposable distributions include many symmetric multivariate distributions such as spherically symmetric distributions, ℓ_1 -norm symmetric distributions, Liouville multivariate distributions and α -symmetric distributions with $\alpha = 1$. It is shown that the inference for the regression parameter β depends only on the distribution of the base component. Some consequences of this fact are also discussed.

1. Introduction. In many applications of the following linear model

$$Y = X\beta + \epsilon,$$

the errors are known to be non-normal. If the errors are i.i.d., one may consider the non-normal univariate distributions such as the Cauchy, Student's t , Laplace's double exponential, logistic, extreme value distributions and the stable-law distribution. If the errors are identically distributed but not independent, multivariate versions of these and other distributions are usually adopted for the error vector. In general, inference for linear models having multivariate error distributions is a thorny problem. Some of these multivariate distributions can be represented as an independent product of a random scalar component and a random vector component. In this case, the multivariate distribution is said to be *radially decomposable*, the scalar variable being called the *radial component* and the vector component the *base component*. In section 3, we shall show that the inference for the regression parameters depends only on the distribution of the base component, but not on that of the radial component. A change in the distribution of the radial component will only affect the inference of the scale parameter of the error term.

AMS 1991 Subject Classifications: Primary 62J05, Secondary 62A25.

Key words and phrases: Linear models, Liouville distributions, ℓ_1 -norm distributions, radially decomposable, spherically symmetric distributions, α -symmetric distributions.

2. Radially decomposable distributions. An $n \times 1$ random vector \mathbf{Z} is said to be *radially decomposable* if there exist a positive random scalar R and an independent $n \times 1$ random vector \mathbf{W} such that \mathbf{Z} and $R\mathbf{W}$ are identically distributed, that is

$$\mathbf{Z} \stackrel{d}{=} R\mathbf{W}, \quad (1)$$

where the symbol $\stackrel{d}{=}$ means both sides are identically distributed. For properties of this operator, see Fang, Kotz and Ng (1989) and Zolotarev (1985). In the decomposition (1), R is called the *radial component* and \mathbf{W} is called the *base component*. We also use these terms for the corresponding distributions of \mathbf{Z} , R and \mathbf{W} respectively. Frequently, \mathbf{W} is confined in a subset, but this is not a condition for the definition.

Many symmetric multivariate distributions are radially decomposable. If \mathbf{W} is distributed uniformly on the unit sphere in the n dimensional space \mathbb{R}^n , the distribution of \mathbf{Z} of (1) is called a spherically symmetric distribution. The family of spherically symmetric distributions is a large family including the multivariate versions of normal, Student, Cauchy, Pearson Type II and Type VII, logistic, Kotz type, Bessel and symmetric stable law, etc. Readers are referred to the first four chapters of Fang, Kotz and Ng (1989) for the properties of the spherically symmetric distributions. If the base component \mathbf{W} is distributed according to a Dirichlet distribution on a closed simplex (i.e. the elements are all positive and sum to one), the distribution of \mathbf{Z} is called a Liouville distribution (see Fang, Kotz and Ng, 1989, chapter 6 and the references cited there). When \mathbf{W} is uniformly distributed on the closed simplex, the distribution of \mathbf{Z} is called a multivariate ℓ_1 -norm symmetric distribution by Fang and Fang (1988, 1989). This is a special case of the Liouville distribution. To provide the final example of families of radially decomposable distributions, we first introduce the so-called α -symmetric distributions of Cambanis, Keener and Simons (1983). A multivariate distribution is said to be α -symmetric if its n -dimensional characteristic function has the following form

$$\phi(\mathbf{t}) = \psi\left(\sum |t_i|^\alpha\right). \quad (2)$$

Cambanis, Keener and Simons (1983) shows that if \mathbf{Z} has an α -symmetric distribution with $\alpha = 1$, \mathbf{Z} can be decomposed into an independent product of (1). The elements of the base component \mathbf{W} can be represented as $W_i = U_i/\sqrt{D_i}$, where (U_1, \dots, U_n) is uniformly distributed on the unit sphere in the n -dimensional space and (D_1, \dots, D_n) has an independent Dirichlet distribution. Note that when $\alpha = 2$, the α -symmetric distribution is a spherically symmetric distribution, also radially decomposable. It is an open conjecture that the α -symmetric distributions are all radially decomposable for $0 < \alpha \leq 2$. See Fang, Kotz and Ng (1989, Chapter 7) for a shorter proof of

the radial decomposability of a 1-symmetric distribution and other properties of an α -symmetric distribution. Note that since the standard Cauchy distribution has c.f. $\exp(-|t|)$, an i.i.d. sample of size n from the standard Cauchy distributions form an n -dimensional α -symmetric distribution with $\alpha = 1$.

