An Alternative Proof For the Minimum Fisher Information of Gaussian Distribution

A. PAK

Abstract

Fisher information is of key importance in estimation theory. It is used as a tool for characterizing complex signals or systems, with applications, e.g. in biology, geophysics and signal processing. The problem of minimizing Fisher information in a set of distributions has been studied by many researchers. In this paper, based on some rather simple statistical reasoning, we provide an alternative proof for the fact that Gaussian distribution with finite variance minimizes the Fisher information over all distributions with the same variance.

Mathematics Subject Classification 2010: 62H12, 62H10, 62F12
Keywords: Fisher information, Gaussian distribution, Minimum risk equivariant estimator.

1. INTRODUCTION

The role of Fisher information as a way of measuring information in a distribution is well established in the literature. Fisher information is used in estimation theory for constructing a basic bound, known as Cramer-Rao lower bound (CRLB), on the variance of an estimator ([Khoolenjani and Alamatsaz (2016)]). Applications of Fisher information in geophysics ([Balasco et al. (2008)]), biology ([Frank (2009)]), analysing complex signals or systems ([Martin et al. (2009)], [Nagy (2003)]), signal processing ([Vignat and Bercher (2003)], [Zivojnovic and Noll (1997)]), computing the asymptotic covariance matrix of the models ([Hussin et al. (2010)] and [Mamun et al. (2013)]) and obtaining performance bounds ([Xu et al. (2008)]) are discussed in the literature. It is also used in statistical physics and biology as a way of inference and understanding ([Frieden (2009)]). Recently, [Dulek and Gezici (2014)] studied the maximization of Fisher information in presence of a constraint on the cost of measurements. [Neri et al. (2013)] studied the theoretical evaluation of the achievable performance using Fisher information.

Gaussian distribution is one of the most well-known and widely applied distributions in many fields such as statistics, engineering and physics. One of the major reasons why Gaussian distribution has become so prominent is because of the Central Limit Theorem (CLT) and the fact that the distribution of noise in numerous

10.2478/jamsi-2018-0008

DE GRUYTER OPEN

engineering systems is well fitted by Gaussian distribution. It is well known that Gaussian distribution minimizes the Fisher information, which equals to the inverse of Cramer-Rao lower bound, (see [Shao (1999)]). This fact is established in [Park et al. (2013)]. Especially, when there is no information about the distribution of observations, Gaussian assumption appears as the most traditional choice. Therefore, optimization of estimation methods based on the CRLB that holds under Gaussian distribution yields the best CRLB-related performance. [Stoica and Babu (2011)] provided a general proof of result that the largest CRLB is achievable by the Gaussian distribution. In this note, we provide a simple alternative proof for the fact that Gaussian distribution yields the minimum Fisher information. Using certain standard statistical reasoning, we believe that there is a value – in general, and also here – in presenting alternative proofs for fundamental theorems. Such alternative proofs can shed new light on the statement being proven, introduce new arguments that can be useful elsewhere, or yield different generalizations and applications.

2. MAIN RESULT

Let us first review the fundamental notions and basic definitions used in the paper.

Suppose that X is a random observable taking on values in a sample space $\mathscr{X}$ according to a probability distribution from the family $\mathscr{F} = \{f(x;\theta) : \theta \in \Theta\}$, in which θ is a deterministic parameter.

DEFINITION 1. The Fisher information $I(F)$ of a distribution F on the real line is defined as

$$I(F) = \int_{-\infty}^{\infty} \left(\frac{d \ln f(x;\theta)}{dx} \right)^2 f(x;\theta) dx, \tag{1}$$

where f denotes the density of F.

DEFINITION 2. An *estimator* is a real-valued function δ defined over the sample space $\mathscr{X}$. It is used to estimate an estimand, $g(\theta)$, a real-valued function of the parameter.

Quite generally, suppose that the consequences of estimating $g(\theta)$ by a value d are measured by $L(\theta,d)$. Of the *loss function* L, we shall assume that $L(\theta,d) \geq 0$ for all θ, d and $L[\theta, g(\theta)] = 0$ for all θ, so that the loss is zero when the correct value is estimated. The accuracy, or rather inaccuracy, of an estimator θ is then measured by the *risk function*

$$R(\theta,\delta) = E_\theta\{L[\theta,\delta(X)]\}. \tag{2}$$

DEFINITION 3. A set of functions $\{g(x) : g \in \mathscr{G}\}$ from the sample space $\mathscr{X}$ onto $\mathscr{X}$ is called a *group of transformations* of $\mathscr{X}$ if

i. (Inverse) For every $g \in \mathscr{G}$ there is a $g' \in \mathscr{G}$ such that $g'(g(x)) = x$ for all $x \in \mathscr{X}$.

ii. (Composition) For every $g \in \mathscr{G}$ and $g' \in \mathscr{G}$ there exists $g'' \in \mathscr{G}$ such that $g'(g(x)) = g''(x)$ for all $x \in \mathscr{X}$.

iii. (Identity) The identity, $e(x)$, defined by $e(x) = x$ is an element of $\mathscr{G}$.

