Cramer-Rao Bound

Konstantinos G. Derpanis

September 8, 2006

The *Cramer-Rao bound* establishes the lower limit on how much "information" about an unknown probability distribution parameter a set of measurements carries. More specifically, the inequality establishes the minimum variance for an unbiased estimator of the underlying parameter, θ, of a probability distribution, $p(x;\theta)$. Three important points must be kept in mind about the Cramer-Rao bound: 1) the bound pertains only to unbiased estimators, biased estimators may violate the lower bound 2) the bound may be unreachable in practice, and 3) *maximum likelihood estimators* achieve the lower bound as the size of the measurement set tends to infinity. The following note outlines the proof of the Cramer-Rao bound for a single parameter (adapted from (Gershenfeld, 1999)).

Before proving the Cramer-Rao bound, let's first establish several key components of the proof.

The score, S is defined as,

$$
\begin{aligned}
S &= \frac{\partial}{\partial\theta}\log p(x;\theta) && (1)\\
&= \frac{\partial p(x;\theta)}{\partial\theta}\frac{1}{p(x;\theta)}. && (2)
\end{aligned}
$$

The expected value of the score, $E\{S\}$, is

$$
\begin{aligned}
E\{S\} &= \int_{-\infty}^{\infty}\left(\frac{\partial p(x;\theta)}{\partial\theta}\frac{1}{p(x;\theta)}\right)p(x;\theta)dx && (3)\\
&= 0. && (4)
\end{aligned}
$$

The variance of the score, $Var\{S\}$, is termed the *Fisher information* (denoted $I(\theta)$).

The score for a set of N independent identically-distributed (i.i.d.) variables is the sum of the respective scores,

$$
\begin{aligned}
S(x_1,x_2,\ldots,x_N) &= \frac{\partial}{\partial\theta}\log p(x_1,x_2,\ldots,x_N) && (5)\\
&= \frac{\partial}{\partial\theta}\log\Big(p(x_1)p(x_2)\cdots p(x_N)\Big) && (6)\\
&= \sum_{i=1}^{N}\frac{\partial}{\partial\theta}\log p(x_i;\theta) && (7)\\
&= \sum_{i=1}^{N}S(x_i). && (8)
\end{aligned}
$$

Similarly, it can be shown that the Fisher information for the set is $NI(\theta)$.

Theorem: The mean square error of an unbiased estimator, g, of a probability distribution parameter, θ, is lower bounded by the reciprocal of the Fisher information, $I(\theta)$, formally,

$$Var(g) \geq \frac{1}{I(\theta)}. \tag{9}$$

This lower bound is known as the Cramer-Rao bound.

Proof: Without loss of generality, the proof considers the bound related to a single measurement. The general case of a set of i.i.d. measurements follows in a straightforward manner. By the *Cauchy-Schwarz inequality*[1] ,

$$\left(E\left\{(S - E\{S\})(g - E\{g\})\right\}\right)^2 \leq E\left\{(S - E\{S\})^2\right\}E\left\{(g - E\{g\})^2\right\}. \tag{10}$$

Further expansion of (10) yields,

$$\left(E\left\{Sg - E\{S\}g - SE\{g\} + E\{S\}E\{g\}\right\}\right)^2 \leq E\left\{S^2 - 2SE\{S\} + E\{S\}^2\right\} Var\{g\}. \tag{11}$$

Since the expected value of the score is zero, $E\{S\} = 0$, (11) simplifies as follows,

$$\left(E\left\{Sg - SE\{g\}\right\}\right)^2 \leq E\{S^2\}\, Var\{g\} \tag{12}$$

$$\left(E\{Sg\} - E\left\{SE\{g\}\right\}\right)^2 \leq I(\theta)\, Var\{g\} \tag{13}$$

$$\left(E\{Sg\} - E\{S\}E\{g\}\right)^2 \leq I(\theta)\, Var\{g\} \tag{14}$$

$$\left(E\{Sg\}\right)^2 \leq I(\theta)\, Var\{g\} \tag{15}$$

$$\left(\int_{-\infty}^{\infty} \left(\frac{\partial p(x;\theta)}{\partial \theta} \frac{1}{p(x;\theta)}\right) g(x)p(x;\theta)dx\right)^2 \leq I(\theta)\, Var\{g\} \tag{16}$$

$$\left(\frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} g(x)p(x;\theta)dx\right)^2 \leq I(\theta)\, Var\{g\} \tag{17}$$

$$\left(\frac{\partial}{\partial \theta} E\{g\}\right)^2 \leq I(\theta)\, Var\{g\}. \tag{18}$$

Since g is an unbiased estimator (i.e., $E\{g\} = \theta$), (18) becomes,

$$\left(\frac{\partial}{\partial \theta}\theta\right)^2 \leq I(\theta)\, Var\{g\} \tag{19}$$

$$1 \leq I(\theta)\, Var\{g\}. \tag{20}$$

Thus,

$$Var(g) \geq \frac{1}{I(\theta)}, \tag{21}$$

as desired.

[1] *Cauchy-Schwarz inequality*: $E\{XY\}^2 \leq E\{X\}E\{Y\}$

References

Gershenfeld, N. (1999). *The Nature of Mathematical Modeling*. New York: Cambridge University Press.