Working Paper 92-34
July 1992

División de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe
Fax: (341) 624-9849

Robust Bayesian Inference in l_q-Spherical Models

Jacek Osiewalski and Mark F.J. Steel*

Abstract

The class of multivariate l_q-spherical distributions is introduced and defined through their isodensity surfaces.

We prove that, under a Jeffreys' type improper prior on the scale parameter, posterior inference on the location parameters is the same for all l_q-spherical sampling models with common q. This gives us perfect inference robustness with respect to any departures from the reference case of independent sampling from the exponential power distribution.

Key words:
Bayesian inference; Exponential power distributions; Inference robustness; l_q-norm; Symmetric multivariate distributions.

*Osiewalski, Department of Econometrics, Academy of Economics, Kraków, Poland; Steel, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid.

1.- Introduction

A new class of multivariate distributions, which we name l_q-spherical, is defined through properties of the density function. In particular, the isodensity surfaces are spheres in l_q-norm for $q \geq 1$. Robustness results obtained in Osiewalski and Steel (1992) for the case $q=2$, are here found to extend to general $q \in (0, \infty]$.

Fang, Kotz and Ng (1990) mention eight different ways of constructing symmetric multivariate distributions, one of which is symmetry of the density function. In their Chapter 5 they induce symmetry on the survival function with an l_1-norm. They start from an exponential distribution (defined on the positive real line), resulting in a density function proportional to that of our l_1-spherical distribution over the positive orthant of the sample space. In their Chapter 7 they use symmetry of the characteristic function to define α-symmetric distributions. For $\alpha=2$ this corresponds to our l_2-spherical distributions, usually referred to in the literature as spherical and discussed in detail in Kelker (1970), Cambanis, Huang and Simons (1981), Dickey and Chen (1985), and Fang, Kotz and Ng (1990, chapters 2-4).

In the same way as independent sampling from a univariate Normal distribution constitutes a reference case for the l_2-spherical family, independent sampling from exponential power distributions [see Box and Tiao (1973, chapter 3)] forms a useful reference class for the entire l_q-spherical family.

Within a Bayesian framework, we prove that under a commonly used diffuse prior on the scale parameter posterior inference on the location vector is fully robust with respect to departures of any l_q-spherical sampling density from its reference case.

2.- Defining l_q-Spherical Distributions

In this section we introduce the class of multivariate l_q-spherical distributions, where symmetry is imposed through the density function. This implies we only consider continuous distributions.

Let us first introduce the following notation, for $a = (a_1, \ldots, a_n)'$:

$$v_q(a) = \begin{cases} \left(\sum_{i=1}^{n} |a_i|^q\right)^{\frac{1}{q}} & \text{if } 0<q<\infty \\ \max_{i=1,\ldots,n} |a_i| & \text{if } q=\infty \;. \end{cases}$$

If we choose q in the range $[1, \infty]$ then $v_q = \| a \|_q$, the l_q-norm of the vector a. In the case $q \in (0,1)$ $v_q(\cdot)$ does not satisfy the triangle inequality. For our purposes, however, the latter is not required.

Definition: For any scalar $0<q\leq\infty$, $x= (x_1, \ldots, x_n)' \in \mathbb{R}^n$ has an l_q-spherical distribution with location $\mu = (\mu_1, \ldots, \mu_n)' \in \mathbb{R}^n$ and scale $\tau^{-1} \in \mathbb{R}_+$, denoted as $x \sim l_q^n(\mu, \tau^{-1} I_n)$, if its density function is given by

$$p(x|\mu,\tau) = \tau^n g_q\left[v_q\{\tau(x-\mu)\}\right], \tag{1}$$

and $g_q(\cdot)$ is a nonnegative function such that $p(x \mid \mu,\tau)$ is a proper density. ■

The isodensity surfaces follow immediately from the Definition as

$$\{x\in\mathbb{R}^n | v_q[\tau(x-\mu)]=\alpha\}, \tag{2}$$

where $\alpha>0$. For $q\geq1$ they could be considered spheres with respect to l_q-norm centered at μ. This fact motivates calling these densities "l_q-spherical".