In each of the three families mentioned above, all member distributions have a common base component. That is, the member distributions are different only by their radial components. This suggests that for such a family, the distribution of a scale-invariant statistic $T(\mathbf{Z})$, $T(a\mathbf{Z}) \stackrel{d}{=} T(\mathbf{Z})$, depends only on the base component. In fact, we can show this for a k -dimensional scale invariant statistic $\mathbf{T}(\mathbf{Z})$, $\mathbf{T}(a\mathbf{Z}) \stackrel{d}{=} \mathbf{T}(\mathbf{Z})$ as follows. By taking conditional expectation as an intermediate step and making use of the scale-invariance, we have

$$\begin{aligned} E(\exp(is'\mathbf{T}(\mathbf{Z}))) &= E(E(\exp(is'\mathbf{T}(R\mathbf{W}))|R)) \\ &= E(E(\exp(is'\mathbf{T}(\mathbf{W}))|R)) = E(\exp(is'\mathbf{T}(\mathbf{W}))). \end{aligned} \tag{3}$$

Therefore $\mathbf{T}(\mathbf{Z})$ and $\mathbf{T}(\mathbf{W})$ have the same characteristic function, hence the distribution of $\mathbf{T}(\mathbf{Z})$ depends only on that of \mathbf{W} . Typical scale-invariant statistics are ratios of linear combinations of the components (or ordered components) of \mathbf{Z} and some ratios of homogeneous functions of equal order.

To appreciate the implication of this property of a scale-invariant statistic of a radially decomposable \mathbf{Z} , we consider again the three families discussed above. Let $\mathbf{T}(\mathbf{Z})$ be a scale-invariant statistic of \mathbf{Z} in the sequel. As long as \mathbf{Z} is spherically symmetric, the distribution of $\mathbf{T}(\mathbf{Z})$ is the same as when the elements of \mathbf{Z} are i.i.d. standard normal. If \mathbf{Z} has a Liouville distribution with Dirichlet parameter $(\alpha_1, \dots, \alpha_n)$, the distribution of $\mathbf{T}(\mathbf{Z})$ is the same as when the elements of \mathbf{Z} are independently $\text{gamma}(\alpha_i, 1)$, $i = 1, \dots, n$. When $\alpha_i = 1$, $i = 1, \dots, n$, the case of a multivariate ℓ_1 -norm symmetric distribution, the distribution of $\mathbf{T}(\mathbf{Z})$ is the same as if the elements of \mathbf{Z} are i.i.d. standard exponential. Finally, if \mathbf{Z} is α -symmetric with $\alpha = 1$, the distribution of $\mathbf{T}(\mathbf{Z})$ can be obtained treating all the elements of \mathbf{Z} as i.i.d. standard Cauchy.

3. Linear models with radially decomposable error distribution.

Let us consider the linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \sigma\mathbf{Z}, \tag{4}$$

where \mathbf{Y} is an $n \times 1$ observable random vector, \mathbf{X} is a given $n \times m$ design matrix of full rank, $\boldsymbol{\beta}$ is an $m \times 1$ vector of unknown parameters without restriction, σ is a positive unknown parameter, and \mathbf{Z} is a random $n \times 1$ vector whose distribution is known to be a member of a family \mathbb{P} of radially decomposable distributions with a common base-component distribution. Note

that this can be a non-parametric family because the radial component can have any continuous distribution on the positive axis. The transformation of \mathbf{Z} to \mathbf{Y} , as given by (1), gives rise to an induced family of distributions, $Q(\theta, \beta, \sigma)$, where θ symbolically indexes the member distributions in \mathbb{P} . The traditional analysis and Bayesian analysis both start with \mathbf{Y} and $Q(\theta, \beta, \sigma)$. Since $Q(\theta, \beta, \sigma)$ is a mixture of parametric and non-parametric family, we shall encounter difficulties in these two approaches.

