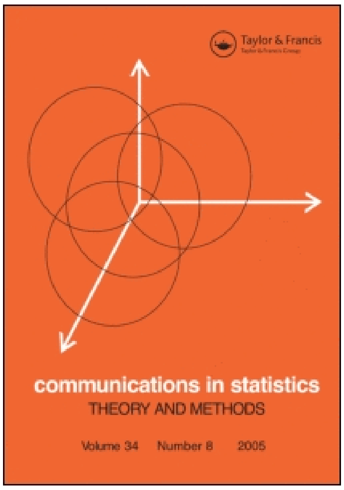

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597238

A Bayesian Test for the Mean of the Power Exponential Distribution

Miguel A. Gómez-Villegas [a]; Javier Portela [b]; Luis Sanz [a]
[a] Departamento de Estadística e Investigación Operativa I, Universidad Complutense de Madrid, Madrid, Spain [b] Departamento de Estadística e Investigación Operativa III, Universidad Complutense de Madrid, Madrid, Spain

Online Publication Date: 01 January 2008

To cite this Article Gómez-Villegas, Miguel A., Portela, Javier and Sanz, Luis(2008)'A Bayesian Test for the Mean of the Power Exponential Distribution',Communications in Statistics - Theory and Methods,37:18,2865 — 2876

To link to this Article: DOI: 10.1080/03610920802162698

URL: http://dx.doi.org/10.1080/03610920802162698

Communications in Statistics—Theory and Methods, 37: 2865–2876, 2008

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920802162698

Taylor & Francis
Taylor & Francis Group

Inference

A Bayesian Test for the Mean of the Power Exponential Distribution

MIGUEL A. GÓMEZ–VILLEGAS[1], JAVIER PORTELA[2], AND LUIS SANZ[1]

[1]Departamento de Estadística e Investigación Operativa I, Universidad Complutense de Madrid, Madrid, Spain
[2]Departamento de Estadística e Investigación Operativa III, Universidad Complutense de Madrid, Madrid, Spain

In this article, we deal with the problem of testing a point null hypothesis for the mean of a multivariate power exponential distribution. We study the conditions under which Bayesian and frequentist approaches can match. In this comparison it is observed that the tails of the model are the key to explain the reconciliability or irreconciliability between the two approaches.

Keywords Mixed prior distributions; Multivariate point null hypothesis; Posterior probability; Power exponential distribution; p-value; Robust Bayesian analysis.

Mathematics Subject Classification 62F15; 62F03.

1. Introduction

1.1. *Multivariate Power Exponential Distribution*

Introduced by Gómez et al. (1998), the multidimensional power exponential distribution is useful to modelize a lot of classes of random phenomena, including those that can be modelized with a normal distribution, but mainly, those phenomena whose distributions have higher or lower tails than the normal distribution. Besides, it can be used to robustify statistical procedures.

Recent examples of the power exponential distribution uses are applications to repeated measurements (see Lindsey, 1999), and an application to obtain robust models for repeated measurements in order to model dependencies among responses

Received July 13, 2006; Accepted March 3, 2008
Address correspondence to Miguel A. Gómez–Villegas, Departamento de Estadística e Investigación Operativa I, Facultad de Ciencias Matemáticas, Universidad Complutense de Madrid, Madrid 28040, Spain; E-mail: ma.gv@mat.ucm.es

(see Lindsey and Lindsey, 2006). It has also been applied in Bayesian networks, as an alternative to the mixture of normal distributions. Some applications in the field of speech recognition and image processing are given in Basu et al. (2001). The use of the multivariate power exponential distribution in linear dynamic models can be seen in Gómez et al. (2002).

As it is defined in Gómez et al. (1998), a p-variate random vector $\boldsymbol{X} = (X_1, \ldots, X_p)'$, $p \geq 1$, is distributed as a p-dimensional power exponential distribution with parameters $\boldsymbol{\mu}$, Σ, and β, being $\boldsymbol{\mu} \in R^n$, Σ a $(p \times p)$ definite positive symmetric matrix and $\beta \in (0, \infty)$, if its density function is

$$f(\boldsymbol{x} \mid \boldsymbol{\mu}, \Sigma, \beta) = k|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}[(\boldsymbol{x}-\boldsymbol{\mu})^t \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})]^\beta\right\}, \tag{1.1}$$

with $k = \frac{p\Gamma(p/2)}{\pi^{p/2}\Gamma[1+p/(2\beta)]2^{1+p/(2\beta)}}$.

We denote this distribution as $PE_p(\boldsymbol{\mu}, \Sigma, \beta)$, where $\boldsymbol{\mu}$ is the location parameter, Σ is the scale matrix, and β is a parameter related to kurtosis that shows the disparity with the normal distribution.

The density in (1.1) represents, when $\beta = 1$, the density of a multivariate normal distribution. The sharpness of the density diminishes as β increases. If $\beta = 1/2$, (1.1) is a multivariate generalization of the double exponential distribution. Besides, when β tends to infinite, (1.1) tends to a multivariate generalization of the uniform distribution.

Finally, it can be observed that (1.1) is the density of an elliptically contoured random vector; see Fang and Zhang (1990).

1.2. *The Problem: Testing Point Null Hypothesis for the Location Parameter*

For the location parameter $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ of a p-dimensional multivariate power exponential distribution, with density (1.1), a point null hypothesis test, without lost of generality, can be represented as:

$$H_0 : \mu_1 = \cdots = \mu_p = 0 \quad \text{versus} \quad H_1 : \text{ at least one } \mu_i \neq 0 \quad \text{for } i = 1, \ldots, p. \tag{1.2}$$

There are many approaches for the univariate two-sided hypothesis test, both in frequentist and Bayesian tests, but not for the multivariate two-sided one. Some exceptions are Oh (1998), who dealt with the multivariate normal distribution, Oh and DasGupta (1999), who explored the relevance of π_0, the prior probability of the sharp null hypothesis, in the difference between the infimum of the posterior probability and the p-value for some classes of priors on the alternative hypothesis, and Gómez–Villegas et al. (2008) who compared the p-value and the posterior probability for some classes of prior distributions when testing for the mean of a multivariate normal distribution.

