Information Sciences 179 (2009) 3832–3842

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

On minimum Fisher information distributions with restricted support and fixed variance

J.-F. Bercher [a,*], C. Vignat [b]

[a] Université Paris-Est, LabInfo-IGM, 5 bd Descartes, 77454 Marne-la-Vallée Cedex 2, France
[b] Université de Marne-la-Vallée, LabInfo-IGM, 5 bd Descartes, 77454 Marne-la-Vallée Cedex 2, France

ARTICLE INFO

Article history:
Received 2 March 2009
Received in revised form 23 June 2009
Accepted 27 July 2009

Keywords:
Fisher information
Minimization problems
Whittaker functions
Differential equations

ABSTRACT

Fisher information is of key importance in estimation theory. It also serves in inference problems as well as in the interpretation of many physical processes. The mean-squared estimation error for the location parameter of a distribution is bounded by the inverse of the Fisher information associated with this distribution. In this paper we look for minimum Fisher information distributions with a restricted support. More precisely, we study the problem of minimizing the Fisher information in the set of distributions with fixed variance defined on a bounded subset $\mathscr{S}$ of $\mathbb{R}$ or on the positive real line. We show that the solutions of the underlying differential equation can be expressed in terms of Whittaker functions. Then, in the two considered cases, we derive the explicit expressions of the solutions and investigate their behavior. We also characterize the behavior of the minimum Fisher information as a function of the imposed variance.

1. Introduction

Importance of Fisher information as a measure of the information in a distribution is well known. It has many implications in estimation theory, as exemplified by the Cramér-Rao bound which is a fundamental limit on the variance of an estimator. Recent applications for computing performance bounds can be found in [2,35]. It is used as a method of inference and understanding in statistical physics and biology, as promoted by Frieden [10–12]. It is also used as a tool for characterizing complex signals or systems, [17,20,30] with applications, e.g. in geophysics [25,23,3,16], in biology [9], in reconstruction [5,6,21] or in signal processing [38,37,29]. Other applications are in random censoring [26], hypothesis testing [19], classification [8]. In robust estimation, minimization of Fisher information has been originally considered by Huber [15], and in the case of scale and location parameters in a Kolmogorov neighborhood of a parent distribution in [33,34]. Fisher information for orthogonal polynomials and special functions have been studied in [24,36]. Connections with the differential equations of Physics have been explored in [11]. It is also interesting to mention that Fisher information associated to a distribution appears in quantum physics under the name of Weiszäcker energy; in this setting, several inequalities for the Fisher information are derived in [22].

It is well known that the distribution with a fixed variance that minimizes the Fisher information on $\mathbb{R}$ is the standard gaussian distribution. However, there are many situations where the variables at hand are known to belong to some subset of $\mathbb{R}$. For instance, the random variable may be known, on physical grounds, to have only non negative outcomes, e.g. the variable represents an energy. Variables may also be known to have a distribution with a support restricted to a given interval: this is the case of normalized variables or of the measurements obtained from a physical device with a (necessarily)

* Corresponding author.
E-mail addresses: jf.bercher@esiee.fr (J.-F. Bercher), vignat@univ-mlv.fr (C. Vignat).

0020-0255/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2009.07.013

finite output range. In such cases, a model of the data probability distribution can be defined by the distribution compatible with the constraints extracted from the data with minimum (Fisher) information. This idea, advocated by Frieden and others, is reminiscent of the idea of maximum entropy distributions, when the measure of information considered is the Fisher information. By the Cramér-Rao bound, the Fisher information serves as a benchmark for estimators.

In the case of the estimation of a location parameter, the distributions with minimum Fisher information correspond to the most difficult estimation cases, and thus it is interesting to look for these distributions. This problem has been considered in the important paper [28], which reading has motivated the present work. This paper presents general results and characterizations of the solution to the minimization of Fisher information on a compact support, subject to a variance constraint. Here, we give the explicit closed-form expressions of the solutions, characterize their behavior and propose alternate simpler proofs. We also extend the results to the case of distributions with support confined to the positive half line (and of course the case of any other semi-infinite support as a straightforward consequence).

Let f denote the probability density of a random variable X. The Fisher information (with respect to a translation parameter) is defined as

$$I[f] = \int_{\mathscr{S}} \left(\frac{d\ln f(x)}{dx}\right)^2 f(x)\mathrm{d}x = \int_{\mathscr{S}} \left(\frac{df(x)}{dx}\right)^2 \frac{1}{f(x)}\mathrm{d}x, \tag{1}$$

where $\mathscr{S} = \mathrm{Supp}[f]$ denotes the support of the density, $f(x)$ is supposed differentiable and both $f(x)$ and its derivative $f'(x)$ are square integrable on $\mathscr{S}$. We note $\mathscr{D}$ the set of functions that verify these hypotheses. It is known [7] that the Fisher information is a strictly convex function of the distribution, that is: for $f(x), g(x)$, and with $\lambda \in (0,1)$, then

$$I[\lambda f + (1-\lambda)g] < \lambda I[f] + (1-\lambda)I[g]. \tag{2}$$

We are here interested in classifying distributions with a given variance, and consider the variational problem

$$I(\sigma^2) = \inf_f \{I[f] : \mathrm{Supp}[f] = \mathscr{S}, \quad f \in \mathscr{D} \text{ and } \mathrm{Var}[f] = \sigma^2\} \tag{3}$$

which consists in finding a distribution with minimum Fisher information on the set of all distributions with support $\mathscr{S}$ and a fixed variance σ^2. The value of the minimum Fisher information obtained for a given variance σ^2 is denoted $I(\sigma^2)$ – the use of the square brackets and parenthesis distinguishes between the functionals of the probability distributions and the functionals of the variance. Although $I[f]$ is a convex functional, the set defined by the constraint $\mathrm{Var}[f] = \sigma^2$ is not convex, so that uniqueness of the solution is not guaranteed. For instance, on $\mathbb{R}$, the normal distribution minimizes Fisher information in the set of distributions on $\mathbb{R}$ with a given variance, but in fact irrespectively of the value of the mean: all normal distributions, whatever their mean, are equivalent solutions. We will obtain in the following that solutions on $\mathbb{R}^+$ or on an interval are in fact unique.

In Section 2 we give the differential equation that is associated with the problem of minimization of Fisher information, and then we underline its relationship with known differential equations. So doing, we exhibit some explicit expressions of the solutions, in terms of Whittaker and parabolic cylinder functions. Then, we examine two particular cases. First, in Section 4 we characterize the solutions with positive support, the behavior of the minimum Fisher information with respect to the variance, and show that the solution is unique and turns out to be a chi distribution. Second, in Section 5 where the support is restricted to an interval, we give the expression of the unique solution to the problem and we study the general behavior of the minimum Fisher information with respect to the variance. Finally, we give the expression of the probability density with minimum Fisher information among all distributions with finite variance defined on an interval.

