 DOI: http://doi.org/10.4038/sljastats.v19i1.7984

Recovering Fisher-Information from the MGF Alone without Requiring Explicit PMF or PDF from a One-Parameter Exponential Family

Nitis Mukhopadhyay
Department of Statistics, U-4120
University of Connecticut, Storrs, CT 06269-4120, U.S.A.
(E-mail: nitis.mukhopadhyay@uconn.edu)

Abstract: It is well-known that a finite *moment generating function* (m.g.f.) corresponds to a unique probability distribution. So, an important question arises: Is it possible to obtain an expression of Fisher-information, $\mathcal{I}_X(\theta)$, using the m.g.f. alone, that is without requiring explicitly a *probability mass function* (p.m.f.) or *probability density function* (p.d.f.), given that the p.m.f. or p.d.f came from a one-parameter exponential family? We revisit the core of statistical inference by developing a clear link (Theorem 1.1) between the m.g.f. and $\mathcal{I}_X(\theta)$. Illustrations are included.

Key-Words: Cramér-Rao inequality; Cramér-Rao lower bound; exponential family; Fisher-information; minimal sufficient statistic; moment generating function; one-parameter exponential.

AMS Subject Classification: 62B10; 60E10; 62A10.

Received: 14th September 2018 / Revised: 04th June 2019 / Accepted: 15th September 2019

1. Introduction

A *moment generating function* (m.g.f.) is widely used in both probability theory and statistics. Suppose that an observation X has its *probability mass function* (p.m.f.) or *probability density function* (p.d.f.) given by $f(x;\theta)$ where $x \in \mathcal{X}$, a subset of the real line. Here, θ is an unknown parameter with a parameter space Θ, *a sub-interval* of the real line. The m.g.f. of X is:

$$M_X(t;\theta) = \int_{\mathcal{X}} e^{tx} f(x;\theta)dx \text{ for } t \in \mathcal{T}, \tag{1.1}$$

where the domain space for t, namely $\mathcal{T}$, is assumed a sub-interval of the real line.

Let us suppose that $f(x;\theta), \mathcal{X}$ and Θ satisfy the customary regularity conditions for Cramér-Rao inequality to hold. A crucial assumption is that the partial derivative $\frac{\partial}{\partial\theta}$ of an integral (over $\mathcal{X}$) is finite and it may be moved inside the integral sign. See Lehmann (1983, p. 125), Rao (1973, p. 329), and Mukhopadhyay (2000, p. 366) for details.

Research in this area remains vibrant and exciting. We add a number of references: Tanaka (2006), Ghosh (1988), Mukhopadhyay and Banerjee (2013), and Mukhopadhyay (2014). Such discourses invariably bring back the notion of available information in an observation X about an unknown parameter θ to the forefront

To be specific, Fisher-information in X about unknown θ is given by:

$$\begin{aligned} \text{Fisher-information: } \mathcal{I}_X(\theta) = E_\theta\left[\left(\tfrac{\partial}{\partial\theta}\log f(X;\theta)\right)^2\right], \text{ or equivalently,} \\ E_\theta\left[-\tfrac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right], \text{ for all } \theta\in\Theta, \end{aligned} \tag{1.2}$$

assuming that the expectations are finite and non-zero.

Now, we mention a very powerful and well-known result that emphasizes the importance of m.g.f. and it is briefly stated as follows:

A finite m.g.f. is associated with a unique probability distribution

One may refer to Casella and Berger (2002, p. 66) and Mukhopadhyay (2000, pp. 190-192). A probability distribution, once accurately identified from an m.g.f., will immediately lead to an expression of Fisher-information. But, that route will nearly demand that we are aware of an exact expression of the p.m.f or p.d.f. which gives rise to a particular m.g.f. on hand assuming the m.g.f. is finite.

1.1. Main Result

Suppose that we are given an expression of $M_X(t;\theta)$ rather than $f(x;\theta)$, but we cannot immediately identify an exact nature of the p.m.f or p.d.f. $f(x;\theta)$. In such a situation, it will be useful if there is a way to obtain $\mathcal{I}_X(\theta)$ using the expression of $M_X(t;\theta)$ *alone*.

A natural query comes up: Can we recover an expression for $\mathcal{I}_X(\theta)$ from $M_X(t;\theta)$ without first identifying the corresponding unique distribution associated with $M_X(t;\theta)$? Theorem 1.1 answers in the affirmative when the associated distribution comes from a one-parameter exponential family.

An important question may be raised: How would one know whether or not a given m.g.f. came from a one-parameter exponential family? This is often an important assumption behind numerous probabilistic models under consideration. An experimenter may not know an exact statistical or probabilistic model that will uniquely give rise to a dataset on hand. But, a one-parameter exponential family has a rich structure with flexible analytical and geometrical features lending useful stochastic models as possibile choices to consider in data analyses.

We quote how Wikipedia (https://en.wikipedia.org/wiki/Exponential_family) explains in its opening paragraph: "In probability and statistics, an exponential family is a set of probability distributions of a certain form ... This special form is chosen for mathematical convenience, on account of some useful algebraic properties, as well as for generality, as exponential families are in a sense very natural sets of distributions to consider. The concept of exponential families is credited to ... Anderson (1970), Pitman and Wishart (1936), Darmois (1935), Koopman (1936). The term exponential class is sometimes used in place of 'exponential family' (Kupperman (1958))".