We assume that Σ and β are known and let us suppose that our prior opinion about $\boldsymbol{\mu}$ is given by the density $\pi(\boldsymbol{\mu})$. Then, from a Bayesian point of view, the prior to test (1.2) will be given by a mixed prior distribution, $\pi^*(\boldsymbol{\mu})$, assigning mass π_0 to H_0 and spreading the remainder, $1 - \pi_0$, over the alternative according to the density $\pi(\boldsymbol{\mu})$,

$$\pi^*(\boldsymbol{\mu}) = \pi_0 I_{\{\boldsymbol{\mu}=\mathbf{0}\}}(\boldsymbol{\mu}) + (1-\pi_0)\pi(\boldsymbol{\mu}) I_{\{\boldsymbol{\mu}\neq\mathbf{0}\}}(\boldsymbol{\mu}). \tag{1.3}$$

What we propose is to use $\pi(\boldsymbol{\mu})$, our prior opinion about $\boldsymbol{\mu}$, and compute π_0 by means of:

$$\pi_0 = \int_{E(\boldsymbol{\mu}=\mathbf{0},\varepsilon)} \pi(\boldsymbol{\mu})d\boldsymbol{\mu}, \tag{1.4}$$

being $E(\boldsymbol{\mu}=\mathbf{0},\varepsilon)$ an ellipsoid centered at $\mathbf{0}$. Any other metric around $\boldsymbol{\mu}=\mathbf{0}$ can be used in (1.4), we use this because of its computational tractability and intuitive appeal.

Although the usual value taken for π_0 is 0.5, several reasons can justify the choice of π_0 as in (1.4). Firstly, in the univariate case, when using (1.3) and (1.4) with suitable small values of ε – in case of normal likelihood $\varepsilon \in (0.1, 0.3)$ – and $\pi(\boldsymbol{\mu})$ in the class of all unimodal and symmetric distributions or in the class of ε-contaminated distributions, a suitable approximation between the posterior probability and the p-value is obtained. These results can be seen in Gómez–Villegas and Gómez (1992), Gómez–Villegas and Sanz (1998, 2000), and Gómez et al. (2002), Gómez–Villegas et al. (2004).

The second reason to use π_0 as in (1.4) is that if $\pi(\boldsymbol{\mu})$ reflects our prior opinion about $\boldsymbol{\mu}$, then the prior probability of $\boldsymbol{\mu}=\mathbf{0}$ is zero, but if we use (1.3), the prior mass assigned to $\boldsymbol{\mu}=\mathbf{0}$ is π_0 and this probability is obtained through $\pi(\boldsymbol{\mu})$.

The third reason arises because, if $\pi(\boldsymbol{\mu})$ reflects our opinion about $\boldsymbol{\mu}$ and $\pi^*(\boldsymbol{\mu})$, given by (1.3), is the solution that we propose to test (1.2), it seems natural that both $\pi(\boldsymbol{\mu})$ and $\pi^*(\boldsymbol{\mu})$ must satisfy

$$\lim_{\varepsilon\to 0} \delta(\pi^* \mid \pi) = 0 \tag{1.5}$$

for some suitable measure of discrepancy, δ. One of the most popular measures of discrepancy is:

$$\delta(\pi^* \mid \pi) = \int \pi(\boldsymbol{\mu}) \ln \frac{\pi(\boldsymbol{\mu})}{\pi^*(\boldsymbol{\mu})} d\boldsymbol{\mu}; \tag{1.6}$$

see, by example, Bernardo and Smith (1994, p. 76). With our approximation, we have:

$$\begin{aligned}
\delta(\pi^* \mid \pi) &= \int \pi(\boldsymbol{\mu}) \ln \left[\frac{\pi_0}{\pi(\boldsymbol{\mu})} I_{\boldsymbol{\mu}=\mathbf{0}}(\boldsymbol{\mu}) + (1-\pi_0) I_{\boldsymbol{\mu}\neq\mathbf{0}}(\boldsymbol{\mu})\right]^{-1} d\boldsymbol{\mu} \\
&= -\int \pi(\boldsymbol{\mu}) \ln \left[\frac{\pi_0}{\pi(\boldsymbol{\mu})} I_{\boldsymbol{\mu}=\mathbf{0}}(\boldsymbol{\mu}) + (1-\pi_0) I_{\boldsymbol{\mu}\neq\mathbf{0}}(\boldsymbol{\mu})\right] d\boldsymbol{\mu} \\
&= -\int_{\boldsymbol{\mu}\neq\mathbf{0}} \pi(\boldsymbol{\mu}) \ln(1-\pi_0) d\boldsymbol{\mu} \\
&= -\ln(1-\pi_0).
\end{aligned} \tag{1.7}$$

We think this is a desirable property. Usually in the literature, at least in the univariate case, the expression (1.3) is used with $\pi_0 = 0.5$. However, for $\pi_0 = 0.5$, (1.7) gives $\delta(\pi^* \mid \pi) = 0.693$ that seems a high discrepancy between these two distributions, π^* and π. Whereas with our approximation, the result (1.5) is verified, because if ε goes to zero then π_0 goes to zero too.

