Cramér–Rao bound

In estimation theory and statistics, the **Cramér–Rao bound (CRB)**, **Cramér–Rao lower bound (CRLB)**, **Cramér–Rao inequality**, **Fréchet–Darmois–Cramér–Rao inequality**, or **information inequality** expresses a lower bound on the variance of unbiased estimators of a deterministic (fixed, though unknown) parameter. This term is named in honor of Harald Cramér,[1] Calyampudi Radhakrishna Rao,[2][3] Maurice Fréchet[4] and Georges Darmois[5] all of whom independently derived this limit to statistical precision in the 1940s.[6][7]

In its simplest form, the bound states that the variance of any unbiased estimator is at least as high as the inverse of the Fisher information. An unbiased estimator which achieves this lower bound is said to be (fully) efficient. Such a solution achieves the lowest possible mean squared error among all unbiased methods, and is therefore the minimum variance unbiased (MVU) estimator. However, in some cases, no unbiased technique exists which achieves the bound. This may occur either if for any unbiased estimator, there exists another with a strictly smaller variance, or if an MVU estimator exists, but its variance is strictly greater than the inverse of the Fisher information.

The Cramér–Rao bound can also be used to bound the variance of *biased* estimators of given bias. In some cases, a biased approach can result in both a variance and a mean squared error that are *below* the unbiased Cramér–Rao lower bound; see estimator bias.

Contents

Statement

The Cramér–Rao bound is stated in this section for several increasingly general cases, beginning with the case in which the parameter is a scalar and its estimator is unbiased. All versions of the bound require certain regularity conditions, which hold for most well-behaved distributions. These conditions are listed later in this section.

Scalar unbiased case

Suppose θ is an unknown deterministic parameter which is to be estimated from measurements x, distributed according to some probability density function $f(x;\theta)$. The variance of any *unbiased* estimator $\hat{\theta}$ of θ is then bounded by the reciprocal of the Fisher information $I(\theta)$:

$$\operatorname{var}(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

where the Fisher information $I(\theta)$ is defined by

$$I(\theta) = \mathrm{E}\left[\left(\frac{\partial \ell(x;\theta)}{\partial \theta}\right)^2\right] = -\mathrm{E}\left[\frac{\partial^2 \ell(x;\theta)}{\partial \theta^2}\right]$$

and $\ell(x;\theta) = \log(f(x;\theta))$ is the natural logarithm of the likelihood function and $\mathbf{E}$ denotes the expected value (over x).

The efficiency of an unbiased estimator $\hat{\theta}$ measures how close this estimator's variance comes to this lower bound; estimator efficiency is defined as

$$e(\hat{\theta}) = \frac{I(\theta)^{-1}}{\operatorname{var}(\hat{\theta})}$$

or the minimum possible variance for an unbiased estimator divided by its actual variance. The Cramér–Rao lower bound thus gives

$$e(\hat{\theta}) \leq 1.$$

General scalar case

A more general form of the bound can be obtained by considering a biased estimator $T(X)$ of a function of the parameter, $\psi(\theta)$. Here, biasedness is understood as stating that $E\{T(X)\} - \theta = \psi(\theta) - \theta$ is not generally equal to 0. In this case, the bound is given by

$$\operatorname{var}(T) \geq \frac{[\psi'(\theta)]^2}{I(\theta)}$$

where $\psi'(\theta)$ is the derivative of $\psi(\theta)$ (by θ), and $I(\theta)$ is the Fisher information defined above.

Bound on the variance of biased estimators

Apart from being a bound on estimators of functions of the parameter, this approach can be used to derive a bound on the variance of biased estimators with a given bias, as follows. Consider an estimator $\hat{\theta}$ with bias $b(\theta) = E\{\hat{\theta}\} - \theta$, and let $\psi(\theta) = b(\theta) + \theta$. By the result above, any unbiased estimator whose expectation is $\psi(\theta)$ has variance greater than or equal to $(\psi'(\theta))^2/I(\theta)$. Thus, any estimator $\hat{\theta}$ whose bias is given by a function $b(\theta)$ satisfies

$$\operatorname{var}\left(\hat{\theta}\right) \geq \frac{[1 + b'(\theta)]^2}{I(\theta)}.$$

The unbiased version of the bound is a special case of this result, with $b(\theta) = 0$.

