Gérard Letac; Hélène Massam

Gaussian approximation of Gaussian scale mixtures

Kybernetika, Vol. 56 (2020), No. 6, 1063–1080

Persistent URL: http://dml.cz/dmlcz/148499

KYBERNETIKA — VOLUME *56* (2020), NUMBER 6, PAGES 1063–1080

GAUSSIAN APPROXIMATION OF GAUSSIAN SCALE MIXTURES

Gérard Letac and Hélène Massam*

For a given positive random variable $V > 0$ and a given $Z \sim N(0,1)$ independent of V, we compute the scalar t_0 such that the distance in the $L^2(\mathbb{R})$ sense between $ZV^{1/2}$ and $Z\sqrt{t_0}$ is minimal. We also consider the same problem in several dimensions when V is a random positive definite matrix.

Keywords: mormal approximation, Gaussian scale mixture, Plancherel theorem

Classification: 62H17, 62H10

1. INTRODUCTION

Let $Z \sim N(0, I_n)$ be a standard Gaussian random variable in $\mathbb{R}^n$. Consider an independent random positive definite matrix V of order n with distribution μ. We call the distribution of $V^{1/2}Z$ a Gaussian scale mixture, where $V^{1/2}$ is the unique positive definite matrix such that $(V^{1/2})^2 = V$. Denote by f the density of $V^{1/2}Z$ in $\mathbb{R}^n$. In many practical circumstances, μ is not very well known, and f is complicated. On the other hand, for $n = 1$, and

$$f(x) = \int_0^\infty e^{-\frac{x^2}{2v}} \frac{\mu(\mathrm{d}v)}{\sqrt{2\pi v}} \tag{1}$$

we note that, as the logarithm of a Laplace transform, $\log f(\sqrt{x})$ is convex and thus the histogram of the symmetric density (1) looks like that of a normal distribution. The central aim of the present paper is to say something of the best normal approximation $N(0, t_0)$ of f in the sense of $L^2(\mathbb{R}^n)$.

In Section 2, we recall some known facts and examples about the pair (f, μ) when $n = 1$. In Section 3, our main result, for $n = 1$, is Theorem 3.1 in which we show the existence of t_0, its uniqueness and the fact that $t_0 < \mathbb{E}(V)$. This theorem also gives the equation, see (11), that has to be solved to obtain t_0 when μ is known. In Section 4 we consider the case $n \geq 2$ and investigate the fact that several distributions of the random positive definite matrix V can give the same Gaussian mixture $V^{1/2}Z$. In Section 5, we consider the problem of the Gaussian approximation of a Gaussian mixture in the more difficult case $n \geq 2$. In that case, t_0 is a positive definite matrix, and in Theorem 5.2,

DOI: 10.14736/kyb-2020-6-1063

*Hélène Massam passed away on August 22d, 2020.

we show the existence of t_0. Proposition 5.3 considers the particular case where V is concentrated on the multiples of I_n. A basic tool we use in this paper is the Plancherel identity.

2. THE UNIDIMENSIONAL CASE: A REVIEW

A probability density f on $\mathbb{R}$ is called a discrete Gaussian scale mixture if there exist numbers $0 < v_1 < \cdots < v_n$ and $p_1, \ldots, p_n > 0$ such that $p_1 + \cdots + p_n = 1$ and

$$f(x) = \sum_{i=1}^{n} p_i \frac{1}{\sqrt{2\pi v_i}} e^{-\frac{x^2}{2v_i}}.$$

It is easy to see that if $V \sim \sum_{i=1}^{n} p_i \delta_{v_i}$ is independent of $Z \sim N(0,1)$ then the density of $ZV^{1/2}$ is f. A way to see this is to observe that for all $s \in \mathbb{R}$ we have

$$\int_{-\infty}^{\infty} e^{sx} f(x)\,\mathrm{d}x = \sum_{i=1}^{n} p_i e^{\frac{s^2}{2} v_i} = \mathbb{E}(\mathbb{E}(e^{sZV^{1/2}}|V)) = \mathbb{E}(e^{sZV^{1/2}}).$$

More generally, we will say that the density f is a *Gaussian scale mixture* if there exists a probability distribution $\mu(\mathrm{d}v)$ on $(0,\infty)$ such that (1) holds. As in the finite mixture case, if $V \sim \mu$ is independent of $Z \sim N(0,1)$ the density of $ZV^{1/2}$ is f. To see this denote

$$L_V(u) = \int_0^{\infty} e^{-uv} \mu(\mathrm{d}v). \tag{2}$$

Then

$$\int_{-\infty}^{\infty} e^{sx} f(x)\,\mathrm{d}x = L_V(-s^2/2) = \mathbb{E}(e^{sZV^{1/2}}). \tag{3}$$

For instance if $a > 0$ and if

$$f(x) = \frac{a}{2} e^{-a|x|} \tag{4}$$

is the double exponential density, then for $|s| < a$ we have

$$\int_{-\infty}^{\infty} e^{sx} f(x)\,\mathrm{d}x = \frac{a^2}{a^2 - s^2} = L_V(-s^2/2)$$

where

$$L_V(u) = \frac{a^2}{a^2 + 2u} = \frac{a^2}{2} \int_0^{\infty} e^{-vu - \frac{a^2}{2} v}\,\mathrm{d}v.$$

This means that the mixing measure $\mu(\mathrm{d}v)$ is an exponential distribution with mean $2/a^2$.

There are other examples of pairs $(f, \mu) \sim (ZV^{1/2}, V)$ in the literature. For instance, [7] offer an interesting list of univariate mixing measures, containing also some examples with $n > 1$. Another such list can be found in [2]. Note that if f is known then the distribution of $\log Z^2 + \log V$ is known and finding the distribution μ or the distribution of $\log V$ is a problem of deconvolution. If its solution exists, it is unique, as shown for instance by (3).

An example of such a deconvolution is given by [9] who extends (4) to $f(x) = Ce^{-a|x|^{2\alpha}}$ where $0 < \alpha < 1$ as follows: he recalls that for $A > 0$ and $0 < \alpha < 1$, see [1], p. 424, there exists a probability density g, called a positive stable law, such that, for $\theta > 0$,

$$\int_0^\infty e^{-t\theta} g(t)\,\mathrm{d}t = e^{-A\theta^\alpha}. \tag{5}$$

If in the equality above we make the change of variable $t \to v = 1/t$, let $\theta = x^2/2$ and define $\mu(\mathrm{d}v) = C\sqrt{2\pi}g(1/v)v^{-3/2}\mathrm{d}v$, where C is such that $\mu(\mathrm{d}v)$ is a probability, we obtain

$$\int_0^\infty e^{-\frac{1}{2}\frac{x^2}{v}} \frac{1}{\sqrt{2\pi v}}\mu(\mathrm{d}v) = Ce^{-2^{-\alpha}A|x|^{2\alpha}}. \tag{6}$$

Integrating both sides of (6) with respect to x from $-\infty$ to $+\infty$, we obtain

$$C = \alpha\frac{A^{\frac{1}{2\alpha}}}{\sqrt{2}}\frac{1}{\Gamma(\frac{1}{2\alpha})}.$$

If $V \sim \mu$, its Laplace transform L_V cannot be computed except for $\alpha = 1/2$. For $\alpha = 1/2$ and A arbitrary, one can verify that (5) is satisfied for

$$g(t) = \frac{A}{2\sqrt{\pi}}t^{-3/2}e^{-\frac{A^2}{4t}}.$$

Then

$$\mu(\mathrm{d}v) = \frac{A^2}{4}e^{-\frac{A^2}{4}v}\mathbf{1}_{(0,+\infty)}(v)\mathrm{d}v,$$

that is the mixing distribution is an exponential distribution again.