Theorem 1.1. *Suppose that a p.m.f. or p.d.f. $f(x;\theta)$ belongs to a one-parameter exponential family. More specifically, we suppose that*:

$$f(x;\theta) = c(\theta)h(x)\exp\{w(\theta)U(x)\} \text{ for } x\in\mathcal{X} \text{ and } \theta\in\Theta, \tag{1.3}$$

where $\mathcal{X}$, $h(x)$, $U(x)$ do not involve θ and Θ, $c(\theta)$, $w(\theta)$ do not involve x. Let us denote:

$$M_U(t;\theta) \equiv \int_{\mathcal{X}} \exp\{tU(x)\}f(x;\theta)dx, \text{ the finite m.g.f. of } U, \tag{1.4}$$

for $t \in \mathcal{T}$, *a sub-interval of the real line. Then,* $\mathcal{I}_X(\theta)$, *given by (1.2), can be recovered from* $M_U(t;\theta)$, *the m.g.f. of the (complete and) minimal sufficient statistic* $U \equiv U(X)$ *for* θ, *as follows*:

$$\mathcal{I}_X(\theta) = \lim_{t\to 0} \frac{\left(\frac{\partial}{\partial\theta} M_U(t;\theta)\right)^2}{\left\{M_U(2t;\theta) - (M_U(t;\theta))^2\right\}}, \tag{1.5}$$

for all $\theta \in \Theta$.

This tool will be especially useful in many situations for obtaining $\mathcal{I}_X(\theta)$. Illustrations will follow in Sections 2 and 3.

1.2. Layout

In Section 2, we illustrate the usefulness of the limiting expression (1.5) with three initial examples. These examples highlight the importance of Theorem 1.1. In Section 3, two additional illustrations with more complexities exhibit the true potential of Theorem 1.1 by finding the expressions of $\mathcal{I}_X(\theta)$ from the given m.g.f.s. One may note that the probability distributions associated with the m.g.f.s from Section 3 may not be readily identifiable from cursory reading.

In Section 4, we begin with some motivations behind Theorem 1.1. Those motivations and a formal proof of Theorem 1.1 are closely tied with each other and hence we keep these in one place. In Section 5, we briefly discuss an ad hoc technique if one can easily identify an expression of $w(\theta)$ in (1.3) by glancing at $M_U(t;\theta)$ and some trial and error under *natural parametrization*. Some concluding thoughts are included in Section 6.

2. Initial Illustrations

We illustrate the limiting expression shown in (1.5) with the help of three concrete examples. The specific m.g.f.s under consideration are associated with frequently used p.m.f.s or p.d.f.s from one-parameter exponential families.

2.1. Illustration 1

Suppose that U has its m.g.f. given by:

$$\begin{gathered} M_U(t;\theta) = \exp\left(t\theta + \tfrac{1}{2}t^2\right) \text{ with } t \in \mathcal{T} = (-\infty,\infty), \theta \in \Theta = (-\infty,\infty), \\ \text{and } \tfrac{\partial}{\partial\theta} M_U(t;\theta) = t\exp\left(t\theta + \tfrac{1}{2}t^2\right). \end{gathered} \tag{2.1}$$

Now, using L'Hôpital's rule repeatedly, the expression on the right-hand side of (1.5) simplifies to:

$$\lim_{t\to 0} \frac{t^2 \exp\left(2t\theta + t^2\right)}{\exp\left(2t\theta + 2t^2\right) - \exp\left(2t\theta + t^2\right)} = \lim_{t\to 0} \exp\left(-t^2\right) = 1,$$

so that, from Theorem 1.1, we have $\mathcal{I}_X(\theta) = 1$ for all $-\infty < \theta < \infty$. ♦

The m.g.f. $M_U(t;\theta)$ from (2.1) clearly corresponds to $U = X$ where X has the $N(\theta, 1)$ distribution with an unknown mean θ. The p.d.f. of X obviously has the same form as in (1.3) with appropriate $c(\theta), h(x), w(\theta), U(x) = x$, and $\mathcal{X} = \Theta = (-\infty,\infty)$.

2.2. Illustration 2

Suppose that $U(X)$ has its m.g.f. given by:

$$M_U(t;\theta) = (1-2t\theta)^{-1/2} \text{ with } t \in \mathcal{T} = (-\infty, \tfrac{1}{2}\theta^{-1}), \theta \in \Theta = (0,\infty), \\ \text{and } \tfrac{\partial}{\partial\theta} M_U(t;\theta) = t\,(1-2t\theta)^{-3/2}. \tag{2.2}$$

Again, using L'Hôpital's rule repeatedly, for $-\infty < t < \frac{1}{4}\theta^{-1}$, the expression on the right-hand side of (1.5) becomes:

$$\lim_{t\to 0} \frac{t^2\,(1-2t\theta)^{-3}}{(1-4t\theta)^{-1/2} - (1-2t\theta)^{-1}} = \theta^{-1} \lim_{t\to 0} \frac{t\,(1-2t\theta)^{-3} + 3\theta t^2\,(1-2t\theta)^{-4}}{(1-4t\theta)^{-3/2} - (1-2t\theta)^{-2}} = \frac{1}{2}\theta^{-2},$$

so that, from Theorem 1.1, we have $\mathcal{I}_X(\theta) = \frac{1}{2}\theta^{-2}$ for all $0 < \theta < \infty$. ♦

Again, the m.g.f. $M_U(t;\theta)$ from (2.2) clearly corresponds to $U \equiv X^2$ where X has the $N(0,\theta)$ distribution having an unknown variance θ. The p.d.f. of X obviously has the same form as in (1.3) with appropriate $c(\theta), h(x), w(\theta), U(x) = x^2$, and $\mathcal{X} = \Theta = (-\infty, \infty)$.

