Communications in Information and Systems
Vol. 7, No. 3, pp. 265-272, 2007

003

WHEN THE CRAMÉR-RAO INEQUALITY PROVIDES NO INFORMATION*

STEVEN J. MILLER†

Abstract. We investigate a one-parameter family of probability densities (related to the Pareto distribution, which describes many natural phenomena) where the Cramér-Rao inequality provides no information.

Key words: Cramér-Rao Inequality, Pareto distribution, power law

1. Cramér-Rao Inequality. One of the most important problems in statistics is estimating a population parameter from a finite sample. As there are often many different estimators, it is desirable to be able to compare them and say in what sense one estimator is better than another. One common approach is to take the unbiased estimator with smaller variance. For example, if $X_1, \dots, X_n$ are independent random variables uniformly distributed on $[0, \theta]$, $Y_n = \max_i X_i$ and $\overline{X} = (X_1 + \cdots + X_n)/n$, then $\frac{n+1}{n} Y_n$ and $2\overline{X}$ are both unbiased estimators of θ but the former has smaller variance than the latter and therefore provides a tighter estimate.

Two natural questions are (1) which estimator has the minimum variance, and (2) what bounds are available on the variance of an unbiased estimator? The first question is very hard to solve in general. Progress towards its solution is given by the Cramér-Rao inequality, which provides a lower bound for the variance of an unbiased estimator (and thus if we find an estimator that achieves this, we can conclude that we have a minimum variance unbiased estimator).

Cramér-Rao Inequality: *Let $f(x;\theta)$ be a probability density function with continuous parameter θ. Let $X_1, \dots, X_n$ be independent random variables with density $f(x;\theta)$, and let $\widehat{\Theta}(X_1, \dots, X_n)$ be an unbiased estimator of θ. Assume that $f(x;\theta)$ satisfies two conditions:*

1. *we have*

$$\frac{\partial}{\partial\theta}\left[\int \cdots \int \widehat{\Theta}(x_1, \dots, x_n) \prod_{i=1}^{n} f(x_i;\theta) \mathrm{d}x_i\right]$$
$$= \int \cdots \int \widehat{\Theta}(x_1, \dots, x_n) \frac{\partial \prod_{i=1}^{n} f(x_i;\theta)}{\partial\theta} \mathrm{d}x_1 \cdots \mathrm{d}x_n; \quad (1.1)$$

*The author would like to thank Aryeh Kontorovich and Alan Landma for many enlightening conversations and the referees for helpful comments. The author was partly supported by NSF grant DMS0600848.

†Department of Mathematics, Brown University, 151 Thayer Street, Providence, RI 02912. E-mail: sjmiller@math.brown.edu

2. *for each θ, the variance of $\widehat{\Theta}(X_1,\dots,X_n)$ is finite.*

Then

$$\operatorname{var}(\widehat{\Theta}) \ \geq \ \frac{1}{n\mathbb{E}\left[\left(\frac{\partial \log f(x;\theta)}{\partial\theta}\right)^2\right]}, \tag{1.2}$$

where $\mathbb{E}$ denotes the expected value with respect to the probability density function $f(x;\theta)$.

For a proof, see for example [CaBe]. The expected value in (1.2) is called the *information number* or the *Fisher information* of the sample.

As variances are non-negative, the Cramér-Rao inequality (equation (1.2)) provides no useful bounds on the variance of an unbiased estimator if the information is infinite, as in this case we obtain the trivial bound that the variance is greater than or equal to zero. We find a simple one-parameter family of probability density functions (related to the Pareto distribution) that satisfy the conditions of the Cramér-Rao inequality, but the expectation (i.e., the information) is infinite. Explicitly, our main result is