Another elegant example of deconvolution is given by [8] and [6] with the logistic distribution

$$f(x) = \frac{e^x}{(1+e^x)^2} = \sum_{n=1}^\infty (-1)^{n+1} n e^{-n|x|}. \tag{7}$$

Using the representation of (4) as an exponential mixture of scale Gaussians, i. e.

$$\frac{a}{2}e^{-a|x|} = \int_0^{+\infty} \frac{e^{-\frac{x^2}{2v}}}{\sqrt{2\pi v}}\frac{a^2}{2}e^{-\frac{a^2 v}{2}}\,\mathrm{d}v$$

and applying it to $a = n$ in (7) above, we obtain

$$f(x) = \sum_{n=1}^\infty (-1)^{n+1} n^2 \int_0^{+\infty} \frac{e^{-\frac{x^2}{2v}}}{\sqrt{2\pi v}} e^{-\frac{n^2 v}{2}}\,\mathrm{d}v \tag{8}$$

and thus, if μ exists here, it must be

$$\mu(\mathrm{d}v) = \left(\sum_{n=1}^\infty (-1)^{n+1} n^2 e^{-\frac{n^2}{2}v}\right)\mathbf{1}_{(0,+\infty)}(v)\mathrm{d}v \tag{9}$$

which indeed exists since this is the Kolmogorov distribution ([4]), also called Kolmogorov–Smirnov distribution. A direct proof that (9) defines a probability on $(0,+\infty)$ relies on the following Jacobi formula (see [3]):

$$\prod_{n=1}^{\infty}(1-q^{2n-1})^2(1-q^{2n}) = \sum_{n=-\infty}^{\infty}(-1)^n q^{n^2}. \tag{10}$$

Taking $q=e^{-x/2}$, (10) yields

$$\prod_{n=1}^{\infty}(1-e^{-(2n-1)x/2})^2(1-e^{-nx}) = \sum_{n=-\infty}^{\infty}(-1)^n e^{-n^2x/2} := F(x).$$

We observe that $F(0)=0$, $F(+\infty)=1$, and F is increasing as the product of increasing positive factors. Moreover,

$$F'(x) = -\frac{1}{2}\sum_{n-\infty}^{+\infty}(-1)^n n^2 e^{-n^2x?2} = \sum_{n=1}^{+\infty}(-1)^{n+1}n^2e^{-n^2x/2}$$

is the density of (9).

3. THE NORMAL APPROXIMATION TO THE GAUSSIAN SCALE MIXTURE

The mixture f as defined in (1) keeps some characteristics of the normal distribution: It is a symmetric density, $f(x)=e^{-\kappa(\frac{x^2}{2})}$ where $u\mapsto\kappa(u)$ is convex since

$$e^{-\kappa(u)} = \int_0^{\infty} e^{-u/v}\frac{\mu(\mathrm{d}v)}{\sqrt{2\pi v}} = \int_0^{\infty} e^{-uw}\nu(\mathrm{d}w)$$

is the Laplace transform of the positive measure $\nu(dw)$ defined as the image of $\frac{\mu(\mathrm{d}v)}{\sqrt{2\pi v}}$ by the map $u\mapsto w=1/v$.

As said in the introduction, in some practical applications, the distribution of V is not very well known, and it is interesting to replace f by the density of an ordinary normal distribution $N(0,t_0)$. The $L^2(\mathbb{R})$ distance is well adapted to this problem. See [5] for an example of the utilisation of this idea. We are going to prove the following result.

Theorem 3.1. If f is defined by (1), then

1. $f\in L^2(\mathbb{R})$ if and only if
$$\mathbb{E}\left(\frac{1}{\sqrt{V+V_1}}\right)<\infty$$
when V and V_1 are independent with the same distribution μ.

2. If $f\in L^2(\mathbb{R})$, there exists a unique $t_0=t_0(\mu)>0$ which minimizes
$$t\mapsto I_V(t) = \int_{-\infty}^{\infty}\left[f(x)-\frac{1}{\sqrt{2\pi t}}e^{-\frac{x^2}{2t}}\right]^2\mathrm{d}x.$$

3. The scalar $y_0 = 1/t_0$ the unique positive solution of the equation

$$\int_0^\infty \frac{\mu(\mathrm{d}v)}{(1+vy)^{3/2}} = \frac{1}{2^{3/2}}. \tag{11}$$

In particular, if μ_λ is the distribution of λV, then $t' = t_0(\mu_\lambda) = \lambda t_0(\mu)$.

4. The value of $I_V(t_0)$ is

$$I_V(t_0) = \sqrt{\frac{2}{\pi}} \left(\mathbb{E}\left(\frac{1}{\sqrt{V+V_1}} \right) - 2\mathbb{E}\left(\frac{1}{\sqrt{V+t_0}} \right) + \frac{1}{\sqrt{2t_0}} \right)$$

and

$$I_{\lambda V}(t') = \frac{1}{\sqrt{\lambda}} I_V(t_0). \tag{12}$$

5. Finally $t_0 \le \mathbb{E}(V)$.

P r o o f. Recall that if $g \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$ and if $\hat{g}(s) = \int_{-\infty}^\infty e^{isx} g(x)\,\mathrm{d}x$, then Plancherel theorem says that

$$\frac{1}{2\pi} \int_{-\infty}^\infty |\hat{g}(s)|^2 ds = \int_{-\infty}^\infty |g(x)|^2\,\mathrm{d}x. \tag{13}$$

Furthermore if $g \in L^1(\mathbb{R})$, then $g \in L^2(\mathbb{R})$ if and only if $\hat{g} \in L^2(\mathbb{R})$.

Let us apply (13) first to $g = f$. From (1) and (3), we have $\hat{f}(s) = L_V(s^2/2)$. Then

$$\begin{aligned} \int_{-\infty}^\infty \hat{f}^2(s)\,\mathrm{d}s &= \int_{-\infty}^\infty L_V^2(s^2/2) ds = \sqrt{2} \int_0^\infty L(u)^2 \frac{\mathrm{d}u}{\sqrt{u}} \\ &= \sqrt{2} \int_0^\infty \mathbb{E}(e^{-u(V+V_1)}) \frac{\mathrm{d}u}{\sqrt{u}} = \sqrt{2\pi}\,\mathbb{E}(\frac{1}{\sqrt{V+V_1}}) \end{aligned}$$

where the last equality is obtained by recalling that $\int_0^{+\infty} e^{-uv} \frac{\mathrm{d}v}{\sqrt{v}} = \frac{\sqrt{\pi}}{\sqrt{u}}$. Thus statement 1. of the theorem is proved.

To prove 2., 3. and 4., we apply (13) to $g(x) = f(x) - \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$ for which $\hat{g}(s) = L(s^2/2) - e^{-ts^2/2}$. As a consequence

$$I_V(t) = \frac{1}{2\pi} \int_{-\infty}^\infty \left[L_V(s^2/2) - e^{-ts^2/2} \right]^2 \mathrm{d}s = \frac{1}{\pi} \int_0^\infty \left[L_V(u) - e^{-tu} \right]^2 \frac{\mathrm{d}u}{\sqrt{2u}}$$

and

$$I'_V(t) = \frac{\sqrt{2}}{\pi} \int_0^\infty \left[L_V(u) - e^{-tu} \right] e^{-tu} \sqrt{u}\,\mathrm{d}u. \tag{14}$$

Since $\int_0^\infty e^{-2tu} \sqrt{u}\,\mathrm{d}u = \frac{\Gamma(3/2)}{(2t)^{3/2}}$ and since

$$\int_0^\infty L_V(u) e^{-tu} \sqrt{u}\,\mathrm{d}u = \int_0^\infty \int_0^\infty e^{-u(v+t)} \sqrt{u}\,\mathrm{d}u\,\mu(\mathrm{d}v) = \Gamma(3/2) \int_0^\infty \frac{\mu(\mathrm{d}v)}{(t+v)^{3/2}},$$

then $I_V'(t) = 0$ if and only if $\int_0^\infty \frac{\mu(\mathrm{d}v)}{(t+v)^{3/2}} = \frac{1}{(2t)^{3/2}}$. We can rewrite this equation in t as $F(1/t) = 1/2^{3/2}$ where $F(y) = \int_0^\infty \frac{\mu(\mathrm{d}v)}{(1+vy)^{3/2}}$. Thus (14) can be rewritten

$$I_V'(t) = \frac{\sqrt{2}}{\pi}\frac{\Gamma(3/2)}{t^{3/2}}\left[F\left(\frac{1}{t}\right) - \frac{1}{2^{3/2}}\right]. \tag{15}$$

Since $0 < 1/2^{3/2} < 1$, $F(0) = 1$, $\lim_{y\to\infty} F(y) = 0$ and

$$F'(y) = -\frac{3}{2}\int_0^\infty \frac{v\mu(\mathrm{d}v)}{(1+vy)^{5/2}} < 0,$$

it follows that I_V' has only one zero t_0 on $(0,\infty)$ and from (15), it is easy to see from the sign of I_V' that I_V reaches its minimum at t_0.

