Caltech Department of Mathematics

Supplement 3: Proof of The Fréchet–Cramér–Rao Lower Bound

S3.1 A lower bound on the variance of an estimator

The Larsen and Marx textbook states the Cramér–Rao Lower Bound [6, Theorem 5.5.1, p. 320], but does not derive it. In this note I present a slight generalization of their statement. The argument is essentially that of B. L. van der Waerden [8, pp. 160–162], who points out that Maurice Fréchet [5] seems to have beaten Harald Cramér [3],[4, § 32.3–32.8, pp. 477–497, esp. p. 480] and C. Radakrishna Rao [7] by a couple of years.

The FCR result puts a lower bound on the variance of estimators. Let $X_1, \dots, X_n$ be independent and identically distributed random variables with parametric density function $f(x, \theta)$. The joint density f_n at $\boldsymbol{x} = (x_1, \dots, x_n)$ is given by

$$f_n(\boldsymbol{x}; \theta) = f(x_1, \theta) f(x_2, \theta) \cdots f(x_n, \theta).$$

This is also the **likelihood function** for θ.

A **statistic** is a random variable T that is a function of $X_1, \dots, X_n$, say

$$T = T(X_1, \dots, X_n).$$

The expectation of T is the multiple integral

$$\boldsymbol{E}_\theta\, T = \int T(\boldsymbol{x}) f_n(\boldsymbol{x}; \theta)\, d\boldsymbol{x}$$

and it depends on the unknown parameter θ. The variance of T is given by

$$\boldsymbol{Var}_\theta\, T = \boldsymbol{E}_\theta \big(T - \boldsymbol{E}_\theta\, T\big)^2.$$

We say that T is a **unbiased estimator of $\boldsymbol{\theta}$** if for each θ

$$\boldsymbol{E}_\theta\, T = \theta.$$

More generally, define the **bias function** of T as

$$b(\theta) = \boldsymbol{E}_\theta\, T - \theta.$$

S3.1.1 Theorem (Fréchet–Cramér–Rao) *Assume f is continuously differentiable with respect to θ, and assume that the support $\{x : f(x; \theta) > 0\}$ does not depend on θ. Let T be an estimator of θ, with differentiable bias function $b(\theta)$. Then $\boldsymbol{Var}_\theta\, T$ is bounded below, and:*

$$\boldsymbol{Var}_\theta\, T \geqslant \frac{\big[1 + b'(\theta)\big]^2}{n\, \boldsymbol{E}_\theta \left[\left(\frac{\partial}{\partial \theta} \log f(X; \theta) \right)^2 \right]}.$$

Proof: By definition of the bias,

$$\theta + b(\theta) = \boldsymbol{E}_\theta\, T = \int T(\boldsymbol{x}) f_n(\boldsymbol{x}; \theta)\, d\boldsymbol{x} \tag{1}$$

Let $f'_n(\boldsymbol{x};\theta)$ indicate the partial derivative with respect to θ. Differentiate both sides of (1) to get (differentiating under the integral sign):

$$\begin{aligned} 1 + b'(\theta) &= \int T(\boldsymbol{x}) f'_n(\boldsymbol{x};\theta)\, d\boldsymbol{x} \\ &= \int T(\boldsymbol{x}) \frac{f'_n(\boldsymbol{x};\theta)}{f_n(\boldsymbol{x};\theta)} f_n(\boldsymbol{x};\theta)\, d\boldsymbol{x}. \end{aligned} \tag{2}$$

Notice that the last term is an expected value. Let L denote the log-likelihood,

$$L(\boldsymbol{x};\theta) = \log f_n(\boldsymbol{x};\theta),$$

and observe that

$$\frac{f'_n(\boldsymbol{x};\theta)}{f_n(\boldsymbol{x};\theta)} = L'(\boldsymbol{x};\theta).$$

Okay, so now we can rewrite (2) as

$$1 + b'(\theta) = \boldsymbol{E}_\theta \big[T(\boldsymbol{x}) L'(\boldsymbol{x};\theta) \big]. \tag{3}$$

Take the fact that

$$1 = \int f_n(\boldsymbol{x};\theta)\, d\boldsymbol{x},$$

and differentiate both sides to get

$$\begin{aligned} 0 &= \int f'_n(\boldsymbol{x};\theta)\, d\boldsymbol{x} = 0 \\ &= \int \frac{f'(\boldsymbol{x};\theta)}{f(\boldsymbol{x};\theta)} f(\boldsymbol{x};\theta)\, d\boldsymbol{x} \\ &= \boldsymbol{E}_\theta\, L'(\boldsymbol{x};\theta). \end{aligned} \tag{4}$$