It's trivial to have a small variance – an "estimator" that is constant has a variance of zero. But from the above equation we find that the mean squared error of a biased estimator is bounded by

$$\mathrm{E}\left((\hat{\theta} - \theta)^2\right) \geq \frac{[1 + b'(\theta)]^2}{I(\theta)} + b(\theta)^2,$$

using the standard decomposition of the MSE. Note, however, that if $1 + b'(\theta) < 1$ this bound might be less than the unbiased Cramér–Rao bound $1/I(\theta)$. For instance, in the example of estimating variance below, $1 + b'(\theta) = \frac{n}{n+2} < 1$.

Multivariate case

Extending the Cramér–Rao bound to multiple parameters, define a parameter column vector

$$\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_d]^T \in \mathbb{R}^d$$

with probability density function $f(x; \boldsymbol{\theta})$ which satisfies the two regularity conditions below.

The Fisher information matrix is a $d \times d$ matrix with element $I_{m,k}$ defined as

$$I_{m,k} = \mathrm{E}\left[\frac{\partial}{\partial \theta_m} \log f(x; \boldsymbol{\theta}) \frac{\partial}{\partial \theta_k} \log f(x; \boldsymbol{\theta})\right] = -\mathrm{E}\left[\frac{\partial^2}{\partial \theta_m \, \partial \theta_k} \log f(x; \boldsymbol{\theta})\right].$$

Let $\boldsymbol{T}(X)$ be an estimator of any vector function of parameters, $\boldsymbol{T}(X) = (T_1(X), \ldots, T_d(X))^T$, and denote its expectation vector $\mathrm{E}[\boldsymbol{T}(X)]$ by $\boldsymbol{\psi}(\boldsymbol{\theta})$. The Cramér–Rao bound then states that the covariance matrix of $\boldsymbol{T}(X)$ satisfies

$$\mathrm{cov}_{\boldsymbol{\theta}}(\boldsymbol{T}(X)) \geq \frac{\partial \boldsymbol{\psi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} [I(\boldsymbol{\theta})]^{-1} \left(\frac{\partial \boldsymbol{\psi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T$$

where

- The matrix inequality $A \geq B$ is understood to mean that the matrix $A - B$ is positive semidefinite, and
- $\partial \boldsymbol{\psi}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ is the Jacobian matrix whose ij element is given by $\partial \psi_i(\boldsymbol{\theta})/\partial \theta_j$.

If $\boldsymbol{T}(X)$ is an unbiased estimator of $\boldsymbol{\theta}$ (i.e., $\boldsymbol{\psi}(\boldsymbol{\theta}) = \boldsymbol{\theta}$), then the Cramér–Rao bound reduces to

$$\mathrm{cov}_{\boldsymbol{\theta}}(\boldsymbol{T}(X)) \geq I(\boldsymbol{\theta})^{-1}.$$

If it is inconvenient to compute the inverse of the Fisher information matrix, then one can simply take the reciprocal of the corresponding diagonal element to find a (possibly loose) lower bound.[8]

$$\mathrm{var}_{\boldsymbol{\theta}}(T_m(X)) = [\mathrm{cov}_{\boldsymbol{\theta}}(\boldsymbol{T}(X))]_{mm} \geq \left[I(\boldsymbol{\theta})^{-1}\right]_{mm} \geq ([I(\boldsymbol{\theta})]_{mm})^{-1}.$$

Regularity conditions

The bound relies on two weak regularity conditions on the probability density function, $f(x; \theta)$, and the estimator $T(X)$:

- The Fisher information is always defined; equivalently, for all x such that $f(x; \theta) > 0$,

$$\frac{\partial}{\partial \theta} \log f(x; \theta)$$

exists, and is finite.