2.3. Illustration 3

Suppose that $U(X)$ has its m.g.f. given by:

$$M_U(t;\theta) = \exp\{\theta\,(\exp(t)-1)\} \text{ with } t \in \mathcal{T} = (-\infty,\infty), \theta \in \Theta = (0,\infty), \\ \text{and } \tfrac{\partial}{\partial\theta} M_U(t;\theta) = (\exp(t)-1)\exp\{\theta\,(\exp(t)-1)\}. \tag{2.3}$$

Using L'Hôpital's rule repeatedly, the expression on the right-hand side of (1.5) becomes:

$$\lim_{t\to 0} \frac{\exp\left\{2\theta\left(e^t-1\right)\right\}\left(e^t-1\right)^2}{\exp\left\{\theta\left(e^{2t}-1\right)\right\} - \exp\left\{2\theta\left(e^t-1\right)\right\}} = \lim_{t\to 0} \frac{1}{\theta e^{\theta}\left(2e^{-\theta}\right) - \theta} = \frac{1}{\theta}.$$

Thus, from Theorem 1.1, we have $\mathcal{I}_X(\theta) = \theta^{-1}$ for all $0 < \theta < \infty$. ♦

The m.g.f. $M_U(t;\theta)$ from (2.3) corresponds to $U \equiv X$ where X has the Poisson(θ) distribution with an unknown mean θ. Also, the p.m.f. of X obviously has the same form as in (1.3) with appropriate $c(\theta), h(x), w(\theta), U(x) = x$, and $\mathcal{X} = \{0,1,2,...\}, \Theta = (0,\infty)$.

3. Illustrations with More Complexities

In this section, two complex illustrations are discussed. In either case, it may be difficult to readily identify an associated population distribution $f(x;\theta)$ simply by looking at the m.g.f., and such a possibility bolsters the importance of Theorem 1.1. A main point is that (1.5) does not require f explicitly to come up with $\mathcal{I}_X(\theta)$.

3.1. Illustration 4

Suppose that $U(X)$ has its m.g.f. given by:

$$M_U(t;\theta) \equiv \Gamma(\theta+t)/\Gamma(\theta) \text{ with } t \in \mathcal{T} = (-\theta,\infty) \text{ and } \theta \in \Theta = (0,\infty). \tag{3.1}$$

We are told that $f(x;\theta), \theta \in \Theta$, comes from a one-parameter exponential family. The m.g.f. from (3.1) may not be very readily familiar, and thus in order to evaluate $\mathcal{I}_X(\theta)$, or equivalently $\mathcal{I}_U(\theta)$, we exploit Theorem 1.1.

Let us denote:

$$\tfrac{d}{d\theta}\Gamma(\theta) = \Gamma'(\theta) \text{ and } \tfrac{d}{d\theta}\Gamma'(\theta) = \Gamma''(\theta).$$

Now, observe:

$$\begin{gathered}\tfrac{\partial}{\partial t}\left\{\tfrac{\partial}{\partial\theta}\left(\tfrac{\Gamma(\theta+t)}{\Gamma(\theta)}\right)\right\}^2 = 2\tfrac{\partial}{\partial\theta}\left(\tfrac{\Gamma(\theta+t)}{\Gamma(\theta)}\right)\tfrac{\partial}{\partial\theta}\left(\tfrac{\Gamma'(\theta+t)}{\Gamma(\theta)}\right), \\ \tfrac{\partial}{\partial t}\left(\Gamma(\theta+2t)\right) = 2\Gamma'(\theta+2t) \text{ and } \tfrac{\partial}{\partial t}\left\{\Gamma(\theta+t)\right\}^2 = 2\Gamma(\theta+t)\Gamma'(\theta+t).\end{gathered} \tag{3.2}$$

Also, we note:

$$\begin{gathered}\tfrac{\partial}{\partial t}\left\{\tfrac{\partial}{\partial\theta}\left(\tfrac{\Gamma(\theta+t)}{\Gamma(\theta)}\right)\tfrac{\partial}{\partial\theta}\left(\tfrac{\Gamma'(\theta+t)}{\Gamma(\theta)}\right)\right\} = \left\{\tfrac{\partial}{\partial\theta}\left(\tfrac{\Gamma'(\theta+t)}{\Gamma(\theta)}\right)\right\}^2 + \tfrac{\partial}{\partial\theta}\left(\tfrac{\Gamma(\theta+t)}{\Gamma(\theta)}\right)\tfrac{\partial}{\partial\theta}\left(\tfrac{\Gamma''(\theta+t)}{\Gamma(\theta)}\right), \\ \tfrac{\partial}{\partial t}\Gamma'(\theta+2t) = 2\Gamma''(\theta+2t), \text{ and } \tfrac{\partial}{\partial t}\left(\Gamma(\theta+t)\Gamma'(\theta+t)\right) = \left(\Gamma'(\theta+t)\right)^2 + \Gamma(\theta+t)\Gamma''(\theta+t).\end{gathered} \tag{3.3}$$

Hence, after repeatedly using L'Hôpital's rule and combining (3.2)-(3.3), the expression on the right-hand side of (1.5) becomes:

$$\lim_{t\to 0}\left(\Gamma(\theta)\right)^2 \frac{\left\{\tfrac{\partial}{\partial\theta}\left(\tfrac{\Gamma(\theta+t)}{\Gamma(\theta)}\right)\right\}^2}{\Gamma(\theta+2t)\Gamma(\theta) - \left\{\Gamma(\theta+t)\right\}^2} = \frac{\left\{\tfrac{d}{d\theta}\left(\tfrac{\Gamma'(\theta)}{\Gamma(\theta)}\right)\right\}^2\left(\Gamma(\theta)\right)^2}{\Gamma''(\theta)\Gamma(\theta) - \left(\Gamma'(\theta)\right)^2},$$

which reduces to $\frac{d}{d\theta}\left(\Gamma'(\theta)/\Gamma(\theta)\right)$. Thus, from Theorem 1.1, we come up with $\mathcal{I}_X(\theta) = \frac{d}{d\theta}\left(\Gamma'(\theta)/\Gamma(\theta)\right)$ for all $0 < \theta < \infty$. ♦