In contrast to the above two approaches, the structural analysis (or error analysis) of the linear model (4) starts with, and focuses at, the source of variation \mathbf{Z} and its family of distributions \mathbb{P} . We take the model (4) literally and interpret the generation of the model output \mathbf{Y} with the following analogy. The elements of \mathbf{Z} (the standardized errors) are like numbers generated by one of the random-number generators — the member distributions of \mathbb{P} . These random numbers are then processed (transformed) by one of the channels indexed by (β, σ) , yielding the elements of \mathbf{Y} . We first ask the question: how much do we know about \mathbf{Z} , the numbers originally generated? Although \mathbf{Z} cannot be observed directly, we can observe a great deal of its characteristics. In theoretical physics, one can have a fresh look at the observational system by making a change of ‘reference frame’. In our terminology, it amounts to changing the coordinates for the input and output variables in equation (4), namely \mathbf{Z} and \mathbf{Y} . This we shall do below.

Originally, \mathbf{Z} and \mathbf{Y} have the rectangular coordinates in \mathbb{R}^n . The m independent columns of \mathbf{X} span an m -dimensional subspace $\mathcal{L}(\mathbf{X})$ in \mathbb{R}^n , so that n is decomposed into $\mathcal{L}(\mathbf{X})$ and its orthogonal complement $\mathcal{L}^\perp(\mathbf{X})$. In the subspace $\mathcal{L}(\mathbf{X})$ we use an ordinary rectangular coordinates $\mathbf{b}(\cdot)$ taking the columns of \mathbf{X} as the basis, while in $\mathcal{L}^\perp(\mathbf{X})$ we use a spherical coordinate system of a radius $r(\cdot)$ and a directional vector $\mathbf{u}(\cdot)$ on the unit sphere in $\mathcal{L}^\perp(\mathbf{X})$. That is, for the input variable \mathbf{Z} as a point in \mathbb{R}^n , we make the change of coordinates $\mathbf{Z} \rightarrow (\mathbf{b}(\mathbf{Z}), r(\mathbf{Z}), \mathbf{u}(\mathbf{Z}))$ as follows

$$\mathbf{b}(\mathbf{Z}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}, \quad (5a)$$

$$r^2(\mathbf{Z}) = \mathbf{Z}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Z}, \quad (5b)$$

$$\mathbf{u}(\mathbf{Z}) = r^{-1}(\mathbf{Z})(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Z}. \quad (5c)$$

Similarly, we obtain the corresponding change of coordinates for the output variable, $\mathbf{Y} \longleftrightarrow (\mathbf{b}(\mathbf{Y}), r(\mathbf{Y}), \mathbf{u}(\mathbf{Y}))$, by substituting \mathbf{Y} for \mathbf{Z} in (4). The expressions of old coordinates in terms of the new ones are:

$$\mathbf{Z} = \mathbf{X}\mathbf{b}(\mathbf{Z}) + r(\mathbf{Z})\mathbf{u}(\mathbf{Z}), \quad (6a)$$

$$\mathbf{Y} = \mathbf{X}\mathbf{b}(\mathbf{Y}) + r(\mathbf{Y})\mathbf{u}(\mathbf{Y}), \quad (6b)$$

Substituting (6a) and (6b) into (4) and comparing the coordinates on both sides,

$$\mathbf{X}\mathbf{b}(\mathbf{Y}) + r(\mathbf{Y})\mathbf{u}(\mathbf{Y}) = \mathbf{X}(\beta + \sigma\mathbf{b}(\mathbf{Z})) + \sigma r(\mathbf{Z})\mathbf{u}(\mathbf{Z}), \quad (7)$$

we obtain an equivalent representation of model (4) in the new coordinates system:

$$\mathbf{b}(\mathbf{Y}) = \boldsymbol{\beta} + \sigma \mathbf{b}(\mathbf{Z}), \tag{8a}$$

$$r(\mathbf{Y}) = \sigma r(\mathbf{Z}), \tag{8b}$$

$$\mathbf{u}(\mathbf{Y}) = \mathbf{u}(\mathbf{Z}), \tag{8c}$$

This representation shows that we can observe the $n - m - 1$ dimensional coordinates $\mathbf{u}(\mathbf{Z})$ of \mathbf{Z} , but not the remaining $m + 1$ dimensional coordinates $(\mathbf{b}(\mathbf{Z}), r(\mathbf{Z}))$.