- The operations of integration with respect to x and differentiation with respect to θ can be interchanged in the expectation of T; that is,

$$\frac{\partial}{\partial \theta} \left[\int T(x) f(x; \theta) \, dx \right] = \int T(x) \left[\frac{\partial}{\partial \theta} f(x; \theta) \right] \, dx$$

whenever the right-hand side is finite.
This condition can often be confirmed by using the fact that integration and differentiation can be swapped when either of the following cases hold:

1. The function $f(x; \theta)$ has bounded support in x, and the bounds do not depend on θ;
2. The function $f(x; \theta)$ has infinite support, is continuously differentiable, and the integral converges uniformly for all θ.

Simplified form of the Fisher information

Suppose, in addition, that the operations of integration and differentiation can be swapped for the second derivative of $f(x; \theta)$ as well, i.e.,

$$\frac{\partial^2}{\partial \theta^2} \left[\int T(x) f(x; \theta) \, dx \right] = \int T(x) \left[\frac{\partial^2}{\partial \theta^2} f(x; \theta) \right] \, dx.$$

In this case, it can be shown that the Fisher information equals

$$I(\theta) = -\,\mathrm{E} \left[\frac{\partial^2}{\partial \theta^2} \log f(X; \theta) \right].$$

The Cramèr–Rao bound can then be written as

$$\mathrm{var}\left(\hat{\theta} \right) \geq \frac{1}{I(\theta)} = \frac{1}{-\,\mathrm{E} \left[\frac{\partial^2}{\partial \theta^2} \log f(X; \theta) \right]}.$$

In some cases, this formula gives a more convenient technique for evaluating the bound.

Single-parameter proof

The following is a proof of the general scalar case of the Cramér–Rao bound described above. Assume that $T = t(X)$ is an unbiased estimator for the value $\psi(\theta)$ (based on the observations X), and so $\mathrm{E}(T) = \psi(\theta)$. The goal is to prove that, for all θ,

$$\mathrm{var}(t(X)) \geq \frac{[\psi'(\theta)]^2}{I(\theta)}.$$

Let X be a random variable with probability density function $f(x; \theta)$. Here $T = t(X)$ is a statistic, which is used as an estimator for $\psi(\theta)$. Define V as the score:

$$V = \frac{\partial}{\partial\theta}\ln f(X;\theta) = \frac{1}{f(X;\theta)}\frac{\partial}{\partial\theta}f(X;\theta)$$

where the chain rule is used in the final equality above. Then the expectation of V, written $\mathbf{E}(V)$, is zero. This is because:

$$\mathbf{E}(V) = \int f(x;\theta)\left[\frac{1}{f(x;\theta)}\frac{\partial}{\partial\theta}f(x;\theta)\right]dx = \frac{\partial}{\partial\theta}\int f(x;\theta)\,dx = 0$$

where the integral and partial derivative have been interchanged (justified by the second regularity condition).

If we consider the covariance $\mathbf{cov}(V,T)$ of V and T, we have $\mathbf{cov}(V,T) = \mathbf{E}(VT)$, because $\mathbf{E}(V) = 0$. Expanding this expression we have

$$\begin{aligned}\mathbf{cov}(V,T) &= \mathbf{E}\left(T\cdot\left[\frac{1}{f(X;\theta)}\frac{\partial}{\partial\theta}f(X;\theta)\right]\right)\\ &= \int t(x)\left[\frac{1}{f(x;\theta)}\frac{\partial}{\partial\theta}f(x;\theta)\right]f(x;\theta)\,dx\\ &= \frac{\partial}{\partial\theta}\left[\int t(x)f(x;\theta)\,dx\right] = \psi'(\theta)\end{aligned}$$

again because the integration and differentiation operations commute (second condition).

