Lecture 5

Let us give one more example of MLE.

Example 3. The uniform distribution $U[0,\theta]$ on the interval $[0,\theta]$ has p.d.f.

$$f(x|\theta) = \begin{cases} \frac{1}{\theta}, & 0 \le x \le \theta, \\ 0, & \text{otherwise} \end{cases}$$

The likelihood function

$$\begin{aligned} \varphi(\theta) = \prod_{i=1}^{n} f(X_i|\theta) &= \frac{1}{\theta^n} I(X_1, \ldots, X_n \in [0,\theta]) \\ &= \frac{1}{\theta^n} I(\max(X_1, \ldots, X_n) \le \theta). \end{aligned}$$

Here the indicator function $I(A)$ equals to 1 if A happens and 0 otherwise. What we wrote is that the product of p.d.f. $f(X_i|\theta)$ will be equal to 0 if at least one of the factors is 0 and this will happen if at least one of X_is will fall outside of the interval $[0,\theta]$ which is the same as the maximum among them exceeds θ. In other words,

$$\varphi(\theta) = 0 \text{ if } \theta < \max(X_1, \ldots, X_n),$$

and

$$\varphi(\theta) = \frac{1}{\theta^n} \text{ if } \theta \ge \max(X_1, \ldots, X_n).$$

Therefore, looking at the figure 5.1 we see that $\hat{\theta} = \max(X_1, \ldots, X_n)$ is the MLE.

5.1 Consistency of MLE.

Why the MLE $\hat{\theta}$ converges to the unkown parameter θ_0? This is not immediately obvious and in this section we will give a sketch of why this happens.

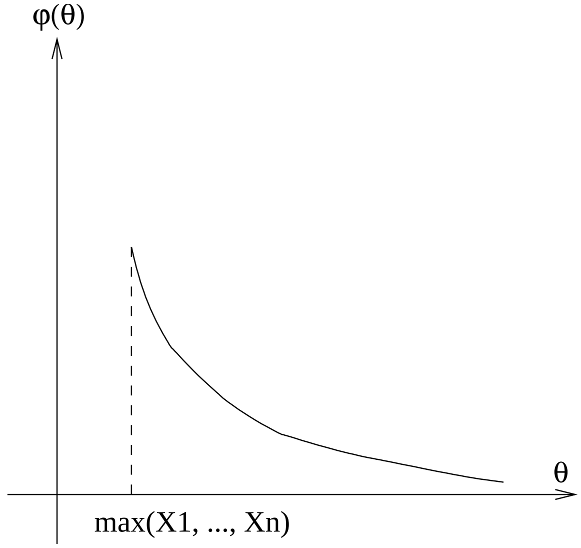

Figure 5.1: Maximize over θ

First of all, MLE $\hat{\theta}$ is a maximizer of

$$L_n\theta = \frac{1}{n}\sum_{i=1}^{n}\log f(X_i|\theta)$$

which is just a log-likelihood function normalized by $\frac{1}{n}$ (of course, this does not affect the maximization). $L_n(\theta)$ depends on data. Let us consider a function $l(X|\theta) = \log f(X|\theta)$ and define

$$L(\theta) = \mathbb{E}_{\theta_0} l(X|\theta),$$

where we recall that θ_0 is the true uknown parameter of the sample $X_1, \ldots, X_n$. By the law of large numbers, for any θ,

$$L_n(\theta) \to \mathbb{E}_{\theta_0} l(X|\theta) = L(\theta).$$

Note that $L(\theta)$ does not depend on the sample, it only depends on θ. We will need the following

Lemma. *We have, for any* θ,

$$L(\theta) \leq L(\theta_0).$$

Moreover, the inequality is strict $L(\theta) < L(\theta_0)$ *unless*

$$\mathbb{P}_{\theta_0}(f(X|\theta) = f(X|\theta_0)) = 1.$$

which means that $\mathbb{P}_\theta = \mathbb{P}_{\theta_0}$.

Proof. Let us consider the difference

$$L(\theta) - L(\theta_0) = \mathbb{E}_{\theta_0}(\log f(X|\theta) - \log f(X|\theta_0)) = \mathbb{E}_{\theta_0} \log \frac{f(X|\theta)}{f(X|\theta_0)}.$$

