Cramer–Rao lower bound was discussed for univariate parameters.

\[ \sqrt{n} \left( \frac{Z_1 Y_1 - \ldots - Z_n Y_n}{Z_1 + \ldots + Z_n} \right) \overset{N(0,1)}{\longrightarrow} \text{N}(0,1) \]

\[ x_1, \ldots, x_n \overset{iid}{\sim} f \left( \frac{x - \mu}{\sigma} \right) \]

\[ x_1, \ldots, x_n \overset{iid}{\sim} \text{Pois}(\lambda), \quad \text{Var}(X) > 0 \]

**Multiparameter case:**

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

Score vector is defined as

\[ u_{\theta}(x) = \left( \frac{\partial}{\partial \theta_1} \log f_{\theta}(x), \ldots, \frac{\partial}{\partial \theta_k} \log f_{\theta}(x) \right) \]

Also we define the information matrix

\[ I(\theta) = (C I_{ij}(\theta))_{i,j=1}^k, \text{ where} \]

\[ I_{ij}(\theta) = E \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f_{\theta}(x) \right] \]

Multiparameter Cramer–Rao lower bound

Suppose \( I(\theta) \) is positive definite and

\[ Q_i = \frac{2}{\theta_i} E_{\theta_i}(\delta(x)) \]
exists and differentiation w.r.t. \( \theta \) can be done under integration w.r.t. \( x \). Then
\[
\text{Var}_\theta(\delta(x)) \geq \alpha^\prime \Gamma(\theta)^{-1} \alpha
\]
where \( \alpha = (\alpha_1, \ldots, \alpha_K) \)

be \( \text{Var}_\theta \), 12-version with one data point
\[
\text{Var}_\theta(\delta(x)) \geq \left( \frac{\partial}{\partial \theta} E_\theta(\delta(x)) \right)^T \Gamma(\theta)^{-1} \left( \frac{\partial}{\partial \theta} E_\theta(\delta(x)) \right)
\]

Method of finding estimators

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

1. Method of moments estimators

Let \[ m_j = \frac{1}{n} \sum_{i=1}^n x_i^j, \quad j = 1, 2, \ldots \]

Let \( m_j' = E[x^j] \). Generally, \( m_j' \) are functions of the parameters \( \theta_1, \ldots, \theta_K \). Let us write \( m_j'(\theta_1, \ldots, \theta_K) \) instead of writing \( m_j' \).

the method of moment estimator is obtained by solving the following equations
\[ m_j = m_j'(\theta_1, \ldots, \theta_K), \quad j = 1, \ldots, K. \]
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 $\mu$ and $\sigma^2$ we will solve the following two equations:

\[
\frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}, \quad \frac{1}{n} \sum_{i=1}^{n} x_i^2 = \bar{x}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{(n-1)\hat{\sigma}^2}{n}
\]

$\Rightarrow \mu = \bar{x}$, $\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{(n-1)\hat{\sigma}^2}{n}$

In this case MOM of $\mu$ in UMVUE, but MOM of $\sigma^2$ is not.

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

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

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

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

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

$E[X] = kp$, $E[X^2] = kp(1-p) + (kp)^2$
MOM of \( k \) and \( p \) satisfy
\[
\frac{1}{n} \sum_{i=1}^{n} x_i = kp \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^{n} x_i^2 = kp (1-p) + (kp)^2
\]
\[
\therefore \quad p = \frac{\bar{x}}{k} \quad \Rightarrow \quad \frac{1}{n} \sum_{i=1}^{n} x_i^2 = k \left[ \frac{\bar{x}}{k} \right] (1-\frac{\bar{x}}{k}) + \bar{x}^2
\]
\[
\Rightarrow \quad k = \frac{\bar{x}}{\bar{x} - \left( \frac{1}{n} \right) \sum_{i=1}^{n} (x_i - \bar{x})^2}
\]

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

this MOM does not make any sense.

Example: \( x_1, \ldots, x_n \sim \mathcal{N}(\theta, \theta^2) \)
\[
\frac{1}{n} \sum_{i=1}^{n} x_i = \theta \quad \text{can be used to find the MOM.}
\]

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 \( \theta \) of \( x = (x_1, \ldots, x_n) \) is given by
\[
L(\theta_1, \ldots, \theta_k) = \prod_{i=1}^{n} f_{\theta_1, \ldots, \theta_k}(x_i) \quad \text{MLE of } \theta, \text{ denoted by } \hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k), \text{ is the value of } \theta \text{ that}
\]
maximizes the likelihood \( L(\theta_1, \ldots, \theta_k) \).
Maximum likelihood estimator is generally calculated by solving the equations

\[ \frac{\partial}{\partial \theta_i} \log L(\theta) = 0 \quad i = 1, \ldots, K. \]

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

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

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

\[ \sum_{i=1}^{n} u_{\theta_j}(x_i) = 0 \]

where \( u_{\theta_j}(x_i) \) is the score function evaluated at \( x_i \).