The Cauchy–Schwarz inequality shows that

$$\sqrt{\mathrm{var}(T)\,\mathrm{var}(V)} \geq |\mathbf{cov}(V,T)| = |\psi'(\theta)|$$

therefore

$$\mathrm{var}(T) \geq \frac{[\psi'(\theta)]^2}{\mathrm{var}(V)} = \frac{[\psi'(\theta)]^2}{I(\theta)}$$

which proves the proposition.

Examples

Multivariate normal distribution

For the case of a d-variate normal distribution

$$\boldsymbol{x} \sim N_d\left(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{C}(\boldsymbol{\theta})\right)$$

the Fisher information matrix has elements[9]

$$I_{m,k} = \frac{\partial\boldsymbol{\mu}^T}{\partial\theta_m}\boldsymbol{C}^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\theta_k} + \frac{1}{2}\mathrm{tr}\left(\boldsymbol{C}^{-1}\frac{\partial\boldsymbol{C}}{\partial\theta_m}\boldsymbol{C}^{-1}\frac{\partial\boldsymbol{C}}{\partial\theta_k}\right)$$

where "tr" is the trace.

For example, let $w[n]$ be a sample of N independent observations with unknown mean θ and known variance σ^2.

$$w[n] \sim \mathbb{N}_N \left(\theta \mathbf{1}, \sigma^2 \mathbf{I} \right).$$

Then the Fisher information is a scalar given by

$$I(\theta) = \left(\frac{\partial \boldsymbol{\mu}(\theta)}{\partial \theta} \right)^T \mathbf{C}^{-1} \left(\frac{\partial \boldsymbol{\mu}(\theta)}{\partial \theta} \right) = \sum_{i=1}^{N} \frac{1}{\sigma^2} = \frac{N}{\sigma^2},$$

and so the Cramér–Rao bound is

$$\mathrm{var}\left(\hat{\theta} \right) \geq \frac{\sigma^2}{N}.$$

Normal variance with known mean

Suppose X is a normally distributed random variable with known mean μ and unknown variance σ^2. Consider the following statistic:

$$T = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n}.$$

Then T is unbiased for σ^2, as $E(T) = \sigma^2$. What is the variance of T?

$$\mathrm{var}(T) = \frac{\mathrm{var}(X - \mu)^2}{n} = \frac{1}{n} \left[E\{(X - \mu)^4\} - \left(E\{(X - \mu)^2\} \right)^2 \right]$$

(the second equality follows directly from the definition of variance). The first term is the fourth moment about the mean and has value $3(\sigma^2)^2$; the second is the square of the variance, or $(\sigma^2)^2$. Thus

$$\mathrm{var}(T) = \frac{2(\sigma^2)^2}{n}.$$

Now, what is the Fisher information in the sample? Recall that the score V is defined as

$$V = \frac{\partial}{\partial \sigma^2} \log L(\sigma^2, X)$$

where L is the likelihood function. Thus in this case,

$$V = \frac{\partial}{\partial \sigma^2} \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(X-\mu)^2/2\sigma^2} \right] = \frac{(X - \mu)^2}{2(\sigma^2)^2} - \frac{1}{2\sigma^2}$$

where the second equality is from elementary calculus. Thus, the information in a single observation is just minus the expectation of the derivative of V, or

$$I = -E\left(\frac{\partial V}{\partial \sigma^2} \right) = -E\left(-\frac{(X - \mu)^2}{(\sigma^2)^3} + \frac{1}{2(\sigma^2)^2} \right) = \frac{\sigma^2}{(\sigma^2)^3} - \frac{1}{2(\sigma^2)^2} = \frac{1}{2(\sigma^2)^2}.$$

Thus the information in a sample of n independent observations is just n times this, or $\dfrac{n}{2(\sigma^2)^2}$.

