1 On Cramer-Rao’s bound and MLE

Consider the likelihood and log-likelihood functions:

\[ L(\theta) = \prod_{i=1}^{n} f_{\theta}(X_i) \quad l(\theta) = \sum_{i=1}^{n} \ln f_{\theta}(X_i) \]

Since \( X_1, \ldots, X_n \) are i.i.d., this is also true for \( Y_1 = \frac{\partial}{\partial \theta} \ln f_{\theta}(X_1), \ldots, Y_n = \frac{\partial}{\partial \theta} \ln f_{\theta}(X_n) \).

The log-likelihood takes its maximum at the zero’s of its derivative, which is called the score function:

\[ S(\theta) = \frac{\partial}{\partial \theta} l(\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \ln f_{\theta}(X_i) = \sum_{i=1}^{n} Y_i \]

The expectation of each \( Y_i \)’s is zero:

\[ E[Y_i] = \int \left( \frac{\partial}{\partial \theta} \ln f_{\theta}(x) \right) f_{\theta}(x) dx = \int \frac{1}{f_{\theta}(x)} \left( \frac{\partial}{\partial \theta} f_{\theta}(x) \right) f_{\theta}(x) dx \]

\[ = \int \frac{\partial}{\partial \theta} f_{\theta}(x) dx = \frac{\partial}{\partial \theta} \int f_{\theta}(x) dx = \frac{\partial}{\partial \theta} 1 = 0 \]

Hence, by linearity of expectation, we have:

\[ E[S(\theta)] = \sum_{i=1}^{n} E[Y_i] = 0 \]

The variance of \( S(\theta) \) is called the Fisher information, and it is the quantity:

\[ I(\theta) = E[S(\theta)^2] \]

It turns out that:

\[ I(\theta) = E[S(\theta)^2] = E[\sum_{i=1}^{n} Y_i^2] \]

\[ = E\left[ \sum_{i=1}^{n} \left( \frac{\partial}{\partial \theta} \ln f_{\theta}(X_i) \right)^2 \right] \]

\[ = n E\left[ \left( \frac{\partial}{\partial \theta} \ln f_{\theta}(X) \right)^2 \right] \]

where \( X \sim f_{\theta} \). **Attention**: some textbooks define \( I(\theta) \) using a single random variable, i.e., as \( E\left[ \left( \frac{\partial}{\partial \theta} \ln f_{\theta}(X) \right)^2 \right] \). In such cases, it must be multiplied by \( n \).

---

1. (1) follows since \( E[Y_i Y_j] = E[Y_i] E[Y_j] \) for independent \( Y_i, Y_j \).
2. (2) follows since \( E[Y_i] = 0 \).
We can now link Fisher information to the Cramér-Rao inequality from [1, Remark 20.2]:

\[
\text{Var}(T) \geq \frac{1}{nE[(\frac{\partial}{\partial \theta} \ln f_\theta(X))^2]} \quad \text{for all } \theta,
\]

by observing that, due to (3), the right-hand side is the inverse of \( I(\theta) \), i.e.:

\[
\text{Var}(T) \geq \frac{1}{nE[(\frac{\partial}{\partial \theta} \ln f_\theta(X))^2]} = \frac{1}{I(\theta)} \quad \text{for all } \theta.
\]

2 Example

The textbook [1] pages 324-325 shows that the MLE estimator of the mean \( \mu \) of a normal distribution \( N(\mu, \sigma^2) \) is \( \bar{X}_n = (X_1 + \ldots + X_n)/n \). Let \( X \sim \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} \).

The Fisher information is:

\[
I(\theta) = nE[\left( \frac{\partial}{\partial \mu} \ln f_\mu(X) \right)^2] \\
= nE[\left( \frac{X-\mu}{\sigma^2} \right)^2] \\
= \frac{n}{\sigma^4} E[(X-\mu)^2] \\
= \frac{n}{\sigma^4} \text{Var}(X) = \frac{n}{\sigma^2} \sigma^2 = \frac{n}{\sigma^2} = \frac{1}{\text{Var}(\bar{X}_n)}
\]

where the last equality follows from the Central Limit Theorem. By taking the reciprocals:

\[
\text{Var}(\bar{X}_n) = \frac{1}{I(\theta)}
\]

we have that the lower bound of the Cramér-Rao inequality is reached, hence \( \bar{X}_n \) is a MVUE (Minimum Variance Unbiased Estimator).

References