The three reasons above are enough, given a prior density $\pi(\boldsymbol{\mu})$, to justify the choice of $\pi^*(\boldsymbol{\mu})$ as in (1.3) with π_0 as in (1.4) for the problem of testing a multivariate point null hypothesis. Anyhow, in this article the results are obtained as a function of π_0 and then can be specified for every π_0 as in (1.4). In particular, it is possible to compute the value of ε that provides $\pi_0 = 0.5$.

In Sec. 2, bounds on posterior probabilities for the class of elliptical distributions are computed and compared with the p-value of the frequentist approach, for the $n = 1$ case. Section 3 presents a frequentist test for the mean and Monte Carlo methods to approach the general case. Section 4 shows two interesting applications. Finally, Sec. 5 develops some final comments and conclusions.

2. Lower Bounds on Posterior Probabilities

In order to make comparisons between the p-value and the posterior probabilities, we will take wide classes of prior distributions and then we compute the infimum of the posterior probabilities over these classes. This is the usual procedure to compare Bayesian and frequentist approaches, because a frequentist should behave like a Bayesian using a large class of priors.

Because of the structure of the problem, it looks reasonable to deal with the class $\Gamma_{EU}(\boldsymbol{\mu}^0, \Sigma^0)$, class of distributions on R^p having probability density functions of the type

$$\pi(\boldsymbol{\mu}) = \psi\big((\boldsymbol{\mu} - \boldsymbol{\mu}^0)'(\Sigma^0)^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}^0)\big)$$

with $\psi(\cdot)$ a decreasing function on $[0, \infty)$, $\boldsymbol{\mu}^0 \in R^p$, and Σ^0 a $(p \times p)$ positive definite matrix. This distributions are called elliptical and are unimodal in the sense of Anderson (1955) and have ellipsoidal contours centered at $\boldsymbol{\mu}^0$ with scale matrix Σ^0. In particular, $\Gamma_{EU}(\boldsymbol{\mu}^0, \Sigma^0)$ contains the spherical distributions on R^p.

Furthermore, if the following additional regularity conditions are imposed: (i) $\psi(r^2) \to 0$ as $r \to \infty$; and (ii) $\psi(r^2)$ is of bounded variation in every finite interval away from the origin, then it can be shown (see Jensen and Good, 1983), that $\pi(\boldsymbol{\mu}) \in \Gamma_{EU}(\boldsymbol{\mu}^0, \Sigma^0)$ if and only if $\pi(\boldsymbol{\mu})$ is a mixture of uniform densities on ellipsoids centered at $\boldsymbol{\mu}^0$, $E(\boldsymbol{\mu}^0, k) = \{\boldsymbol{\mu} \mid (\boldsymbol{\mu} - \boldsymbol{\mu}^0)'(\Sigma^0)^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}^0) \leq k^2\}$.

Then, to find the infimum of the posterior probability of the point null hypothesis over the class $\Gamma_{EU}(\boldsymbol{\mu}^0, \Sigma^0)$, it is sufficient to find it over the much smaller class of the uniforms, see Casella and Berger (1987, Lemma 3.1). Without loss of generality, it can be supposed $\boldsymbol{\mu}^0 = \mathbf{0}$. Appropriate data translations can be used to reduce applications to this particular case.

To test (1.2), the posterior probability of H_0 is given by:

$$P(H_0 \mid \boldsymbol{x}) = \left(1 + \frac{1 - \pi_0}{\pi_0}\frac{1}{B}\right)^{-1}, \tag{2.1}$$

where π_0 is given by (1.4) and B is the Bayes factor in favour of H_0, given by:

$$B = \frac{f(\boldsymbol{x} \mid \boldsymbol{\mu} = \mathbf{0})}{\int_{E(\mathbf{0},k)} f(\boldsymbol{x} \mid \boldsymbol{\mu})\pi(\boldsymbol{\mu})d\boldsymbol{\mu}}$$

with $\pi(\boldsymbol{\mu}) = \frac{\Gamma(p/2+1)}{\pi^{p/2}k^p} I_{E(\boldsymbol{\mu}^0=\mathbf{0},k)}(\boldsymbol{\mu})$ a uniform density over $E(\boldsymbol{\mu}^0 = \mathbf{0}, k)$.

Then, as $\pi_0 = \int_{E(\boldsymbol{\mu}^0=\mathbf{0},\varepsilon)} \pi(\boldsymbol{\mu})d\boldsymbol{\mu} = \varepsilon^p/k^p$, the posterior probability of H_0 is:

$$P(H_0 \mid \boldsymbol{x}) = \left[1 + \left(\frac{1}{\varepsilon^p} - \frac{1}{k^p}\right)\frac{\Gamma(p/2+1)}{\pi^{p/2}} \int_{E(\mathbf{0},k)} \frac{f(\boldsymbol{x} \mid \boldsymbol{\mu})}{f(\boldsymbol{x} \mid \boldsymbol{\mu} = \mathbf{0})}\right]^{-1} \tag{2.2}$$

expression which is nonincreasing in k, so the infimum is obtained as k goes to infinity. Then:

$$\inf_{\pi \in \Gamma_{EU}} P(H_0 \mid \boldsymbol{x}) = \left[1 + \frac{\Gamma(p/2+1)}{\pi^{p/2}\varepsilon^p} \int_{R^p} \frac{f(\boldsymbol{x} \mid \boldsymbol{\mu})}{f(\boldsymbol{x} \mid \boldsymbol{\mu} = \mathbf{0})} d\boldsymbol{\mu}\right]^{-1}. \tag{2.3}$$

If we look for the values of ε so that this infimum of the posterior probability agrees with the p-value, we equal this expression to the p-value and solve for ε:

$$\varepsilon = \left[\frac{p(\boldsymbol{x})}{1-p(\boldsymbol{x})}\frac{\Gamma(p/2+1)}{\pi^{p/2}} \int_{R^p} \frac{f(\boldsymbol{x} \mid \boldsymbol{\mu})}{f(\boldsymbol{x} \mid \boldsymbol{\mu} = \mathbf{0})} d\boldsymbol{\mu}\right]^{1/p}, \tag{2.4}$$

where $p(\boldsymbol{x})$ is the p-value of the observed data.