Theorem. *Let*

$$f(x;\theta) \ = \ \begin{cases} a_\theta\ x^{-\theta}\ \log^{-3} x & \text{if } x \geq e \\ 0 & \text{otherwise,} \end{cases} \tag{1.3}$$

where a_θ is chosen so that $f(x;\theta)$ is a probability density function. The information is infinite when $\theta = 1$. Equivalently, the Cramér-Rao inequality yields the trivial (and useless) bound that $\operatorname{Var}(\widehat{\Theta}) \geq 0$ *for any unbiased estimator $\widehat{\Theta}$ of θ when $\theta = 1$.*

In §2 we analyze the density in our theorem in great detail, deriving needed results about a_θ and its derivatives as well as discussing how $f(x;\theta)$ is related to important distributions used to model many natural phenomena. We show the information is infinite when $\theta = 1$ in §3, which proves our theorem. We also discuss there properties of estimators for θ. While it is not clear whether or not this distribution has an unbiased estimator, there is (at least for θ close to 1) an asymptotically unbiased estimator rapidly converging to θ as the sample size tends to infinity. By examining the proof of the Cramér-Rao inequality we see that we may weaken the assumption of an unbiased estimator. While typically there is a cost in such a generalization, as our information is infinite there is no cost in our case. We may therefore conclude that arguments such as those used to prove the Cramér-Rao inequality cannot provide any information for estimators of θ from this distribution.

2. An Almost Pareto Density. Consider

$$f(x;\theta) \;=\; \begin{cases} a_\theta/(x^\theta \log^3 x) & \text{if } x \ge e \\ 0 & \text{otherwise,} \end{cases} \tag{2.1}$$

where a_θ is chosen so that $f(x;\theta)$ is a probability density function. Thus

$$\int_e^\infty a_\theta \frac{\mathrm{d}x}{x^\theta \log^3 x} \;=\; 1. \tag{2.2}$$

We chose to have $\log^3 x$ in the denominator to ensure that the above integral converges, as does $\log x$ times the integrand; however, the expected value (in the expectation in (1.2)) will not converge.

For example, $1/x \log x$ diverges (its integral looks like $\log\log x$) but $1/x\log^2 x$ converges (its integral looks like $1/\log x$); see pages 62–63 of [Rud] for more on close sequences where one converges but the other does not. This distribution is close to the Pareto distribution (or a power law). Pareto distributions are very useful in describing many natural phenomena; see for example [DM, Ne, NM]. The inclusion of the factor of $\log^{-3} x$ allows us to have the exponent of x in the density function equal 1 *and have the density function defined for arbitrarily large x*; it is also needed in order to apply the Dominated Convergence Theorem to justify some of the arguments below. If we remove the logarithmic factors then we obtain a probability distribution only if the density vanishes for large x. As $\log^3 x$ is a very slowly varying function, our distribution $f(x;\theta)$ may be of use in modeling data from an unbounded distribution where one wants to allow a power law with exponent 1, but cannot as the resulting probability integral would diverge. Such a situation occurs frequently in the Benford Law literature; see [Hi, Rai] for more details.

We study the variance bounds for unbiased estimators $\widehat{\Theta}$ of θ, and in particular we show that when $\theta = 1$ then the Cramér-Rao inequality yields a useless bound.

Note that it is not uncommon for the variance of an unbiased estimator to depend on the value of the parameter being estimated. For example, consider again the uniform distribution on $[0,\theta]$. Let $\overline{X}$ denote the sample mean of n independent observations, and $Y_n = \max_{1\le i\le n} X_i$ be the largest observation. The expected value of $2\overline{X}$ and $\frac{n+1}{n}Y_n$ are both θ (implying each is an unbiased estimator for θ); however, $\mathrm{Var}(2\overline{X}) = \theta^2/3n$ and $\mathrm{Var}(\frac{n+1}{n}Y_n) = \theta^2/n(n+1)$ both depend on θ, the parameter being estimated (see, for example, page 324 of [MM] for these calculations).