At this point we emphasize that the choice of this coordinate system is not essential but merely for convenience and familiarity in manifesting the observable and unobservable parts of the source of variability. The analysis of the n -dimensional \mathbf{Z} into a *maximally observable* $n - m - 1$ dimensional component and the remaining unobservable $m + 1$ dimensional component can be carried out coordinates-free, using the theory of an $m + 1$ dimensional transformation group acting on an n -dimensional space. In fact, the analysis in its most general form does not limit to model (4), but can be applied to other models where some parameters are representing a group of transformations, see Fraser (1968, 1979). For a detailed treatment of the analysis using the theory of an exact Lie group of diffeomorphisms on an open set of \mathbb{R}^n , see Ng (1975, §1-2, §1-3 and the Appendix to Ch. 1).

Rewriting (8a)–(8c), we have

$$(\mathbf{b}(\mathbf{Y}) - \boldsymbol{\beta})/\sigma = \mathbf{b}(\mathbf{Z}), \tag{9a}$$

$$r(\mathbf{Y})/\sigma = r(\mathbf{Z}), \tag{9b}$$

$$\mathbf{u}(\mathbf{Y}) = \mathbf{u}(\mathbf{Z}). \tag{9c}$$

One more change of variables in (9a) and (9b), $(\mathbf{b}(\mathbf{Z}), r(\mathbf{Z})) \longleftrightarrow (\mathbf{t}(\mathbf{Z}), r(\mathbf{Z}))$, gives the following equivalent representation of model (4):

$$(\mathbf{b}(\mathbf{Y}) - \boldsymbol{\beta})/r(\mathbf{Y}) = \mathbf{b}(\mathbf{Z})/r(\mathbf{Z}) \equiv \mathbf{t}(\mathbf{Z}), \tag{10a}$$

$$r(\mathbf{Y})/\sigma = r(\mathbf{Z}), \tag{10b}$$

$$\mathbf{u}(\mathbf{Y}) = \mathbf{u}(\mathbf{Z}). \tag{10c}$$

That is, the unobservable $m + 1$ dimensions of \mathbf{Z} are now represented by $(\mathbf{t}(\mathbf{Z}), r(\mathbf{Z}))$ and the observable $n - m - 1$ dimensions are still represented by $\mathbf{u}(\mathbf{Z})$. For inference on $\boldsymbol{\beta}$ the relevant equation is clearly (10a). But there are two modes of inference for $\boldsymbol{\beta}$, depending on whether the unconditional distribution of $\mathbf{t}(\mathbf{Z})$ or the conditional distribution of $\mathbf{t}(\mathbf{Z})$ given the known value of $\mathbf{u}(\mathbf{Z})$ is used as the inferential distribution. In either mode, a hypothesized value of $\boldsymbol{\beta}$ corresponds, as $\mathbf{b}(\mathbf{Y})$ is given, to a hypothesized value of $\mathbf{t}(\mathbf{Z})$ which is then tested against the distribution of $\mathbf{t}(\mathbf{Z})$. Conversely, a

tolerance region for $\mathbf{t}(\mathbf{Z})$ leads to a confidence region for β , and simultaneous tolerance intervals for the components of $\mathbf{t}(\mathbf{Z})$ induce simultaneous intervals of the components of β . A point 'prediction' of $\mathbf{t}(\mathbf{Z})$ yields a point estimate of β . The results of the unconditional mode correspond to that of Fisher's approach of treating the left-hand side of (10a) as a so-called 'pivotal quantity'. From the analysis of \mathbf{Z} , the authors believe the conditional mode is more appropriate because we should not ignore the given value of $\mathbf{u}(\mathbf{Z})$. And for separate inferences of the components of β , we can simply use the component equations of (10a) together with the corresponding component distributions. For example, a separate confidence interval of β_i can be constructed using the equation