Let us now consider the conditions on $g_q(\cdot)$ imposed by our Definition. For finite q, we define $r= v_q\{\tau(x-\mu)\}$, and $z_i=\left(\tau\frac{|x_i-\mu_i|}{r}\right)^q$, $i= 1, \ldots, n$. Transforming x into $(z_1,\ldots,z_{n-1}, r)$, we derive from $p(x \mid \mu,\tau)$ a product of an n-variate Dirichlet density on $z= (z_1,\ldots,z_n)'$ with parameters q^{-1} and

$$p(r|z,\mu,\tau)=p(r)=\frac{\left[2\Gamma\left(1+\frac{1}{q}\right)\right]^n}{\Gamma\left(\frac{n}{q}\right)}\ q\ r^{n-1}g_q(r), \tag{3}$$

which is a proper density function over $\mathbf{R}_+$ if and only if

$$\int_0^\infty u^{n-1}g_q(u)\ du=\frac{\Gamma\left(\frac{n}{q}\right)}{q\left[2\Gamma\left(1+\frac{1}{q}\right)\right]^n}\equiv c_q\ . \tag{4}$$

Extending the analysis in Dickey and Chen (1985), we can represent x in terms of three independent random quantities

$$\tau(x-\mu)=r\ (s\ \times\ z^{\frac{1}{q}}),$$

where $z^{1/q}$ denotes a coordinatewise power, $s \times z^{1/q}$ is a coordinatewise product of vectors and the n elements of s independently take the value 1 or -1 with probability 1/2.

This allows us to derive the following moments for $m_i = 0, 1, 2, \ldots$:

$$E\left[\prod_{i=1}^n\{\tau(x_i-\mu_i)\}^{m_i}\right]=\begin{cases}\dfrac{\Gamma\left(\frac{n}{q}\right)}{\left[\Gamma\left(\frac{1}{q}\right)\right]^n}\dfrac{\Pi\,\Gamma\left(\frac{m_i+1}{q}\right)}{\Gamma\left(\sum\frac{m_i+1}{q}\right)}E\left(r^{\sum m_i}\right) & \\ \qquad\qquad if\ \ all\ \ m_i\ \ are\ even, & \\ 0 \qquad\qquad otherwise, & \end{cases} \tag{5}$$

which exist if and only if $E(r^{\sum m_i})$ is finite. In particular, if $E(r^2) < \infty$,

$$E(x|\mu,\tau)=\mu \tag{6}$$

$$var(x|\mu,\tau)=\frac{\Gamma\left(\frac{n}{q}\right)\Gamma\left(\frac{3}{q}\right)}{\Gamma\left(\frac{1}{q}\right)\Gamma\left(\frac{n+2}{q}\right)}\tau^{-2}E(r^2)I_n. \tag{7}$$

The case of infinite q requires a separate treatment. Now we define

$$r=v_\infty\{\tau(x-\mu)\} \quad and \quad z_i=\frac{w_i-w_{i-1}}{r},$$

$i= 1, .., n$ where $w_0 = 0$ and w_i $(i\geq 1)$ is the ith ordered value of $\tau \mid x_j - \mu_j \mid$, $j=1,\ldots,n$. The transformation from x to $(z_1,\ldots, z_{n-1}, r)$ gives us a Dirichlet density for z with parameters 1 and

$$p(r \mid z,\mu,\tau) = p(r) = n\, 2^n r^{n-1} g_\infty(r), \tag{8}$$

from which the condition on $g_\infty(\cdot)$ implicit in the Definition becomes

$$\int_0^\infty u^{n-1}g_\infty(u)\, du= \frac{1}{n}\, 2^{-n}\equiv c_\infty . \tag{9}$$

Since w_i can now be represented as

$$w_i=r\sum_{j=1}^{i} z_j$$

and $\sum_{j=1}^{i} z_j$ possesses a Beta distribution with parameters i and n-i, $E(w_i^m)$ exists if and only if $E(r^m)$ is finite when $i<n$. Trivially, $E(w_n^m) = E(r^m)$ as $w_n = r$.

Using the exchangeability property of x, we know that $E[\tau(x_i-\mu_i)]^m=\frac{1}{n}\sum_{j=1}^{n} E(w_j)^m$ for even $m>0$, from which we derive

$$E[\tau(x_i-\mu_i)]^m = \begin{cases} \dfrac{m+n}{n(m+1)} E(r^m) & \text{if } m \text{ is even,} \\ 0 & \text{if } m \text{ is odd .} \end{cases} \quad (10)$$

This immediately leads to

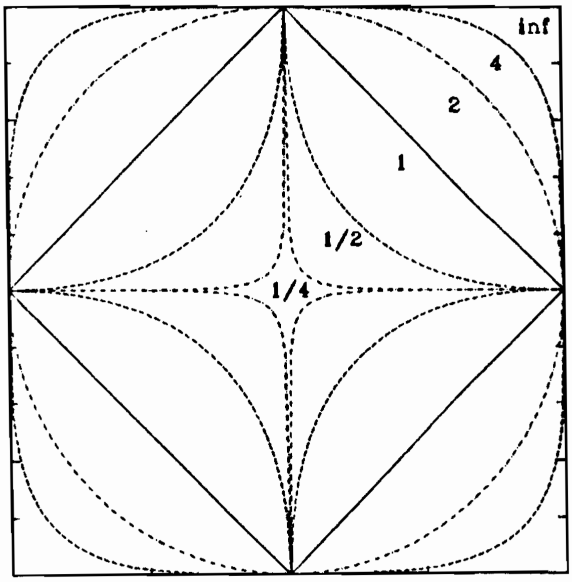
$$E\,(x|\mu,\tau)=\mu \quad (11)$$

$$var(x|\mu,\tau)=\frac{n+2}{3n}\tau^{-2}E(r^2)I_n \quad (12)$$

for finite $E(r^2)$.