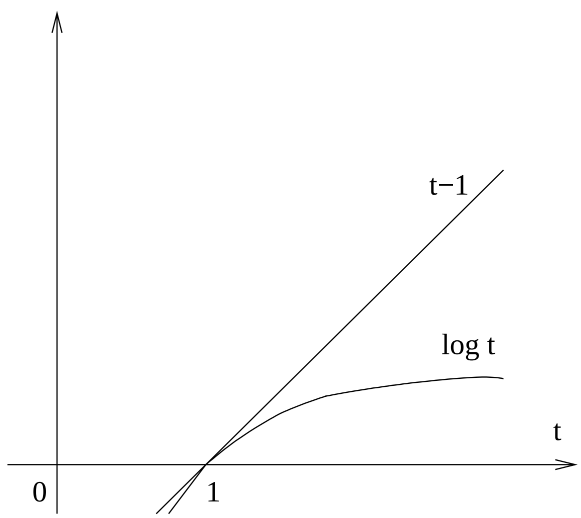

Figure 5.2: Diagram $(t-1)$ vs. $\log t$

Since $(t-1)$ is an upper bound on $\log t$ (see figure 5.2) we can write

$$\begin{aligned}\mathbb{E}_{\theta_0} \log \frac{f(X|\theta)}{f(X|\theta_0)} &\leq \mathbb{E}_{\theta_0}\Big(\frac{f(X|\theta)}{f(X|\theta_0)} - 1\Big) = \int \Big(\frac{f(x|\theta)}{f(x|\theta_0)} - 1\Big) f(x|\theta_0) dx \\ &= \int f(x|\theta)dx - \int f(x|\theta_0)dx = 1 - 1 = 0.\end{aligned}$$

Both integrals are equal to 1 because we are integrating the probability density functions. This proves that $L(\theta) - L(\theta_0) \leq 0$. The second statement of Lemma is also clear.

□

We will use this Lemma to sketch the consistency of the MLE.

Theorem: *Under some regularity conditions on the family of distributions, MLE $\hat{\theta}$ is consistent, i.e. $\hat{\theta} \to \theta_0$ as $n \to \infty$.*

The statement of this Theorem is not very precise but but rather than proving a rigorous mathematical statement our goal here to illustrate the main idea. Mathematically inclined students are welcome to come up with some precise statement.

Proof.

We have the following facts:
1. $\hat{\theta}$ is the maximizer of $L_n(\theta)$ (by definition).
2. θ_0 is the maximizer of $L(\theta)$ (by Lemma).
3. $\forall\theta$ we have $L_n(\theta) \to L(\theta)$ by LLN.

This situation is illustrated in figure 5.3. Therefore, since two functions L_n and L are getting closer, the points of maximum should also get closer which exactly means that $\hat{\theta} \to \theta_0$.

□

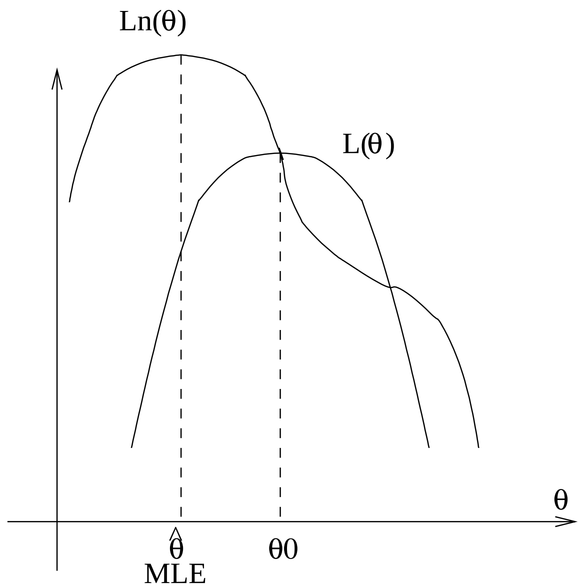

Figure 5.3: Lemma: $L(\theta) \le L(\theta_0)$

5.2 Asymptotic normality of MLE. Fisher information.

We want to show the asymptotic normality of MLE, i.e. that

$$\sqrt{n}(\hat{\theta} - \theta_0) \to^d N(0, \sigma^2_{MLE}) \text{ for some } \sigma^2_{MLE}.$$

Let us recall that above we defined the function $l(X|\theta) = \log f(X|\theta)$. To simplify the notations we will denote by $l'(X|\theta), l''(X|\theta)$, etc. the derivatives of $l(X|\theta)$ with respect to θ.

Definition. (Fisher information.) Fisher Information of a random variable X with distribution $\mathbb{P}_{\theta_0}$ from the family $\{\mathbb{P}_\theta : \theta \in \Theta\}$ is defined by

$$I(\theta_0) = \mathbb{E}_{\theta_0}(l'(X|\theta_0))^2 \equiv \mathbb{E}_{\theta_0}\Big(\frac{\partial}{\partial\theta}\log f(X|\theta_0)\Big)^2.$$

Next lemma gives another often convenient way to compute Fisher information.