To show 5., we will apply Jensen inequality $f(\mathbb{E}(X)) \leq \mathbb{E}(f(X))$ to the convex function $f(x) = x^{-3/2}$ and to the random variable $X = 1 + y_0 V$. From

$$\frac{1}{(1+y_0\mathbb{E}(V))^{3/2}} \leq \mathbb{E}\left(\frac{1}{(1+y_0V)^{3/2}}\right) = \frac{1}{2^{3/2}}$$

it follows that $2 \leq 1 + y_0\mathbb{E}(V)$ and $t_0 = 1/y_0 < \mathbb{E}(V)$. □

Example 1. Suppose that $\Pr(V = 1) = \Pr(V = 2) = 1/2$. Let us compute t_0 and $I(t_0)$. With the help of Mathematica, we see that the solution of

$$\frac{1}{2(1+t)^{3/2}} + \frac{1}{2(2+t)^{3/2}} = \frac{1}{(2t)^{3/2}}$$

is $t_0 = 1.39277$. Finally

$$I_V(t_0) = \sqrt{\frac{2}{\pi}}\left(\frac{1}{4\sqrt{2}} + \frac{1}{2\sqrt{3}} + \frac{1}{8} - \frac{1}{\sqrt{1+t_0}} - \frac{1}{\sqrt{2+t_0}} + \frac{1}{\sqrt{2t_0}}\right) = 0.00019,$$

which is very small.

Example 2. Suppose that V is uniform on $(0,1)$ Then

$$t_0 = 0.36678,\ I_V(t_0) = 0.0182.$$

If V is uniform on $[0,a]$, then from Part 4 of Proposition 3.1, we have $t_0 = a \times 0.36678$.

Example 3. If V follows the standard exponential distribution with density $f(v) = e^{-v}\mathbf{1}_{(0,+\infty)}(v)$, then

$$t_0 = 0.524,\ I_V(t_0) = 0.0207.$$

4. SCALE MIXTURES IN THE EUCLIDEAN CASE AND NON IDENTIFIABILITY

Denote by $\mathcal{S}$ the linear space of symmetric real matrices of dimension n equipped with the scalar product $\langle s, s_1\rangle = \text{trace}\,(ss_1)$ and by $\mathcal{P}$ the convex cone of real positive definite matrices of order n. Thus the norm of s is $||s|| = \sqrt{\text{trace}\, s^2}$. We denote by $\mathrm{d}v$ the Lebesgue measure on $\mathcal{S}$ associated to its Euclidean structure, namely such that the mass of a unit cube is one.

We use the symbol a^* for the transposed matrix of any matrix a. As said before, if $v \in \mathcal{P}$ we denote by $v^{1/2}$ the unique element of $\mathcal{P}$ whose square is v. It is sometimes considered that any non singular matrix a such that $v = aa^*$ should be called a generalized square root of v. The Cholesky decomposition $v = tt^*$ of v into a product of a upper triangular matrix t with positive coefficients on the diagonal with its transposed matrix t^* offers an example of such a generalized square root. It can be remarked that in practice the calculation of t is easier than the calculation of $v^{1/2}$. We denote by $\mathbb{O}(n)$ the orthogonal group of $n \times n$ matrices u such that $u^*u = I_n$.

In this section we define the scale mixtures of the standard normal distribution in $\mathbb{R}^n$ and we observe the phenomena of non identifiability: that is, different distributions of V can give the same mixture.

4.1. Scale mixtures of the normal distribution in $\mathbb{R}^n$.

A scaled Gaussian mixture f on $\mathbb{R}^n$ is the density of a random variable X on $\mathbb{R}^n$ of the form $X = V^{1/2}Z$ where $V \sim \mu$ is a random matrix in $\mathcal{P}$ independent of the standard random Gaussian variable $Z \sim N(0, I_n)$. In the following proposition, we give properties of a mixture of the form $X = V^{1/2}Z$ where Z is invariant by rotation but not necessarily Gaussian.

Proposition 4.1. Let A be a random nonsingular square matrix of order n, independent of $Z \in \mathbb{R}^n \setminus \{0\}$ and such that $uZ \sim Z$ for all $u \in \mathbb{O}(n)$. Let $V = AA^*$. Then the following holds.

1. $AZ \sim V^{1/2}Z$, that is, if we replace $V^{1/2}$ by any generalized square root A of V, the distribution of AZ remains the same.
2. If $AZ \sim Z$ then $\Pr(V = I_n) = 1$. In other terms, $AZ \sim Z$ if and only if $\Pr(AA^* = I_n) = 1$, i.e $A \in \mathbb{O}(n)$ almost surely.

Proof. To prove 1., observe that $U = V^{-1/2}A$ is in the orthogonal group $\mathbb{O}(n)$. Let $\mu(dv)K(v, du)\nu(dz)$ denote the joint distribution of (V, U, Z).

Then if h is a bounded function on $\mathbb{R}^n$,

$$\begin{aligned}
\mathbb{E}(h(AZ)) &= \mathbb{E}(h(V^{1/2}UZ)) \\
&= \int_{\mathcal{P}} \mu(\mathrm{d}v) \int_{\mathbb{O}(n)} K(v, du) \int_{\mathbb{R}^n} h(v^{1/2}uz)\nu(\mathrm{d}z) \\
&= \int_{\mathcal{P}} \mu(\mathrm{d}v) \int_{\mathbb{O}(n)} K(v, \mathrm{d}u) \int_{\mathbb{R}^n} h(v^{1/2}z_1)\nu(\mathrm{d}z_1) \qquad (16)
\end{aligned}$$

$$= \int_{\mathcal{P}} \mu(\mathrm{d}v) \int_{\mathbb{R}^n} h(v^{1/2} z_1)\nu(\mathrm{d}z_1) = \mathbb{E}(h(V^{1/2}Z)), \tag{17}$$

where in (16), $z_1 = uz$, and (17) follows from $\int_{\mathbb{O}(n)} K(v, \mathrm{d}u) = 1$.

To prove 2., consider also $\varphi(s) = \mathbb{E}(e^{i\langle s,Z\rangle})$. Since $uZ \sim Z$ for all $u \in \mathbb{O}(n)$ there exists a real function g defined on $[0, \infty)$ such that $\varphi(s) = g(\|s\|^2)$. Since $Z \sim AZ$ we can write

$$g(\|s\|^2) = \mathbb{E}(g(s^* V s)) \ . \tag{18}$$

Next, let us show that if $R \geq 0$ is independent of $Z = (Z_1, \dots, Z_n)$ and if $Z_1 R \sim Z_1$ then $\Pr(R = 1) = 1$. Indeed, for $t \geq 0$ we have that $\mathbb{E}(|Z_1|^{it}) = \mathbb{E}(|Z_1|^{it})\mathbb{E}(R^{it})$. Since there exists $0 < t_0 \leq \infty$ such that $\mathbb{E}(|Z_1|^{it}) \neq 0$ for $0 \leq t < t_0$, it holds that $\mathbb{E}(R^{it}) = 1$ for $0 \leq t < t_0$. This implies that $\Pr(R > 0) = 1$ and $0 = 1 - \Re(\mathbb{E}(R^{it})) = \mathbb{E}(1 - \cos(t \log R))$ or $\Pr(t \log R \in 2\pi\mathbb{Z}) = 1$ for $0 \leq t < t_0$. We deduce easily that $\Pr(R = 1) = 1$.