2. The differential equation associated with the minimization of Fisher information

In the following, it is convenient to introduce the transformation $f(x) = u(x)^2$ and to work with $u(x)$ instead of $f(x)$. With this notation, the Fisher information becomes

$$I[f] = 4\int_{\mathscr{S}} u'(x)^2\mathrm{d}x, \tag{4}$$

where $u'(x)$ denotes the first order derivative of $u(x)$. The variational problem (3) can be restated as follows:

$$I(\sigma^2) = \inf_{u:u^2\in\mathscr{D}} \left\{4\int_{\mathscr{S}} u'(x)^2\mathrm{d}x : \int_{\mathscr{S}} u(x)^2\mathrm{d}x = 1 \quad \text{and } Var[u^2] = \sigma^2\right\}. \tag{5}$$

The problem above can also be completed by some additional conditions on the boundaries of the domain: for instance, when $\mathscr{S} = [0:+\infty)$ we need $u(+\infty) = 0$ so as to ensure a proper integrable density. At the left endpoint, we shall ensure continuity and set $u(0) = 0$ in order to keep the Fisher information finite. Indeed, the Fisher information associated with a distribution is $+\infty$ when the distribution is not continuous on the domain $\mathscr{S}$. Similarly, when the support $\mathscr{S}$ is restricted to $\mathscr{S} = [-1,1]$, and the density is set to 0 outside of this domain, we shall take $u(1) = u(-1) = 0$ in order to ensure continuity at these boundary points.

The Lagrangian functional associated with the problem (5) is

$$L(u;\alpha,\beta) = \int_{\mathcal{S}} u'(x)^2 \mathrm{d}x + \alpha\left(\int_{\mathcal{S}} u(x)^2 \mathrm{d}x - 1\right) + \beta\left(\int_{\mathcal{S}} (x-\mu)^2 u(x)^2 \mathrm{d}x - \sigma^2\right), \tag{6}$$

where α and β are the Lagrange parameters associated with the normalization and variance constraints respectively. The mean is denoted by μ. The minimum of the Lagrangian functional is obtained by standard calculus of variations [13] which asserts that a minimizer of (6) is necessary a solution of the Euler–Lagrange equation

$$-\frac{d}{dx}\frac{\partial L}{\partial u'} + \frac{\partial L}{\partial u} = 0 \tag{7}$$

which leads to

$$u''(x) - (\alpha + \beta(x-\mu)^2)u(x) = 0. \tag{8}$$

Of course, with the simple change of variable $z = x - \mu$, the differential equation reduces to

$$u''(z) - (\alpha + \beta z^2)u(z) = 0 \tag{9}$$

which is a parabolic differential equation. Interestingly, the minimum Fisher information can be written in terms of the constraints and of the Lagrange parameters associated to these constraints.

Proposition 1. *The minimum Fisher information in (5) can be expressed as*

$$-\frac{1}{4}I(\sigma^2) = \alpha + \beta\sigma^2 \tag{10}$$

with $u(a) = u(b) = 0$ and where a and b denote the left and right endpoints of the support $\mathcal{S}$.

Note that the Lagrange parameters are (complicated) functions of the constraints, so that the right hand side of (10) is not an affine function in σ^2.

Proof. By integration by parts,

$$-\frac{1}{4}I(\sigma^2) = -\int_{\mathcal{S}} u'(x)^2 \mathrm{d}x = -[u(x)u'(x)]_a^b + \int_a^b u(x)u''(x)\mathrm{d}x. \tag{11}$$

Using the boundary conditions and the differential Eq. (8), we then obtain

$$-\frac{1}{4}I(\sigma^2) = \int_a^b (\alpha + \beta(x-\mu)^2)u^2(x)\mathrm{d}x. \tag{12}$$

which reduces to (10) taking into account the values of the constraints. $\square$

3. Solutions to the differential equation

Let us consider the parabolic differential Eq. (9). Using the change of variable $x = \sqrt{2\sqrt{\beta}}z$, together with the substitution $d(x) = u\left((2\sqrt{\beta})^{-\frac{1}{2}}x\right)$, the differential equation becomes

$$d''(x) + \left(-\frac{\alpha}{2\sqrt{\beta}} - \frac{1}{4}x^2\right)d(x) = 0 \tag{13}$$

which is the Weber differential equation, whose standard form is

$$d''(x) + \left(v + \frac{1}{2} - \frac{1}{4}x^2\right)d(x) = 0. \tag{14}$$

Here we simply have $v = -a/2\sqrt{(b)} - 1/2$. The solutions of the Weber equation can be expressed as a linear combination of the parabolic cylinder function $D_v(x)$ and $D_{-v-1}(ix)$ [32]:

$$d(x) = c_1 D_v(x) + c_2 D_{-v-1}(ix) \tag{15}$$

But the Weber equation above can also be converted into the Whittaker equation using the substitution $d(x) = \frac{1}{\sqrt{x}}w\left(\frac{x^2}{2}\right)$ and $z = x^2/2$, which leads to

$$w''(z) + \left[\frac{3}{16z^2} + \frac{v+1/2}{2z} - \frac{1}{4}\right]w(z) \tag{16}$$

that has the form of the Whittaker differential equation

$$w''(z) + \left[\frac{1/4-\mu^2}{z^2} + \frac{\lambda}{z} - \frac{1}{4}\right]w(z) = 0 \tag{17}$$

with $\lambda = \nu/2 + 1/4$ and $\mu = 1/4$. But the Whittaker differential equation can also be obtained directly from the initial differential Eq. (9) with the substitution $u(z) = \frac{1}{\sqrt{z}} w\left(\frac{\xi z^2}{2}\right)$ and with $x = \xi z^2/2$. Then, one readily obtains

$$w''(x) + \left[\frac{3}{16x^2} - \frac{\alpha}{2\xi x} - \frac{\beta}{\xi^2}\right] w(x) = 0, \tag{18}$$

which reduces to the Whittaker differential Eq. (17) with $\xi = 2\sqrt{\beta}$, $\lambda = -\frac{\alpha}{4\sqrt{\beta}}$, and $\mu = \frac{1}{4}$. The general solutions of the Whittaker differential equation can be expressed in terms of the two linearly independent Whittaker functions $M_{\lambda,\mu}(z)$ and $M_{\lambda,-\mu}(z)$, or with the help of the Whittaker $W_{\lambda,\mu}(z)$ function defined by

$$W_{\lambda,\mu}(z) = \frac{\Gamma(-2\mu)}{\Gamma\left(\frac{1}{2} - \mu - \lambda\right)} M_{\lambda,\mu}(z) + \frac{\Gamma(2\mu)}{\Gamma\left(\frac{1}{2} + \mu + \lambda\right)} M_{\lambda,-\mu}(z) \tag{19}$$

for $2m \notin \mathbb{N}$. In these expressions, the Whittaker $M_{\lambda,\mu}(z)$ function can be expressed as a simple function of the confluent hypergeometric function, or Kummer function, $M\left(\frac{1}{2} + \lambda - \mu, 1 + 2\lambda, z\right)$ according to [1, Eq. 13.1.32]

$$M_{\lambda,\mu}(z) = z^{\lambda+\frac{1}{2}} e^{-\frac{z}{2}} M\left(\frac{1}{2} + \lambda - \mu, 1 + 2\lambda, z\right). \tag{20}$$