3.2. Illustration 5

Suppose that $U(X)$ has the following m.g.f.:

$$\begin{gathered}M_U(t;\theta) = \exp\left\{-\tfrac{1}{2}\theta\left(1-(1-2t\theta)^{-1}\right)\right\}(1-2t\theta)^{-1/2} \\ \text{for } t \in \mathcal{T} = (-\infty, \tfrac{1}{2}\theta^{-1}), \text{ and } \Theta = (0,\infty).\end{gathered} \tag{3.4}$$

Again, the parent population distribution $f(x;\theta), \theta \in \Theta$ comes from a one-parameter exponential family. The m.g.f. from (3.4) may not be very familiar and hence the population distribution f associated with it may not be immediately obvious. In order to evaluate $\mathcal{I}_X(\theta)$, or equivalently $\mathcal{I}_U(\theta)$, we rely upon Theorem 1.1.

For $t \in (-\infty, \frac{1}{4}\theta^{-1})$, we observe:

$$\begin{gathered}\tfrac{\partial}{\partial\theta}M_U(t;\theta) = \exp\left\{-\tfrac{1}{2}\theta\left(1-\tfrac{1}{1-2t\theta}\right)\right\}\left(-\tfrac{1}{2}+\tfrac{1}{2(1-2t\theta)}+\tfrac{\theta}{(1-2t\theta)^2}t\right) \\ \times\left\{(1-2t\theta)^{-1/2} + t(1-2t\theta)^{-3/2}\right\},\end{gathered} \tag{3.5}$$

and denote:

$$g(t,\theta) \equiv \tfrac{\partial}{\partial t}\left(\tfrac{\partial}{\partial\theta}M_U(t;\theta)\right)^2, h(t,\theta) \equiv \tfrac{\partial}{\partial t}\left(M_U(2t,\theta) - \left(M_U(t,\theta)\right)^2\right). \tag{3.6}$$

After substantial manipulations, from (3.5) and (3.6), we arrive at the following expressions:

$$\begin{aligned} g(t,\theta) &= 2\left\{\left(-\tfrac{1}{2}+\tfrac{1}{2-4t\theta}+\tfrac{\theta}{(1-2t\theta)^2}t\right)\frac{\exp\left(-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right)}{\sqrt{(1-2t\theta)}}+\frac{\exp\left(-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right)}{\left(\sqrt{(1-2t\theta)}\right)^3}t\right\} \\ &\quad\times\left\{\left(\tfrac{4}{(2-4t\theta)^2}\theta+4\tfrac{\theta^2}{(1-2t\theta)^3}t+\tfrac{\theta}{(1-2t\theta)^2}\right)\frac{\exp\left(-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right)}{\sqrt{(1-2t\theta)}}\right. \\ &\quad+\left(-\tfrac{1}{2}+\tfrac{1}{2-4t\theta}+\tfrac{\theta}{(1-2t\theta)^2}t\right)\frac{\theta^2}{\left(\sqrt{(1-2t\theta)}\right)^5}\exp\left\{-\tfrac{1}{2}\theta\left(1-\tfrac{1}{1-2t\theta}\right)\right\} \\ &\quad+\left(-\tfrac{1}{2}+\tfrac{1}{2-4t\theta}+\tfrac{\theta}{(1-2t\theta)^2}t\right)\frac{\exp\left(-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right)}{\left(\sqrt{(1-2t\theta)}\right)^3}\theta+\frac{\theta^2\exp\left(-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right)}{\left(\sqrt{(1-2t\theta)}\right)^7}t \\ &\quad\left.+3\frac{\exp\left(-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right)}{\left(\sqrt{(1-2t\theta)}\right)^5}t\theta+\frac{\exp\left(-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right)}{\left(\sqrt{(1-2t\theta)}\right)^3}\right\}, \end{aligned} \tag{3.7}$$

and

$$\begin{aligned} h(t,\theta) &= 2\frac{\theta^2}{\left(\sqrt{(1-4t\theta)}\right)^5}\exp\left\{-\tfrac{1}{2}\theta\left(1-\tfrac{1}{1-4t\theta}\right)\right\}+2\frac{\exp\left\{-\frac{1}{2}\theta\left(1-\frac{1}{1-4t\theta}\right)\right\}}{\left(\sqrt{(1-4t\theta)}\right)^3}\theta \\ &\quad-2\frac{\exp^2\left\{-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right\}}{(1-2t\theta)^3}\theta^2-2\frac{\exp^2\left\{-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right\}}{(1-2t\theta)^2}\theta. \end{aligned} \tag{3.8}$$

Next, we evaluate the limiting expression found on the right-hand side of (1.5) and in order to do so, again we repeatedly apply L'Hôpital's rule. The limiting value from (1.5) becomes:

$$\lim_{t\to 0}\frac{g(t,\theta)}{h(t,\theta)},$$

but we note that $g(0,\theta)=h(0,\theta)=0$ for all θ. Let us denote $a(t,\theta)=\partial g(t,\theta)/\partial t$ and $b(t,\theta)=\partial h(t,\theta)/\partial t$. From (3.7) and (3.8), after substantial calculations, we arrive at the following expressions:

