CRB

Cramer Rao lower bound in the last class.

Cramer-Rao lower bound was discussed for univariate parameters.

<u>Qn. 4:</u> $\dfrac{\sqrt{n}(Z_1Y_1 + \cdots + Z_nY_n)}{Z_1^2 + \cdots + Z_n^2} \longrightarrow N(0,1)$

<u>Qn: 12:</u> $X_1, \ldots, X_n \overset{iid}{\sim} \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$

<u>Qn. 18:</u> $X_1, \ldots, X_n \overset{iid}{\sim} Pois(\lambda)$, $\quad Var(s^2) > Var(\bar{X})$

<u>Multiparameter case:</u>

Let $X \sim f_{\underline{\theta}}(x)$ $\quad \underline{\theta} = (\theta_1, \ldots, \theta_K)$. In the multiparameter case, we define a score vector rather than a scalar score.

Score vector is defined as

$$U_{\underline{\theta}}(x) = \left(\frac{\partial}{\partial\theta_1} \log f_{\underline{\theta}}(x), \ldots, \frac{\partial}{\partial\theta_K} \log f_{\underline{\theta}}(x)\right)$$

Also we define the information matrix

$$I(\underline{\theta}) = ((I_{ij}(\underline{\theta})))_{i,j=1}^{K}, \text{ where}$$

$$I_{ij}(\underline{\theta}) = E\left[\frac{\partial}{\partial\theta_i} \log f_{\underline{\theta}}(x) \frac{\partial}{\partial\theta_j} \log f_{\underline{\theta}}(x)\right]$$

<u>Multiparameter Cramer-Rao lower bound</u>

Suppose $I(\underline{\theta})$ is positive definite and $\alpha_i = \frac{\partial}{\partial\theta_i} E_{\underline{\theta}}(\delta(\underline{X}))$

exists and differentiation w.r.t. θ_i can be done under integration w.r.t. x. Then

$$Var_{\underline{\theta}}(\delta(\underline{X})) \geq \underline{\alpha}' I(\underline{\theta})^{-1} \underline{\alpha}$$

where $\underline{\alpha} = (\alpha_1, \dots, \alpha_k)$

~~Var~~ 1d-version with one data point

$$Var_\theta(\delta(X)) \geq \frac{\left(\frac{d}{d\theta} E_\theta(\delta(X))\right)^2}{I(\theta)}$$

Method of finding estimators

1. Method of moments
2. Maximum likelihood estimators
3. Bayes and minimax estimators.

1. Method of moments estimators

Let $m_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j$, $j = 1, 2, \dots\dots$

Let $\mu_j' = E[X^j]$. Generally μ_j's are functions of the parameters $\theta_1, \dots, \theta_k$. Let us write $\mu_j'(\theta_1, \dots, \theta_k)$ instead of writing μ_j'.

the method of moment estimator is obtained by solving the following equations

$$m_j = \mu_j'(\theta_1, \dots, \theta_k), \; j = 1, \dots, k.$$

Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

$E[X] = \mu$, $E[X^2] = \mu^2 + \sigma^2$

to get Method of moment (MOM) estimators of μ and σ^2 we will solve the following two equations

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \mu \quad , \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mu^2 + \sigma^2$$

$$\Rightarrow \mu = \bar{X}, \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{(n-1)S^2}{n}$$

In this case MOM of μ is UMVUE, but MOM of σ^2 is not.

Example: $X_1, \ldots, X_n \overset{iid}{\sim} DE(\mu, \sigma)$

$$f(x|\mu, \sigma) = \frac{1}{2\sigma} e^{-\frac{|x-\mu|}{\sigma}}, \quad -\infty < x < \infty$$

$E[X] = \mu \Rightarrow$ the MOM estimator of μ is

$$\frac{1}{n}\sum_{i=1}^{n} X_i$$

in this case MOM is not the UMVUE.

First issue: MOM only uses the information of the population moments. It does not take into account the entire distribution.

Example: $X_1, \ldots, X_n \overset{iid}{\sim} Bin(k, p)$, where k and p are both parameters.