LEMMA 2.1. *As a function of $\theta \in [1,\infty)$, a_θ is a strictly increasing function and $a_1 = 2$. It has a one-sided derivative at $\theta = 1$, and $\frac{\mathrm{d}a_\theta}{\mathrm{d}\theta} \in (0,\infty)$.*

Proof. We have

$$a_\theta \int_e^\infty \frac{\mathrm{d}x}{x^\theta \log^3 x} \;=\; 1. \tag{2.3}$$

When $\theta = 1$ we have

$$a_1 \;=\; \left[\int_e^\infty \frac{\mathrm{d}x}{x\log^3 x}\right]^{-1}, \tag{2.4}$$

which is clearly positive and finite. In fact, $a_1 = 2$ because the integral is

$$\int_e^\infty \frac{\mathrm{d}x}{x\log^3 x} \;=\; \int_e^\infty \log^{-3} x\,\frac{\mathrm{d}\log x}{\mathrm{d}x} \;=\; \frac{-1}{2\log^2 x}\bigg|_e^\infty \;=\; \frac{1}{2}; \tag{2.5}$$

though all we need below is that a_1 is finite and non-zero, we have chosen to start integrating at e to make a_1 easy to compute.

It is clear that a_θ is strictly increasing with θ, as the integral in (2.4) is strictly decreasing with increasing θ (because the integrand is decreasing with increasing θ).

We are left with determining the one-sided derivative of a_θ at $\theta = 1$, as the derivative at any other point is handled similarly (but with easier convergence arguments). It is technically easier to study the derivative of $1/a_\theta$, as

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{1}{a_\theta} \;=\; -\frac{1}{a_\theta^2}\frac{\mathrm{d}a_\theta}{\mathrm{d}\theta} \tag{2.6}$$

and

$$\frac{1}{a_\theta} \;=\; \int_e^\infty \frac{\mathrm{d}x}{x^\theta \log^3 x}. \tag{2.7}$$

The reason we consider the derivative of $1/a_\theta$ is that this avoids having to take the derivative of the reciprocals of integrals. As a_1 is finite and non-zero, it is easy to pass to $\frac{\mathrm{d}a_\theta}{\mathrm{d}\theta}|_{\theta=1}$. Thus we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{1}{a_\theta}\bigg|_{\theta=1} \;&=\; \lim_{h\to 0^+}\frac{1}{h}\left[\int_e^\infty \frac{\mathrm{d}x}{x^{1+h}\log^3 x} - \int_e^\infty \frac{\mathrm{d}x}{x\log^3 x}\right]\\ &=\; \lim_{h\to 0^+}\int_e^\infty \frac{1-x^h}{h}\frac{1}{x^h}\frac{\mathrm{d}x}{x\log^3 x}.\end{aligned} \tag{2.8}$$

We want to interchange the integration with respect to x and the limit with respect to h above. This interchange is permissible by the Dominated Convergence Theorem (see Appendix A for details of the justification). Note

$$\lim_{h\to 0^+}\frac{1-x^h}{h}\frac{1}{x^h} \;=\; -\log x; \tag{2.9}$$

one way to see this is to use the limit of a product is the product of the limits, and then use L'Hospital's rule, writing x^h as $e^{h\log x}$. Therefore

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{1}{a_\theta}\bigg|_{\theta=1} \;=\; -\int_e^\infty \frac{\mathrm{d}x}{x\log^2 x}; \tag{2.10}$$

as this is finite and non-zero, this completes the proof and shows $\frac{\mathrm{d}a_\theta}{\mathrm{d}\theta}|_{\theta=1} \in (0,\infty)$. $\square$

Remark 2.2. *We see now why we chose* $f(x;\theta) = a_\theta/x^\theta \log^3 x$ *instead of* $f(x;\theta) = a_\theta/x^\theta \log^2 x$. *If we only had two factors of* $\log x$ *in the denominator, then the one-sided derivative of* a_θ *at* $\theta = 1$ *would be infinite.*