Since the Kummer function has an exact series representation, we also have

$$M_{\lambda,\mu}(z) = z^{\mu+\frac{1}{2}} e^{-\frac{z}{2}} \sum \frac{\left(\frac{1}{2} + \lambda - \mu\right)_n}{n!(2\lambda+1)_n} z^n, \tag{21}$$

where $()_n$ denotes the Pochhammer symbol. As far as the parabolic cylinder function is concerned, it can also be written as a Whittaker function [32, p. 347]:

$$D_\nu(x) = 2^{\frac{\nu}{2}+\frac{1}{4}} z^{-\frac{1}{2}} W_{\frac{\nu}{2}+\frac{1}{4},-\frac{1}{4}}\left(\frac{1}{2}z^2\right). \tag{22}$$

According to this discussion and the relationships between parabolic cylinder functions, Whittaker M and W functions, we find that the solutions of the differential equation associated with the problem of minimum Fisher information can be expressed as various equivalent linear combinations of the functions

$$D_{-\frac{\alpha}{2\sqrt{\beta}}-\frac{1}{2}}\left(\left(2\sqrt{\beta}\right)^{\frac{1}{2}} z\right) \text{ and } D_{\frac{\alpha}{2\sqrt{\beta}}-\frac{1}{2}}\left(\left(2\sqrt{\beta}\right)^{\frac{1}{2}} iz\right),$$

or

$$\frac{1}{\sqrt{z}} M_{-\frac{\alpha}{4\sqrt{\beta}},\frac{1}{4}}\left(\sqrt{\beta}z^2\right), \quad \frac{1}{\sqrt{z}} M_{-\frac{\alpha}{4\sqrt{\beta}},-\frac{1}{4}}\left(\sqrt{\beta}z^2\right), \quad \text{and } \frac{1}{\sqrt{z}} W_{-\frac{\alpha}{4\sqrt{\beta}},\frac{1}{4}}\left(\sqrt{\beta}z^2\right)$$

Some equivalent expressions of the solutions are given below

$$u(z) = c_1 D_{-\frac{\alpha}{2\sqrt{\beta}}-\frac{1}{2}}\left(\left(2\sqrt{\beta}\right)^{\frac{1}{2}} z\right) + c_2 D_{\frac{\alpha}{2\sqrt{\beta}}-\frac{1}{2}}\left(\left(2\sqrt{\beta}\right)^{\frac{1}{2}} iz\right) \tag{23}$$

$$= c_1' \frac{1}{\sqrt{z}} M_{-\frac{\alpha}{4\sqrt{\beta}},-\frac{1}{4}}\left(\sqrt{\beta}z^2\right) + c_2' \frac{1}{\sqrt{z}} M_{-\frac{\alpha}{4\sqrt{\beta}},\frac{1}{4}}\left(\sqrt{\beta}z^2\right) \tag{24}$$

$$= c_1'' \frac{1}{\sqrt{z}} M_{-\frac{\alpha}{4\sqrt{\beta}},\frac{1}{4}}\left(\sqrt{\beta}z^2\right) + c_2'' \frac{1}{\sqrt{z}} W_{-\frac{\alpha}{4\sqrt{\beta}},\frac{1}{4}}\left(\sqrt{\beta}z^2\right). \tag{25}$$

In these formulas, the values of the constants in the linear combinations will be determined according to auxiliary constraints, e.g. the boundary values. We now turn to the characterization of solutions on $\mathbb{R}^+$ and then on an interval.

4. Solutions defined on the positive real line

We consider the expression of the solution in terms of the parabolic cylinder functions, as given in (23). A first point is to check that the solution is bounded on $\mathbb{R}^+$. Since the Weber equation has only one irregular singularity at $z = +\infty$, it is sufficient to examine the behavior of the solution for $z \to +\infty$. The asymptotic expansion of the Weber function $D_\nu(z)$, with $\arg z \leqslant 3\pi/4$ is [32, p. 347],[1, Eq. 19.8.1]:

$$D_\nu(z) \sim e^{-\frac{1}{4}z^2} z^\nu \left(1 - \frac{\nu(\nu-1)}{2z^2} + \frac{\nu(\nu-1)(\nu-2)(\nu-3)}{24z^4} - \cdots\right). \tag{26}$$

For $z \to +\infty$, we see at once from the asymptotic expansion that $D_\nu(z) \to 0$. Replacing ν by $-\nu - 1$ and z by iz, we can also observe that $D_{-\nu-1}(iz)$ tends to infinity when $z \to +\infty$. Therefore, since the solution must correspond to a proper integrable density, the second parabolic cylinder function in (23) must be discarded, with $c_2 = 0$, and the solution becomes

$$u_{\alpha,\beta}(z) = \frac{D_{-\frac{\alpha}{2\sqrt{\beta}}-\frac{1}{2}}\left(\left(2\sqrt{\beta}\right)^{\frac{1}{2}}z\right)}{\left(\int_0^{+\infty}\left(D_{-\frac{\alpha}{2\sqrt{\beta}}-\frac{1}{2}}\left(\left(2\sqrt{\beta}\right)^{\frac{1}{2}}z\right)\right)^2 \mathrm{d}z\right)^{\frac{1}{2}}} \tag{27}$$

where the denominator has been introduced in order to ensure the normalization of the probability density $f(z) = u(z)^2$.

As a consequence of the simple form obtained above, it is possible to characterize the general behavior of the minimum Fisher information as a function of the variance. This is given by the following

Proposition 2. *The Fisher information of the minimum information probability density with positive support and fixed variance σ^2, corresponding to the solution (27), verifies*

$$I_\nu(\sigma^2) = \frac{K_\nu}{\sigma^2}, \tag{28}$$

where $\nu = -\alpha/2\sqrt{\beta} - 1/2$ and K_ν is a constant.