Now denote $V = (V_{ij})_{1\leq i,j\leq n}$ and apply the above observation to $R = \sqrt{V_{11}}$ by taking $s = (t, 0, \dots, 0)$ in (18). We obtain

$$\mathbb{E}(e^{itZ_1}) = \varphi((t, 0, \dots, 0)) = g(t^2) = \mathbb{E}(g(t^2 V_{11})) = \mathbb{E}(e^{it\sqrt{V_{11}}Z_1})$$

which implies $Z_1 \sim V_{11} Z_1$ and $\Pr(V_{11} = 1) = 1$. Similarly $\Pr(V_{ii} = 1) = 1$ for all $i = 2, \cdots, n$.

Finally, we consider $R = \sqrt{1 + V_{12}}$ and we take $s = (t/\sqrt{2}, t/\sqrt{2}, \dots, 0)$ in (18). Using the fact that $(Z_1 + Z_2)/\sqrt{2} \sim Z_1$ we write

$$\begin{aligned} \mathbb{E}(e^{itZ_1}) &= \mathbb{E}(e^{it(Z_1+Z_2)/\sqrt{2}}) = \varphi((t/\sqrt{2}, t/\sqrt{2}, \dots, 0)) \\ &= \mathbb{E}(g(\frac{1}{2}\, t^2(V_{11} + V_{22} + 2V_{12})) = \mathbb{E}(g(t^2(1 + V_{12})) \\ &= \mathbb{E}(e^{itZ_1\sqrt{1+V_{12}}}) \end{aligned}$$

and we get $\Pr(V_{12} = 0) = 1$. Similarly $\Pr(V_{ij} = 0) = 1$ for $i \neq j$ and finally $\Pr(V = I_n) = 1$ as desired. $\square$

4.2. Nonidentifiability

In Example 4 below, we show that for $n \geq 2$, the measure μ which generates a given f as in (1) may not be unique. Theorem 4.2 gives a more general result. We denote by ω the uniform probability, or Haar probability, on $\mathbb{O}(n)$ and by $\mathcal{D}$ the set of diagonal matrices $b = \mathrm{diag}(b_1, \dots, b_n)$ such that $0 < b_1 \leq b_2 \leq \dots \leq b_n$. It is a well known fact that if $V = U^* B U$ with $U \in \mathbb{O}(n)$ and $B \in \mathcal{D}$ then $u^* V u \sim V$ for all $u \in \mathbb{O}(n)$ if and only if $U \sim \omega$ and B are independent (in this case, the distribution of V is determined by the distribution of its set of eigenvalues determined by B). While the 'if' part is clear, a short proof of the 'only if ' part is as follows: consider $\alpha(db)K(b, du) \sim (B, U)$ and $\mu \sim V$. For any h bounded continuous on $\mathcal{P}$ and any $u_0 \in \mathbb{O}(n)$ we write

$$\int_{\mathcal{P}} h(v)\mu(\mathrm{d}v) \quad = \quad \int_{\mathcal{P}} h(u_0^* v u_0)\mu(\mathrm{d}v)$$

$$= \int_{\mathcal{D}} \left(\int_{\mathbb{O}(n)} h(u_0^* u^* b u u_0) K(b, \mathrm{d}u) \right) \alpha(db)$$

$$= \int_{\mathcal{D}} \left(\int_{\mathbb{O}(n)} h(u^* b u) K(b, \mathrm{d}(u u_0^*)) \right) \alpha(\mathrm{d}b)$$

This shows that, α almost surely, the probability $K(b, du)$ on $\mathbb{O}(n)$ is invariant by $u \mapsto uu_0^*$ for all $u_0 \in \mathbb{O}(n)$ and is equal to ω by uniqueness of the Haar probability on $\mathbb{O}(n)$.

Finally, for $a_1, \dots a_n > 0$ given, we recall the definition of the Dirichlet distribution $D(a_1, \dots, a_n))$ of the variable $(X_1, \dots, X_n)$ on the simplex

$$T_n = \{(x_1, \dots, x_n) \in (0, \infty)^n \ ; \ x_1 + \dots + x_n = 1\} :$$

the density of $(X_2, \dots, X_n)$ is proportional to

$$(1 - (x_2 + \dots + x_n)^{a_1 - 1} x_2^{a_2 - 1} \dots x_n^{a_n - 1}.$$

Theorem 4.2. Suppose that a probability $\mu(dv)$ on $\mathcal{P}$ is invariant by the transformations $v \mapsto uvu^*$ for any $u \in \mathbb{O}(n)$. Then we have the following.

1. Let $V \sim \mu$. Then there exists a unique probability $\nu_\mu(d\lambda)$ on $(0, \infty)$ such that if $\Lambda \sim \nu_\mu$ and if V and Λ are independent of $Z \sim N(0, I_n)$, then

$$V^{1/2} Z = \Lambda^{1/2} Z.$$

2. In the special case where $b = \mathrm{diag}(b_1 \dots, b_n) \in \mathcal{D}$ is fixed let μ_b be the distribution in $\mathcal{P}$ of $U^* b U$ where $U \sim \omega$. For $(X_1, \dots, X_n) \sim D(\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2})$, denote by $\rho_b(d\lambda)$ the distribution of $b_1 X_1 + \dots + b_n X_n$. Then

$$\rho_b = \nu_{\mu_b}. \tag{19}$$

3. If $\alpha(db)$ is a probability on $\mathcal{D}$, denote by μ the distribution of $V = U^* B U$ where $B \sim \alpha$ and $U \sim \omega$ are independent. Then

$$\nu_\mu(d\lambda) = \int_{\mathcal{D}} \alpha(\mathrm{d}b) \rho_b(\mathrm{d}\lambda). \tag{20}$$

Proof. We begin with a remark. Consider the Fourier transform of $V^{1/2} Z$ defined for $s \in \mathbb{R}^n$ by $\varphi(s) = \mathbb{E}(e^{is^* V^{1/2} Z}) = \mathbb{E}(e^{-\frac{1}{2} s^* V s})$. For $u \in \mathbb{O}(n)$ the fact that $u^* V u \sim V$ implies that $\varphi(us) = \varphi(s)$. This implies in turn that $\varphi(s)$ is a function of $\|s\|$ only, or that there exists a function L such that $\varphi(s) = L(\frac{1}{2} \|s\|^2)$. Recall that we intent to show the existence of a positive random variable Λ such that $L(\frac{1}{2} \|s\|^2) = \mathbb{E}(e^{-\frac{1}{2} \Lambda \|s\|^2})$ that is, that L is a Laplace transform. Actually this point is not immediate, and we start the proof of the theorem by showing (19) first.