DEFINITION 4. Let $\mathscr{G}$ be a group of transformations of the sample space $\mathscr{X}$. Then, the family $\mathscr{F} = \{f(x;\theta) : \theta \in \Theta\}$ is *invariant under the group* $\mathscr{G}$ if for every $\theta \in \Theta$ there exists a unique $\theta' \in \Theta$ such that $Y = g(X)$ has the distribution $f(y;\theta')$ if X has the distribution $f(x;\theta)$.

The θ' uniquely determined by θ is denoted by $\bar{g}(\theta)$.

DEFINITION 5. An estimation problem (Θ,δ,L) is said to be *invariant under the group* $\mathscr{G}$ if the family $\mathscr{F} = \{f(x;\theta) : \theta \in \Theta\}$ of distributions is invariant under $\mathscr{G}$ and if the loss function is invariant under $\mathscr{G}$ in the sense that for every $g \in \mathscr{G}$ and every δ in the class of estimators D, there exists a unique $\delta^* \in D$ such that

$$L(\theta,\delta) = L(\bar{g}(\theta),\delta^*) \quad \forall \theta \in \Theta. \tag{3}$$

The δ^* uniquely determined by g and δ is denoted by $\tilde{g}(\delta)$.

In an invariant estimation problem, an estimator δ is said to be *equivariant* if for all $g \in \mathscr{G}$

$$\delta(g(x)) = \tilde{g}(\delta(x)). \tag{4}$$

If an equivariant estimator exists and minimizes the risk function, it is called the *minimum risk equivariant* (MRE) estimator.

THEOREM 1. Let $\mathbf{Y} = (Y_1,\ldots,Y_n)$ be distributed as

$$f(\mathbf{y}-\theta) = f(y_1-\theta,\ldots,y_n-\theta),$$

$X_i = Y_i - Y_n$ and $\mathbf{X} = (X_1,\ldots,X_{n-1})$. Suppose that the loss function is given by $L(\theta,W) = (W-\theta)^2$ and that there exists a location invariant estimator δ_0 of θ with finite risk. Then, the minimum risk equivariant estimator of θ exists and is given by

$$\delta^*(\mathbf{Y}) = \delta_0(\mathbf{Y}) - E_0[\delta_0(\mathbf{Y}) \mid \mathbf{x}]$$

PROOF. See [Lehmann and Casella (1998)].

THEOREM 2. (Cramer-Rao Inequality) Let $Y_1, ..., Y_n$ be independent random variables with a common probability density $f_\theta(y)$ and $W(Y_1, ..., Y_n)$ be an unbiased estimator of θ. Then, under the regularity conditions we have

$$Var(W) \geq \frac{1}{nI(F)}. \tag{5}$$

PROOF. See [Lehmann and Casella (1998)].

THEOREM 3. Among all densities with mean θ and finite variance σ^2, Fisher information is minimized by Gaussian density.

PROOF. Let F be a univariate distribution with density f and fixed finite variance σ^2 and $Y_1, ..., Y_n$ be independently identically distributed random variables with density $f_\theta(y)$, where $\theta = E(Y_i)$ is a location parameter. Assume that $s_n(F)$ is the risk of the minimum risk equivariant estimator of θ under squared error loss $L(\theta, W) = (W - \theta)^2$. For Gaussian distribution with mean θ and finite variance σ^2, if we let $\delta_0 = \bar{Y}$ in Theorem 1, it follows that δ_0 is independent of $\mathbf{X}$ and hence $E_0[\bar{Y} \mid \mathbf{x}] = 0$. Thus, the minimum risk equivariant estimator of θ becomes $\bar{Y}$ with risk $E_\theta(\bar{Y} - \theta)^2 = \frac{\sigma^2}{n}$. On the other hand, we obtain Fisher information in the Gaussian case as

$$\begin{aligned}
I(N) &= \int_{-\infty}^{\infty} \left(\frac{d \ln f(y;\theta)}{dy} \right)^2 f(y;\theta) dy \\
&= \int_{-\infty}^{\infty} \left(\frac{d}{dy} \ln \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y-\theta)^2} \right)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y-\theta)^2} dy \\
&= \frac{1}{\sigma^4} \int_{-\infty}^{\infty} \frac{(y-\theta)^2}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y-\theta)^2} dy \\
&= \frac{1}{\sigma^2}.
\end{aligned} \tag{6}$$