$E[X] = kp$, $E[X^2] = kp(1-p) + (kp)^2$

MOM of k and p satisfy

$$\frac{1}{n}\sum_{i=1}^{n} x_i = kp \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} x_i^2 = kp(1-p) + (kp)^2$$

$$p = \frac{\bar{X}}{k} \Rightarrow \frac{1}{n}\sum_{i=1}^{n} x_i^2 = k\frac{\bar{X}}{k}\left(1-\frac{\bar{X}}{k}\right) + \bar{X}^2$$

$$\Rightarrow k = \frac{\bar{X}}{\bar{X} - \left(\frac{1}{n}\right)\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

the R.H.S. might not be a positive integer.

this MOM does not make any sense.

Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \theta^2)$

$\frac{1}{n}\sum_{i=1}^{n} x_i = \theta$ can be used to find the MOM.

however, if one uses two equations

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \theta, \quad \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \theta^2 + \theta^2$$

they are not consistent equations.

Maximum likelihood estimator

$f_{\theta_1, \ldots, \theta_k}(x)$ be the p.d.f. of X. The likelihood function of $\underline{X} = (X_1, \ldots, X_n)$ is given by $L(\theta_1, \ldots, \theta_k) = \prod_{i=1}^{n} f_{\theta_1, \ldots, \theta_k}(x_i)$. MLE of $\underline{\theta}$, denoted by $\hat{\underline{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$, is the value of $\underline{\theta}$ that maximizes the likelihood $L(\theta_1, \ldots, \theta_k)$.

Maximum likelihood estimator is generally calculated by solving the equations

$$\frac{\partial}{\partial \theta_i} \log L(\underline{\theta}) = 0 \qquad i = 1, \ldots, K.$$

$$\log L(\underline{\theta}) = \sum_{i=1}^{n} \log f_{\theta_1, \ldots, \theta_K}(x_i)$$

$$\Rightarrow \frac{\partial}{\partial \theta_j} \sum_{i=1}^{n} \log f_{\theta_1, \ldots, \theta_K}(x_i) = 0 \quad \forall\ j = 1, \ldots, K$$

$$\Rightarrow \sum_{i=1}^{n} \frac{\partial}{\partial \theta_j} \log f_{\theta_1, \ldots, \theta_K}(x_i) = 0 \quad \forall\ j = 1, \ldots, K.$$

$$\Rightarrow \sum_{i=1}^{n} u_{\underline{\theta}}(x_i) = \underline{0}.$$

where $u_{\underline{\theta}}(x_i)$ is the score function evaluated at x_i.

Also the Maximum Likelihood Estimator (MLE) should satisfy $\left.\frac{\partial^2 L(\underline{\theta})}{\partial \underline{\theta}}\right|_{\theta = \hat{\theta}}$ is negative definite.

Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, 1)$, $\theta > 0$. Find the MLE of θ.

if $\overline{X} > 0$ then $\theta = \overline{X}$ is the MLE.

if $\overline{X} < 0$.

$$L(\theta) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\sum_{i=1}^n \frac{(x_i-\theta)^2}{2}\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^n (x_i-\bar{x}+\bar{x}-\theta)^2\right]\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^n (x_i-\bar{x})^2\right\} \exp\left\{-\frac{n(\bar{x}-\theta)^2}{2}\right\}$$

$\bar{x} = -L$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \qquad \exp\left\{-\frac{n(-L-\theta)^2}{2}\right\}$$

$$= \qquad \exp\left\{-\frac{n(L+\theta)^2}{2}\right\}$$

$L > 0$, also $\theta > 0$

$\Rightarrow \exp\left\{-\frac{n(L+\theta)^2}{2}\right\}$ decreasing as a function of θ for $\theta > 0$. Hence maximum is achieved at $\theta = 0$.

therefore MLE of $\theta = \bar{x}$ if $\bar{x} > 0$

$\qquad\qquad\qquad\; = 0$ if $\bar{x} < 0$

$\Rightarrow$ MLE of $\theta = \text{Max}\{\bar{x}, 0\}$.

Example: MLE, if exists, is a function of a sufficient statistic.

back ①

think about the factorization theorem.