Let $V = U^*bU$ with $U \sim \omega$ and consider the Fourier transform $\varphi(s)$ of $V^{1/2}Z$, namely

$$\varphi(s) = \mathbb{E}(e^{-\frac{1}{2}\,(Us)^*bUs}) = \mathbb{E}(e^{-\frac{1}{2}\,(b_1(Us)_1^2+\dots+b_n(Us)_n^2)}) \tag{21}$$

where $Us = ((Us)_1, \dots, (Us)_n)$. Now we observe that $(Us)/\|s\|$ is uniformly distributed on the unit sphere of $\mathbb{R}^n$. If $Y = (Y_1, \dots, Y_n) \sim N(0, I_n)$ then $Y/\|Y\|$ is also uniformly distributed on the sphere and it is a classical fact that

$$(X_1, \dots, X_n) = \frac{(Y_1^2, \dots, Y_n^2)}{Y_1^2 + \dots + Y_n^2} \sim D(\frac{1}{2}, \dots, \frac{1}{2}).$$

Therefore

$$\frac{1}{\|s\|^2}(Us)^*b(Us) \sim b_1X_1 + \dots + b_nX_n \sim \rho_b$$

and $\varphi(s) = \int_0^\infty e^{-\frac{1}{2}\,\|s\|^2\lambda}\rho_b(\mathrm{d}\lambda)$, which is a reformulation of (19). Note that in this particular case where $V = U^*bU$ then L is the Laplace transform of ρ_b.

To prove 3., we simply condition by B and use (19) to obtain

$$\varphi(s) = \mathbb{E}(e^{-\frac{1}{2}\,(Us)^*B(Us)}) = \int_{\mathcal{D}} \left(\int_0^\infty e^{-\frac{1}{2}\,\|s\|^2\lambda}\rho_b(\mathrm{d}\lambda)\right)\alpha(\mathrm{d}b)$$

which proves (20).

Recall that any random variable V on $\mathcal{P}$ such that $u^*Vu \sim V$ for all $u \in \mathbb{O}(n)$ has the above form U^*BU where $B \sim \alpha(db)$ is random and independent of $U \sim \omega$. This shows that 3. implies 1. $\square$

Corollary 4.3. If $V \sim uVu^*$ for any $u \in \mathbb{O}(n)$ and has distribution μ then the density f of $V^{1/2}Z$ where $Z \sim N(0, I_n)$ is independent of V has the form $f(x) = L_1(\|x\|^2/2)$. More specifically

$$f(x) = \int_0^\infty e^{-\frac{\|x\|^2}{2\lambda}}\frac{\nu_\mu(\mathrm{d}\lambda)}{\sqrt{2\pi\lambda}}. \tag{22}$$

Remarks.

1. Note that in Corollary 4.3 the function L_1 is the Laplace transform of the image $m(dy)$ of the measure $\frac{\nu_\mu(d\lambda)}{\sqrt{2\pi\lambda}}$ by the map $\lambda \mapsto y = 1/2\lambda$. Since in general (20) is not easy to apply, this offers, in some cases, a way to compute $\nu_\mu(d\lambda)$, when f and L_1 are known, and when m is obvious. Example 4 below will be obtained by this technique with $L_1(s) = (1 + 2s)^{-p}$ with $p > n/2$.

2. For $n \geq 3$ it is difficult to give the density of $\rho_b(d\lambda)$ explicitly. For $n = 2$ it is the image of the beta distribution on $(0, 1)$ with parameters (1/2,1/2) by the affinity $t \mapsto \lambda = (1 - t)b_1 + tb_2$:

$$\rho_b(d\lambda) = \frac{1}{\pi\sqrt{(b_2 - \lambda)(\lambda - b_1)}}1_{(b_1,b_2)}(\lambda)d\lambda.$$

For instance if $\alpha(db_1, db_2) = \alpha_1(db_1)K(b_1, db_2)$ is the joint distribution of $B = \mathrm{diag}(B_1, B_2)$, formula (20) implies $\nu_\mu(d\lambda)$ has density

$$\frac{1}{\pi}\int_0^\lambda \left(\int_\lambda^\infty \frac{K(b_1, \mathrm{d}b_2)}{\sqrt{b_2-\lambda}}\right)\frac{\alpha_1(\mathrm{d}b_1)}{\sqrt{\lambda - b_1}}.$$

3. Another approach to formula (19) is possible using zonal polynomials.

   Indeed for any symmetric matrices a and b of order n we can write

$$\int_{\mathbb{O}(n)} e^{\mathrm{trace}\, u^*bua}\omega(\mathrm{d}u) = \sum_\kappa \frac{C_\kappa(a)C_\kappa(b)}{|\kappa|!C_\kappa(I_n)}.$$

   Equality (21) suggests to apply this identity to the matrices $a = -ss^*/2$ and $b \in \mathcal{D}$. Fortunately the zonal polynomials are simple when computed on a, a matrix of rank one. More specifically $C_\kappa(a) = 0$ except when $\kappa = (m, 0, 0, \dots, 0)$ where m is a non negative integer. In this case, by a reasoning similar to that in the proof of (19), we have

$$\frac{C_\kappa(a)}{|\kappa|!C_\kappa(I_n)} = \frac{(-1)^m}{2^m m!}\int_{\mathbb{O}(n)} (us)_1^{2m}\omega(\mathrm{d}u) = \frac{(-1)^m\|s\|^{2m}}{2^m m!}\mathbb{E}(X_1^m)$$

   where $X_1 \sim \beta(\frac{1}{2}, \frac{1}{2}(n-1))$. However, the computation of

$$c_m(b_1, \dots, b_n) = C_{(m,0,0,\dots,0)}(\mathrm{diag}(b_1, \dots, b_n))$$

   is the real difficulty and using the Pochhammer symbol $(x)_n = \Gamma(n+x)/\Gamma(x)$, one can only write

$$\mathbb{E}(e^{-\frac{1}{2}(Us)^*bUs}) = \sum_{m=0}^\infty \frac{(-1)^m\|s\|^{2m}(1/2)_m}{2^m(n/2)_m m!}c_m(b_1, \dots, b_n).$$

4. An interesting question is the following: suppose that more generally $V \sim \mu$ and $V_1 \sim \mu_1$ in $\mathcal{P}$ are such that $V^{1/2}Z \sim V_1^{1/2}Z$ with $Z \sim N(0, I_n)$ independent of V and V_1. We do not assume here that μ and μ_1 are invariant by $\mathbb{O}(n)$. Consider the Laplace transforms $L_\mu(a) = \int_{\mathcal{P}} e^{-\mathrm{trace}\,(av)}\mu(\mathrm{d}v)$ and L_{μ_1} defined at least on the closed convex cone $\overline{\mathcal{P}}$ of the semi positive definite matrices of order n. Then $V^{1/2}Z \sim V_1^{1/2}Z$ implies that for any $s \in \mathbb{R}^n$ we have

$$L_\mu(\frac{1}{2}ss^*) = L_{\mu_1}(\frac{1}{2}ss^*)$$

   which means that L_μ and L_{μ_1} coincide on the matrices $a \in \overline{\mathcal{P}}$ of rank one. As we have just seen in Theorem 4.2 it does not imply $\mu = \mu_1$. This raises the following problem: given μ, describe the extreme points of the convex set of probabilities μ_1 such that L_μ and L_{μ_1} coincide on the matrices $a \in \overline{\mathcal{P}}$ of rank one.

4.3. An explicit example of non identifiability.

We will now give an example of two different measures μ_1 and μ_2 giving the same scale mixture of Gaussian variables.