To test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}^0$, being $\boldsymbol{Y} \sim PE_p(\boldsymbol{\mu}, \Sigma, \beta)$, with β and Σ known, we use the transformed vector $\boldsymbol{X}$ such that $\boldsymbol{X} = \Sigma^{-1/2}(\boldsymbol{Y} - \boldsymbol{\mu}^0)$.

We can derive the exact distribution of the test statistic defined as $T = (\boldsymbol{X}'\boldsymbol{X})^{\beta}$. As $\boldsymbol{Y}$ comes from an elliptical distribution, it can be shown, applying Theorem 2.5.5 from Fang and Zhang (1990), that $Q = \boldsymbol{X}'\boldsymbol{X}$ has, under H_0, the following density function:

$$f(q) = \frac{p}{\Gamma\left(1+\frac{p}{2\beta}\right)2^{1+\frac{1}{2\beta}}} q^{\frac{p}{2}-1} e^{\frac{1}{2}q^{\beta}}, \quad q > 0. \tag{2.5}$$

From (2.5), it follows that, under H_0, $T = Q^{\beta}$ is distributed as a Gamma variate $\Gamma(1/2, p/(2\beta))$. Then, a frequentist test for the point null hypothesis can be based on the p-value $p(t) = P(T > t)$, with T distributed as $\Gamma(1/2, p/(2\beta))$ and $t = (\boldsymbol{x}^t\boldsymbol{x})^{\beta}$.

In order to obtain values of ε that make equal the infimum of the posterior probability and the frequentist p-value, we use (2.4), where we have

$$\int_{R^p} \frac{f(\boldsymbol{x} \mid \boldsymbol{\mu})}{f(\boldsymbol{x} \mid \boldsymbol{\mu} = \mathbf{0})} d\boldsymbol{\mu} = e^{\frac{1}{2}t}\left[\Gamma\left(1+\frac{1}{2\beta}\right)2^{1+\frac{1}{2\beta}}\right]^p,$$

then the value of ε should be:

$$\varepsilon^* = \left[\frac{p(\boldsymbol{x})}{1-p(\boldsymbol{x})}\Gamma\left(\frac{p}{2}+1\right)e^{\frac{1}{2}t}\right]^{1/p} \frac{1}{\sqrt{\pi}}\Gamma\left(1+\frac{1}{2\beta}\right)2^{1+\frac{1}{2\beta}}, \tag{2.6}$$

being $p(\boldsymbol{x})$ the frequentist p-value. In the multivariate normal case ($\beta = 1$), the value of ε^* is given by (2.6) and the p-value is obtained through a χ^2_p test.

Table 1 shows the values of ε^* for some dimensions p, some significant p-values, and some values of β. It can be observed that the values of ε^* are increasing functions of the dimension p. This is a reasonable behaviour, because it means that as great the uncertainty over H_0 is, as great the radius of the ellipsoid must be to make equal both approximations, Bayesian and frequentist.

Table 1
Values of ε^* to obtain agreement for some β, p, and $p(\boldsymbol{x})$

	$\beta = 0.5$		$\beta = 1$		$\beta = 1.5$	
p-value	$p = 2$	$p = 5$	$p = 2$	$p = 5$	$p = 2$	$p = 5$
0.1	5.22	9.25	1.48	2.93	0.97	2.05
0.05	5.46	9.91	1.45	3.01	0.92	2.07
0.01	6.26	11.56	1.42	3.23	0.85	2.13
0.001	7.22	13.89	1.41	3.51	0.79	2.22

Figure 1 shows the infimum of the posterior probability of H_0, given by (2.3) for fixed values of ε based on the range observed in Table 1, and the p-value. It can be seen how for some reasonable range of values of ε, for each dimension p, and parameter β we obtain suitable approximations between frequentist and Bayesian approaches. Dashed line represents the p-value, while dotted lines represent the infimum of the posterior probability of H_0 over the class $\pi \in \Gamma_{EU}$ for the values of ε given below.

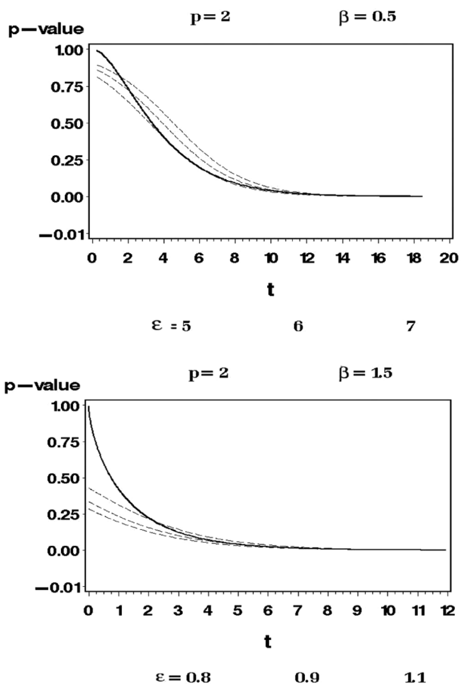

Figure 1. p-value and infimum of the posterior probability for some values of ε. Dashed line: p-value, dotted lines: infimum of the posterior probability of the point null hypothesis, H_0, infimum taken over the class of priors Γ_{EU}, from bottom to top for the values of ε pointed out in each case.

