CRAMER RAO LOWER BOUND

HAOCHEN

1. Efficiency

How to evaluate my estimator, say $\hat{\theta}$, after we derive it. A common way to describe the 'goodness' of estimator is the metric *mean square error(MSE)*, in statistical world we often call it the quadratic risk:

$$risk = \mathbb{E}[(\hat{\theta} - \theta)^2]$$

There is a basic equality between risk and $\mathbb{E}[\hat{\theta}]$, $Var[\hat{\theta}]$.

$$risk = \mathbb{E}[(\hat{\theta} - \theta)^2] = (bias[\hat{\theta}])^2 + Var[\hat{\theta}]$$

proof skipped.

If $\hat{\theta}$ is unbiased, $\mathbb{E}[\hat{\theta}] = 0$, and then:

$$risk = Var[\hat{\theta}]$$

We say an estimator is more efficient if it has lower risk(or *MSE*). Notice that in the case of comparing unbiased estimators, the better one has small variance.

Often, we want to search for optimal estimator which means minimum *MSE,* so we may ask that does that minimum exist, or does *MSE* has a lower bound. In fact, we have the following theorem.

2. Cramer Rao lower bound

Theorem 1. *Suppose $X_1, X_2, ..., X_n$ are samples iid with pdf $f(\cdot;\theta)$, and $\Theta \in \mathbb{R}$, and $T = T(X_1, X_2, ..., X_n)$ is an unbiased estimator of θ,under the smoothness assumption on pdf, we have:*

$$Var[T] \geq \frac{1}{nI(\theta)}$$

where $I(\theta)$ is fisher information.

So if derived etimator has much information, then its lower bound is small, which means small variance. Now lets prove it, we first construct another random variable Z.

Proof. For later use, we define $Z = \frac{\partial l'(\theta)}{\partial \theta}$,where $l'(\theta) = \sum_i log(f(x_i|\theta))$, we can write

$$\begin{aligned} Z &= \sum_{i=1}^{n} \frac{\partial log(f(x_i|\theta))}{\partial \theta} \\ &= \sum_{i=1}^{n} \frac{\frac{\partial}{\partial \theta} f(x_i|\theta)}{f(x_i|\theta)} \end{aligned}$$

Then Z is a random variable with mean 0 and variance $nI(\theta)$.(proof skipped)

Now for Z and T, we have the following obvious inequality:

$$-1 \leq Corr(Z,T) \leq 1$$

thus,

$$Corr(Z,T)^2 \leq 1$$
$$\frac{Cov(Z,T)^2}{Var[Z]Var[T]} \leq 1$$
$$Cov(Z,T)^2 \leq Var[Z]Var[T]$$

Also we know $Var[Z] = nI(\theta)$(proof skipped), so

$$Cov(Z,T)^2 \leq nI(\theta)Var[T]$$

Now the lower bound appears, we finally want to show $Cov(Z,T) = 1$. We know $E[T] = 0$ since T is unbiased. So we have

$$\begin{aligned} Cov(Z,T) &= E[ZT] \\ &= \int ... \int [\sum_{i=1}^{n} \frac{\frac{\partial}{\partial \theta} f(x_i|\theta)}{f(x_i|\theta)}] \cdot T(x_1, ..., x_n) \cdot \prod_{i=1}^{n} f(x_i|\theta) dx_1 ... dx_n \end{aligned} \tag{2.1}$$

Noting that:

$$[\sum_{i=1}^{n} \frac{\frac{\partial}{\partial \theta} f(x_i|\theta)}{f(x_i|\theta)}] \cdot \prod_{i=1}^{n} f(x_i|\theta) = \frac{\partial}{\partial \theta} \prod_{i=1}^{n} f(x_i|\theta)$$

so 2.1 written as

$$\begin{aligned} &\int ... \int T(x_1, ..., x_n) \cdot \frac{\partial}{\partial \theta} \prod_{i=1}^{n} f(x_i|\theta) dx_1 ... dx_n \\ &= \frac{\partial}{\partial \theta} \int ... \int T(x_1, ..., x_n) \cdot \prod_{i=1}^{n} f(x_i|\theta) dx_1 ... dx_n \\ &= \frac{\partial}{\partial \theta} E[T] \\ &= \frac{\partial}{\partial \theta} \theta \\ &= 1 \end{aligned}$$

□

3. Example: Poisson

We know MLE of poisson is $\bar{X}$,and $I(\theta) = \frac{1}{\lambda}$, so by the theorem, $Var[\bar{X}] \geq \frac{n}{\lambda}$, actually, we can computer

$$\begin{aligned} Var[\bar{X}] &= \frac{Var[X_1]}{n} \\ &= \frac{1/\lambda}{n} \\ &= \frac{n}{\lambda} \end{aligned}$$

so this estimator achieve minimum variance(largest information).