$$\begin{aligned}
a(t,\theta) &= 2\Bigg\{\left(\tfrac{4}{(2-4t\theta)^2}\theta+4\tfrac{\theta^2}{(1-2t\theta)^3}t+\tfrac{\theta}{(1-2t\theta)^2}\right)\tfrac{\exp\left\{-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right\}}{\sqrt{(1-2t\theta)}}\\
&+\left(-\tfrac{1}{2}+\tfrac{1}{2-4t\theta}+\tfrac{\theta}{(1-2t\theta)^2}t\right)\tfrac{\theta^2}{\left(\sqrt{(1-2t\theta)}\right)^5}\exp\left\{-\tfrac{1}{2}\theta\left(1-\tfrac{1}{1-2t\theta}\right)\right\}\\
&+\left(-\tfrac{1}{2}+\tfrac{1}{2-4t\theta}+\tfrac{\theta}{(1-2t\theta)^2}t\right)\tfrac{\exp\left\{-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right\}}{\left(\sqrt{(1-2t\theta)}\right)^3}\theta\\
&+\tfrac{\theta^2}{\left(\sqrt{(1-2t\theta)}\right)^7}\exp\left\{-\tfrac{1}{2}\theta\left(1-\tfrac{1}{1-2t\theta}\right)\right\}t\\
&+3\tfrac{\exp\left\{-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right\}}{\left(\sqrt{(1-2t\theta)}\right)^5}t\theta+\tfrac{\exp\left\{-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right\}}{\left(\sqrt{(1-2t\theta)}\right)^3}\Bigg\}^2\\
&+2\left\{\left(-\tfrac{1}{2}+\tfrac{1}{2-4t\theta}+\tfrac{\theta}{(1-2t\theta)^2}t\right)\tfrac{\exp\left\{-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right\}}{\sqrt{(1-2t\theta)}}+\tfrac{\exp\left\{-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right\}}{\left(\sqrt{(1-2t\theta)}\right)^3}t\right\}\\
&\times\Bigg\{\left(\tfrac{32}{(2-4t\theta)^3}\theta^2+24\tfrac{\theta^3}{(1-2t\theta)^4}t+8\tfrac{\theta^2}{(1-2t\theta)^3}\right)\tfrac{\exp\left\{-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right\}}{\sqrt{(1-2t\theta)}}\\
&+2\left(\tfrac{4}{(2-4t\theta)^2}\theta+4\tfrac{\theta^2}{(1-2t\theta)^3}t+\tfrac{\theta}{(1-2t\theta)^2}\right)\tfrac{\theta^2}{\left(\sqrt{(1-2t\theta)}\right)^5}\exp\left\{-\tfrac{1}{2}\theta\left(1-\tfrac{1}{1-2t\theta}\right)\right\}\\
&+2\left(\tfrac{4}{(2-4t\theta)^2}\theta+4\tfrac{\theta^2}{(1-2t\theta)^3}t+\tfrac{\theta}{(1-2t\theta)^2}\right)\tfrac{\exp\left\{-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right\}}{\left(\sqrt{(1-2t\theta)}\right)^3}\theta\\
&+6\left(-\tfrac{1}{2}+\tfrac{1}{2-4t\theta}+\tfrac{\theta}{(1-2t\theta)^2}t\right)\tfrac{\theta^3}{\left(\sqrt{(1-2t\theta)}\right)^7}\exp\left\{-\tfrac{1}{2}\theta\left(1-\tfrac{1}{1-2t\theta}\right)\right\}\\
&+\left(-\tfrac{1}{2}+\tfrac{1}{2-4t\theta}+\tfrac{\theta}{(1-2t\theta)^2}t\right)\tfrac{\theta^4}{\left(\sqrt{(1-2t\theta)}\right)^9}\exp\left\{-\tfrac{1}{2}\theta\left(1-\tfrac{1}{1-2t\theta}\right)\right\}\\
&+3\left(-\tfrac{1}{2}+\tfrac{1}{2-4t\theta}+\tfrac{\theta}{(1-2t\theta)^2}t\right)\tfrac{\theta^2}{\left(\sqrt{(1-2t\theta)}\right)^5}\exp\left\{-\tfrac{1}{2}\theta\left(1-\tfrac{1}{1-2t\theta}\right)\right\}\\
&+10\tfrac{\theta^3}{\left(\sqrt{(1-2t\theta)}\right)^9}\exp\left\{-\tfrac{1}{2}\theta\left(1-\tfrac{1}{1-2t\theta}\right)\right\}t+\tfrac{\theta^4}{\left(\sqrt{(1-2t\theta)}\right)^{11}}\\
&\times\exp\left\{-\tfrac{1}{2}\theta\left(1-\tfrac{1}{1-2t\theta}\right)\right\}t+2\tfrac{\theta^2}{\left(\sqrt{(1-2t\theta)}\right)^7}\exp\left\{-\tfrac{1}{2}\theta\left(1-\tfrac{1}{1-2t\theta}\right)\right\}\\
&+15\tfrac{\theta^2}{\left(\sqrt{(1-2t\theta)}\right)^7}\exp\left\{-\tfrac{1}{2}\theta\left(1-\tfrac{1}{1-2t\theta}\right)\right\}t+6\tfrac{\exp\left\{-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right\}}{\left(\sqrt{(1-2t\theta)}\right)^5}\theta\Bigg\}
\end{aligned}\tag{3.9}$$

and

$$\begin{aligned}
b(t,\theta) &= 24\tfrac{\theta^3}{\left(\sqrt{(1-4t\theta)}\right)^7}\exp\left\{-\tfrac{1}{2}\theta\left(1-\tfrac{1}{1-4t\theta}\right)\right\}+4\tfrac{\theta^4}{\left(\sqrt{(1-4t\theta)}\right)^9}\\
&\times\exp\left\{-\tfrac{1}{2}\theta\left(1-\tfrac{1}{1-4t\theta}\right)\right\}+12\tfrac{\theta^2}{\left(\sqrt{(1-4t\theta)}\right)^5}\exp\left\{-\tfrac{1}{2}\theta\left(1-\tfrac{1}{1-4t\theta}\right)\right\}
\end{aligned}$$

$$\begin{aligned}
&-4\tfrac{\exp^2\left\{-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right\}}{(1-2t\theta)^5}\theta^4-16\tfrac{\exp^2\left\{-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right\}}{(1-2t\theta)^4}\theta^3\\
&-8\tfrac{\exp^2\left\{-\frac{1}{2}\theta\left(1-\frac{1}{1-2t\theta}\right)\right\}}{(1-2t\theta)^3}\theta^2.
\end{aligned}\tag{3.10}$$