3. A Simulation Approach for the General Case

For samples with more than one observation ($n > 1$), (2.3) and (2.4) become, respectively,

$$\inf_{\pi\in\Gamma_{EU}} P(H_0 \mid \boldsymbol{x}) = \left[1 + \frac{\Gamma(p/2+1)}{\pi^{p/2}\varepsilon^p} \int_{R^p} \frac{f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \mid \boldsymbol{\mu})}{f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \mid \boldsymbol{\mu} = \boldsymbol{0})} d\boldsymbol{\mu}\right]^{-1} \tag{3.1}$$

and

$$\varepsilon = \left[\frac{p(\boldsymbol{x})}{1 - p(\boldsymbol{x})} \frac{\Gamma(p/2+1)}{\pi^{p/2}} \int_{R^p} \frac{f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \mid \boldsymbol{\mu})}{f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \mid \boldsymbol{\mu} = \boldsymbol{0})} d\boldsymbol{\mu}\right]^{1/p}. \tag{3.2}$$

Let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ be n independent observations from a multivariate Exponential Power distribution $PE_p(\boldsymbol{\mu}, \Sigma, \beta)$. Let $\boldsymbol{X}_i = \Sigma^{-1/2}(\boldsymbol{Y}_i - \boldsymbol{\mu}_0)$ be the components of the transformed vector. As $T_i = (\boldsymbol{X}_i^t\boldsymbol{X}_i)^\beta$ is distributed, under $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$, as a $\Gamma(1/2, p/(2\beta))$ variable, it follows that $T = \sum_{i=1}^n (\boldsymbol{X}_i^t\boldsymbol{X}_i)^\beta$ is distributed as a $\Gamma(1/2, (np)/(2\beta))$ variable, since it is the sum of n independent Gamma variables. The frequentist p-value for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ will be computed as $p(\boldsymbol{x}) = P\{T > t\}$, where T is distributed as $\Gamma(1/2, (np)/(2\beta))$ and t is the sample value of the statistic, $t = \sum_{i=1}^n (\boldsymbol{x}_i^t\boldsymbol{x}_i)^\beta$.

Now, for the multivariate Exponential Power distribution we must compute:

$$\int_{R^p} \frac{f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \mid \boldsymbol{\mu})}{f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \mid \boldsymbol{\mu} = \boldsymbol{0})} d\boldsymbol{\mu}$$
$$= \exp\left\{\frac{1}{2}\sum_{i=1}^n (\boldsymbol{x}_i^t\boldsymbol{x}_i)^\beta\right\} \int_{R^p} \exp\left\{-\frac{1}{2}\sum_{i=1}^n \left[(\boldsymbol{x}_i - \boldsymbol{\mu})^t(\boldsymbol{x}_i - \boldsymbol{\mu})\right]^\beta\right\} d\boldsymbol{\mu}. \tag{3.3}$$

For the particular case $\beta = 1$ (Normal case), it can be shown that (3.3) depends on the value of the well-known statistic $q = n\overline{\boldsymbol{X}}^t\overline{\boldsymbol{X}}$, so different samples that give the same value of q lead to the same value of ε. In this case, (2.4) becomes:

$$\varepsilon = \left[\frac{p(\boldsymbol{x})}{1 - p(\boldsymbol{x})} \left(\frac{2}{n}\right)^{p/2} \Gamma\left(\frac{p}{2} + 1\right) e^{q/2}\right]^{1/p},$$

where the p-value $p(\boldsymbol{x})$ is computed as $P\{\chi_p^2 > q)$. Result obtained by Gómez–Villegas et al. (2004).

For $\beta \neq 1$, the integral (3.3) can not be simplified, as it would be in the multivariate normal case. Furthermore, when $\beta \neq 1$, for each p-value we have infinite samples that lead to the same value of the statistic t associated to this p-value. Computing the integral (3.3) leads to different results for each of these samples and then for a different ε for each sample.

In order to understand the behavior of ε when $\beta \neq 1$, depending on the values of p, n, β, and t, we have developed the following procedure:

1. The p-value (and hence t), n, β, and p are fixed.
2. A large base of samples $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ such that $t = \sum_{1=1}^n (\boldsymbol{x}_i^t\boldsymbol{x}_i)^\beta$ for each of them is built. This is done by giving values in a large grid covering a big range of values to the p components of n vectors, $\boldsymbol{y}_i$, and then applying the transformation $\boldsymbol{x}_i = [t/\sum_{i=1}^n (\boldsymbol{y}_i^t\boldsymbol{y}_i)^\beta]^{1/(2\beta)}\boldsymbol{y}_i$. Since this application is surjective, the original values

$\boldsymbol{y}_i$ chosen in a large grid lead to a base of $\boldsymbol{x}_i$ that covers roughly the image space of $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$ such that $t = \sum_{i=1}^{n} (\boldsymbol{x}_i^t \boldsymbol{x}_i)^\beta$.