Example 4. Let $p > n/2$ and consider the probability on $\mathbb{R}^n$ with density

$$f(x) = \frac{C}{(1+||x||^2)^p}, \tag{23}$$

where C will be computed below. Then consider two probability measures μ_1 and μ_2. The first is

$$\mu_1(dv) \quad = \quad \frac{(\det(v))^{-p+\frac{1}{2}-\frac{n+1}{2}}}{2^{n(p-\frac{1}{2})}\Gamma_{\mathcal{P}}(p-\frac{1}{2})}\exp\{-\frac{1}{2}\,\mathrm{trace}\,(v^{-1})\}\mathbf{1}_{\mathcal{P}}dv, \tag{24}$$

where $\Gamma_{\mathcal{P}}(t) = (2\pi)^{\frac{1}{2}\,n(n-1)}\prod_{j=1}^{d}\Gamma(t-\frac{j-1}{2})$. Therefore V^{-1} follows a Wishart distribution with shape parameter $p-\frac{1}{2}$. The second is defined by $\mu_2(dv) \sim \Lambda I_n$ where Λ has density

$$\frac{\lambda^{-p+\frac{n}{2}-1}}{2^{p-\frac{n}{2}}\Gamma(p-\frac{n}{2})}e^{-\frac{1}{2\lambda}}\mathbf{1}_{(0,+\infty)}(\lambda),$$

i.e. Λ^{-1} follows a Gamma distribution, with shape parameter $p-\frac{1}{2}\,n$. For $x\in\mathbb{R}^n$ and $i=1,2$, we now show that

$$\int_{\mathcal{P}}\frac{e^{-\frac{x^*v^{-1}x}{2}}}{(2\pi)^{n/2}(\det v)^{1/2}}\mu_i(\mathrm{d}v) = f(x) \tag{25}$$

where f is defined by (23). For $i=1$, making the change of variable $y=v^{-1}$, the left-hand side of (25) becomes

$$\begin{aligned}
&\int_{\mathcal{P}}\frac{(\det y)^{1/2}e^{-\frac{x^*yx}{2}}}{(2\pi)^{n/2}}\frac{(\det(y))^{p-\frac{1}{2}-\frac{n+1}{2}}}{2^{n(p-\frac{1}{2})}\Gamma_{\mathcal{P}}(p-\frac{1}{2})}\exp\{-\frac{1}{2}\,\mathrm{trace}\,y\}\mathrm{d}y\\
=&\int_{\mathcal{P}}\frac{(\det(y))^{p-\frac{n+1}{2}}}{(2\pi)^{n/2}2^{n(p-\frac{1}{2})}\Gamma_{\mathcal{P}}(p-\frac{1}{2})}e^{-\frac{1}{2}\,\mathrm{trace}\,(y,I_n+xx^*)}\mathrm{d}y\\
=&\ \frac{2^{np}\Gamma_{\mathcal{P}}(p)}{(2\pi)^{n/2}2^{n(p-\frac{1}{2})}\Gamma_{\mathcal{P}}(p-\frac{1}{2})}\det(I_n+xx^*)^{-p}\\
=&\ \frac{2^{np}\Gamma_{\mathcal{P}}(p)}{(2\pi)^{n/2}2^{n(p-\frac{1}{2})}\Gamma_{\mathcal{P}}(p-\frac{1}{2})}\frac{1}{(1+||x||^2)^p}\\
=&\ \frac{1}{(2\pi)^{n/2}}\frac{\Gamma(p)}{2^{-\frac{n}{2}}\Gamma(p-\frac{n}{2})}\frac{1}{(1+||x||^2)^p},
\end{aligned}$$

yielding $C = \frac{1}{(2\pi)^{n/2}}\frac{\Gamma(p)}{2^{-\frac{n}{2}}\Gamma(p-\frac{n}{2})}$. For $i=2$, making the change of variable $y=\frac{1}{\lambda}$, the left-hand side of (25) becomes

$$\begin{aligned} & \int_0^{+\infty} \frac{e^{-\frac{x^t x}{2\lambda}}}{(2\pi)^{n/2}\lambda^{\frac{n}{2}}} \frac{\lambda^{-p+\frac{n}{2}-1}}{2^{p-\frac{n}{2}}\Gamma(p-\frac{n}{2})} e^{-\frac{1}{2\lambda}} \mathbf{1}_{(0,+\infty)}(\lambda) \\ = & \frac{1}{(2\pi)^{n/2}\Gamma(p-\frac{n}{2})} \int_0^{+\infty} \frac{\lambda^{-p-1}}{2^{p-\frac{n}{2}}} e^{-\frac{1}{2\lambda}(1+||x||^2)} \mathrm{d}\lambda \\ = & \frac{1}{(2\pi)^{n/2}\Gamma(p-\frac{n}{2})} \int_0^{+\infty} \frac{y^{p-1}}{2^{p-\frac{n}{2}}} e^{-\frac{y}{2}(1+||x||^2)} \mathrm{d}y \\ = & \frac{\Gamma(p)}{(2\pi)^{n/2} 2^{-\frac{n}{2}}\Gamma(p-\frac{n}{2})} \frac{1}{(1+||x||^2)^p} \end{aligned} \tag{26}$$

Therefore, with the notation of Theorem 4.2 we have proved that if $\mu_2 \sim \Lambda I_n$ then $\Lambda \sim \nu_{\mu_1}$.

5. EXISTENCE OF THE BEST NORMAL APPROXIMATION IN THE EUCLIDEAN CASE

In this section, we study the conditions that the distribution $\mu(dv)$ on $\mathcal{P}$ must satisfy to garantee that the density f of $V^{1/2}Z$ is in $L^2(\mathbb{R}^n)$ when $V \sim \mu$ and $Z \sim N(0, I_n)$ are independent. We also find a Gaussian law $N(0, t_0)$ on $\mathbb{R}^n$ which is the closest to f in the $L^2(\mathbb{R}^n)$ sense. We consider also the particular case where $V^{1/2}Z = \Lambda^{1/2}Z$ where Λ is a random scalar.

5.1. Best approximation

We first recall two simple formulas.

Lemma 5.1. Let $A \in \mathcal{P}$. Then

$$\int_{\mathbb{R}^n} e^{-\frac{1}{2}s^* As}\mathrm{d}s = \frac{(2\pi)^{n/2}}{\sqrt{\det A}}, \quad \int_{\mathbb{R}^n} e^{-\frac{1}{2}s^* As} ss^* \mathrm{d}s = \frac{(2\pi)^{n/2}}{\sqrt{\det A}} A^{-1}.$$

Proof. Without loss of generality, we may assume that A is diagonal, and the proof is obvious in this particular case. □

We next state that there exists a matrix $v = t_0$ such that the L^2 distance between the multivariate Gaussian mixture $f(x)$ and the Gaussian distribution $N(0, t_0)$ is minimum.

Theorem 5.2. Let $\mu(dv)$ be a probability distribution on the convex cone $\mathcal{P}$. Let $f(x)$ denote the density of the random variable $X = V^{1/2}Z$ of $\mathbb{R}^n$ where $V \sim \mu$ is independent of $Z \sim N(0, I_n)$. Then

1. $f \in L^2(\mathbb{R}^n)$ if and only if $\mathbb{E}\left(\frac{1}{\det\sqrt{V+V_1}}\right) < \infty$ where V and V_1 are independent with the same distribution μ.

2. For $f \in L^2(\mathbb{R}^n)$, consider the function I defined on $\mathcal{P}$ by

$$t \mapsto I(t) = \int_{\mathbb{R}^n} \left[f(x) - \frac{1}{\sqrt{(2\pi)^n \det t}} e^{-\frac{1}{2}\, x^* t^{-1} x} \right]^2 \mathrm{d}x. \tag{27}$$

Then I reaches its minimum at some t_0, and this t_0 is a solution in $\mathcal{P}$ of the following equation in $t \in \mathcal{P}$:

$$\int_{\mathcal{P}} \frac{(v+t)^{-1}}{\sqrt{\det(v+t)}} \mu(\mathrm{d}v) = \frac{1}{2^{1+\frac{1}{2}\, n}} \frac{t^{-1}}{\sqrt{\det t}}. \tag{28}$$

Proof. We have

$$\hat{f}(s) = \int_{\mathbb{R}^n} e^{i\langle s,x\rangle} f(x)\, \mathrm{d}x = \mathbb{E}(e^{i\langle V^{1/2}Z,s\rangle}) = \mathbb{E}(e^{-\frac{1}{2}\, s^* V s}). \tag{29}$$

Now using Plancherel Theorem and Lemma 5.1, we prove part 1. as follows:

$$\int_{\mathbb{R}^n} f^2(x)dx = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{f}(s)^2\, \mathrm{d}s = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \mathbb{E}(e^{-\frac{1}{2}\, s^*(V+V_1)s})\, \mathrm{d}s$$
$$= \frac{1}{(2\pi)^{n/2}} \mathbb{E}\left(\frac{1}{\det\sqrt{V+V_1}}\right).$$

