

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

$$\begin{aligned} \mathbb{E}[Y_i] &= \int \left(\frac{\partial}{\partial \theta} \ln f_{\theta}(x) \right) f_{\theta}(x) dx = \int \frac{1}{f_{\theta}(x)} \left(\frac{\partial}{\partial \theta} f_{\theta}(x) \right) f_{\theta}(x) dx \\ &= \int \frac{\partial}{\partial \theta} f_{\theta}(x) dx = \frac{\partial}{\partial \theta} \int f_{\theta}(x) dx = \frac{\partial}{\partial \theta} 1 = 0 \end{aligned}$$

Hence, by linearity of expectation, we have:

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

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

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

It turns out¹² that:

$$\begin{aligned} I(\theta) = \mathbb{E}[S(\theta)^2] &= \mathbb{E}\left[\left(\sum_{i=1}^n Y_i\right)\left(\sum_{j=1}^n Y_j\right)\right] \\ &= \mathbb{E}\left[\sum_{i=1}^n Y_i^2 + \sum_{i=1}^n \sum_{j=1, j \neq i}^n Y_i Y_j\right] \\ &= \mathbb{E}\left[\sum_{i=1}^n Y_i^2\right] + \sum_{i=1}^n \sum_{j=1, j \neq i}^n \mathbb{E}[Y_i] \mathbb{E}[Y_j] \end{aligned} \quad (1)$$

$$= \mathbb{E}\left[\sum_{i=1}^n Y_i^2\right] + 0 \quad (2)$$

$$\begin{aligned} &= \mathbb{E}\left[\sum_{i=1}^n \left(\frac{\partial}{\partial \theta} \ln f_{\theta}(X_i)\right)^2\right] \\ &= n \mathbb{E}\left[\left(\frac{\partial}{\partial \theta} \ln f_{\theta}(X)\right)^2\right] \end{aligned} \quad (3)$$

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

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

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

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

$$\text{Var}(T) \geq \frac{1}{n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f_{\theta}(X)\right)^2\right]} \quad \text{for all } \theta,$$

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

$$\text{Var}(T) \geq \frac{1}{n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f_{\theta}(X)\right)^2\right]} = \frac{1}{I(\theta)} \quad \text{for all } \theta.$$

2 Example

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

The Fisher information is:

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

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

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