Next, combining (3.9) and (3.10), the limiting expression found on the right-hand side of (1.5) becomes:

$$\lim_{t\to 0}\frac{a(t,\theta)}{b(t,\theta)} = \frac{2\,(2\theta+1)^2}{8\theta^3+4\theta^2} = \frac{(2\theta+1)}{2\theta^2}, \tag{3.11}$$

which shows that $\mathcal{I}_X(\theta) = \frac{1}{2}(2\theta+1)\theta^{-2}$. ♦

In complicated cases, one may find it bothersome to write down correct expressions of the successive partial derivatives (with respect to t) for both the numerator and denominator while implementing Theorem 1.1. But, in this day and age of computers, that should not cause any hindrance. Indeed, the lengthy expressions in (3.5) and (3.7)-(3.11) were obtained using MAPLE.

Remark 3.1. Here is one other complicated illustration. Suppose that the distribution of X belongs to a one-parameter exponential family, and $U = U(X)$ has its m.g.f. given by:

$$M_U(t;\theta) = (\exp(\theta)-1)^{-1}\left[\exp\left(\theta\exp(t)\right)-1\right], t\in(-\infty,\infty), \theta\in(0,\infty).$$

What is $\mathcal{I}_X(\theta)$? One may exploit Theorem 1.1 and check:

$$\mathcal{I}_X(\theta) = \frac{\left(1-e^{-\theta}-\theta e^{-\theta}\right)^2}{\left(-1+e^{-\theta}\right)^2\theta\left(1-e^{-\theta}-\theta e^{-\theta}\right)} \tag{3.12}$$

All other details are left out.

3.3. Some Comments

A natural temptation would be to check the expressions of $\mathcal{I}_X(\theta)$ by utilizing the associated p.d.f.s of U which, by the way, ought to be *proportional* to $f(x;\theta)$. In Illustration 4, the form of the m.g.f. from (3.1) and our subsequent derivation of $\mathcal{I}_X(\theta)$ may supply hints to guess the corresponding $f(x;\theta)$. In that case, one will be able to check easily that $\mathcal{I}_X(\theta)$ so found was indeed correct.

Illustration 5 is quite different from the other four illustrations. The m.g.f. from (3.4) is not something that one handles everyday! While it is difficult to come up with the corresponding $f(x;\theta)$, it may remain out of sight. Exactly right there lies the usefulness of Theorem 1.1.

4. Motivation and Proof of Theorem 1.1

We suppose that $f(x;\theta)$, $\mathcal{X}$, and Θ satisfy the regularity conditions that are customarily assumed for the Cramér-Rao inequality to hold. A crucial point is that the partial derivative $\frac{\partial}{\partial\theta}$ of an integral over $\mathcal{X}$ may be moved inside the integral sign. In the case when $f(x;\theta)$ belongs to a one-parameter exponential family, such requirements would be necessarily satisfied. See, for example, Lehmann (1983, p. 125), Lehmann and Casella (1998, pp. 115-118), Rao (1973, p. 329), and Mukhopadhyay (2000, p. 366).

One may genuinely ask why one should expect Theorem 1.1 to hold in an exponential family? Here are some motivations. We recall (1.3) and denote:

$$c^{'}(\theta) = \frac{\partial}{\partial\theta}c(\theta) \text{ and } w^{'}(\theta) = \frac{\partial}{\partial\theta}w(\theta).$$

We immediately have:

$$\frac{\partial}{\partial\theta}\log f(x;\theta) = w^{'}(\theta)\left\{\frac{c^{'}(\theta)}{c(\theta)w^{'}(\theta)} + U(x)\right\}. \tag{4.1}$$

Now, since $f(x;\theta)$ is a p.m.f. or p.d.f., we have:

$$c(\theta)\int_{\mathcal{X}} h(x)\exp\left\{w(\theta)U(x)\right\}dx = 1,$$

which gives:

$$\begin{aligned} 0 = c(\theta)w^{'}(\theta)\textstyle\int_{\mathcal{X}} h(x)U(x)\exp\left\{w(\theta)U(x)\right\}dx \\ +c^{'}(\theta)\textstyle\int_{\mathcal{X}} h(x)\exp\left\{w(\theta)U(x)\right\}dx = \frac{c^{'}(\theta)}{c(\theta)} + w^{'}(\theta)E_\theta\left[U(X)\right] \\ \Rightarrow E_\theta\left[U(X)\right] = -\frac{c^{'}(\theta)}{c(\theta)w^{'}(\theta)} = d(\theta), \text{ say,} \end{aligned} \tag{4.2}$$

by interchanging the integral with respect to x and the partial derivative with respect to θ.