Therefore, we have

$$E_\theta(\bar{Y} - \theta)^2 = \frac{1}{nI(N)}. \tag{7}$$

We know that for any distribution F, $\bar{Y}$ is an unbiased estimator of θ with risk given in (7). So, the risk of the minimum risk equivariant estimator for any distribution F must be less than $1/(nI(N))$. Now, let b be the constant bias of the MRE estimator δ^*. Then, $\delta_1(y) = \delta^*(y) - b$ is a location invariant estimator of θ and the risk of δ_1 under squared error loss becomes

$$R_{\delta_1} = E[\delta^*(y) - b - \theta]^2 = Var(\delta^*) \leq Var(\delta^*) + b^2 = R_{\delta^*}.$$

Since δ^* is the MRE estimator, $b = 0$, i.e., δ^* is unbiased (see [Shao (1999)], p. 215). Therefore, by using Theorem 2, we have

$$s_n(F) \geq \frac{1}{nI(F)}.$$

Thus, $I(N) \leq I(F)$ and the proof is complete.

3. CONCLUSION

This paper focuses on deriving an alternative proof for the fact that the Fisher information is minimized by Gaussian distribution. The risk of the sample mean in Gaussian density is used to obtain an upper bound for the risk of the minimum risk equivariant estimator for any other distribution F. Then, applying Cramer-Rao inequality, a lower bound is obtained for the risk of the minimum risk equivariant estimator, regardless of F. Combining these two bounds, the result is concluded.

REFERENCES

Balasco, M., Lapenna, V., Lovallo, M., Romano, G., Siniscalchi, A. and Telesca, L. (2008). Fisher information measure analysis of earth's apparent resistivity. *International Journal of Nonlinear Science* **5**(3) 230–236.

Dulek, B. and Gezici, S. (2014). Average Fisher information maximisation in presence of cost-constrained measurements. *Electronics Letters* **47**(11) 654–656.

Frank, S.A. (2009). Natural selection maximizes Fisher information. *Journal of Evolutionary Biology* **22**(2) 231–244.

Frieden, B.R. (2009). Fisher information, disorder, and the equilibrium distributions of physics. *Physical Review* **A-41**(8) 4265–4276.

Hussin, A.G., Abuzaid, A., Zulkifili, F. and Mohamed I. (2010). Asymptotic covariance and detection of influential observations in a linear functional relationship model for circular data with application to the measurements of wind directions. Science Asia, 36, 249-253.

Khoolenjani, N.B. and Alamatsaz, M.H. (2016). A De Bruijn's identity for dependent random variables based on copula theory. *Probability in the Engineering and Informational Sciences*, 30(1), 125-140.

Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation, 2nd ed*, New York: Springer.

Mamun, S.M.A., Hussin, G.A., Zubairi, Y.Z. and Imon, R.A.H.M. (2013). Maximum likelihood estimation of linear structural relationship model parameters assuming the slope is known. Science Asia, 39, 561-565.

Martin, M.T., Pennini, F. and Plastino, A. (2009). Fisher's information and the analysis of complex signals. *Physics Letters* **A-256**(2-3) 173–180.

Nagy, A. (2003). Fisher information in density functional theory. *The Journal of Chemical Physics* **119**(18) 9401–9405.

Neri, A., Carli, M. and Battisti, F. (2013). Maximum likelihood estimation of depth field for trinocular images. *Electronics Letters* **49**(6) 394–396.

Park, S., Serpedin, E. and Qaraqe, K. (2013). Gaussian assumption: The least favorable but the most useful. *IEEE Signal Processing Magazine* **30** 183–186.

Shao, J. (1999). *Mathematical statistics*, New York: Springer-Verlag.

Stoica, P. and Babu, P. (2011). The Gaussian data assumption leads to the largest Cramer-Rao bound. *IEEE Signal Processing Magazine* **28** 132–133.

Vignat, C. and Bercher, J.F. (2003). On Fisher information inequalities and score functions in non-invertible linear systems. *Journal of Inequalities in Pure and Applied Mathematics* **4**(4) 71.

Xu, B., Chen, Q., Wu, Z. and Wang, Z. (2008). Analysis and approximation of performance bound for two-observer bearings-only tracking. *Information Sciences* **178**(8) 2059–2078.

Zivojnovic, V. and Noll, D. (1997). Minimum Fisher information spectral analysis. *ICASSP-97, IEEE International Conference on Acoustics, Speech, and Signal Processing* **5**.

Abbas Pak
Department of Computer Sciences,
Faculty of Mathematical Sciences,
Shahrekord University,
P. O. Box 115, Shahrekord, Iran.
email: abbas.pak1982@gmail.com