Remark 2.3. *Though the actual value of* $\frac{\mathrm{d}a_\theta}{\mathrm{d}\theta}|_{\theta=1}$ *does not matter, we can compute it quite easily. By* (2.10) *we have*

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{1}{a_\theta}\bigg|_{\theta=1} &= -\int_e^\infty \frac{\mathrm{d}x}{x\log^2 x} \\
&= -\int_e^\infty \log^{-2} x\ \frac{\mathrm{d}\log x}{\mathrm{d}x} \\
&= \frac{1}{\log x}\bigg|_e^\infty \ = \ -1. \qquad (2.11)
\end{aligned}$$

Thus by (2.6), *and the fact that* $a_1 = 2$ *(Lemma 2.1), we have*

$$\frac{\mathrm{d}a_\theta}{\mathrm{d}\theta}\bigg|_{\theta=1} \ = \ -a_1^2 \cdot \frac{\mathrm{d}}{\mathrm{d}\theta}\frac{1}{a_\theta}\bigg|_{\theta=1} \ = \ 4. \qquad (2.12)$$

3. Computing the Information. We now compute the expected value, $\mathbb{E}\left[\left(\frac{\partial \log f(x;\theta)}{\partial\theta}\right)^2\right]$; showing it is infinite when $\theta = 1$ completes the proof of our main result. Note

$$\begin{aligned}
\log f(x;\theta) &= \log a_\theta - \theta\log x + \log\log^{-3} x \\
\frac{\partial \log f(x;\theta)}{\partial\theta} &= \frac{1}{a_\theta}\frac{\mathrm{d}a_\theta}{\mathrm{d}\theta} - \log x. \qquad (3.1)
\end{aligned}$$

By Lemma 2.1 we know that $\frac{\mathrm{d}a_\theta}{\mathrm{d}\theta}$ is finite for each $\theta \geq 1$. Thus

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial \log f(x;\theta)}{\partial\theta}\right)^2\right] &= \mathbb{E}\left[\left(\frac{1}{a_\theta}\frac{\mathrm{d}a_\theta}{\mathrm{d}\theta} - \log x\right)^2\right] \\
&= \int_e^\infty \left(\frac{1}{a_\theta}\frac{\mathrm{d}a_\theta}{\mathrm{d}\theta} - \log x\right)^2 \cdot a_\theta \frac{\mathrm{d}x}{x^\theta \log^3 x}. \qquad (3.2)
\end{aligned}$$

If $\theta > 1$ then the expectation is finite and non-zero. We are left with the interesting case when $\theta = 1$. As $\frac{\mathrm{d}a_\theta}{\mathrm{d}\theta}|_{\theta=1}$ is finite and non-zero, for x sufficiently large (say $x \geq x_1$ for some x_1, though by Remark 2.3 we see that we may take any $x_1 \geq e^4$) we have

$$\left|\frac{1}{a_1}\frac{\mathrm{d}a_\theta}{\mathrm{d}\theta}\bigg|_{\theta=1}\right| \ \leq \ \frac{\log x}{2}. \qquad (3.3)$$

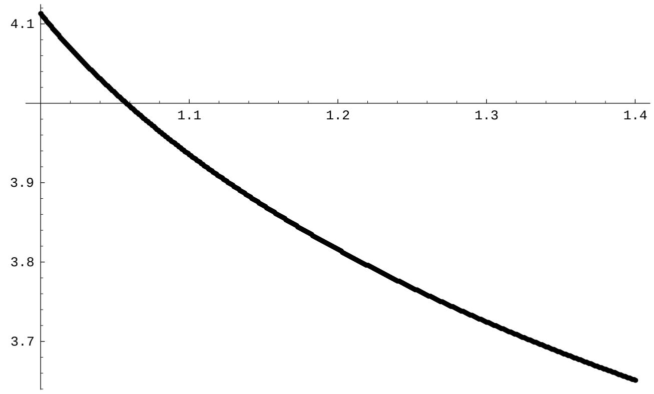

FIG. 1. *Plot of the median $\tilde{\mu}_\theta$ of $f(x;\theta)$ as a function of θ ($\tilde{\mu}_1 = e^{\sqrt{2}}$).*