$$(b_i(\mathbf{Y}) - \beta_i)/r(\mathbf{Y}) = t_i(\mathbf{Z}), \quad i = 1, \dots, m \quad (11)$$

and the conditional distribution of $t_i(\mathbf{Z})$ given $\mathbf{u}(\mathbf{Z})$ or the unconditional distribution of $t_i(\mathbf{Z})$, depending on one's preference of the two modes of inference. Confidence regions or simultaneous confidence intervals for a subset of components of β , say $\beta^{(2)} = (\beta_2, \dots, \beta_m)$, typically of interest in a regression situation where the first column of \mathbf{X} consists of ones, shall be based on the equation

$$(\mathbf{b}^{(2)}(\mathbf{Y}) - \beta^{(2)})/r(\mathbf{Y}) = \mathbf{t}^{(2)}(\mathbf{Z}) \quad (12)$$

and the corresponding distribution of $\mathbf{t}^{(2)}(\mathbf{Z})$, where $\mathbf{b}^{(2)}(\mathbf{Y}) = (b_2(\mathbf{Y}), \dots, b_m(\mathbf{Y}))$ and $\mathbf{t}^{(2)}(\mathbf{Z}) = (t_2(\mathbf{Z}), \dots, t_m(\mathbf{Z}))$.

Suppose one is interested in the unconditional mode of inference. Note that $\mathbf{t}(\mathbf{Z})$ is scale-invariant. The discussion in §2 implies that for all member distributions in the family \mathbb{P} , the (unconditional) distribution of $\mathbf{t}(\mathbf{Z})$ is the same. Therefore the unconditional inference for β will be identical for all member distributions in \mathbb{P} . In other words, as far as β is concerned, we may select a member of \mathbb{P} which is the most convenient for deriving the distribution of $\mathbf{t}(\mathbf{Z})$.

For the conditional mode of inference on β , we note that $(\mathbf{t}(\mathbf{Z}), \mathbf{u}(\mathbf{Z}))$ is also scale-invariant, so that the distribution of $(\mathbf{t}(\mathbf{Z}), \mathbf{u}(\mathbf{Z}))$ is the same for all members of \mathbb{P} . This implies that the conditional distribution of $\mathbf{t}(\mathbf{Z})$ given $\mathbf{u}(\mathbf{Z})$ is the same for all members of \mathbb{P} . Therefore, the conditional inference of β given $\mathbf{u}(\mathbf{Z})$ is also identical for all members of \mathbb{P} and we may select the most convenient member of \mathbb{P} to obtain the conditional distribution of $\mathbf{t}(\mathbf{Z})$ given $\mathbf{u}(\mathbf{Z})$.

Unfortunately, the inference for σ , which is based on equation (10b) and the conditional or unconditional distribution of $r(\mathbf{Z})$, does depend on individual members of \mathbb{P} . In other words, if we have no idea about the true distribution of \mathbf{Z} , making inference on σ is extremely difficult.

In principle, the information on the true distribution of \mathbf{Z} is contained in $\mathbf{u}(\mathbf{Z})$, because $\mathbf{u}(\mathbf{Z})$ represents the maximally observable part of \mathbf{Z} . And if the family has only one member, $\mathbf{u}(\mathbf{Z})$ would be used for assessing the aptness of

that assumed distribution. But since the distribution of $\mathbf{u}(\mathbf{Z})$ is the same for all members of \mathbb{P} , due to scale-invariance, we have no discriminant information on the members of \mathbb{P} .