We remark that as $q \to \infty$ the right-hand side of (4) converges to that of (9), and the constant in (5) for one nonzero $m_i = m$ converges to the constant in (10). Clearly, such convergence also applies to the isodensity surfaces in (2), which are graphically displayed in Figure 1 for n=2 and selected values of q.

Fig.1. Isodensity lines for n=2.



(values for q in the graph)

For given α and any n, the isodensity surfaces for all $q \in (0, \infty]$ coincide when $x = \mu \pm e$ where e has only one nonzero element (namely $\alpha\tau^{-1}$), and for other values of x the isodensity surfaces move further from μ as q grows.

3.- Exponential Power Distributions

An interesting special class of l_q-spherical distributions is generated by independent sampling from exponential power distributions [Box and Tiao (1973, chapter 3)]:

$$p(x_i|\mu,\tau)=d_q\tau\ \exp\left[-\frac{\tau^q}{2}|x_i-\mu_i|^q\right], \tag{13}$$

with

$$d_q=\left[2^{1+\frac{1}{q}}\ \Gamma(1+\frac{1}{q})\right]^{-1},$$

for $0<q<\infty$, *and from*

$$p(x_i|\mu,\tau)=\frac{\tau}{2}\ I\ (\mu_i-\tau^{-1}<x_i<\mu_i+\tau^{-1}) \tag{14}$$

for $q=\infty$, where $I(\cdot)$ denotes the indicator function. Thus, sampling from (13) and (14) corresponds to particular choices of $g_q(\cdot)$ in the Definition. For $0<q<\infty$ and $u\in\mathbb{R}_+$, we choose $g_q(u)=d_q^n\ \exp\left(-\frac{1}{2}u^q\right)$, from which, using (3), we obtain

$$p(r)\ =2^{-\frac{n}{q}}\left[\Gamma\left(\frac{n}{q}\right)\right]^{-1}q\ r^{n-1}\ \exp-\frac{1}{2}\ r^q,$$

i.e. the q^{th} power of the l_q-radius is Gamma distributed with parameters n/q and 1/2. This implies that

$$E\ (r^m)=2^{\frac{m}{q}}\ \frac{\Gamma\left(\frac{m+n}{q}\right)}{\Gamma\left(\frac{n}{q}\right)},$$

which can be inserted into (5), and indicates that all moments of r exist for $m>-n$, and, thus, all positive moments of x in (5) are finite. For infinite q, we implicitly choose

$$g_\infty\ (u)=2^{-n}\ I\ (0<u<1),$$

leading to a Beta distribution for r with parameters n and 1. Therefore, existence of moments is the same as in the case of finite q.

Due to the independence, the n-variate distribution considered in this Section forms a convenient reference case within the class of l_q-spherical distributions. For $q=1$ we obtain the case of independent sampling from a double-exponential or Laplace distribution with variance $8\tau^{-2}$, for $q=2$ it corresponds to sampling from a Normal with variance τ^{-2}, and for $q=\infty$ we sample from a uniform distribution with variance $\tau^{-2}/3$.

Box and Tiao (1973, chapter 3) consider exponential power distributions for $q \geq 1$, and parameterize instead in terms of the sampling variance, which is given by

$$var\ (x_i|\mu,\tau) = 2^{\frac{2}{q}} \frac{\Gamma\left(\frac{3}{q}\right)}{\Gamma\left(\frac{1}{q}\right)} \tau^{-2}$$

for $0<q<\infty$, and its limit for the case $q = \infty$.

4.- Posterior Inference

We now focus on conducting inference on the location vector μ when the sampling distribution is l_q-spherical. In many practical cases, μ will be parameterized in terms of a lower-dimensional parameter. However, such a regression context will not explicitly be considered here.