As $a_1 = 2$, we have

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial \log f(x;\theta)}{\partial\theta}\right)^2\right]\Bigg|_{\theta=1} &\ge \int_{x_1}^{\infty}\left(\frac{\log x}{2}\right)^2 a_1 \frac{\mathrm{d}x}{x\log^3 x}\\
&= \int_{x_1}^{\infty}\frac{\mathrm{d}x}{2x\log x}\\
&= \frac{1}{2}\int_{x_1}^{\infty}\log^{-1}x\ \frac{\mathrm{d}\log x}{\mathrm{d}x}\\
&= \frac{1}{2}\log\log x\Big|_{x_1}^{\infty}\\
&= \infty. \qquad (3.4)
\end{aligned}$$

Thus the expectation is infinite. Let $\widehat{\Theta}$ be *any* unbiased estimator of θ. If $\theta = 1$ then the Cramér-Rao inequality gives

$$\mathrm{var}(\widehat{\Theta}) \ \ge\ 0, \qquad (3.5)$$

which provides no information as variances are always non-negative. This completes the proof of our theorem. □

We now discuss estimators for θ for our distribution $f(x;\theta)$. If $X_1,\dots,X_n$ are n independent random variables with common distribution $f(x;\theta)$, then as $n\to\infty$ the sample median converges to the population median $\tilde{\mu}_\theta$ (if $n = 2m+1$ then the sample median converges to being normally distributed with median $\widetilde{\mu}_\theta$ and variance $1/8mf(\tilde{\mu}_\theta;\theta)^2$; see for example Theorem 8.17 of [MM]). For θ close to 1 we see in Figure 1 that the median $\tilde{\mu}_\theta$ of $f(x;\theta)$ is strictly decreasing with increasing θ, which implies that there is an inverse function g such that $g(\tilde{\mu}_\theta) = \theta$. We obtain an estimator to θ by applying g to the sample median. This estimator is a consistent estimator (as the sample size tends to infinity it will tend to θ) and should be asymptotically unbiased.

The proof of the Cramér-Rao inequality starts with

$$0 \;=\; \mathbb{E}\left[\int \cdots \int \left(\widehat{\Theta}(x_1,\dots,x_n)-\theta\right) h(x_1;\theta)\cdots h(x_n;\theta)dx_1\cdots dx_n\right], \tag{3.6}$$

where $\widehat{\Theta}(x_1,\dots,x_n)$ is an unbiased estimator of θ depending only on the sample values $x_1,\dots,x_n$. In our case (when each $h(x;\theta) = f(x;\theta)$) we may not have an unbiased estimator. If we denote this expectation by $\mathcal{F}(\theta)$, for our investigations all that we require is that $d\mathcal{F}(\theta)/d\theta$ is finite (which is easy to show). Going through the proof of the Cramér-Rao inequality shows that the effect of this is to replace the factor of 1 in (1.2) with $(1 + d\mathcal{F}(\theta)/d\theta)^2$; thus the generalization of the Cramér-Rao inequality for our estimator is

$$\operatorname{var}(\widehat{\Theta}) \;\geq\; \left(1+\frac{d\mathcal{F}(\theta)}{d\theta}\right)^2 \;\Big/\; n\mathbb{E}\left[\left(\frac{\partial \log f(x;\theta)}{\partial\theta}\right)^2\right]. \tag{3.7}$$

As our variance is infinite for $\theta = 1$ we see that, no matter what 'nice' estimator we use, we will not obtain any useful information from such arguments.

Appendix A. Applying the Dominated Convergence Theorem.

We justify applying the Dominated Convergence Theorem in the proof of Lemma 2.1. See, for example, [SS] for the conditions and a proof of the Dominated Convergence Theorem.

