The Cramér–Rao inequality for multidimensional parameter spaces

The Cramér–Rao inequality (under some regularity conditions) gives unified lower bound for the covariance matrix of every unbiased estimator (for a given parameter function) based merely on a quantity, called Fisher information matrix, which can be calculated from the underlying distribution as a function of the parameter.

Let $(\Omega, \mathcal{A}, \mathcal{P})$ be dominated, identifiable, parametric statistical space, and $X_1, \dots, X_n$ be i.i.d. (univariate or multivariate) sample from the $\mathbb{P}_\theta$ distribution, where $\theta \in \Theta \subset \mathbb{R}^k$ is the parameter space. Suppose, we want to estimate the function $\psi(\theta) = (\psi_1(\theta), \dots, \psi_k(\theta))$ of the parameter, where $\psi : \mathbb{R}^k \to \mathbb{R}^k$ is one-to-one function (usually it is the identity, when the parameter θ itself is to be estimated).

1. Based on the above i.i.d. sample, construct the k-dimensional statistic $\mathbf{T}(X_1, \dots, X_n)$, briefly $\mathbf{T} = (T_1, \dots T_k)$ which is an unbiased estimator of $\psi(\theta)$ in the sense that

$$\mathbb{E}_\theta \mathbf{T} = \psi(\theta), \quad \forall \theta \in \Theta,$$

where the expectation of the random vector $\mathbf{T}$ is the vector $(\mathbb{E}_\theta T_1, \dots, \mathbb{E}_\theta T_k)'$ and $'$ denotes the transposition so that the vectors be column vectors.

2. Based on the $\mathbb{P}_\theta$ distribution itself, calculate the Fisher information matrix of the 1-element sample:

$$\mathbf{I}_1(\theta) = \mathbb{E}_\theta \left(\frac{\partial}{\partial \theta} \ln f_\theta(X) \right) \left(\frac{\partial}{\partial \theta} \ln f_\theta(X) \right)' = \mathrm{Var}_\theta \left(\frac{\partial}{\partial \theta} \ln f_\theta(X) \right),$$

where $X \sim \mathbb{P}_\theta$ and f_θ is the p.d.f. of the $\mathbb{P}_\theta$ distribution (if it is absolutely continuous, and use p_θ for the p.m.f. if $\mathbb{P}_\theta$ is discrete). Here under Var of a random vector its covariance matrix is understood, whereas under the derivative with respect to the vector θ of the scalar valued function g the column vector of the derivatives with respect to the components of θ is understood, that is $\frac{\partial}{\partial \theta} g = (\frac{\partial}{\partial \theta_1} g, \dots, \frac{\partial}{\partial \theta_k} g)'$. The information matrix of the n-element sample is defined by $\mathbf{I}_n(\theta) = n\mathbf{I}_1(\theta)$, provided the regularity condition

$$\mathbb{E}_\theta \left(\frac{\partial}{\partial \theta} \ln f_\theta(\mathbf{X}) \right) = \mathbf{0}$$

holds, where $\mathbf{0}$ is the zero vector. In fact, this condition follows from some simpler ones (we can "differentiate through" the $\int$ or $\sum$). In this case, $\mathbf{I}_n(\theta)$ is the $k \times k$ covariance matrix of the k-dimensional random vector $\frac{\partial}{\partial \theta} \ln \prod_{i=1}^n f_\theta(X_i)$.

3. The Cramér–Rao inequality states the following relation between 1 and 2. Denote by $\mathbf{S} = (s_{ij})$ the $k \times k$ matrix of entries $s_{ij} = \frac{\partial}{\partial \theta_j} \psi_i(\theta)$. Then for the covariance matrix of any unbiased estimator (for $\psi(\theta)$) the inequality

$$\mathrm{Var}_\theta \mathbf{T} \geq \frac{1}{n} \mathbf{S}\mathbf{I}_1^{-1}(\theta)\mathbf{S}' = \mathbf{S}\mathbf{I}_n^{-1}(\theta)\mathbf{S}'$$

holds, where the inequality means that the difference of the left and right hand sides is positive semidefinite. Here $\mathbf{S}$ also depends on θ, however, if θ itself is estimated, then $\mathbf{S}$ is the identity matrix and does not appear in the formula above.