3. For each one of the samples, $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$, we compute (3.3). A Monte-Carlo approach can here be used. We choose the sample $\boldsymbol{x}_{\min}$, such that its Euclidean distance to the origin is the smallest, so that it guarantees that the $\boldsymbol{\mu}$ values are to be simulated approximately from the null hypothesis distribution, covering the most reasonable range of values. Then, the integral (3.3) can be written as:

$$\begin{aligned}
&\int_{R^p} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} [(\boldsymbol{x}_i - \boldsymbol{\mu})^t (\boldsymbol{x}_i - \boldsymbol{\mu})]^\beta \right\} d\boldsymbol{\mu} \\
&\quad = \frac{1}{k} \int_{R^p} \exp\left\{ -\frac{1}{2} \sum_{i=1, i\neq \min}^{n} [(\boldsymbol{x}_i - \boldsymbol{\mu})^t (\boldsymbol{x}_i - \boldsymbol{\mu})]^\beta \right\} k \\
&\quad\quad \times \exp\left\{ -\frac{1}{2} [(\boldsymbol{x}_{\min} - \boldsymbol{\mu})^t (\boldsymbol{x}_{\min} - \boldsymbol{\mu})]^\beta \right\} d\boldsymbol{\mu}. \qquad (3.4)
\end{aligned}$$

The second term in (3.4) is a multivariate exponential power distribution with mean $\boldsymbol{x}_{\min}$, where k is the integration constant. Samples are taken from this distribution (an algorithm to generate these values can be seen in Gómez et al., 1998). Then we obtain, by averaging, a Monte Carlo estimate of the integral.

4. Finally, for each one of the samples, we compute the mean or median of the values of ε obtained using (3.2).

Table 2 shows the results for different values of n, β, and p, fixing the p-value as $p(\boldsymbol{x}) = 0.10$. For $\beta = 1$, values of ε are exact for each n, p, and β; for $\beta \neq 1$, the presented values of ε are the median over the grid created in the procedure above.

In Table 2, we can see that for $\beta < 1$, the values of ε that make equal the infimum of the posterior probability of H_0 and the frequentist p-value are increasing (while not monotonous increasing) as the sample size increases, yielding high values and tending to infinite as n grows. This means that for a high kurtosis likelihood distribution (heavy tails), the prior mass on the point null hypothesis should be higher as n increases to make numerically equal the frequentist p-value and the posterior probability of H_0 for the point null testing problem.

Anyhow, we have presented just the median of the optimum value of ε over a grid of samples that lead to the same frequentist p-value. The range of values of ε in these samples for $\beta < 1$ is variable: it can happen that, given two samples leading to the same statistic t (and then leading to the same frequentist p-value), we need an $\varepsilon = 1.20$ in order to agree frequentist and Bayesian procedure in the first sample,

Table 2
Optimum values of ε for $p(\boldsymbol{x}) = 0.10$

	$\beta = 0.5$		$\beta = 1$		$\beta = 1.5$	
n	$p = 2$	$p = 5$	$p = 2$	$p = 5$	$p = 2$	$p = 5$
2	3.89	9.45	1.05	2.06	0.29	0.27
5	19.3	32.11	0.66	1.30	0.05	0.01
10	1602.1	4510	0.47	0.92	0.01	0.005

and $\varepsilon = 1,200$ in the second sample. This heavy tails effect has also been found in the univariate testing problem (see Gómez–Villegas and Sanz, 1998) and leads to an inefficient procedure if we wish to fix ε, before realizing the test, based only on the values of n, p, and β. This shows, once again, the difficulty of reaching an agreement for heavy tails distributions between frequentist and Bayesian testing.

Alternatively, for $\beta \geq 1$, ε tends to 0 as n tends to infinite. As β gets higher, this convergence seems to be faster. Then for likelihoods with low kurtosis and light tails, it seems that less prior mass on the null hypothesis point should be set to have similar results to the frequentist test, as n increases. Here we obtain an easier agreement between frequentist and Bayesian procedures within a reasonable range of ε.

4. Applications

In this section, we develop two applications in order to show the behavior of the infimum of the posterior probability of the point null hypothesis with respect to the frequentist p-value, in the multivariate context presented, through the analysis of the values of ε that lead to the agreement. In the first application high kurtosis data is used, while the second application is based on low kurtosis data. These are examples where the point null hypothesis testing problem has a really practical sense for the researcher.

4.1. *Response Time to Visual Stimulus*

Crowder and Hand (1990) analyzed the response to visuals flashes of the left and right eyes equipped with different lenses, for seven subjects. We pose the problem of determining if the vector of differences (y_1, y_2) = (left eye difference of response, right eye difference of response) between the two lenses has mean $\boldsymbol{\mu}^0 = (0, 0)$. This would mean that response time of reaction would be identical for both lenses.

We assume that the vectors of differences, $\boldsymbol{y}_i$, are distributed as independent $PE_p(\boldsymbol{\mu}, \Sigma, \beta)$. In an empirical Bayesian approach, Σ is approximated by the sample covariance matrix, Σ^*, and, as we want to test $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$, we pose $\boldsymbol{x}_i = (\sum^*)^{-1/2}\boldsymbol{y}_i$. The transformed vectors of differences are then $(-2.29, 0.11)$, $(-1.35, -0.10)$, $(-0.76, -0.16)$, $(-0.17, -0.22)$, $(-1.53, 0.27)$, $(0.79, 0.10)$, $(-1.02, -2.60)$. To choose β, the posterior distribution mode for β is estimated, letting a noninformative prior for β, and using (1.1). This leads to a maximization problem: it is needed to maximize the posterior distribution $p(\beta \mid \boldsymbol{x}_1, \ldots, \boldsymbol{x}_7)$, with

$$p(\beta \mid \boldsymbol{x}_1, \ldots, \boldsymbol{x}_7) \propto \left[\Gamma\left(1 + \frac{p}{2\beta}\right)2^{1+p/(2\beta)}\right]^{-1} \int_{R^p} \exp\left\{-\frac{1}{2}\sum_{i=1}^{7}[(\boldsymbol{x}_i - \boldsymbol{\mu})^t(\boldsymbol{x}_i - \boldsymbol{\mu})]^\beta\right\} d\boldsymbol{\mu}, \tag{4.1}$$

where the integral is estimated via Monte Carlo as explained in Sec. 3, and maximization is reached by means of a simple bisection algorithm. This process leads to an estimated posterior mode of $\hat{\beta} = 0.75$, as a consequence of data's high kurtosis. We will assume this $\hat{\beta}$ value fixed for the testing procedure.

