

STATS300B – Lecture 7

Julia Palacios
Scribed by Michael Howes

01/25/22

Contents

1	Motivation	1
2	Random functions	1
3	Kullback–Leibler information	2
4	Asymptotic normality	3

1 Motivation

1. We want to prove consistency of the MLE.
2. The MLE maximizes the log-likelihood which is an empirical average. In particular, the log-likelihood is a random function.
3. We will thus use the weak law for random functions.

2 Random functions

Recall the following,

Theorem 1. Let K be a compact set and suppose X_1, X_2, \dots are i.i.d. and $W_i(t) = h(t, X_i) \in C(K)$ for all values of X_i . Let $\mu(t) = \mathbb{E}[W_1(t)]$ and assume that $\mathbb{E}[\|W\|_\infty] < \infty$, then $\|\bar{W}_n - \mu\|_\infty \xrightarrow{P} 0$.

We will also use the following theorem about the optimizers of random functions.

Theorem 2. Let G_n be random functions in $C(K)$ where K is compact. Let $g \in C(K)$ be a deterministic function. Suppose that $\|G_n - g\|_\infty \xrightarrow{P} 0$, then

1. If $t_n \xrightarrow{P} t^*$, then $G_n(t_n) \xrightarrow{P} g(t^*)$.
2. Let t_n be a random variable maximizing G_n . If g achieves its maximum at a unique value t^* , then $t_n \xrightarrow{P} t^*$.
3. Suppose that $K \subseteq \mathbb{R}$ and t_n is a random variable solving $G_n(t_n) = 0$. If t^* is the unique value in K such that $g(t^*) = 0$, then $t_n \xrightarrow{P} t^*$.

3 Kullback–Leibler information

Definition 1. Let P and Q be probability measures with densities p and q with respect to a common σ -finite measure μ . The *Kullback–Leibler information* of P and Q is defined to be,

$$K(P, Q) = \mathbb{E}_P \left[\log \left(\frac{p(X)}{q(X)} \right) \right].$$

Proposition 1. For all probability distributions, $K(P, Q) \geq 0$ and $K(P, Q) = 0$ if and only if $P = Q$.

Proof. Recall that $-\log$ is convex. Thus, by Jensen's inequality,

$$\begin{aligned} \mathbb{E}_P \left[\log \left(\frac{p(X)}{q(X)} \right) \right] &= \mathbb{E}_P \left[-\log \left(\frac{q(X)}{p(X)} \right) \right] \\ &\geq -\log \left(\mathbb{E}_P \left[\frac{q(X)}{p(X)} \right] \right) \\ &= -\log \left(\int_{p(x) \neq 0} q(x) dx \right) \\ &\geq -\log(1) \\ &= 0. \end{aligned}$$

Furthermore, we have equality when $\frac{q(X)}{p(X)}$ is constant P -almost everywhere. But since p and q are densities this happens only when $P = Q$. \square

Kullback–Leibler distant relates to the log-likelihood in the following way,

Lemma 1. Consider a model $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$ and consider the following assumptions,

1. If $\theta \neq \theta^*$, then $P_\theta \neq P_{\theta^*}$.
2. There exists a measure μ such that for every θ , P_θ has a density p_θ with respect to μ .
3. The support of P_θ , $\{x : p_\theta(x) > 0\}$ does not depend on θ .

Then, if X_1, \dots, X_n are i.i.d. P_{θ^*} , then

$$\frac{1}{n} \left[\frac{L_n(\theta^*)}{L_n(\theta)} \right] \xrightarrow{a.s.} K(P_{\theta^*}, P_\theta),$$

where $L_n(\theta)$ is the log-likelihood evaluated at θ . In particular, for all $\theta \neq \theta^*$,

$$\mathbb{P}(L_n(\theta^*|X) \geq L_n(\theta|X)) \rightarrow 1.$$

Proof. The proof is a simple application of the strong law of large numbers. \square

We are now ready to state the consistency theorem for the MLE.

Theorem 3. Suppose that the three assumption of the previous lemma hold and that Ω is compact. Fix $\theta^* \in \Omega$ and define $h : \Omega \times \mathcal{X} \rightarrow \mathbb{R}$ as the function,

$$h(\theta, x) = \log \left[\frac{p_\theta(X)}{p_{\theta^*}(x)} \right].$$

Let X_0, X_1, X_2, \dots be i.i.d. samples from P_{θ^*} and define $W_i(\theta) = h(\theta, X_i)$. Suppose that h is continuous in θ for every x and,

$$\mathbb{E}_{\theta^*}[\|W_0\|_\infty] < \infty.$$

Define $\hat{\theta}_n$ to be the MLE given X_1, \dots, X_n . Then, under P_{θ^*} , $\hat{\theta}_n \xrightarrow{P} \theta^*$.

4 Asymptotic normality

We now know that the MLE is consistent, but what can we say about the limiting distribution?

Theorem 4. Let $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$ be a model for $X \in \mathcal{X}$. Suppose that $\Omega \subseteq \mathbb{R}$ and that P_θ has a density p_θ with respect to a common base measure μ . Suppose that the following hold,

1. The support of p_θ does not depend on θ .
2. For every $x \in \mathcal{X}$, $\frac{\partial^2}{\partial \theta^2} p_\theta(x)$ exists and is continuous in θ .
3. If $l(\theta) = \log p_\theta(x)$, then the Fisher information exists, is finite and can be calculated as either

$$I(\theta) = \mathbb{E}_\theta \left[\left(\frac{\partial}{\partial \theta} l(\theta) \right)^2 \right] \quad \text{or} \quad I(\theta) = -\mathbb{E}_\theta \left[\frac{\partial^2}{\partial \theta^2} l(\theta) \right],$$

and $\mathbb{E} \left[\frac{\partial}{\partial \theta} l(\theta) \right] = 0$.

4. For every θ in the interior of Ω , there exists $\varepsilon > 0$ such that $\mathbb{E}_\theta \left\| \mathbf{1}_{[\theta-\varepsilon, \theta+\varepsilon]} \frac{\partial^2}{\partial \theta^2} l(\theta) \right\| < \infty$.
5. The MLE $\hat{\theta}_n$ is consistent.

Then for any θ^* in the interior of Ω , if $X_i \stackrel{\text{iid}}{\sim} P_{\theta^*}$, then

$$\sqrt{n} \left(\hat{\theta}_n - \theta \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{1}{I(\theta)} \right).$$