Cramér–Rao lower bound

The Cramér–Rao lower bound is a classic result in statistics. I provide below an outline of a proof in the multi-parameter setting. See for example [1] chapter 8.

Consider an unbiased estimator \( \hat{\theta}(x) \) for an unknown parameter vector \( \theta \) with likelihood \( f_X(x) \) at sample \( x \).

\[
0 = \mathbb{E} \left[ \hat{\theta}(X) - \theta \right] = \int \left( \hat{\theta}(x) - \theta \right) f_X(x) \, dx
\]

If we take the (vector) derivative with respect to the parameters, and we are allowed to distribute it, we get

\[
0 = \int \left( \hat{\theta}(x) - \theta \right) \frac{\partial f_X(x)}{\partial \theta} \, dx - I \int f_X(x) \, dx
\]

or, with some manipulation,

\[
\int \left( \left( \hat{\theta}(x) - \theta \right) \sqrt{f_X(x)} \right) \left( \frac{\partial \log f_X(x)}{\partial \theta} \sqrt{f_X(x)} \right) \, dx = I
\]

Consider any vectors \( a \) and \( b \) in parameter space. The previous result means

\[
\int \left( a' \left( \hat{\theta}(x) - \theta \right) \sqrt{f_X(x)} \right) \left( \frac{\partial \log f_X(x)}{\partial \theta} \sqrt{f_X(x)} b \right) \, dx = a'b
\]

This can be thought of as an inner product in the Hilbert space \( L^2 \), which means we can apply Cauchy-Schwarz to get

\[
a' \left( \int \left( \hat{\theta}(x) - \theta \right) \left( \hat{\theta}(x) - \theta \right)' f_X(x) \, dx \right) a
\]

\[
b' \left( \int \frac{\partial \log f_X(x)}{\partial \theta'} \frac{\partial \log f_X(x)}{\partial \theta} f_X(x) \, dx \right) b \geq (a'b)^2
\]
Fisher Information

Define the Fisher information to be

\[ \mathcal{I}(\theta) \overset{\triangle}{=} \mathbb{E} \left[ \frac{\partial \log f_X(X)}{\partial \theta} \frac{\partial \log f_X(X)}{\partial \theta} \right] \]

\[ = \text{cov} \left[ \frac{\partial \log f_X(X)}{\partial \theta} \right] \]

\[ = \mathbb{E} \left[ -\frac{\partial^2}{\partial \theta \partial \theta'} \log f_X(X) \right] \]

if the log-likelihood is twice differentiable on its support in the last instance.

With \( b \overset{\triangle}{=} \mathcal{I}^{-1}(\theta) a \), the previous result translates to

\[ \left( a' \text{cov} \left[ \hat{\theta}(X) \right] a \right) (a' \mathcal{I}^{-1}(\theta) a) \geq (a' \mathcal{I}^{-1}(\theta) a)^2 \]

So we can conclude that

\[ a' \left( \text{cov} \left[ \hat{\theta}(X) \right] - \mathcal{I}^{-1}(\theta) \right) a \geq 0 \]

for all vectors \( a \).

This conforms with the definition of a positive semi-definite matrix, and can be written as

\[ \text{cov} \left[ \hat{\theta}(X) \right] \geq \mathcal{I}^{-1}(\theta) \]

(1)

Note that the Cramér–Rao lower bound is a special case of the Kullback inequality about the relative entropy of one measure with respect to another.

References