Statistics

Luc Veldhuis

November 2016

Optimalitiy theory

Criteria

- MSE
- UMVU

Cramer-Rao lower bound

For some unbiased estimators T of $g(\theta)$ there exists a lower bound on a $\mathbb{V}_\theta(T)$ (=MSE(T)) if T unbiased

Definition

The Fisher information in the data $\underline{X}$ is $I_\theta = \mathbb{V}_\theta(\dot{l}_\theta(\underline{X}))$ with $\dot{l}_\theta(\underline{X}) = \frac{\delta}{\delta\theta} log(p_\theta(X))$ with p_θ the joint distrubution of $\underline{X}$

Optimalitiy theory

Remark

If $\underline{X} = (X_1, \ldots, X_n)$ is independently identically distributed, then $I_\theta = n i_\theta$ with $i_\theta = \mathbb{V}_\theta(\dot{l}(X_1))$

Theorem

Assume $\theta \mapsto p_\theta$ is differentiable for each x. Then, under regularity conditions, each unbiased estimator T of $g(\theta) \in \mathbb{R}$ satisfies:

$$\mathbb{V}_\theta(T) \geq \frac{(g'(\theta))^2}{I_\theta}$$

with g' the derivative of g

Optimalitiy theory

Example

$X \sim Bin(n, p)$ Determine the Fisher information of the data.

$$
\begin{aligned}
I_p = i_p = \mathbb{V}_p(\dot{l}_p(X)) &== \mathbb{V}_p(\frac{\delta}{\delta p} log(\binom{n}{p} p^x (1-p)^{n-x})) \\
&= \mathbb{V}_p(\frac{\delta}{\delta p} x log(p) + (n-x) log(1-p) + C) \\
&= \mathbb{V}_p(\frac{x}{p} - \frac{n-p}{1-p}) = \mathbb{V}_p(\frac{x-np}{p(1-p)}) \\
&= \mathbb{V}_p(\frac{x}{p(1-p)}) = \frac{n}{p(1-p)}
\end{aligned}
$$

As $\hat{p}_{ML} = \frac{x}{n}$ with $\mathbb{V}_p(\hat{p}_M L) = \frac{p(1-p)}{n} = \frac{1}{i_\theta} = \frac{1}{I_\theta}$.
Thus the Cramer-Rao lower bound is sharp (achieved by the estimator). Hence, $\hat{p}_{ML}$ is UMVU.

Optimalitiy theory

Estimators that achieve the lower bound are called efficient:
They achieve the smallest possible MSE among all unbiased estimators.
The distance between the variance of the estimator and the lower bound is a measure of efficiency. This is used to distinguish between estimators.
The distance tells us how much the estimator can be improved.

Remark

If $\underline{X} = (X_1, \ldots, X_n)$ is a i.i.d sample, then the Cramer-Rao lower bound is:

$$\frac{(g'(\theta))^2}{n i_\theta}$$

Optimalitiy theory

Theorem

Let T be unbiased estimator of $g(\theta)$ and $\mathbb{V}_\theta(T) = \frac{(g'(\theta))^2}{ni_\theta}$ then automatically, T is UMVU.
In particular, for $g(\theta) = \theta$ and $\mathbb{V}_\theta(T) = \frac{1}{ni_\theta}$, then T is UMVU.
Previously $\sqrt{n}(\hat{\theta}_{ML} - \theta) \rightsquigarrow N(0, \frac{1}{i_\theta})$, that is, for large n:
$\hat{\theta}_{ML} \approx\sim N(\theta, \frac{1}{ni_\theta})$. Thus, ML estimators are asymptotically UMVU.

Definition

Asymptotically: when $n \to \infty$

Optimalitiy theory

Remark

Also median of posterior distribution converges to θ and (under some conditions) the Bayes estimator $\hat{\theta}_B$ is asymptotically normally distributed and UMVU: CRB asymptotically.
Moment estimators generally are not UMVU (asymptotically).

Note

Distribution is symmetric: mean $=$ median

Optimalitiy theory

Example

$X_1, \ldots, X_n \sim U[0,\theta]$, $\frac{n+1}{n}X_{(n)}$ is unbiased for θ, $\theta > 0$ and
$\mathbb{V}_\theta(\frac{n+1}{n}X_{(n)}) = \frac{\theta^2}{n(n+2)}$
This variance is smaller than the Cramer-Rao lower bound and the theorems seems invalidated. However, $\theta \mapsto p_\theta$ is not defferentiable as it is discountinuous as $\theta = X_{(n)}$

Remark

The class of biased estimators may harbour estimators with a smaller MSE than the unbiased ones.

Attention

Optimality of tests is skipped §6.4