The Cramer–Rao bound states that

$$\operatorname{var}(T) \geq \frac{1}{I}.$$

In this case, the inequality is saturated (equality is achieved), showing that the estimator is efficient.

However, we can achieve a lower mean squared error using a biased estimator. The estimator

$$T = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{n + 2}.$$

obviously has a smaller variance, which is in fact

$$\operatorname{var}(T) = \frac{2n(\sigma^2)^2}{(n+2)^2}.$$

Its bias is

$$\left(1 - \frac{n}{n+2}\right)\sigma^2 = \frac{2\sigma^2}{n+2}$$

so its mean squared error is

$$\operatorname{MSE}(T) = \left(\frac{2n}{(n+2)^2} + \frac{4}{(n+2)^2}\right)(\sigma^2)^2 = \frac{2(\sigma^2)^2}{n+2}$$

which is clearly less than the Cramér–Rao bound found above.

When the mean is not known, the minimum mean squared error estimate of the variance of a sample from Gaussian distribution is achieved by dividing by $n + 1$, rather than $n - 1$ or $n + 2$.

See also

- Chapman–Robbins bound
- Kullback's inequality
- Brascamp–Lieb inequality

References and notes

1. Cramér, Harald (1946). *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press. ISBN 0-691-08004-6. OCLC 185436716 (https://www.worldcat.org/oclc/185436716).
2. Rao, Calyampudi Radakrishna (1945). "Information and the accuracy attainable in the estimation of statistical parameters". *Bulletin of the Calcutta Mathematical Society*. **37**: 81–89. MR 0015748 (https://www.ams.org/mathscinet-getitem?mr=0015748).
3. Rao, Calyampudi Radakrishna (1994). S. Das Gupta (ed.). *Selected Papers of C. R. Rao*. New York: Wiley. ISBN 978-0-470-22091-7. OCLC 174244259 (https://www.worldcat.org/oclc/174244259).

4. Fréchet, Maurice (1943). "Sur l'extension de certaines évaluations statistiques au cas de petits échantillons". *Rev. Inst. Int. Statist*. **11**: 182–205.

5. Darmois, Georges (1945). "Sur les limites de la dispersion de certaines estimations". *Rev. Int. Inst. Statist*. **13**: 9–15.

6. Gart, John J. (1958). "An extension of the Cramér–Rao inequality". *Ann. Math. Stat*. **29**: 367–380.

7. Malécot, Gustave (1999). "Statistical methods and the subjective basis of scientific knowledge" (http://www.gsejournal.org/content/31/3/269) [translated from Année X 1947 by Daniel Gianola]. *Genet. Sel. Evol*. **31**: 269–298.

8. For the Bayesian case, see eqn. (11) of Bobrovsky; Mayer-Wolf; Zakai (1987). "Some classes of global Cramer–Rao bounds". *Ann. Stat*. **15** (4): 1421–38.

9. Kay, S. M. (1993). *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall. p. 47. ISBN 0-13-042268-1.

Further reading

- Amemiya, Takeshi (1985). *Advanced Econometrics* (https://books.google.com/books?id=0bzGQE14CwEC&pg=PA14). Cambridge: Harvard University Press. pp. 14–17. ISBN 0-674-00560-0.
- Bos, Adriaan van den (2007). *Parameter Estimation for Scientists and Engineers*. Hoboken: John Wiley & Sons. pp. 45–98. ISBN 0-470-14781-4.
- Kay, Steven M. (1993). *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall. ISBN 0-13-345711-7.. Chapter 3.
- Shao, Jun (1998). *Mathematical Statistics*. New York: Springer. ISBN 0-387-98674-X.. Section 3.1.3.

External links

- FandPLimitTool (https://archive.is/20121215055105/http://www4.utsouthwestern.edu/wardlab/fandplimittool.asp) a GUI-based software to calculate the Fisher information and Cramer-Rao Lower Bound with application to single-molecule microscopy.