1 On Cramér-Rao's bound and MLE

Consider the likelihood and log-likelihood functions:

$$L(\theta) = \prod_{i=1}^{n} f_\theta(X_i) \qquad \ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f_\theta(X_i)$$

Since $X_1, \ldots, X_n$ are i.i.d., this is also true for $Y_1 = \frac{\partial}{\partial\theta} \log f_\theta(X_1)$, …, $Y_n = \frac{\partial}{\partial\theta} \log f_\theta(X_n)$. The log-likelihood takes its maximum at the zero's of its derivative, which is called the *score function*:

$$S(\theta) = \frac{\partial}{\partial\theta}\ell(\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f_\theta(X_i) = \sum_{i=1}^{n} Y_i$$

The expectation of each Y_i's is zero (use Leibniz integral rule):

$$\begin{aligned} \mathrm{E}\big[Y_i\big] &= \int (\frac{\partial}{\partial\theta} \log f_\theta(x)) f_\theta(x) dx = \int \frac{1}{f_\theta(x)} (\frac{\partial}{\partial\theta} f_\theta(x)) f_\theta(x) dx \\ &= \int \frac{\partial}{\partial\theta} f_\theta(x) dx = \frac{\partial}{\partial\theta} \int f_\theta(x) dx = \frac{\partial}{\partial\theta} 1 = 0 \end{aligned}$$

Hence, by linearity of expectation, we have:

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

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

$$I(\theta) = \mathrm{E}\big[S(\theta)^2\big]$$

It turns out[1][2] that:

$$\begin{aligned} I(\theta) = \mathrm{E}\big[S(\theta)^2\big] &= \mathrm{E}\big[(\sum_{i=1}^{n} Y_i)(\sum_{j=1}^{n} Y_j)\big] \\ &= \mathrm{E}\big[\sum_{i=1}^{n} Y_i^2 + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} Y_i Y_j\big] \\ &= \mathrm{E}\big[\sum_{i=1}^{n} Y_i^2\big] + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathrm{E}\big[Y_i\big]\mathrm{E}\big[Y_j\big] \qquad (1) \\ &= \mathrm{E}\big[\sum_{i=1}^{n} Y_i^2\big] + 0 \qquad (2) \\ &= \mathrm{E}\big[\sum_{i=1}^{n} (\frac{\partial}{\partial\theta} \log f_\theta(X_i))^2\big] \\ &= n\mathrm{E}\big[(\frac{\partial}{\partial\theta} \log f_\theta(X))^2\big] \qquad (3) \end{aligned}$$

where $X \sim f_\theta$. **Important**: some textbooks define $I(\theta)$ using a single random variable, i.e., as $\mathrm{E}\big[(\frac{\partial}{\partial\theta} \log f_\theta(X))^2\big]$. In such cases, it must be multiplied by n whenever it is used.

[1] (1) follows since $\mathrm{E}\Big[Y_i Y_j\Big] = \mathrm{E}\Big[Y_i\Big]\mathrm{E}\Big[Y_j\Big]$ for independent Y_i, Y_j.

[2] (2) follows since $\mathrm{E}\Big[Y_i\Big] = 0$.

We can now link Fisher information to the Cramér-Rao inequality from [1, Remark 20.2]:

$$\mathrm{Var}(T) \geq \frac{1}{n\mathrm{E}\big[(\frac{\partial}{\partial\theta}\log f_\theta(X))^2\big]} \qquad \text{for all } \theta,$$

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

$$\mathrm{Var}(T) \geq \frac{1}{n\mathrm{E}\big[(\frac{\partial}{\partial\theta}\log f_\theta(X))^2\big]} = \frac{1}{I(\theta)} \qquad \text{for all } \theta.$$

2 Example

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

$$\begin{aligned} I(\theta) &= n\mathrm{E}\big[(\frac{\partial}{\partial\mu}\log f_\mu(X))^2\big] \\ &= n\mathrm{E}\big[(\frac{X-\mu}{\sigma^2})^2\big] \\ &= \frac{n}{\sigma^4}\mathrm{E}\big[(X-\mu)^2\big] \\ &= \frac{n}{\sigma^4}\mathrm{Var}(X) = \frac{n}{\sigma^4}\sigma^2 = \frac{n}{\sigma^2} = \frac{1}{\mathrm{Var}(\bar{X}_n)} \end{aligned}$$

where the last equality follows because for i.i.d. random variables $\mathrm{Var}(\bar{X}_n) = \sigma^2/n$. By taking the reciprocals:

$$\mathrm{Var}(\bar{X}_n) = \frac{1}{I(\theta)}$$

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

[1] F.M. Dekking, C. Kraaikamp, H.P. Lopuhaä, and L.E. Meester. *A Modern Introduction to Probability and Statistics*. Springer, 2005.