In other words, $U \equiv U(X)$ is an unbiased estimator of a parametric function $d(\theta)$ defined by (4.2). Then, from (4.1), we can conclude that the variance of $U(X)$ must attain the Cramér-Rao lower bound, that is,

$$V_\theta[U(X)] = \frac{\left\{\frac{\partial}{\partial\theta}E_\theta\left(U(X)\right)\right\}^2}{\mathcal{I}_X(\theta)} \Leftrightarrow \mathcal{I}_X(\theta) = \frac{\left\{\frac{\partial}{\partial\theta}E_\theta\left(U(X)\right)\right\}^2}{V_\theta[U(X)]}. \tag{4.3}$$

See Rao (1973, p. 325) and Lehmann and Casella (1998, p. 121).

4.1. Proof of Theorem 1.1

At this point, we turn around to note that $\exp\left\{tU(X)\right\}$ is an unbiased estimator of $M_U(t;\theta)$ for all fixed $t \in \mathcal{T}$. Now, the Cramér-Rao inequality will imply:

$$V_\theta\left[\exp\left\{tU(X)\right\}\right] \geq \frac{\left[\frac{\partial}{\partial\theta}E_\theta\left\{\exp\left(tU(X)\right)\right\}\right]^2}{\mathcal{I}_X(\theta)} \Leftrightarrow \mathcal{I}_X(\theta) \geq \frac{\left(\frac{\partial}{\partial\theta}M_U(t;\theta)\right)^2}{V_\theta\left[\exp\left\{tU(X)\right\}\right]}, \tag{4.4}$$

for all fixed $t \in \mathcal{T}, \theta \in \Theta$.

Clearly, $V_\theta\left[\exp\left\{tU(X)\right\}\right]$ is equivalently expressed as:

$$E_\theta\left[\exp\left\{2tU(X)\right\}\right] - \left[E_\theta\left\{\exp\left(tU(X)\right)\right\}\right]^2 = M_U(2t;\theta) - \left(M_U(t;\theta)\right)^2,$$

which happens to be the denominator of the expression on the right-hand side of (1.5).

Next, when can one expect the equality to hold in (4.4)? The clue is hidden inside (4.3). If we pick t very small, then the associated unbiased estimator $\exp\left\{tU(X)\right\}$ for $M_U(t;\theta)$ would behave "nearly" as an estimator that is linear in U. This may be validated as follows:

We note that both $\left(\frac{\partial}{\partial\theta}M_U(t;\theta)\right)^2$ and $M_U(2t;\theta) - \left(M_U(t;\theta)\right)^2$ tend to zero as t is made small. Let us denote:

$$q_1(t;\theta) = \frac{\partial}{\partial\theta}\left\{E_\theta\left(U(X)e^{tU(X)}\right)\right\}, q_2(t;\theta) = \frac{\partial}{\partial\theta}\left\{E_\theta\left(U^2(X)e^{tU(X)}\right)\right\}. \tag{4.5}$$

Next, we apply L'Hôpital's rule repeatedly and interchange the partial derivatives with respect to t and θ to write:

$$\begin{aligned}
&\lim_{t\to 0}\frac{\left(\frac{\partial}{\partial\theta}M_U(t;\theta)\right)^2}{M_U(2t;\theta)-\left(M_U(t;\theta)\right)^2}\\
&= \lim_{t\to 0}\frac{\left(\frac{\partial}{\partial\theta}M_U(t;\theta)\right)\frac{\partial}{\partial\theta}\left\{E_\theta\left(U(X)e^{tU(X)}\right)\right\}}{E_\theta\left[U(X)e^{2tU(X)}\right]-M_U(t;\theta)E_\theta\left[U(X)e^{tU(X)}\right]}\\
&= \lim_{t\to 0}\frac{\left(\frac{\partial}{\partial\theta}M_U(t;\theta)\right)\left[\{q_1(t;\theta)\}^2+q_2(t;\theta)\right]}{2q_2(2t;\theta)-\left[\{q_1(t;\theta)\}^2+M_U(t;\theta)q_2(t;\theta)\right]},\text{ using (4.5)} \qquad (4.6)\\
&= \frac{\left(\frac{\partial}{\partial\theta}E_\theta[U(X)]\right)^2}{E_\theta\left[U^2(X)\right]-E_\theta^2\left[U(X)\right]}\\
&= \frac{\left(\frac{\partial}{\partial\theta}E_\theta[U(X)]\right)^2}{V_\theta\left[U(X)\right]}.
\end{aligned}$$

On the other hand, in view of (4.1) and (4.3), for such a linear unbiased estimator of $M_U(t;\theta)$, that is for small t, its variance must attain the Cramér-Rao lower bound. Hence, we have the following result:

$$\mathcal{I}_X(\theta)=\lim_{t\to 0}\frac{\left(\frac{\partial}{\partial\theta}M_U(t;\theta)\right)^2}{V_\theta\left[\exp\left(tU(X)\right)\right]},$$

which coincides with the limiting expression on the right-hand side of (1.5). The proof is complete. ■

5. Natural Parametrization?

Now, we briefly discuss a parallel idea of an ad hoc nature. Given how Section 4 was laid out, one may feel tempted to cast Theorem 1.1 in the light of Theorem 7.2 of Lehmann (1983, p. 127) or equivalently Theorem 5.4 of Lehmann and Casella (1998, p. 116) or the formula from (2.46) in Barndorff-Nielsen and Cox (1994, p. 27). One may consider other appropriate sources too. But, then, one should first identify $w(\theta)$ used in (1.3) readily from a given expression of $M_U(t;\theta)$ alone.