Lemma. *We have,*

$$\mathbb{E}_{\theta_0} l''(X|\theta_0) \equiv \mathbb{E}_{\theta_0}\frac{\partial^2}{\partial\theta^2}\log f(X|\theta_0) = -I(\theta_0).$$

Proof. First of all, we have

$$l'(X|\theta) = (\log f(X|\theta))' = \frac{f'(X|\theta)}{f(X|\theta)}$$

and

$$(\log f(X|\theta))'' = \frac{f''(X|\theta)}{f(X|\theta)} - \frac{(f'(X|\theta))^2}{f^2(X|\theta)}.$$

Also, since p.d.f. integrates to 1,

$$\int f(x|\theta)dx = 1,$$

if we take derivatives of this equation with respect to θ (and interchange derivative and integral, which can usually be done) we will get,

$$\int \frac{\partial}{\partial\theta} f(x|\theta)dx = 0 \text{ and } \int \frac{\partial^2}{\partial\theta^2} f(x|\theta)dx = \int f''(x|\theta)dx = 0.$$

To finish the proof we write the following computation

$$\begin{aligned}
\mathbb{E}_{\theta_0} l''(X|\theta_0) &= \mathbb{E}_{\theta_0}\frac{\partial^2}{\partial\theta^2}\log f(X|\theta_0) = \int (\log f(x|\theta_0))'' f(x|\theta_0)dx \\
&= \int \Big(\frac{f''(x|\theta_0)}{f(x|\theta_0)} - \Big(\frac{f'(x|\theta_0)}{f(x|\theta_0)}\Big)^2\Big) f(x|\theta_0)dx \\
&= \int f''(x|\theta_0)dx - \mathbb{E}_{\theta_0}(l'(X|\theta_0))^2 = 0 - I(\theta_0 = -I(\theta_0).
\end{aligned}$$

□

We are now ready to prove the main result of this section.

Theorem. (Asymptotic normality of MLE.) *We have,*

$$\sqrt{n}(\hat{\theta} - \theta_0) \to N\Big(0, \frac{1}{I(\theta_0)}\Big).$$

Proof. Since MLE $\hat{\theta}$ is maximizer of $L_n(\theta) = \frac{1}{n}\sum_{i=1}^n \log f(X_i|\theta)$ we have,

$$L_n'(\hat{\theta}) = 0.$$

Let us use the Mean Value Theorem

$$\frac{f(a) - f(b)}{a - b} = f'(c) \text{ or } f(a) = f(b) + f'(c)(a - b) \text{ for } c \in [a, b]$$

with $f(\theta) = L_n'(\theta), a = \hat{\theta}$ and $b = \theta_0$. Then we can write,

$$0 = L_n'(\hat{\theta}) = L_n'(\theta_0) + L_n''(\hat{\theta}_1)(\hat{\theta} - \theta_0)$$

for some $\hat{\theta}_1 \in [\hat{\theta}, \theta_0]$. From here we get that

$$\hat{\theta} - \theta_0 = -\frac{L_n'(\theta_0)}{L_n''(\hat{\theta}_1)} \text{ and } \sqrt{n}(\hat{\theta} - \theta_0) = -\frac{\sqrt{n}L_n'(\theta_0)}{L_n''(\hat{\theta}_1)}. \tag{5.1}$$

Since by Lemma in the previous section θ_0 is the maximizer of $L(\theta)$, we have

$$L'(\theta_0) = \mathbb{E}_{\theta_0} l'(X|\theta_0) = 0. \tag{5.2}$$

Therefore, the numerator in (5.1)

$$\begin{aligned} \sqrt{n}L_n'(\theta_0) &= \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n l'(X_i|\theta_0) - 0\Big) \qquad (5.3) \\ &= \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n l'(X_i|\theta_0) - \mathbb{E}_{\theta_0} l'(X_1|\theta_0)\Big) \to N\Big(0, \mathrm{Var}_{\theta_0}(l'(X_1|\theta_0))\Big) \end{aligned}$$

converges in distribution by Central Limit Theorem.

Next, let us consider the denominator in (5.1). First of all, we have that for all θ,

$$L_n''(\theta) = \frac{1}{n}\sum l''(X_i|\theta) \to \mathbb{E}_{\theta_0} l''(X_1|\theta) \text{ by LLN.} \tag{5.4}$$

Also, since $\hat{\theta}_1 \in [\hat{\theta}, \theta_0]$ and by consistency result of previous section $\hat{\theta} \to \theta_0$, we have $\hat{\theta}_1 \to \theta_0$. Using this together with (5.4) we get

$$L_n''(\hat{\theta}_1) \to \mathbb{E}_{\theta_0} l''(X_1|\theta_0) = -I(\theta_0) \text{ by Lemma above.}$$

Combining this with (5.3) we get

$$-\frac{\sqrt{n}L_n'(\theta_0)}{L_n''(\hat{\theta}_1)} \to N\Big(0, \frac{\mathrm{Var}_{\theta_0}(l'(X_1|\theta_0))}{(I(\theta_0))^2}\Big).$$

Finally, the variance,

$$\mathrm{Var}_{\theta_0}(l'(X_1|\theta_0)) = \mathbb{E}_{\theta_0}(l'(X|\theta_0))^2 - (\mathbb{E}_{\theta_0} l'(x|\theta_0))^2 = I(\theta_0) - 0$$

where in the last equality we used the definition of Fisher information and (5.2).

□