Theorem: For any $q \in (0,\infty]$, if $x \sim l_q^n(\mu, \tau^{-1} I_n)$ and we assume the improper prior structure

$$p(\mu,\tau) = p(\mu)\ p(\tau) = k\tau^{-1}\ p(\mu), \tag{15}$$

$k>0$, then

$$p(x,\mu) = k\ p(\mu)\ c_q\ [v_q(x-\mu)]^{-n}. \tag{16}$$

Proof: We need to integrate out τ from the joint density $p(x,\mu,\tau)$, which is the product of (1) and (15). Using the one-to-one transformation from (x,μ,τ) to (x,μ,r) where $r= v_q[\tau(x-\mu)]=$ $= \tau\ v_q(x-\mu)$ and integrating out r through (4) for finite and (9) for infinite q, the result follows. ■

The prior structure in (15) is the product of an improper Jeffreys' type prior on τ, which is widely used to represent a lack of prior information about the scale, and any prior on μ, either proper or improper. Combining (15) with any $l_q^n(\mu,\tau^{-1}I_n)$ sampling density on the observables x results in the improper marginal density (16), which leads to a proper posterior density $p(\mu \mid x)$ provided $p(x)=\int_{\Re^n} p(x,\mu)\, d\mu < \infty$. If part of x is not actually observed (e.g. missing data or forecasting), then the out-of-sample predictive density can easily be obtained from p(x).

The main implication of the Theorem can be deduced from the functional form of (16), where $g_q(\cdot)$ no longer appears. In other words, for a given q any choice of $g_q(\cdot)$ which leads to a proper sampling density in (1) will produce the same inference on μ and the unobserved part of x. Therefore, the prior in (15) is sufficient for complete robustness of posterior (on μ) and predictive inference within the entire class of l_q-spherical sampling distributions with q fixed. As stressed in Section 3, the case obtained by independent sampling from an exponential power distribution is a particularly convenient member of the l_q-spherical class. Our Theorem thus tells us that, assuming (15), the posterior results of Box and Tiao (1973, chapter 3), obtained under independent sampling from (13) or (14), remain unaffected by any departure within the class of multivariate l_q-spherical sampling densities with the same q.

Such robustness does not hold for the scale parameter τ; in fact, the entire influence of the choice of labelling function $g_q(\cdot)$ is summarized in the conditional posterior of τ

$$p(\tau \mid x,\mu) = c_q^{-1}\{v_q(x-\mu)\}^n \tau^{n-1} g_q[v_q\{\tau(x-\mu)\}], \qquad (17)$$

which is proper by (4) and (9). Clearly, τ given μ is updated by the sample information and its posterior does not preserve the functional form in (15). Thus, the Theorem does not hold for more than one independent vector observation from $l_q^n(\mu,\tau^{-1}I_n)$, and achieving robustness seems a hopeless task in this case.

5.- Extensions

We can easily generalize the Theorem in the previous section to cases where we sample from $y = Ax$ with $x \sim l_q^n(\mu, \tau^{-1}I_n)$ and A is a nonsingular matrix function of parameters η. Under the prior structure $p(\mu,\eta,\tau) = k\,\tau^{-1}p(\mu,\eta)$, $k>0$, we obtain

$$p(y,\mu,\eta) = k\ p(\mu,\eta)\ c_q\ \ |\det(A)|^{-1}\ [v_q(A^{-1}y-\mu)]^{-n},$$

which extends the robustness results to the posterior of η. For $q=2$ the sampling distribution on y is called elliptical or ellipsoidal, a case which was treated in Osiewalski and Steel (1992) within a regression context.

Finally, our Theorem formally suggests a possible extension to a wider class of sampling densities $p(x \mid \mu,\tau) = \tau^n g\,[v\,\{\tau(x-\mu)\}]$, where we require $v(\cdot) \geq 0$, $v(\alpha a) = \alpha v(a)$ for any positive scalar α, and we choose $g(\cdot)$ in the class G of nonnegative functions such that $\int_0^\infty u^{n-1}g(u)\,du = c < \infty$ and $p(x \mid \mu,\tau)$ is proper (in the case of l_q-sphericity, the last two conditions coincide). Then, the prior in (15) will result in (16) with c and $v(\cdot)$ instead of c_q and $v_q(\cdot)$. If, in addition, c is the same for any member of some $G_* \subseteq G$, then robustness for inference on the location μ (and on unobserved elements of x) holds within the entire class G_*.

Acknowledgements: We benefitted from useful discussions with Carmen Fernández, Jean-Pierre Florens and Michel Mouchart. Of course, the usual disclaimer applies. The second author gratefully acknowledges the hospitality of the Academy of Economics, Kraków.

References

Box, G.E.P. & Tiao, G.C. (1973). Bayesian Inference in Statistical Analysis. Reading, Mass.: Addison-Wesley.

Cambanis, S., Huang, S. & Simons, G. (1981). On the theory of elliptically contoured distributions. J. Multivar. Anal. 11, 368-85.

Dickey, J.M. & Chen, C.H. (1985). Direct subjective-probability modelling using ellipsoidal distributions. In Bayesian Statistics 2, Ed. J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, pp. 157-82, Amsterdam: North-Holland.

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

Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale generalization. Sankhyā A 32, 419-30.

Osiewalski, J. & Steel, M.F.J., (1992). Robust Bayesian inference in elliptical regression models. J. Econometrics, forthcoming.