To prove part 2, we use Plancherel theorem again for the function

$$g(x) = f(x) - \frac{e^{-\frac{x^* t^{-1} x}{2}}}{(2\pi)^{n/2} (\det t)^{1/2}}$$

and obtain

$$I(t) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \left[\hat{f}(s) - h_t(s)\right]^2 \mathrm{d}s,$$

where $h_t = e^{-\frac{1}{2}\, s^* t s}$. From Lemma 4.1 applied to $A = 2t$ we have $\|h_t\|^2 = \pi^{n/2}/\sqrt{\det t}$. Expanding the square in $I(t)$ we obtain

$$(2\pi)^n I(t) - ||\hat{f}||^2 = \frac{(\pi)^{n/2}}{\sqrt{\det t}} - 2\langle \hat{f}, h_t\rangle := I_1(t),$$

where $h_t = e^{-\frac{1}{2}\, s^* t s}$. We now want to show that the minimum of $I_1(t)$ is reached at some $t_0 \in \mathcal{P}$.

We show that

$$K_1 = \{y \in \mathcal{P}; I_1(y^{-1}) \leq 0\}$$

is non empty and compact. Writing

$$I_2(y) = \langle \hat{f}, h_{y^{-1}}\rangle \frac{1}{(2\pi)^{n/2}\sqrt{\det y}},$$

we see that $y \in K_1$, i.e. $I_1(y^{-1}) \leq 0$ if and only if $\frac{1}{2^{1+\frac{1}{2}n}} \leq I_2(y)$. From (29), the definition of $h_t(s)$ and Lemma 4.1, we have that

$$\begin{aligned} I_2(y) &= \frac{1}{(2\pi)^{n/2}\sqrt{\det y}} \int_{\mathbb{R}^n} \mathbb{E}(e^{-\frac{s^*Vs}{2}}) e^{-\frac{s^*y^{-1}s}{2}}\, \mathrm{d}s = \frac{1}{(2\pi)^{n/2}\sqrt{\det y}} \mathbb{E}\Big(\int_{\mathbb{R}^n} e^{\frac{s^*(V+y^{-1})s}{2}}\, \mathrm{d}s\Big) \\ &= \frac{1}{\sqrt{\det y}} \mathbb{E}\Big(\frac{1}{\sqrt{\det(V+y^{-1})}}\Big) = \int_{\mathcal{P}} \frac{\mu(\mathrm{d}v)}{\sqrt{\det(I_n+vy)}}. \end{aligned}$$

For $0 < C \leq 1$ let us show that

$$K_2 = \{y \in \mathcal{P}; I_2(y) \geq C\}$$

is compact. Note that $K_1 = K_2$ for $C = 1/2^{1+\frac{1}{2}n}$. Since I_2 is continuous, K_2 is closed. The set K_2 is not empty since $I_2(y) \geq 1$. Let us prove that K_2 is bounded. Recall $\|y\| = (\operatorname{trace} y^2)^{1/2}$. Suppose that $y^{(k)} \in K_2$ is such that $\|y^{(k)}\| \to_{k\to\infty} \infty$ and let us show that for such a $y^{(k)}$, $I_2(y^{(k)}) \to 0$, which is a contradiction.

Indeed, $\operatorname{trace}(vy^{(k)}) \to_{k\to\infty} \infty$ if $v \in \mathcal{P}$. To see this, assume that $v = \operatorname{diag}(v_1, \ldots, v_n)$. Then

$$\begin{aligned} \operatorname{trace}(vy^{(k)}) &= v_1 y_{11}^{(k)} + \cdots + v_n y_{nn}^{(k)} \\ &\geq \operatorname{trace}(y^{(k)}) \times \min_i v_i \geq \|y^{(k)}\| \times \min_i v_i \to_{k\to\infty} \infty, \end{aligned}$$

where the last inequality is due to the fact that if $\lambda_1, \ldots, \lambda_n$ are positive, then $\sqrt{\lambda_1^2 + \ldots, +\lambda_n^2} \leq \lambda_1 + \ldots + \lambda_n$. Moreover, if $(\lambda_1 \ldots, \lambda_n)$ are the eigenvalues of $vy^{(k)}$,

$$\det(I_n + vy^{(k)}) = (1+\lambda_1)\ldots(1+\lambda_n) \geq 1 + \lambda_1 + \cdots + \lambda_n = 1 + \operatorname{trace}(vy^{(k)}) \to_{k\to\infty} \infty$$

By dominated convergence, it follows that $I_2(y^{(k)}) \to_{k\to\infty} 0$ and this proves that K_2 is bounded. We have therefore shown that K_1 is compact. This proves that the minimum of $I_1(t)$ and thus of $I(t)$ is reached at some point t_0 of $\mathcal{P}$.

The last task is to show that t_0 is a solution of equation (28). Since $I(t)$ is differentiable and reaches its minimum on the open set $\mathcal{P}$, the differential of $I(t)$ must cancel at t_0. The differential of I is the following linear form on $\mathcal{S}$

$$h \in \mathcal{S} \mapsto I'(t)(h) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \Big[\hat{f}(s) - e^{-\frac{1}{2}s^*ts}\Big] e^{-\frac{1}{2}s^*ts} s^*hs\, \mathrm{d}s.$$

The equality $I'(t) = 0$ is equivalent to

$$\int_{\mathbb{R}^n} \hat{f}(s) e^{-\frac{1}{2}s^*ts} ss^* ds = \int_{\mathbb{R}^n} e^{-s^*ts} ss^*\, \mathrm{d}s.$$

Using the second formula in Lemma 4.1 and the fact that $\hat{f}(s) = \mathbb{E}(e^{-\frac{1}{2}s^*Vs})$, we obtain

$$\int_{\mathcal{P}} \frac{(v+t)^{-1}}{\sqrt{\det(v+t)}} \mu(\mathrm{d}v) = \frac{(2t)^{-1}}{\sqrt{\det(2t)}} = \frac{1}{2^{1+\frac{1}{2}n}} \frac{t^{-1}}{\sqrt{\det t}},$$

which proves (28). □

Remarks.

1. We note that (28) can also be written in terms of $y = t^{-1}$ as
$$\int_{\mathcal{P}} \frac{(1+vy)^{-1}}{\sqrt{\det(1+vy)}}\mu(\mathrm{d}v) = \frac{1}{2^{1+\frac{n}{2}}} I_n.$$
2. While it is highly probable that the value t_0 at which $I(t)$ reaches its minimum is unique, it is difficult to show for $n \geq 2$ that equation (28) has a unique solution: there is no reason to think that the function $t \mapsto I(t)$ is convex. However a case of uniqueness is proved in Proposition 5.3 below.

5.2. Best approximation for a scalar mixture.

Proposition 5.3. Let $\nu(d\lambda)$ be a probability on $(0,\infty)$ such that
$$\mathbb{E}((\Lambda+\Lambda_1)^{-n/2}) < \infty$$
where Λ and Λ_1 are independent with distribution ν, and let μ be the distribution of $V = \Lambda I_n$. Then $t \mapsto I(t)$ defined in (27) reaches its minimum at a unique point t_0. Furthermore t_0 is a multiple of I_n.

