Statistics and Probability Letters 129 (2017) 167–170

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Minimizing Fisher information with absolute moment constraints

Philip A. Ernst

Department of Statistics, Rice University, Houston, TX 77005, USA

ARTICLE INFO

Article history:
Received 25 November 2016
Received in revised form 16 May 2017
Accepted 19 May 2017
Available online 29 May 2017

MSC 2010:
primary 62B10
secondary 94A17

Keywords:
Fisher information
Absolute moments
Asymmetric moments

ABSTRACT

Given $p > 0$, we minimize the Fisher information of the probability measure μ with density g with given first moment $\int xg(x)dx = m$ and given absolute pth moment $\int |x|^p g(x)\,dx = c^p$.

1. Introduction

The goal of this article is to solve the following problem: given $p > 0$, minimize the Fisher information of the probability measure μ with density g with given first moment $\int xg(x)dx = m$ and given *absolute* pth moment, $\int |x|^p g(x)\,dx = c^p$.

We first offer a brief summary of related literature on the minimization of Fisher information. Wu (1992) studied the problem of finding the distributions minimizing Fisher information for scale over ϵ-contamination neighborhoods of distribution functions G satisfying certain mild conditions. Uhrmann-Klingen (1995) classified the probability distributions minimizing Fisher information with fixed variance and with support on $[-1, 1]$. Uhrmann-Klingen's results motivated the work of Bercher and Vignat (2009), which solved the problem of minimizing Fisher information among distributions with fixed variance defined either on a bounded subset S of $\mathbb{R}$ or on the positive real line. The work of Živojnović (1998) concerned minimizing Fisher information higher-order moment constraints (but not higher-order *absolute* moment constraints). Landsman (2000) illustrated the relationship between the singular Sturm–Liouville problem and minimizing Fisher information for the scale parameter. A special case of our problem of interest was studied by Huber (1981), who examined minimizing Fisher information subject to $g = (1 - a)n + a * h$, where g is a density, n is the Gaussian density, a is a constant, and h is arbitrary function. For a comprehensive overview of the pivotal role of Fisher information in both mathematical statistics and information theory, we refer the reader to Fisher (1956), Kagan et al. (1973), Bickel and Collins (1983), DasGupta (2008), Lehmann (2011), and Bickel and Doksum (2015).

The work most immediate to our problem is that of Kagan (1986), whose results overlap with ours when minimizing Fisher information *given only the first and second moments*. Kagan (1986) considers the distribution of an exponential family in the following canonical form

$$f(x; \boldsymbol{\theta}) = \exp\left(C_0(\boldsymbol{\theta}) + \sum_{i=1}^{m} C_j(\boldsymbol{\theta})\varphi_j(x) + \varphi_{m+1}(x)\right), \tag{1}$$

E-mail address: philip.ernst@rice.edu.

http://dx.doi.org/10.1016/j.spl.2017.05.015
0167-7152/

where $C_j(\boldsymbol{\theta})$ is the natural parameter and $\varphi_j(x)$ is the sufficient statistic. The key result in Kagan (1986) is that (under suitable conditions) the exponential family minimizes the Fisher information $I(\boldsymbol{\theta})$ given the following constraints.

1. The probability measure is absolutely continuous with respect to some measure μ.
2. For $i, j = 0, 1, \ldots, m$ (i can be equal to j),

$$\mathbb{E}\left[\varphi_i(x)\varphi_j(x)\right] = a_{ij}, \tag{2}$$

where the a_{ij} denote given constants.

When $m = 1$ and $\varphi_1(x) = x$, Eq. (2) can be written as

$$\mathbb{E}[X] = a_1, \qquad \mathbb{E}\left[X^2\right] = a_2 \quad a_1, a_2 \in \mathbb{R},$$

which corresponds to the task of minimizing Fisher information given the mean and the variance. Kagan (1986) concludes by showing that the normal distribution minimizes Fisher information if the support of X is $\mathbb{R}$ and by showing that the gamma distribution minimizes the Fisher information if the support of X is $\mathbb{R}^+$. This extends the earlier work of Klebanov and Melamed (1978).

The purpose of this note is to both complement Kagan (1986) and to offer more general results than those provided in Živojnović (1998). The organization of the manuscript is as follows. Corollary 2.2 in Section 2 provides a minimizing equation for the minimization of Fisher information with absolute moment constraints. Section 3 considers asymmetric moments. The author's hope is that this work will enable readers to solve applied problems of interest involving Fisher information with absolute moment constraints.

2. Main results

If μ is a probability measure on $(-\infty, \infty)$, the Fisher information $I(\mu)$ is defined as

$$I(\mu) = \int_{-\infty}^{\infty} \frac{\left(g'(x)\right)^2}{g(x)} dx. \tag{3}$$

Note that Eq. (3) holds only when the integral converges, μ is absolutely continuous with respect to Lebesgue measure, and μ has a density g which is differentiable. Otherwise, $I(\mu) = \infty$. We wish to calculate

$$\inf_{\mu}\left[I(\mu) : \int_{-\infty}^{\infty} \phi_i(x)g(x)dx = c_i, i = 0, \ldots, n\right], \tag{4}$$

where the ϕ_i are n given functions and $\phi_0 \equiv 1$. The c_i are the moment constraints.

It is well known (see, for example, Cohen 1968) that I is a convex functional of μ, i.e., that

$$I\left(\frac{\mu_1 + \mu_2}{2}\right) \leq \left(\frac{I(\mu_1) + I(\mu_2)}{2}\right),$$

for all μ_1 and μ_2 with strict equality holding when $\mu_1 = \mu_2$. By the convexity of I, there is a tangent hyperplane to the surface $(I(\mu), \mu)$ at every point $(I(\mu_0), \mu_0)$. The surface lies over the tangent hyperplane at μ_0. Any hyperplane must be of the form:

$$\int_{-\infty}^{\infty} \psi(x)\mu(dx) = \int_{-\infty}^{\infty} \psi(x)\mu_0(dx),$$

for some function ψ depending on μ_0. Let $g = g_0$ be the density of μ_0. To determine ψ, we first consider varying the density $g = g_0$ of μ_0 to $g + \epsilon h$, where $\epsilon > 0$, $h = h(x)$, and $h(x) \geq 0$ if $g_0(x) = 0$. For $i = 1, \ldots, n$, h is considered to be a legitimate "direction of variation" if h satisfies

$$\int_{-\infty}^{\infty} h(x)dx = \int_{-\infty}^{\infty} h(x)\phi_i(x)dx = 0 \tag{5}$$

for arbitrary $\phi_i(x)$ orthogonal to h.

Now let us assume that g is a minimizing density (either globally, or at the very least a local minimum). Then,

$$I(g + \epsilon h) \geq I(g).$$

Since g gives a critical point,

$$\frac{d}{d\epsilon} I(g + \epsilon h)\Big|_{\epsilon=0} = 0.$$

By straightforward calculus, we obtain

$$\frac{d}{d\epsilon} \int_{-\infty}^{\infty} \frac{\left(g'(x) + \epsilon h'(x)\right)^2}{g(x) + \epsilon h(x)} dx\Big|_{\epsilon=0} = \int_{-\infty}^{\infty} \frac{2g'(x)h'(x)}{g(x)} dx - \int_{-\infty}^{\infty} \left(\frac{g'(x)}{g(x)}\right)^2 h(x)dx.$$