In Table 3, we present the frequentist p-value, the tabulated value ε^*_{tab} that would have been used for these application, with the methods shown in Sec. 2 with $p = 2$, $n = 7$, $\beta = 0.75$, and the value of $\underline{P}(H_0 \mid \boldsymbol{x}, \varepsilon^*_{tab}) = \inf_{\pi \in \Gamma_{EU}} P(H_0 \mid \boldsymbol{x}, \varepsilon^*_{tab})$ obtained for this ε^*_{tab} using (2.6). Since setting a prior mass of $\pi_0 = 0.5$ is habitual

Table 3
P-value and infimum of posterior probabilities of H_0 over Γ_{EU}

ε^*_{sample}	ε^*_{tab}	$\underline{P}(H_0 \mid \boldsymbol{x}, \varepsilon^*_{sample}) = p\text{-}value$	$\underline{P}(H_0 \mid \boldsymbol{x}, \varepsilon^*_{tab})$	$\underline{P}(H_0 \mid \boldsymbol{x}, \pi_0 = 0.5)$
1.33	42	0.0348	0.30	0.24

in the Bayesian point null hypothesis testing problem, we also show the value of $\underline{P}(H_0 \mid \boldsymbol{x}, \pi_0 = 0.5) = \inf_{\pi \in \Gamma_{EU}} P(H_0 \mid \boldsymbol{x}, \pi_0 = 0.5)$ obtained with this setting, letting a prior distribution $\pi(\boldsymbol{\mu})$ uniform over a certain range of reasonable values $(-4, 4) \times (-4, 4)$, and computing (3.1).

With the automatic procedure, the value proposed for agreement, ε^*_{tab}, computed before the sample is drawn, gives a value of $\varepsilon^*_{tab} = 42$. Using (3.1), a value of $\underline{P}\{H_0 \mid \boldsymbol{x}, \varepsilon^*_{tab}\} = 0.30$ is obtained. As the p-value for this sample is $p(\boldsymbol{x}) = 0.0348$, frequentist and Bayesian procedures do not reach the agreement. The automatic procedure is inefficient in this case, because of high tails effect model. Also, the usual setting $\pi_0 = 0.5$ gives a considerably high value to the infimum of the posterior probability with respect to the p-value, due to the high prior mass given to the null hypothesis. Therefore, if the model is of high tails, namely $\beta < 1$, it is not possible, with our procedure, an agreement between Bayesian and frequentist approaches in testing a point null hypothesis. This behavior has been also observed in the univariate case; see Gómez–Villegas and Sanz (1998). By the way, it happens the same in the one-sided case; if the model has high tails; see Casella and Berger (1987).

4.2. *Archaeological Data*

In archaeological science, sometimes it is interesting to determine the geographic center of a settlement, in order to decide about the main excavation place. We can have observations that represent significative findings in an area, represented by their plane or polar coordinates (x, y) with respect the null hypothesis center, and we want to test whether the geographic center, that is the mean of the observations, is some $\boldsymbol{\mu}^0$.

Lizee and Plunkett (2002) used data collected in Farmington River, Connecticut, in order to illustrate sampling strategies in archaeology. In a limited area, there are 12 places where prehistoric artifacts were found. If we assume that findings follow an elliptical distribution and therefore are more likely to exist in places near the mean of the distribution, it is interesting to obtain an estimate of this center point. Furthermore, if we have some beliefs about where this center could be, based on qualitative information, we can determine by means of an hypothesis testing procedure how strongly the data assets this theoretical belief. In the example data a $\boldsymbol{y}_i - \boldsymbol{\mu}^0$ translation is used, where $\boldsymbol{\mu}^0$ is the belief center, so that a $H_0 : \boldsymbol{\mu} = (0, 0)$ test can be used over the transformed data.

As in the previous example, the transformation $\boldsymbol{x}_i = (\Sigma^*)^{-1/2}\mathbf{y}_i$ is applied, where Σ^* is the sample covariance matrix. Figure 2 shows the observations transformed set. These (x_1, x_2) points are $(-0.88, 1.57)$, $(0.13, 1.46)$, $(-0.26, 0.94)$, $(-1.07, 0.10)$, $(-0.11, 0.31)$, $(0.52, 0.96)$, $(-0.29, 0.38)$, $(-1.09, -1.10)$, $(-1.72, -1.11)$, $(1.42, -0.03)$, $(0.83, -0.87)$, $(1.34, -1.05)$. A spline-smoothed graphical estimate of the (x_1, x_2) distribution is also displayed. The test $H_0 : \boldsymbol{\mu} = (0, 0)$ is developed as in the previous application. The sample kurtosis coefficients for x_1 and x_2 are low: -0.90

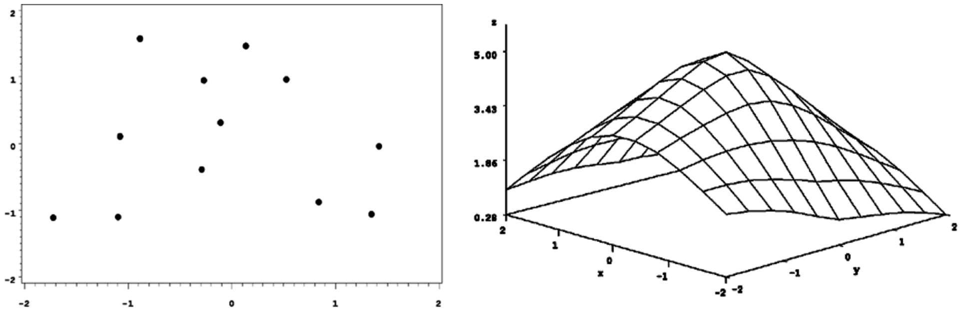

Figure 2. Archaeological findings in an area of interest. Smoothed graphical estimated of (x_1, x_2) distribution.

and −1.42, respectively. The β estimated posterior mode, using (4.1), is then higher than 1, about $\hat{\beta} = 1.25$.