Multiply both sides of this by $\boldsymbol{E}_\theta\, T$ and subtract it from (3) to get

$$1 + b'(\theta) = \boldsymbol{E}_\theta \big[\big(T(\boldsymbol{x}) - \boldsymbol{E}_\theta\, T \big) L'(\boldsymbol{x};\theta) \big]. \tag{5}$$

The right-hand side is the expectation of a product, so we can use the Schwarz Inequality (Lemma S3.2.1) below to get a bound on it. Square both sides of (5) to get

$$\big(1 + b'(\theta)\big)^2 = \big\{ \boldsymbol{E}_\theta \big[\big(T(\boldsymbol{x}) - \boldsymbol{E}_\theta\, T \big) L'(\boldsymbol{x};\theta) \big] \big\}^2 \leqslant \underbrace{\boldsymbol{E}_\theta \big(T - \boldsymbol{E}_\theta\, T \big)^2}_{= \boldsymbol{Var}_\theta\, T} \boldsymbol{E}_\theta (L'^2).$$

Rearranging this gives

$$\boldsymbol{Var}_\theta\, T \geqslant \frac{\big[1 + b'(\theta)\big]^2}{\boldsymbol{E}_\theta \left[\big(\frac{\partial}{\partial\theta} \log f_n(\boldsymbol{x};\theta) \big)^2 \right]}. \tag{6}$$

The joint density f_n is a product, so

$$\frac{\partial}{\partial\theta} \log f_n(\boldsymbol{x};\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(x_i;\theta) \tag{7}$$

Now the same argument as in (4) shows that $E_\theta \frac{\partial}{\partial\theta} \log f(X_i;\theta) = 0$, so (7) is a sum of n independent mean zero variables. Thus its variance is just n times the expected square of any one of them. That is, (6) can be rewritten as

$$\boldsymbol{Var}_\theta\, T \geqslant \frac{\big[1 + b'(\theta)\big]^2}{n\, \boldsymbol{E}_\theta \left[\big(\frac{\partial}{\partial\theta} \log f(X;\theta) \big)^2 \right]}.$$

When the bias is always zero, then $b'(\theta) = 0$, and this reduces to Theorem 5.5.1 in Larsen and Marx [6].

∎

I did play fast and loose with some of the math. In particular, I assumed I could differentiate under the integral sign, and I assumed that the bias was differentiable. See [2] for when this is permissible. I also assumed that the denominator above was nonzero, but all we need to do is restrict attention the support.

S3.2 Schwarz Inequality

You know this result, but the proof that van der Waerden gave was so pretty, I reproduced it here.

S3.2.1 Lemma (Schwarz Inequality) *If Y and Z are random variables with finite second moments, then*

$$(\boldsymbol{E}\, YZ)^2 \leqslant (\boldsymbol{E}\, Y^2)(\boldsymbol{E}\, Z^2).$$

Proof: (van der Waerden [8, p. 161]) The quadratic form in (a, b) defined by

$$\boldsymbol{E}(aY + bZ)^2 = (\boldsymbol{E}\, Y^2)a^2 + 2(\boldsymbol{E}\, YZ)ab + (\boldsymbol{E}\, Z^2)b^2$$

is positive semidefinite, so its determinant is nonnegative (see, e.g., [1]). That is,

$$(\boldsymbol{E}\, Y^2)(\boldsymbol{E}\, Z^2) - (\boldsymbol{E}\, YZ)^2 \geqslant 0.$$

■

Bibliography

[1] K. C. Border. 2001. More than you wanted to know about quadratic forms. On-line note. http://www.hss.caltech.edu/~kcb/Notes/QuadraticForms.pdf

[2] ——— . 2013. Differentiating an integral. On-line note. http://www.hss.caltech.edu/~kcb/Notes/DifferentiatingAnIntegral.pdf

[3] H. Cramér. 1946. A contribution to the theory of statistical estimation. *Skandinavisk Aktuarietidskrift* 29:85–94.

[4] ——— . 1946. *Mathematical methods of statistics*. Number 34 in Princeton Mathematical Series. Princeton, New Jersey: Princeton University Press. Reprinted 1974.

[5] M. Fréchet. 1943. Sur l'extension de certaines evaluations statistiques au cas de petits echantillons. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute* 11(3/4):182–205. http://www.jstor.org/stable/1401114

[6] R. J. Larsen and M. L. Marx. 2012. *An introduction to mathematical statistics and its applications*, fifth ed. Boston: Prentice Hall.

[7] C. R. Rao. 1945. Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society* 37(3):81–91. http://bulletin.calmathsoc.org/article.php?ID=B.1945.37.14

[8] B. L. van der Waerden. 1969. *Mathematical statistics*. Number 156 in Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete. New York, Berlin, and Heidelberg: Springer–Verlag. Translated by Virginia Thompson and Ellen Sherman from *Mathematische Statistik*, published by Springer-Verlag in 1965, as volume 87 in the series Grundlerhen der mathematischen Wissenschaften.