$X_1, \ldots, X_n \overset{iid}{\sim} \theta$ $(X_1, \ldots, X_n) \sim f_\theta(\underset{\sim}{x})$

then by factorization theorem

$f_\theta(\underset{\sim}{x}) = g(T(\underset{\sim}{x}), \theta)\, h(\underset{\sim}{x})$

MLE is obtained by solving

$\frac{\partial}{\partial\theta} \log f_\theta(\underset{\sim}{x}) = 0 \iff \frac{\partial}{\partial\theta} \log g(T(\underset{\sim}{x}), \theta) = 0$

$\Rightarrow$ the solution of θ should be a function of $T(\underset{\sim}{x})$. $\Rightarrow$ MLE is a function of a sufficient statistic.

<u>Example:</u> $X_1, \ldots, X_n \overset{iid}{\sim} U[\theta_1, \theta_2]$

$f_{\theta_1,\theta_1}(x_1, \ldots, x_n) = \frac{1}{(\theta_2-\theta_1)^n} I(\theta_1 \le x_1 \le \theta_2, \ldots, \theta_1 \le x_n \le \theta_2)$

$= \frac{1}{(\theta_2-\theta_1)^n} I(x_{(n)} \le \theta_2, \ \theta_1 \le x_{(1)})$

if you choose $\theta_1 = x_{(1)}$ and $\theta_2 = x_{(n)}$ then the interval (θ_1, θ_2) can be made the smallest even after maintaining $f_{\theta_1,\theta_2}(x_1, \ldots, x_n) \ge 0$.

<u>Example:</u> $X_1, \ldots, X_n \overset{iid}{\sim} U(\theta-1, \theta+1)$

$f_\theta(x_1, \ldots, x_n) = \frac{1}{2^n} I(\theta-1 < x_1 < \theta+1, \ldots, \theta-1 < x_n < \theta+1)$

$= \frac{1}{2^n} I(\theta-1 < x_{(1)}, \ x_{(n)} < \theta+1)$

$= \frac{1}{2^n} I(x_{(n)} - 1 < \theta < x_{(1)}+1)$

back ②

the likelihood function is constant $\frac{1}{2^n}$ for any $\theta \in (X_{(n)}-1, X_{(1)}+1)$

Any point in this interval can be an MLE.

$\Rightarrow$ MLE is not unique.

Invariance property: If $\hat{\theta}$ is the MLE of θ, then for any function $\tau(\theta)$, the MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$.

"MLE is asymptotically most efficient". What do we mean by the above statement?

Asymptotic distribution of MLE:

When all regularity conditions are satisfied, MLE $\hat{\theta}$ of θ has the following asymptotic distribution,

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\right)$$

Let $g(\theta)$ be any function of θ. By delta theorem

$$\sqrt{n}(g(\hat{\theta})-g(\theta)) \xrightarrow{d} N\left(0, \frac{(g'(\theta))^2}{I(\theta)}\right)$$

When n is very very large.

$$\text{var}(\sqrt{n}\, g(\hat{\theta})) \simeq \frac{(g'(\theta))^2}{I(\theta)} \Rightarrow \text{Var}(g(\hat{\theta})) \simeq \frac{(g'(\theta))^2}{n I(\theta)}$$

and $E[\sqrt{n}(g(\hat{\theta})-g(\theta))] \simeq 0 \Rightarrow E[g(\hat{\theta})] \simeq g(\theta)$.

back ③

When n is very very large MLE is unbiased and variance of the MLE is approximately equal to the Cramer-Rao lower bound. In this sense, MLE is asymptotically (i.e. when $n \to \infty$) UMVUE.

<u>Example</u>: When the regularity conditions do not hold, the asymptotic distribution of MLE can be very different from Normal.

$$X_1, \dots, X_n \overset{iid}{\sim} U[0, \theta],$$

$$f_\theta(x_1, \dots, x_n) = \frac{1}{\theta^n} I(0 \le x_1 \le \theta, \dots, 0 \le x_n \le \theta)$$

$$= \frac{1}{\theta^n} I(x_{(n)} \le \theta,\ x_{(1)} \ge 0)$$

$\theta = X_{(n)}$ is the MLE.

What is the asymptotic distribution of $X_{(n)}$.

$$n(-X_{(n)} + \theta) \xrightarrow{d} \mathcal{E}xp(1).$$

Here asymptotic dist. is different from Normal as the regularity conditions do not hold.