Proof. Let $\nu = -\alpha/2\sqrt{\beta} - 1/2$ and $\xi = 2\sqrt{\beta}$, and consider ν fixed. Then, ξ acts as a scaling factor: more precisely, denoting $V_{\nu,\xi}$ the variance associated with the solution with parameters ν and ξ, we readily have $V_{\nu,\xi} = \frac{1}{\xi}V_{\nu,1}$ and $I_{\nu,\xi} = \xi I_{\nu,1}$, where $I_{\nu,\xi}$ is the Fisher information. Therefore, the product $I_{\nu,\xi}V_{\nu,\xi} = I_{\nu,1}V_{\nu,1}$ does not depend on ξ and

$$I_{\nu,\xi} = \frac{I_{\nu,1}V_{\nu,1}}{V_{\nu,\xi}}.$$

Moreover, from the equality $V_{\nu,\xi} = \frac{1}{\xi}V_{\nu,1}$ and the fact that $V_{\nu,1} > 0$, we deduce that the function $\xi \mapsto V_{\nu,\xi}$ maps $\mathbb{R}^+$ to $\mathbb{R}^+$ so that it is always possible to find a value ξ such that $V_{\nu,\xi} = \sigma^2$; thus, the Fisher information $I_\nu(\sigma^2)$ of the probability density associated with the solution $u(z)$ in (27) follows the Eq. (28), with $K_\nu = I_{\nu,1}V_{\nu,1}$. □

Let ν^* be the value that minimizes $I_\nu(\sigma^2)$. For that value and any distribution f on $\mathbb{R}^+$ with same variance σ^2, we always have

$$I[f] \geqslant I_{\nu^*}(\sigma^2) = \frac{K_{\nu^*}}{\sigma^2}, \tag{29}$$

which refines the Cramér-Rao inequality. We will check below that K_{ν^*} is of course bigger than one.

Actually, we know that the Fisher information is infinite in the case of a non differentiable density. It is thus important here to ensure continuity and differentiability at the origin. In order to ensure that the Fisher information of the probability distribution $f(x) = u^2(x)$ remains finite, we need to impose $u(0) = 0$. This implies a condition on ν so that $D_\nu(0) = 0$. It is easy to check that

$$D_\nu(0) = \frac{\sqrt{\pi}}{2^{\frac{\nu}{2}}\Gamma\left(\frac{1}{2} - \frac{1}{2}\nu\right)}. \tag{30}$$

Therefore, $D_\nu(0) = 0$ if and only if $\Gamma\left(\frac{1}{2} - \frac{1}{2}\nu\right) \to +\infty$, that is if and only if ν is an odd positive integer. In such a case, the Weber functions can also be expressed in terms of Hermite polynomials $H_n(x)$:

$$D_n(x) = 2^{-\frac{n}{2}}e^{-\frac{x^2}{4}}H_n(x). \tag{31}$$

Then, the determination of the optimum value ν^* of ν such that $I_\nu(\sigma^2)$ is minimum amounts to minimize K_ν with ν integer. This leads to the following result.

Proposition 3. *For a given variance σ^2, the distribution on $\mathbb{R}^+$ which minimizes the Fisher information is obtained for $\nu^* = 1$, and is*

$$f_\xi(x) = \sqrt{\frac{2}{\pi}}\xi^{\frac{3}{2}}x^2 \exp\left(-\frac{\xi x^2}{2}\right), \tag{32}$$

which is the chi-distribution with three degrees of freedom, and where the parameter ξ is given by $\xi = \left(3 - \frac{8}{\pi}\right)/\sigma^2$. Its Fisher information, according to (45), is $I_{1,\xi} = 3\xi$. Then the minimum Fisher-variance product is $K_1 = 9 - 24/\pi \approx 1.3606$.

Proof. In the present case, it is possible to obtain closed-form formulas for the variance and information associated with (27), even for non integer values of ν.

The following relationships, which can be derived from integral representations of parabolic cylinder functions, see [31, 6.2] are useful:

$$\frac{d}{dx}D_\nu(x) = (-1/2)xD_\nu(x) + \nu D_{\nu-1}(x) \tag{33}$$

$$= (1/2)xD_\nu(x) - D_{\nu+1}(x) \tag{34}$$

Adding the two equalities and taking the square, we have

$$4\left(\frac{d}{dx}D_\nu(x)\right)^2 = \nu^2 D_{\nu-1}(x)^2 - 2\nu D_{\nu-1}(x)D_{\nu+1}(x) + D_{\nu+1}(x)^2. \tag{35}$$

Subtracting the Eq. (33) and (34), we obtain

$$xD_\nu(x) = D_{\nu+1}(x) - \nu D_{\nu-1}(x), \tag{36}$$

from which we deduce

$$xD_\nu(x)^2 = D_{\nu+1}(x)D_\nu(x) - \nu D_{\nu-1}(x)D_\nu(x), \tag{37}$$

$$x^2 D_\nu(x)^2 = D_{\nu+1}(x)^2 - 2\nu D_{\nu+1}(x)D_{\nu-1}(x) + \nu^2 D_{\nu-1}(x)^2, \tag{38}$$

The second ingredient of the calculation are the formulas [14, Eqs. 7.711.2 and 7.711.3], from which we define the functional $S(\mu,\nu)$:

$$S(\mu,\nu) = \int_0^{+\infty} D_\mu(x)D_\nu(x)\mathrm{d}x = \frac{\pi 2^{\mu+\nu+1}}{\mu-\nu}\left[\frac{1}{\Gamma\left(\frac{1}{2}-\frac{1}{2}\mu\right)\Gamma\left(-\frac{1}{2}\nu\right)} - \frac{1}{\Gamma\left(\frac{1}{2}-\frac{1}{2}\nu\right)\Gamma\left(-\frac{1}{2}\mu\right)}\right] \tag{39}$$

for $\mu \neq \nu$, and with

$$S(\nu,\nu) = \int_0^{+\infty} D_\nu(x)^2 \mathrm{d}x = \pi^{\frac{1}{2}} 2^{-\frac{3}{2}} \frac{\Psi\left(\frac{1}{2}-\frac{1}{2}\nu\right) - \Psi\left(-\frac{1}{2}\nu\right)}{\Gamma(-\nu)} \tag{40}$$

where $\Psi(x)$ is the digamma function. When μ or ν are integers, these formulas reduce to

$$S(2p, 2p+k) = -\frac{\pi 2^{2p+\frac{1}{2}k+\frac{1}{2}}}{k}\frac{1}{\Gamma\left(\frac{1}{2}-p\right)\Gamma\left(-p-\frac{1}{2}k\right)} \tag{41}$$