P r o o f. From Theorem 4.2, I reaches its minimum at least at one point $t_0 \in \mathcal{P}$. Without loss of generality by choosing a suitable orthonormal basis of $\mathbb{R}^n$, we can assume that $t_0 = \mathrm{diag}(\lambda_1^0, \ldots, \lambda_n^0)$. We are going to show that $\lambda_1^0 = \ldots = \lambda_n^0$. Consider the restriction I^* of I to the set of diagonal matrices with positive entries, namely
$$I^*(t_1,\ldots,t_n) = I^*(\mathrm{diag}(t_1,\ldots,t_n)).$$
Of course $(t_1,\ldots,t_n) \mapsto I^*(t_1,\ldots,t_n)$ reaches its minimum on $(\lambda_1^0,\ldots,\lambda_n^0)$. By a computation which imitates the proof of Theorem 4.2 we consider
$$\begin{aligned} I_1^*(t_1,\ldots,t_n) &= (2\pi)^n I^*(t_1,\ldots,t_n) - \|\hat{f}\|^2 \\ &= \frac{\pi^{n/2}}{\sqrt{t_1\ldots t_n}} - 2\int_0^\infty \frac{\nu(\mathrm{d}\lambda)}{\prod_{i=1}^n (t_i+\lambda)^{1/2}}. \end{aligned}$$
Since $I_1^*(t_1,\ldots,t_n)$ reaches its minimum at t_0, its gradient is zero at $(\lambda_1^0,\ldots,\lambda_n^0)$. We have
$$\frac{\partial}{\partial t_j} I_1^*(t_1,\ldots,t_n) = -\frac{\pi^{n/2}}{2t_j\sqrt{t_1\ldots t_n}} + \int_0^\infty \frac{\nu(\mathrm{d}\lambda)}{(t_j+\lambda)\prod_{i=1}^n (t_i+\lambda)^{1/2}}$$
and as a consequence, for all $j = 1,\ldots,n$
$$\int_0^\infty \frac{\lambda_j^0}{\lambda_j^0+\lambda} \times \frac{\nu(\mathrm{d}\lambda)}{\prod_{i=1}^n (\lambda_i^0+\lambda)^{1/2}} = \frac{\pi^{n/2}}{2\sqrt{\lambda_1^0\ldots\lambda_n^0}}. \tag{30}$$
The important point of (30) is the fact that the right hand side does not depend on j. Suppose now that there exists j_1 and j_2 such that $\lambda_{j_1}^0 < \lambda_{j_2}^0$. This implies that for all $\lambda > 0$ we have
$$\frac{\lambda_{j_1}^0}{\lambda_{j_1}^0+\lambda} < \frac{\lambda_{j_2}^0}{\lambda_{j_2}^0+\lambda}$$

and the left hand sides of (30) cannot be equal for $j = j_1$ and $j = j_2$. As a consequence $t_0 = \lambda^0 I_n$ for some $\lambda^0 > 0$.

To see that λ^0 is unique, we imitate the proof of Theorem 3.1. We omit the details here. □

We will finish by giving an example of a scalar Gaussian mixture, actually built on the univariate Kolmogorov–Smirnov measure (9) with density

$$k_1(\lambda) = \sum_{n=1}^{+\infty} (-1)^{n+1} n^2 e^{-\frac{n^2\lambda}{2}} \mathbf{1}_{(0,+\infty)}(\lambda).$$

Example 5. Let us verify first that

$$g_n(x) = C_n \frac{e^{||x||}}{(1+e^{||x||})^2},$$

where C_n is the normalizing constant, is a density in $\mathbb{R}^n$. Indeed, using polar coordinates in $\mathbb{R}^n$ with $r = ||x||$, we have $\frac{1}{C_n} = S_{n-1} J(n-1)$ where $S_{n-1} = n\pi^{n/2}/\Gamma(1+\frac{n}{2})$ is the area of the unit sphere in $\mathbb{R}^n$ and where

$$J(t) = \int_0^{+\infty} \frac{e^{-r} r^t}{(1+e^{-r})^2} \mathrm{d}r.$$

Of course $J(0) = 1/2$ and by integration by part $J(1) = \log 2$. For $t > 1$ we have

$$\begin{aligned} J(t) &= \sum_{k=1}^{\infty} (-1)^{k-1} k \int_0^{\infty} e^{-kr} r^t \,\mathrm{d}r = \Gamma(t+1) \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k^t} \\ &= \Gamma(t+1)(1-2^{1-t})\zeta(t), \end{aligned}$$

where $\zeta(t) = \sum_{k=1}^{\infty} \frac{1}{k^t}$ is the Riemann function and the last equality is a well-known formula. Thus for instance

$$C_1 = 1, C_2 = 1/(2\pi \log 2),\ C_3 = 3/(2\pi^3).$$

Next, writing

$$k_n(\lambda) = C_n (2\pi\lambda)^{\frac{n-1}{2}} k_1(\lambda) \mathbf{1}_{(0,+\infty)}(\lambda)$$

let us show that k_n is a density such that

$$\int_0^{+\infty} \frac{e^{-\frac{||x||^2}{2\lambda}}}{(2\pi\lambda)^{n/2}} k_n(\lambda) \mathrm{d}\lambda = g_n(x). \tag{31}$$

This means, of course, that g_n is a scale mixture of multivariate normal $N(0, \lambda I_n)$ distributions. We have

$$1 = \int_{\mathbb{R}^n} g_n(x) dx \quad = \quad C_n \int_{\mathbb{R}^n} \int_0^{+\infty} \frac{e^{-\frac{||x||^2}{2\lambda}}}{(2\pi\lambda)^{n/2}} (2\pi\lambda)^{(n-1)/2} k_1(\lambda) \,\mathrm{d}\lambda \,\mathrm{d}x$$

$$
\begin{aligned}
&= C_n \int_0^{+\infty} (2\pi\lambda)^{(n-1)/2} k_1(\lambda) \left(\int_{\mathbb{R}^n} \frac{e^{-\frac{||x||^2}{2\lambda}}}{(2\pi\lambda)^{n/2}} \,\mathrm{d}x \right) \mathrm{d}\lambda \\
&= C_n \int_0^{+\infty} (2\pi\lambda)^{(n-1)/2} k_1(\lambda) \,\mathrm{d}\lambda.
\end{aligned}
$$

ACKNOWLEDGEMENT

This work was partially supported by an NSERC Discovery Grant for H. Massam. The authors would like to thank the referee for a careful reading of the paper and the suggestion to expand Section 4.

(Received March 20, 2019)

References

[1] W. Feller: An Introduction to Probability Theory and its Applications, vol. 2. Wiley, New York 1966.

[2] T. Gneiting: Normal scale mixture and probability densities. J. Stat. Comput. Simul. *59* (1997), 375–384. DOI:10.1080/00949659708811867

[3] G. H. Hardy and E. M. Wright: An Introduction to the Theory of Numbers. Sixth edition. Oxford University Press, London 2008. DOI:10.1017/s0025557200007464

[4] A. N. Kolmogorov: Sulla determinazióne empìrica di una légge di distribuzióne. G. Inst. Ital. Attuari *4* (1933), 83–91.

[5] G. Letac, H. Massam, and R. Mohammadi: The ratio of normalizing constants for Bayesian Gaussian model selection. arXiv 1706. 04416 (2017), revision Oct. 12th 2018.

[6] J. F. Monahan and L. A. Stefanski: Normal Scale Mixture Approximations to $F^*(z)$ and Computation of the Logistic-Normal Integral. Handbook of the Logistic Distribution (N. Balakrishnan, ed.), Marcel Dekker, New York 1992.

[7] J. A. Palmer, K. Kreutz-Delgado, and S. Maleig: Dependency models based on generalized Gaussian scale mixtures. DRAFT UCSD-SCCN v1.0, Sept 7 (2011).

[8] L. A. Stefanski: A normal scale mixture representation of the logistic distribution. Stat. Probab. Lett- *11* (1991), 69–70. DOI:10.1016/0167-7152(91)90181-p

[9] M. West: On scale mixture of normal distributions. Biometrika *74* (1987), 646–648. DOI:10.1093/biomet/74.3.646

Gérard Letac, Institut Mathématique de Toulouse, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse. France.
e-mail: gerard.letac@math.univ-toulouse.fr

Hélène Massam, Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto, M3J 1P3. Canada.
e-mail: massamh@yorku.ca