4. Some consequences. We shall consider some consequences of the analysis in §3. If the family of \mathbb{P} distributions for the error \mathbf{Z} consists of spherically symmetric distributions, the inference of β in the conditional or unconditional mode should be the same as in the special situation when \mathbf{Z} consists of i.i.d. standard normal components. This may partly explain the often-reported robustness of normal inference when the error distribution is not extremely skewed. In the i.i.d. standard normal case, $\mathbf{b}(\mathbf{Z})$, $r(\mathbf{Z})$ and $\mathbf{u}(\mathbf{Z})$, defined in (5a)–(5c), are independent. The distribution of $\mathbf{b}(\mathbf{Z})$ is multinormal $N_m(\mathbf{0}, (\mathbf{X}'\mathbf{X})^{-1})$, $r^2(\mathbf{Z})$ has a central chi-squared distribution of $n-m$ degree of freedom, and $\mathbf{u}(\mathbf{Z})$ is uniformly distributed on the unit sphere in the $n-m$ dimensional space \mathbb{R}^{n-m} . Therefore the distribution of $\mathbf{t}(\mathbf{Z})/\sqrt{n-m}$, where $\mathbf{t}(\mathbf{Z})$ is defined in (10a), is an m -variate centered Student distribution with $n-m$ degree of freedom and quadratic-form parameter matrix $\mathbf{X}'\mathbf{X}$. This implies that the quadratic form $\mathbf{t}'(\mathbf{Z})\mathbf{X}'\mathbf{X}\mathbf{t}(\mathbf{Z})/(n-m)$ has an $F(m, n-m)$ distribution. Note that $\mathbf{t}(\mathbf{Z})$ is independent of $\mathbf{u}(\mathbf{Z})$, so the two modes of inference should give the same conclusion about β . Since the distribution of $\mathbf{t}(\mathbf{Z})/(n-m)$ has ellipsoidal contours of constant p.d.f., a $100(1-\alpha)\%$ tolerance region for $\mathbf{t}(\mathbf{Z})$ having *minimum-volume* is given by the ellipsoid

$$\{\mathbf{t}(\mathbf{Z}) : \mathbf{t}'(\mathbf{Z})\mathbf{X}'\mathbf{X}\mathbf{t}(\mathbf{Z})/(n-m) \leq F_\alpha(m, n-m)\}, \quad (13)$$

where $F_\alpha(m, n-m)$ is the $(1-\alpha)$ percentile of an $F(m, n-m)$ distribution. This gives the *most-concentrated* $100(1-\alpha)\%$ confidence region for β — a classical result for the normal inference,

$$\{\beta : (\mathbf{b}(\mathbf{Y}) - \beta)' \mathbf{X}' \mathbf{X} (\mathbf{b}(\mathbf{Y}) - \beta) / r^2(\mathbf{Z}) \leq (n-m) F_\alpha(m, n-m)\}. \quad (14)$$

Other terminal inferences such as testing values of β and simultaneous or separate confidence intervals for components of β can be derived as discussed in §3. As for σ , we cannot say much unless we are given a particular member in the family \mathbb{P} since the conditional distribution of $r(\mathbf{Z})$ given $\mathbf{u}(\mathbf{Z})$ and the unconditional distribution both depend on the distribution of the radial component of (1). In fact, the conditional and the unconditional inferences for σ may or may not be identical, depending on the particular spherical distribution. The distribution for $\mathbf{u}(\mathbf{Z})$, however, remains the same for all spherical distributions. Note that $\mathbf{u}(\mathbf{Z})$ corresponds to the set of normed residuals (i.e. residuals divided by the square root of their sum of squares) of \mathbf{Y} in the traditional approach of fitting the linear model by least-squares. Therefore any ‘test of normality’ which is a function of $\mathbf{u}(\mathbf{Y})$ is actually a test for the spherical symmetry of the error term. It seems that this fact is not

well known. Of course, if 'normality' means the validity of the inference on the parameter of primary interest, β , then a 'test of normality' still makes sense in view of the above discussion. The situation when \mathbb{P} is a parametric family admitting p.d.f. has been addressed in Fraser and Ng (1980).

If the family \mathbb{P} consists of α -symmetric distributions having $\alpha = 1$, the analysis in §3 implies that we may assume the components of \mathbf{Z} being i.i.d. standard Cauchy in making inference on β . The case of i.i.d. Cauchy errors does not allow analytic presentation for the conditional distribution of $t(\mathbf{Z})$ given $\mathbf{u}(\mathbf{Z})$ or the unconditional distribution. Numerical integration shall be needed to find tolerance intervals for components of $t(\mathbf{Z})$ in order to construct confidence intervals for the components of β .