Table 4 shows that the automatic procedure behaves reasonably well in this case, obtaining a value of $\underline{P}(H_0 \mid \boldsymbol{x}, \varepsilon^*_{tab}) = \inf_{\pi \in \Gamma_{EU}} P(H_0 \mid \boldsymbol{x}, \varepsilon^*_{tab})$ similar to the p-value. The prior mass $\pi_0 = 0.5$ increases the posterior probability giving extremely high posterior probability to the null hypothesis with respect to the p-value. That is to say, with our procedure, it is possible to reach agreement between frequentist and Bayesian approaches if the model is of low tails.

5. Conclusions

The multivariate power exponential model offers more possibilities than the normal model, basically when the amount of observations is not great. By the way, this family of distributions includes the normal distribution, this is the case if the parameter $\beta = 1$.

In testing a point null hypothesis concerning the mean, the procedure set in Sec. 1.2 can be applied to symmetric models around the mean, as it is the case of the multivariate power exponential distribution.

This procedure allows us to assign mass to the point null hypothesis through a prior density which is, on the other hand, the only source of information that we use. Furthermore, if the tails of the model distribution are low, that is to say $\beta \geq 1$, it is possible to reach agreement between the frequentist and Bayesian approaches. This is not the case if we put a $\pi_0 = 0.5$ mass over the point null hypothesis. If the tails of the model are high, that is to say $\beta < 1$, we cannot reach the agreement between the two approaches. More research must be carry out in this setting.

Table 4
P-value and infimum of posterior probabilities of H_0 over Γ_{EU}

ε^*_{sample}	ε^*_{tab}	$\underline{P}(H_0 \mid \boldsymbol{x}, \varepsilon^*_{sample}) = p(\boldsymbol{x})$	$\underline{P}(H_0 \mid \boldsymbol{x}, \varepsilon^*_{tab})$	$\underline{P}(H_0 \mid \boldsymbol{x}, \pi_0 = 0.5)$
0.084	0.062	0.0209	0.018	0.73

References

Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* 6:170–176.

Basu, S., Miccheli, C. A., Olsen, P. (2001). Power exponential densities for the training and classification of acoustic feature vectors in speech recognition. *J. Computat. Graphic. Statist.* 10(1):158–184.

Bernardo, J. M., Smith, A. F. M. (1994). *Bayesian Theory*. Chichester: Wiley.

Casella, G., Berger, R. L. (1987). Reconciling Bayesian and frequentist evidence in the one-sided testing problem, (with discussions). *J. Amer. Statist. Assoc.* 82:106–135.

Crowder, M. J., Hand, D. J. (1990). *Analysis of Repeated Measures*. London: Chapman and Hall.

Fang, K., Zhang, Y. (1990). *Generalized Multivariate Analysis*. Beijing: Springer-Verlag.

Gómez–Villegas, M. A., Gómez, E. (1992). Bayes factor in testing precise hypotheses. *Commun. Statist. Theor. Meth.* 21:1707–1715.

Gómez, E., Gómez–Villegas, M. A., Marín, J. M. (1998). A multivariate generalization of the power exponential family of distributions. *Commun. Statist. Theor. Meth.* 27:589–600.

Gómez, E., Gómez–Villegas, M. A., Marín, J. M. (2002). Continuous elliptical and exponential power linear dynamic models. *J. Multivariate Anal.* 83:22–36.

Gómez–Villegas, M. A., Maín, P., Sanz, L. (2002). A suitable Bayesian approach in testing point null hypothesis: some examples revisited. *Commun. Statist. Theor. Meth.* 31:201–217.

Gómez–Villegas, M. A., Maín, P., Navarro, H., Sanz, L. (2004). Asymptotic relationships between posterior probabilities and p-values using the hazard rate. *Statist. Probab. Lett.* 66:59–66.

Gómez–Villegas, M. A., Maín, P., Sanz, L. (2008). A Bayesian analysis for the multivariate point null testing problem. Submitted.

Gómez–Villegas, M. A., Sanz, L. (1998). Reconciling Bayesian and frequentist evidence in the point null testing problem. *Test* 7:207–216.

Gómez–Villegas, M. A., Sanz, L. (2000). ε-contaminated priors in testing point null hypothesis: a procedure to determine the prior probability. *Statist. Probab. Lett.* 47:53–60.

Jensen, D. R., Good, I. J. (1983). A representation for ellipsoidal distributions. *Siam J. Appl. Math.* 43:1194–1200.

Lindsey, J. K. (1999). Multivariate elliptically–contoured distributions for repeated measurements. *Biometrics* 55:1277–1280.

Lindsey, J. K., Lindsey, P. J. (2006). Multivariate distributions with correlation matrices for repeated measurements. *Computat. Statist. Data Anal.* 50:720–732.

Lizee, J., Plunkett, T. (2002). Archaeological sampling strategies. http://Archnet.asu.edu.

Oh, M. S. (1998). A Bayes test for simple versus one-sided hypothesis on the mean vector of a multivariate normal distribution. *Commun. Statist. Theor. Meth.* 27:2371–2389.

Oh, H. S., DasGupta, A. (1999). Comparison of the p-value and posterior probability. *J. Statist. Plann. Infer.* 76:93–107.