$$S(2p+1, 2p+1+k) = \frac{\pi 2^{2p+\frac{1}{2}k+\frac{3}{2}}}{k}\frac{1}{\Gamma\left(-\frac{1}{2}-p\right)\Gamma\left(-p-\frac{1}{2}k\right)} \tag{42}$$

$$\text{and } S(m,m) = (2\pi)^{\frac{1}{2}}\frac{m!}{2} \tag{43}$$

So doing, using the expression of the solution (27), the definition of the Fisher information (4), equality (35) and the definitions (39) and (40), we obtain the expression

$$I_{\nu,\xi} = \xi(\nu^2 S(\nu-1,\nu-1) - 2\nu S(\nu-1,\nu+1) + S(\nu+1,\nu+1))/S(\nu,\nu), \tag{44}$$

which reduces to the very simple expression

$$I_{m,\xi} = (2m+1)\xi \tag{45}$$

in the integer case. Let us mention that a similar expression is reported in [36] for the case of Hermite functions, up to a factor 2 which is due to a different definition of Hermite functions. In the same way, the integration of equalities (37) and (38) gives the first and second order moment, so the variance is

$$V_{\nu,\xi} = \frac{1}{\xi}(S(\nu+1,\nu+1) - 2\nu S(\nu+1,\nu-1) + \nu^2 S(\nu-1,\nu-1))/S(\nu,\nu) - ((S(\nu+1,\nu) - \nu S(\nu-1,\nu))/S(\nu,\nu))^2 \tag{46}$$

Figs. 1 and 2 present the evolution of the variance and information, as computed in (44) and (46). Fig. 3 gives the information-variance product K_ν, as a function of ν. Actually, we know that the true Fisher information is only finite for positive odd values of ν (otherwise the density is discontinuous at the origin and the Fisher information is infinite).

Considering Fig. 3, we read that the positive odd integer which minimizes the product $I_{\nu,1} \cdot V_{\nu,1}$ is $\nu^* = 1$. Accordingly, we obtain that the solution $f(x) = u(x)^2$ is (32). The values of its variance and Fisher information follow by direct computation. □

Note that disregarding the differentiability requirement at the origin would lead to select the parameter $\nu = 0.1065$, corresponding to a variance $\sigma^2 = 0.38661$ and a "Fisher information" $I = 0.91886$. So doing one would obtain a product $I \cdot \sigma^2 = 0.35524$, which would break the Cramer-Rao inequality. From the estimation theory point of view, it is clear that if the density has a bounded support and a discontinuity at the left endpoint of this support, then the variance of the estimate of the location parameter will be asymptotically zero, which corresponds to an infinite Fisher information in the Cramér-Rao bound. Clearly, the estimator defined as the minimum of the experimental data converges to the value of the left endpoint of the support and, in turn, provides an estimate of the value of the location parameter with asymptotically zero-variance.

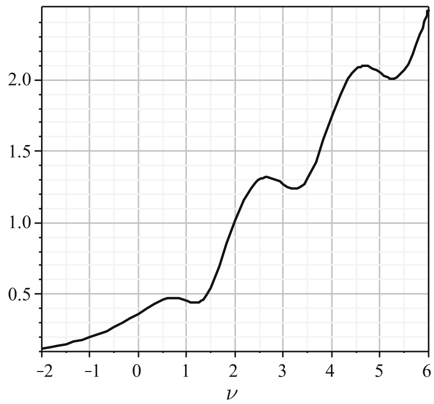

Fig. 1. Evolution of the variance $V_{\nu,1}$ in (44) with respect to ν.

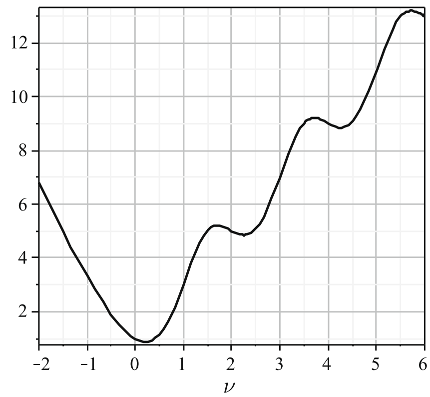

Fig. 2. Evolution of the information $I_{\nu,1}$ in (46) with respect to ν. Actually, the true Fisher information is only finite for positive odd values of ν.

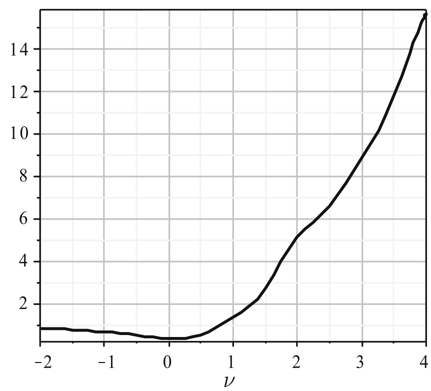

Fig. 3. The information-variance product K_ν as a function of ν.

5. Solutions with compact support

We know that the Fisher information associated with a distribution is invariant by translation of this distribution. Furthermore, a scaling of the distribution $f(x)$ according to $g(x) = \frac{1}{|a|} f\left(\frac{x}{a}\right)$ with $a \neq 0$ yields a scaling of the Fisher information as $I[g] = \frac{1}{|a|^2} I[f]$. Hence, it is possible to restrict our study to any particular interval, and without loss of generality, we choose the interval $[-1, 1]$.

We consider the same problem as before, the minimization of the Fisher information subject to a variance constraint, and add the boundaries conditions $u(1) = u(-1) = 0$, which reads

$$I(\sigma^2) = \inf_{u:u^2 \in \mathscr{D}} \left\{ 4\int_{-1}^{1} u'(x)^2 \mathrm{d}x : \int_{-1}^{1} u(x)^2 \mathrm{d}x = 1 \text{ with } Var(u^2) = \sigma^2 \text{ and } u(1) = u(-1) = 0 \right\} \tag{47}$$

The problem is invariant by the symmetry $x \mapsto -x$ since it is clear that if $u(x)$ is solution, then $v(x) = u(-x)$, which has the same variance and Fisher information, is also a solution. In fact, the distribution with minimum Fisher information for a given variance is unique and even. The general proof of this fact is rather involved and given in [27,28]. But since we have here the general expression of solutions to the underlying differential equation, it is not difficult to characterize this optimum solution.