The Liouville distribution is not often used as an error distribution in a linear model. But if a Liouville distribution with parameter $(\alpha_1, \dots, \alpha_n)$ is assumed for \mathbf{Z} , the conditional and unconditional inference for will be the same as that assuming \mathbf{Z} consists of independent gamma($\alpha_i, 1$) components, $i = 1, \dots, n$. One such application is in the location-scale model

$$\mathbf{Y} = \mu \mathbf{1} + \sigma \mathbf{Z}, \quad (15)$$

where $\mathbf{1}' = (1, \dots, 1)$ and \mathbf{Z} has the multivariate ℓ_1 -norm symmetric distribution. In this case, $\mathbf{X} = \mathbf{1}$, β reduces to μ , $\mathbf{b}(\mathbf{Z})$ becomes the sample mean \bar{Z} of Z_1, \dots, Z_n , $r(\mathbf{Z}) = \sqrt{\Sigma(Z_i - \bar{Z})^2}$, $u_i(\mathbf{Z}) = (Z_i - \bar{Z})/r(\mathbf{Z})$ and $t(\mathbf{Z}) = \bar{Z}/r(\mathbf{Z})$. The inference on μ is the same as when (Z_1, \dots, Z_n) are i.i.d. standard exponential. Therefore, the conditional p.d.f. of $t = \bar{Z}/r(\mathbf{Z})$ given $\mathbf{u}(\mathbf{Z}) = \mathbf{u}(\mathbf{Y}) \equiv \mathbf{u}$ is

$$f(t) = \begin{cases} (n-1)a^{n-1}/t^n, & a < t < \infty; \\ 0, & t \leq a, \end{cases} \quad (16)$$

where $a = -\min(u_1, \dots, u_n)$. Since $f(t)$ is decreasing and concave, the smallest interval with $(1 - \alpha)$ probability content is $a < t < a\alpha^{-1/(n-1)}$. Solving the double inequality after the substitution $t = (\bar{Y} - \mu)/r(\mathbf{Y})$, we obtain the smallest $100(1 - \alpha)\%$ confidence interval for μ :

$$Y_{(1)} \cdot \alpha^{-1/(n-1)} - \bar{Y}(\alpha^{-1/(n-1)} - 1) < \mu < Y_{(1)}, \quad (17)$$

where $Y_{(1)} = \min(Y_1, \dots, Y_n)$, \bar{Y} is the sample mean.

Acknowledgement. This research was supported in part by a Hong Kong UPGC research grant. The authors would like to thank the referees for valuable comments and pointing out some errors in the draft.

References

- CAMBANIS, S., KEENER, R., and SIMONS, G. (1983). On α -symmetric distributions, *J. Multivariate Analysis* **13**, 213–33.
- FANG, K. T., KOTZ, S. and NG, K. W. (1989). *Symmetric Multivariate and Related Distributions*, London and New York: Chapman and Hall.
- FANG, K. T., and FANG, B. Q. (1988). Some families of multivariate symmetric distributions related to exponential distribution, *J. Mult. Anal.* **24**, 109–122.
- FANG, K. T. and FANG, B. Q. (1989). A characterization of multivariate ℓ_1 -norm symmetric distributions, *Statistics and Probability Letters* **7**, 297–299.
- FRASER, D. A. S. (1968). *The Structure of Inference*, John Wiley & Sons, Inc.
- FRASER, D. A. S. (1979). *Inference and Linear Models*, McGraw-Hill Inc.
- FRASER, D. A. S., and NG, K. W. (1980). Multivariate regression: analysis with spherical error, in *Multivariate Analysis V* (ed. P.R. Krishnaiah), pp.369-86. New York: North-Holland.
- JOHNSON, N. L., and KOTZ, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*, New York: Wiley.
- NG, K. W. (1975). *Inference for Component Parameters in Multivariate Models*. University of Toronto Ph.D. thesis.
- ZOLOTAREV, V. M. (1985). *One-dimensional Stable Distributions*, Translations of Mathematical Monographs, Vol. 65. Providence, R.I.: American Mathematical Society.

DEPARTMENT OF STATISTICS
UNIVERSITY OF HONG KONG, HONG KONG

DEPARTMENT OF MATHEMATICS AND STATISTICS
YORK UNIVERSITY
NORTH YORK, CANADA