LEMMA A.1. *For each fixed $h > 0$ and any $x \geq e$, we have*

$$\left|\frac{1-x^h}{h}\frac{1}{x^h}\right| \;\leq\; e\log x, \tag{A.1}$$

and $\frac{e\log x}{x\log^3 x}$ is positive and integrable, and dominates each $\frac{1-x^h}{h}\frac{1}{x^h}\frac{1}{x\log^3 x}$.

Proof. We first prove (A.1). As $x \geq e$ and $h > 0$, note $x^h \geq 1$. Consider the case of $1/h \leq \log x$. Since $|1-x^h| < 1 + x^h \leq 2x^h$, we have

$$\frac{|1-x^h|}{hx^h} \;<\; \frac{2x^h}{hx^h} \;\leq\; \frac{2}{h} \;\leq\; 2\log x. \tag{A.2}$$

We are left with the case of $1/h > \log x$, or $h\log x < 1$. We have

$$\begin{aligned} |1-x^h| &= |1-e^{h\log x}| \\ &= \left|1-\sum_{n=0}^{\infty}\frac{(h\log x)^n}{n!}\right| \\ &= h\log x\sum_{n=1}^{\infty}\frac{(h\log x)^{n-1}}{n!} \\ &< h\log x\sum_{n=1}^{\infty}\frac{(h\log x)^{n-1}}{(n-1)!} \;=\; h\log x\cdot e^{h\log x}. \end{aligned} \tag{A.3}$$

This, combined with $h \log x < 1$ and $x^h \ge 1$ yields

$$\frac{|1-x^h|}{hx^h} \quad < \quad \frac{eh\log x}{h} \quad = \quad e\log x. \tag{A.4}$$

It is clear that $\frac{e\log x}{x\log^3 x}$ is positive and integrable, and by L'Hospital's rule (see (2.9)) we have that

$$\lim_{h\to 0^+} \frac{1-x^h}{h}\frac{1}{x^h}\frac{1}{x\log^3 x} \quad = \quad -\frac{1}{x\log^2 x}. \tag{A.5}$$

Thus the Dominated Convergence Theorem implies that

$$\lim_{h\to 0^+} \int_e^\infty \frac{1-x^h}{h}\frac{1}{x^h}\frac{\mathrm{d}x}{x\log^3 x} \quad = \quad -\int_e^\infty \frac{\mathrm{d}x}{x\log^2 x} \quad = \quad -1 \tag{A.6}$$

(the last equality is derived in Remark 2.3). $\square$

REFERENCES

[CaBe] G. Casella and R. Berger, *Statistical Inference*, 2nd edition, Duxbury Advanced Series, Pacific Grove, CA, 2002.

[DM] D. Devoto and S. Martinez, *Truncated Pareto Law and oresize distribution of ground rocks*, Mathematical Geology, 30:6(1998), pp. 661–673.

[Hi] T. Hill, *A statistical derivation of the significant-digit law*, Statistical Science, 10(1996), pp. 354–363.

[MM] I. Miller and M. Miller, *John E. Freund's Mathematical Statistics with Applications*, seventh edition, Prentice Hall, 2004.

[Ne] M. E. J. Newman, *Power laws, Pareto distributions and Zipfs law*, Contemporary Physics, 46:5(2005), pp. 323-351.

[NM] M. Nigrini and S. J. Miller, *Benford's Law applied to hydrology data – results and relevance to other geophysical data*, Mathematical Geology, 39:5(2007), pp. 469–490.

[Rai] R. A. Raimi, *The first digit problem*, Amer. Math. Monthly, 83:7(1976), pp. 521–538.

[Rud] W. Rudin, *Principles of Mathematical Analysis*, third edition, International Series in Pure and Applied Mathematics, McGraw-Hill Inc., New York, 1976.

[SS] E. Stein and R. Shakarchi, *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*, Princeton University Press, Princeton, NJ, 2005.