Proposition 4. *The unique solution to the problem (47) is the non-negative function*

$$u(x) = \frac{\frac{1}{\sqrt{x}} M_{-\frac{\alpha}{4\sqrt{\beta}},-\frac{1}{4}}(\sqrt{\beta}x^2)}{\left(\int_{-1}^{1} \left(\frac{1}{\sqrt{x}} M_{-\frac{\alpha}{4\sqrt{\beta}},-\frac{1}{4}}(\sqrt{\beta}x^2) \right)^2 \mathrm{d}x \right)^{\frac{1}{2}}}, \tag{48}$$

where $\sqrt{\beta}$ is the first zero of the function $M_{-\frac{\alpha}{4\sqrt{\beta}},-\frac{1}{4}}$.

Proof. We have already established that the general solution of the differential Eq. (9) can be expressed as a linear combination of two Whittaker M functions which are two linearly independent solutions:

$$u(x) = c_1 \frac{1}{\sqrt{x}} M_{-\frac{\alpha}{4\sqrt{\beta}},-\frac{1}{4}}\left(\sqrt{\beta}x^2\right) + c_2 \frac{1}{\sqrt{x}} M_{-\frac{\alpha}{4\sqrt{\beta}},\frac{1}{4}}\left(\sqrt{\beta}x^2\right), \tag{49}$$

where c_1 and c_2 shall be chosen such that the boundaries and normalization conditions are satisfied. In this last equation, the first function is even while the second is odd, as it can be observed from the series development (21). Let us note them $S_{\text{even}}(x)$ and $S_{\text{odd}}(x)$, respectively. The boundaries conditions $u(1) = u(-1) = 0$ then imply that

$$c_1 S_{\text{even}}(1) + c_2 S_{\text{odd}}(1) = 0, \tag{50}$$
$$c_1 S_{\text{even}}(1) - c_2 S_{\text{odd}}(1) = 0. \tag{51}$$

The only solution is $c_1 = c_2 = 0$, except if $S_{\text{odd}}(1)$ and $S_{\text{even}}(1)$ are simultaneously equal to zero. But it is easy to check that these two functions have no common zero. Therefore, the general solution can not include both terms. There shall be only one term, either $S_{\text{even}}(x)$ or $S_{\text{odd}}(x)$, in the general expression of the solution with its parameters adjusted so as to ensure the boundaries constraints.

We deduce that, since $u(x)$ is either odd or even, the associated density $f(x) = u^2(x)$ is even and has zero mean. In such a case, the set defined by the constraint $\mathrm{Var}[f] = \sigma^2$ becomes a convex set and consequently the solution to the minimization of the Fisher information on this convex set is unique. From the fact both $u(x)$ and $v(x) = u(-x)$ are solution, we deduce that the unique solution is necessary the even one, that is (48).

Finally, if $u(x)$ is solution, so is $w(x) = |u(x)|$ because $w'(x)^2 = u'(x)^2$ and $w(x)^2 = u(x)^2$. By uniqueness of the solution we obtain that $u(x) \geqslant 0$ for $x \in (-1,1)$. Therefore the condition $u(1) = 0$, the fact that all zeros of $M_{-\frac{\alpha}{4\sqrt{\beta}},-\frac{1}{4}}(x)$ are zero-crossings, and the non-negativity requirement, yield that the point $\sqrt{\beta}$ is the first zero of $M_{-\frac{\alpha}{4\sqrt{\beta}},-\frac{1}{4}}$. □

In Fig. 4, we give some examples of solutions for several values $\sigma^2 \in [0,1]$ of the variance. We see that for low variances the solution is unimodal while it is bimodal for higher variances. The Fisher information is large for low and high variances and much lower for intermediate variances. In fact, the minimum Fisher information is a convex function of the variance, cf [28, Theorem 6.1].

Proposition 5. *The minimum Fisher information $I(\sigma^2)$ is a strictly convex function of σ^2.*

Proof. Let $u(x)^2$ and $v(x)^2$ be two distributions with minimum informations $I(\sigma_u^2)$ and $I(\sigma_v^2)$ respectively. Since the Fisher information is strictly convex, with $\epsilon \in (0,1)$, we have

$$I[\epsilon u^2 + (1-\epsilon)v^2] < \epsilon I[u^2] + (1-\epsilon)I[v^2] = \epsilon I(\sigma_u^2) + (1-\epsilon)I(\sigma_v^2) \tag{52}$$

Since the distributions u^2 and v^2 have zero mean, we have

$$Var[\epsilon u^2 + (1-\epsilon)v^2] = \epsilon\sigma_u^2 + (1-\epsilon)\sigma_v^2.$$

Then, there exists a distribution with the same variance and minimum Fisher information such that

$$I[\epsilon u^2 + (1-\epsilon)v^2] \geqslant I(\epsilon\sigma_u^2 + (1-\epsilon)\sigma_v^2). \tag{53}$$

Finally, combination of (52) and (53) yields

$$I(\epsilon\sigma_u^2 + (1-\epsilon)\sigma_v^2) < \epsilon I(\sigma_u^2) + (1-\epsilon)I(\sigma_v^2). \qquad \square \tag{54}$$

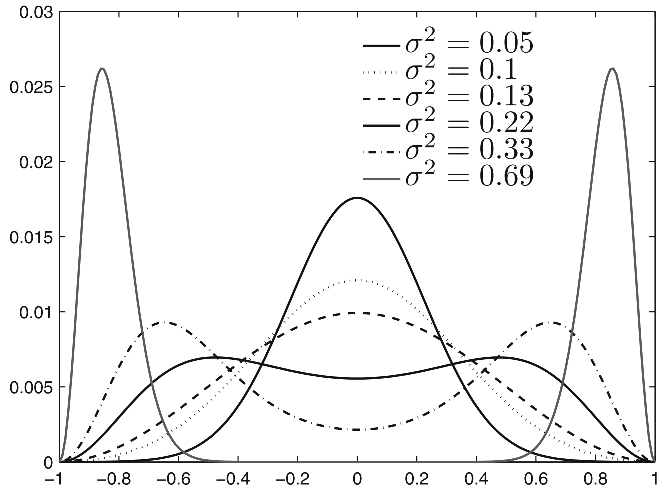

Fig. 4. Minimum Fisher information distributions with compact (0,1) support for several values of variances. For variances lower than $\sigma_*^2 = \frac{1}{3} - \frac{2}{\pi^2} \approx 0.1307$, the distributions are unimodal while they become bimodal when $\sigma^2 > \sigma_*^2$.

In the case of the $\mathbb{R}^+$ support, we obtained explicit expressions of both the variance and Fisher information, and as a result the exact behavior of $I(\sigma^2)$. The present case is more delicate. Indeed, although it is possible to obtain from (21) an exact series expansion of $u^2(x)$, where the coefficients depend of hypergeometric functions, the difficulty here is that we have no analytical expression for the argument of the first zero of the solution, which gives the value of β. Only approximations of this value are available [1, p. 510]. Therefore, we have to resort to a numerical determination of the parameters of the function. Similarly, we cannot give a close form formula for $I(\sigma^2)$. However, we can still characterize its general behavior.