Also the Maximum Likelihood Estimator (MLE) should satisfy \( \frac{\partial^2 L(\theta)}{\partial \theta^2} \bigg|_{\theta = \theta} \) in a negative definite.

Example: \( x_1, \ldots, x_n \overset{iid}{\sim} N(\theta, 1), \theta > 0 \). Find the MLE of \( \theta \).

If \( \bar{x} > 0 \) then \( \hat{\theta} = \bar{x} \) in the MLE.

If \( \bar{x} < 0 \). 

6
\[ L(\theta) = \left( \frac{1}{\sqrt{2\pi}} \right)^n \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n} (x_i - \bar{x} + \bar{x} - \theta)^2 \right\} \]

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

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

\[ \bar{x} = \frac{\sum x_i}{n} \]

\[ L > 0, \text{ also } \theta > 0 \]

\[ \exp \left\{ -\frac{n(\bar{x} - \theta)^2}{2} \right\} \text{ decreasing as a function of } \theta \text{ for } \theta > 0. \text{ Hence maximum is achieved at } \theta = 0. \]

Therefore, MLE of \( \theta \) = \( \bar{x} \) if \( \bar{x} > 0 \)

\[ = 0 \text{ if } \bar{x} < 0 \]

\[ \Rightarrow \text{ MLE of } \theta = \text{ Max} \{ \bar{x}, 0 \} \]

Example: MLE, if exists, is a function of a sufficient statistic.
think about the factorization theorem

\[ f_\theta(x) = g(T(x), \theta) \cdot h(x) \]

then by factorization theorem

\[ f_\theta(x) = g(T(x), \theta) \cdot h(x) \]

MLE is obtained by solving

\[ \frac{\partial}{\partial \theta} \log f_\theta(x) = 0 \iff \frac{\partial}{\partial \theta} \log g(T(x), \theta) = 0 \]

\[ \Rightarrow \text{the solution of } \theta \text{ should be a function of } T(x). \Rightarrow \text{MLE is a function of } \theta \text{ a sufficient statistic.} \]

Example: \( X_1, \ldots, X_n \overset{iid}{\sim} U[\theta_1, \theta_2] \)

\[ f_{\theta_1, \theta_1}(x_1, \ldots, x_n) = \frac{1}{(\theta_2 - \theta_1)^n} I(\theta_1 \leq x_1 \leq \theta_2, \ldots, \theta_1 \leq x_n \leq \theta_2) \]

\[ = \frac{1}{(\theta_2 - \theta_1)^n} I(x_{(n)} \leq \theta_2, \theta_1 \leq x_{(1)}) \]

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

Example: \( X_1, \ldots, X_n \overset{iid}{\sim} U(\theta-1, \theta+1) \)

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

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

\[ = \frac{1}{2^n} I(\theta_{(n)} - 1 < \theta < \theta_{(1)} + 1) \]

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

Any point in this interval can be an MLE.

\( \Rightarrow \) MLE is not unique.

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

"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 \( \theta \) has the following asymptotic distribution,

\[
\sqrt{n} (\hat{\theta} - \theta) \xrightarrow{d} N(0, \frac{1}{I(\theta)})
\]

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

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

When \( n \) is very very large,

\[
\text{Var}(\sqrt{n} g(\hat{\theta})) \approx \left(\frac{g'(\theta)}{I(\theta)}\right)^2 \Rightarrow \text{Var}(g(\hat{\theta})) \approx \left(\frac{g'(\theta)}{n I(\theta)}\right)^2
\]

and \( E[\sqrt{n} (g(\hat{\theta}) - g(\theta))] \approx 0 \Rightarrow E[g(\hat{\theta})] \approx g(\theta) \).
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 in
asymptotically (i.e. when $n \to \infty$) UMVUE.

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

$x_1, \ldots, x_n \simiid U[0, \theta],$

$f_\theta(x_1, \ldots, x_n) = \frac{1}{\theta^n} \Theta I(0 \leq x_1 \leq \theta, \ldots, 0 \leq x_n \leq \theta)
\quad = \frac{1}{\theta^n} I(x_{(n)} \leq \theta, x_{(1)} > 0)$

$\theta = x_{(n)}$ in the MLE.

What is the asymptotic distribution of $x_{(n)}$?

$n (x_{(n)} + \theta) \xrightarrow{d} \text{Exp}(1)$.

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