If that is indeed the case, then one may consider obtaining $\mathcal{I}_X^*(w(\theta))$, Fisher-information under the *mean-value parametrization* or the so-called *natural parametrization* where $\mathcal{I}_X^*(w(\theta))$ can be rewritten as:

$$\mathcal{I}_X^*(w(\theta))=V_{w(\theta)}^{-1}\left[U(X)\right]. \qquad (5.1)$$

From (5.1), we can express $\mathcal{I}_X(\theta)$ as:

$$\begin{aligned}
\mathcal{I}_X(\theta)&=\left[\tfrac{d}{d\theta}w(\theta)\right]^2V_{w(\theta)}^{-1}[U(X)],\text{ but}\\
V_\theta[U(X)]&=\text{Coefficient of }\tfrac{1}{2}t^2\text{ in }M_U(t;\theta).
\end{aligned} \qquad (5.2)$$

Clearly, $w(\theta)$ must satisfy the following relationship:

$$\begin{gathered}
M_U(t;\theta)=g\left(t+w(\theta)\right)/g\left(w(\theta)\right)\text{ for some }g(.),\\
\text{and for all }t\in\mathcal{T}\text{ and }\theta\in\Theta.
\end{gathered} \qquad (5.3)$$

We note that even if one is not readily able to fully identify f by looking at $M_U(t;\theta)$ alone, through *trial and error* on a case by case basis, one may occasionally come up with an explicit expression of $w(\theta)$ satisfying (5.3).

Illustration 1 (Continued): $M_U(t;\theta) = \exp\left(t\theta + \frac{1}{2}t^2\right)$ and (5.3) leads to $g(x) = \exp\left(x^2/2\right)$ and $w(\theta) = \theta$. Then, formula (5.2) leads to $\mathcal{I}_X(\theta) = 1$.

Illustration 2 (Continued): $M_U(t;\theta) = (1-2t\theta)^{-1/2}$ and (5.3) leads to $g(x) = x^{-1/2}$ and $w(\theta) = -(2\theta)^{-1}$. Then, formula (5.2) leads to

$$\mathcal{I}_X(\theta) = \left[\frac{1}{2\theta^2}\right]^2 \left[2\left(-(2\theta)^{-1}\right)^2\right]^{-1} = \frac{1}{2}\theta^{-2}.$$

The formula from (5.2) will work well as long as one can easily identify $w(\theta)$ by proceeding on a case by case basis. We are not aware of a precise analytical method for determining $w(\theta)$ satisfying (5.3) in a general situation.

6. Concluding Thoughts

In view of (4.6), Fisher-information from (1.5) is easily rewritten as:

$$\mathcal{I}_X(\theta) = \sigma^{-2}(\theta)\left[\frac{\partial}{\partial\theta}\mu(\theta)\right]^2, \tag{6.1}$$

where we denote $E_\theta[U] \equiv \mu(\theta)$ and $V_\theta[U] \equiv \sigma^2(\theta)$.

The expression in (6.1) may appear simpler than that on the right-hand side of (1.5). If one can not immediately guess $f(x;\theta)$ from the expression of $M_U(t;\theta)$ alone, then the chance is probably very slim of one's quoting $\mu(\theta)$ and $\sigma^2(\theta)$ right away from thin air. Even though equation (6.1) is correct, this particular form may not be readily useful. There is no additional advantage in overreaching (6.1) beyond (1.5) unless one just happens to identify $\mu(\theta)$ and $\sigma^2(\theta)$ readily.

A direct extension of Theorem 1.1 in the case of a multiparameter exponential family appears straightforward and hence it is omitted for brevity.

Acknowledgements

Professors Masafumi Akahira, Kjell Doksum, Erich Lehmann, Michael Perlman, and Andrew Rukhin showed great enthusiasm after reading preliminary notes on Theorem 1.1. Their encouragements convinced me to prepare this material. The executive editor and a reviewer gave helpful comments. I thank them.

References

[1] Andersen, Erling (1970). Sufficiency and exponential families for discrete sample spaces, *Journal of American Statistical Association*, **65**, 1248–1255.

[2] Barndorff-Nielsen, O. and Cox, D.R. (1994). *Inference and Asymptotics*, London: Chapman & Hall.

[3] Casella, G. and Berger, R.L. (2002). *Statistical Inference*, 2nd edition, Pacific Grove: Duxbury.

[4] Darmois, G. (1935). Sur les lois de probabilites a estimation exhaustive, *C. R. Acad. Sci. Paris* (in French), **200**, 1265–1266.

[5] Ghosh, J.K. (1988). *Statistical Information and Likelihood: A Collection of Critical Essays by Dr. D. Basu*, edited volume, Lecture Notes in Statistics **45**, New York: Springer.

[6] https://en.wikipedia.org/wiki/Exponential_family.

[7] Koopman, B (1936). On distribution admitting a sufficient statistic, *Transactions of American Mathematical Society*, **39**, 399–409.

[8] Kupperman, M. (1958). Probabilities of hypotheses and information-statistics in sampling from exponential-class populations, *Annals of Mathematical Statistics*, **9**, 571–575.

[9] Lehmann, E.L. (1983). *Theory of Point Estimation*, New York: Springer.

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

[11] Mukhopadhyay, N. (2000). *Probability and Statistical Inference*, New York: Dekker (Taylor & Francis).

[12] Mukhopadhyay, N. (2014). On rereading D. Basu's jointly sufficient statistic example made up of two ancillaries and miscellany, *Sankhya, Series A*, **76**, 280-287.

[13] Mukhopadhyay, N. and Banerjee, S. (2013). Sufficiency, Fisher information, and ancillarity: some clarifications, *Metron*, **71**, 33-38.

[14] Pitman, E. and Wishart, J. (1936). Sufficient statistics and intrinsic accuracy, *Mathematical Proceedings of Cambridge Philosophical Society*, **32**, 567–579.

[15] Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*, 2nd edition, New York: Wiley.

[16] Tanaka, H. (2006). Location and scale parameter family of distributions attaining the Bhattacharyya bound, *Commun. Statist. Theor. Meth.,* **35**, 1611-1628.