Since $I(\sigma^2)$ is a convex function, it has a unique minimum, say σ_*^2. The following result shows that this minimum is obtained for $\beta = 0$. Furthermore, this value discriminates two regimes for the solutions: in the case $\sigma < \sigma_*$, we have $\beta > 0$ while in the case $\sigma > \sigma_*$ we have $\beta < 0$ and the corresponding solution is the Whittaker M function with imaginary arguments. These facts have already been noticed in [28, Lemmas 5.2 and 5.3], but we provide here an alternate proof.

Proposition 6. *f σ_*^2 is the minimizer of $I(\sigma^2)$, then*

(a) *for* $\sigma < \sigma_*, \quad \beta > 0$
(b) *for* $\sigma > \sigma_*, \quad \beta < 0$
(c) *and finally for* $\sigma = \sigma_*, \quad \beta = 0$.

Proof. Let $u(x)^2$ and $v(x)^2$ be two distributions with minimum informations $I(\sigma_u^2)$ and $I(\sigma_v^2)$ respectively, and let us define $g_\epsilon(x) = \epsilon u^2(x) + (1-\epsilon)v^2(x)$. By Proposition 1, the Fisher information $I(\sigma^2)$ can be expressed as $I(\sigma^2) = -4(\alpha_v + \beta_v \sigma_v^2)$. We use the expansion proved in [27, Satz 7.2, p. 90]:

$$I[g_\epsilon] = I[v^2] - 4\beta_v \epsilon(\sigma_u^2 - \sigma_v^2) + o(\epsilon^2). \tag{55}$$

Strict convexity gives

$$I[g_\epsilon] < \epsilon I[u^2] + (1-\epsilon)I[v^2] = \epsilon I(\sigma_u^2) + (1-\epsilon)I(\sigma_v^2). \tag{56}$$

Let us now take $\sigma_u = \sigma_*$. Since $I(\sigma_*^2)$ is the minimum Fisher information, we have the majorization

$$\epsilon I(\sigma_*^2) + (1-\epsilon)I(\sigma_v^2) \leqslant I(\sigma_v^2), \tag{57}$$

and therefore $I[g_\epsilon] < I(\sigma_v^2)$. As a consequence, from (55) and the previous inequality, we obtain

$$\beta_v \epsilon(\sigma_*^2 - \sigma_v^2) > 0, \tag{58}$$

which gives cases (a) and (b) in the Proposition. The case (c), $\beta = 0$, follows by continuity. □

Hence, the solution which realizes the minimum $I(\sigma_*^2)$ of the Fisher information corresponds to $\beta = 0$. This means that this solution satisfies the differential equation $u''(x) = \alpha u(x)$, with $u(1) = u(-1) = 0$. This problem has as solutions $u(x) = \cos(k\pi x/2)$, with $\alpha = -k^2\pi^2$ and k integer. Since we know that the solution of the Fisher minimization is non-negative, there is only one possibility and

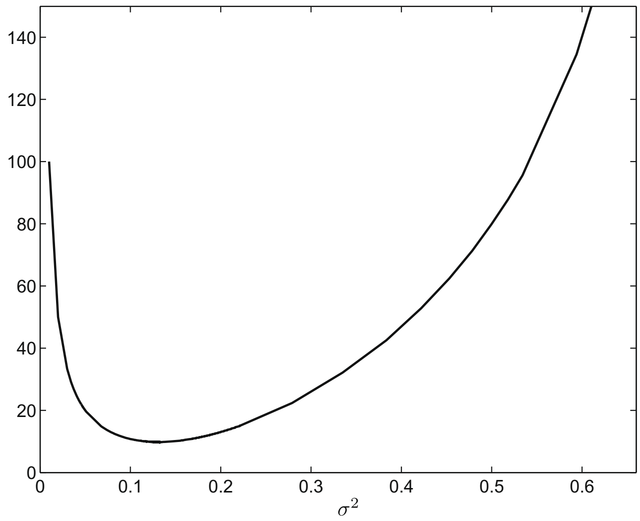

Fig. 5. Evolution of the minimum Fisher information $I(\sigma^2)$ in the case of a distribution with bounded support $(-1,1)$.

Proposition 7. *The probability density function $f(x) = u(x)^2$ defined on $[-1,1]$ with minimum Fisher information is*

$$f(x) = \cos^2(\pi x/2) \tag{59}$$

$$\text{with } \sigma_*^2 = \frac{1}{3} - \frac{2}{\pi^2} \text{ and } I(\sigma_*^2) = \pi^2. \tag{60}$$

Fig. 5 reports the behavior of the Fisher information $I(\sigma^2)$. The Fisher information tends to infinity for $\sigma \to 0$ and $\sigma \to +\infty$ and its minimum is attained for $\sigma^2 = \sigma_*^2$. For $\sigma \to 0$, we obtain by [1, Eqs. 13.1.32 and 13.5.5] that $u(x) \propto \exp\left(-\beta^{\frac{1}{2}}x^2/2\right)$, that is $u(x)$ converges to a Gaussian distribution with variance $\beta^{-\frac{1}{2}}$. Hence, for small values of the variance, the solution has the form of a Gaussian distribution concentrated on the origin, and, as a normal distribution, its Fisher information decreases as $1/\sigma^2$. When σ^2 increases towards 1, the two modes become more and more pronounced and probability accumulates on ± 1, and the probability density function tends to two mass functions, as shown in Fig. 4, and the Fisher information tends to infinity.

6. Conclusion

In this paper, we have solved the problem of minimizing the Fisher information on restricted supports with a fixed variance. The problem has been stated under the form of a general second order linear differential equation. We have shown that the general form of the solutions involves Whittaker functions. We have derived the explicit expressions of the solutions on $\mathbb{R}^+$ and on an interval. We have first studied the set of solutions on $\mathbb{R}^+$ and shown that the distribution with minimum Fisher information is a scaled chi distribution. On the interval $[-1,+1]$, we have characterized the solutions, investigated their behavior, and shown that the distribution with minimum Fisher information is a squared cosine function.

Future work will consider the extension of these results in the multivariate case. We also intend to investigate some interesting generalized versions of Fisher information [18,4] as suggested by one of the referees of this paper.

Acknowledgements

The authors greatly acknowledge Elke Uhrmann-Klingen for friendly correspondence and for communicating us her manuscript [27]. C.V. acknowledges interesting conversations with R. Frieden and B. Soffer during the 3rd International Workshop on Fisher Information held in Tucson, April 2008.

References

[1] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables, ninth Dover printing, tenth GPO printing ed., Dover Publications, 1964.
[2] M.M. Ali, S. Nadarajah, Information matrices for normal and Laplace mixtures, Information Sciences 177 (3) (2007) 947–955.
[3] M. Balasco, V. Lapenna, M. Lovallo, G. Romano, A. Siniscalchi, L. Telesca, Fisher information measure analysis of earth's apparent resistivity, International Journal of Nonlinear Science 5 (3) (2008) 230–236.
[4] D. Boekee, An extension of the Fisher information measure, in: I. Csiszár, P. Elias (eds.), Topics in Information Theory Proc. 2nd Colloquium on Information Theory; Keszthely, Hungary, 25–29 August 1975, vol. 16, János Bolyai Mathematical Society and North-Holland, Keszthely, Hungary, 1977.

[5] J.M. Borwein, A.S. Lewis, M.N. Limber, D. Noll, Maximum entropy reconstruction using derivative information part 2: computational results, Numerische Mathematik 69 (3) (1995) 243–256.
[6] J.M. Borwein, A.S. Lewis, D. Noll, Maximum entropy reconstruction using derivative information, part 1: Fisher information and convex duality, Mathematics of Operations Research 21 (1996) 442–468.
[7] M. Cohen, The Fisher information and convexity, IEEE Transactions on Information Theory 14 (4) (1968) 591–592.
[8] A. Dasgupta, Stein's identity, Fisher information, and projection pursuit: a triangulation, Journal of Statistical Planning and Inference 137 (11) (2007) 3394–3409.
[9] S.A. Frank, Natural selection maximizes Fisher information, Journal of Evolutionary Biology 22 (2) (2009) 231–244.
[10] B.R. Frieden, Fisher information, disorder, and the equilibrium distributions of physics, Physical Review A 41 (8) (1990) 4265–4276.
[11] B.R. Frieden, Science from Fisher Information: A Unification, Cambridge University Press, 2004.
[12] B.R. Frieden, P.M. Binder, Physics from Fisher information: a unification, American Journal of Physics 68 (2000) 1064.
[13] I.M. Gelfand, S.V. Fomin, Calculus of Variations, Dover Publications, 2000.
[14] I. Gradshteyn, I. Ryzhik, Table of Integrals, Series and Products, Academic P, 1966.
[15] P.J. Huber, Robust statistics: a review, The Annals of Mathematical Statistics 43 (3) (1972) 1041–1067.
[16] M. Lovallo, F. Marchese, N. Pergola, L. Telesca, Fisher information measure of temporal fluctuations in satellite advanced very high resolution radiometer (AVHRR) thermal signals recorded in the volcanic area of Etna (Italy), Communications in Nonlinear Science and Numerical Simulation 14 (1) (2009) 174–181.
[17] M.T. Martin, F. Pennini, A. Plastino, Fisher's information and the analysis of complex signals, Physics Letters A 256 (2-3) (1999) 173–180.
[18] A. Mathai, Basic Concepts in Information Theory and Statistics: Axiomatic Foundations and Applications, Wiley, New York, 1975.
[19] M. Montenegro, M.R. Casals, M.A. Lubiano, M. Ángeles Gil, Two-sample hypothesis tests of means of a fuzzy random variable, Information Sciences 133 (1-2) (2001) 89–100.
[20] A. Nagy, Fisher information in density functional theory, The Journal of Chemical Physics 119 (18) (2003) 9401–9405.
[21] S. Nordebo, M. Gustafsson, B. Nilsson, Fisher information analysis for two-dimensional microwave tomography, Inverse Problems 23 (3) (2007) 859–877.
[22] E. Romera, J.S. Dehesa, Weizsäcker energy of many-electron systems, Physical Review A 50 (1) (1994) 256–266.
[23] K.D. Sen, J. Antolín, J.C. Angulo, Fisher–Shannon analysis of ionization processes and isoelectronic series, Physical Review A 76 (3) (2007) 32502.
[24] J. Sánchez-Ruiz, J.S. Dehesa, Fisher information of orthogonal hypergeometric polynomials, Journal of Computational and Applied Mathematics 182 (1) (2005) 150–164.
[25] L. Telesca, V. Lapenna, M. Lovallo, Fisher information analysis of earthquake-related geoelectrical signals, Natural Hazards and Earth System Sciences 5 (4) (2005) 561–564.
[26] C. Tsairidis, K. Ferentinos, T. Papaioannou, Information and random censoring, Information Sciences 92 (1–4) (1996) 159–174.
[27] E. Uhrmann-Klingen, Fisher-Minimale Dichten auf Kompacten intervallen (in German), Ph.D. thesis, Essen, 1992.
[28] E. Uhrmann Klingen, Minimal Fisher information distributions with compact-supports, Sankhya: The Indian Journal of Statistics 57 (3) (1995) 360–374.
[29] C. Vignat, J.F. Bercher, Analysis of signals in the Fisher–Shannon information plane, Physics Letters A 312 (1-2) (2003) 27–33.
[30] C. Vignat, J.-F. Bercher, On Fisher information inequalities and score functions in non-invertible linear systems, Journal of Inequalities in Pure and Applied Mathematics 4 (4) (2003) 71.
[31] Z.X. Wang, D.R. Guo, Special Functions, World Scientific Pub. Co. Inc., 1989.
[32] E.T. Whittaker, G.N. Watson, A Course of Modern Analysis, 4th ed., Cambridge University Press, 1927.
[33] E.K.H. Wu, Distributions minimizing Fisher information for scale in [var epsilon]-contamination neighbourhoods, Statistics & Probability Letters 14 (5) (1992) 373–383.
[34] E.K.H. Wu, P.S. Chan, Distributions minimizing Fisher information for location in Kolmogorov neighbourhoods, Annals of the Institute of Statistical Mathematics 49 (3) (1997) 541–554.
[35] B. Xu, Q. Chen, Z. Wu, Z. Wang, Analysis and approximation of performance bound for two-observer bearings-only tracking, Information Sciences 178 (8) (2008) 2059–2078.
[36] R.J. Yáñez, P. Sánchez-Moreno, A. Zarzo, J.S. Dehesa, Fisher information of special functions and second-order differential equations, Journal of Mathematical Physics 49 (2008) 082104.
[37] V. Zivojnovic, Minimum Fisher information of moment-constrained distributions with application to robust blind identification, Signal Processing 65 (2) (1998) 297–313.
[38] V. Zivojnovic, D. Noll, Minimum Fisher information spectral analysis, in: ICASSP-97, IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5, 1997.