Caltech Department of Mathematics

Ma 3/103 | KC Border
Introduction to Probability and Statistics | Winter 2017

Supplement 5: On the Consistency of MLE

This supplement fills in the details and addresses some of the issues addressed in Section 17.13⋆ on the consistency of Maximum Likelihood Estimators.

S5.1 General issues with maximization

The strategy of Section 17.13⋆ is to show that if θ is not the "true" θ_0, then $L(\theta; x) < L(\theta_0; x)$ with high probability. This raises some issues about maximizers that are completely independent of the probabilistic issues. We consider them here.

S5.1.1 Question *Let $g\colon \Theta \to \boldsymbol{R}$. Assume that θ^* is the unique maximizer of g over Θ. If $g(\theta_n) \to g(\theta^*)$, must it be true that $\theta_n \to \theta^*$?*

The answer to this is No. Here are a couple of examples of what can go wrong.

S5.1.2 Example Let $\Theta = [0, 1]$, and define $g\colon \Theta \to \boldsymbol{R}$ by

$$g(\theta) = \begin{cases} 0 & \theta = 0 \\ 1 - \theta & 0 < \theta < 1 \\ 1 & \theta = 1. \end{cases}$$

See Figure 17.6. Then $\theta^* = 1$ maximizes g over $\Theta = [0, 1]$, and $g(\theta^*) = 1$. Let $\theta_n = 1/n$. Then

$$g(\theta_n) = 1 - (1/n) \xrightarrow[n\to\infty]{} 1 = g(\theta^*),$$

but

$$\theta_n \to 0 \text{ and } g(0) = 0 \neq 1 = g(\theta^*).$$

The problem here is that g is not continuous. □

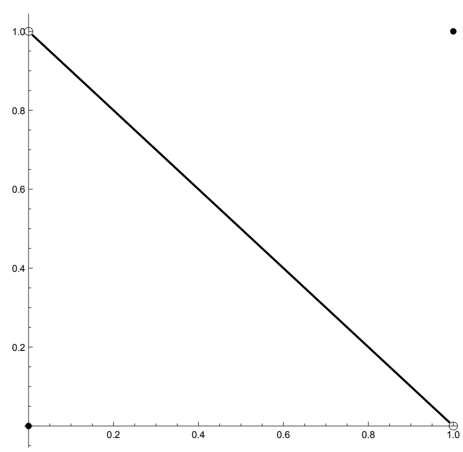

Figure 17.6. Discontinuity.

S5.1.3 Example Let $\Theta = [0, \infty)$, and define $g\colon \Theta \to \boldsymbol{R}$ by

$$g(\theta) = \begin{cases} 1 - \theta & 0 \leqslant \theta \leqslant 1 \\ 1 - (1/\theta) & \theta \geqslant 1. \end{cases}$$

See Figure 17.7.

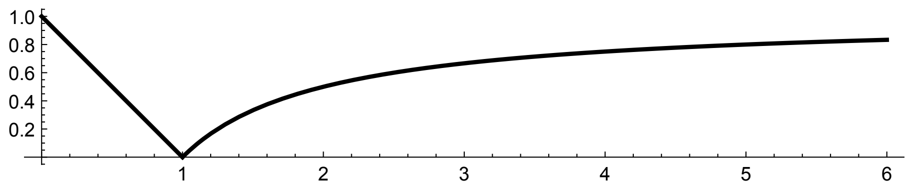

Figure 17.7. Noncompactness.

Then $\theta^* = 0$ maximizes g over $\Theta = [0, \infty)$, and $g(\theta^*) = 1$. Let $\theta_n = n$. Then

$$g(\theta_n) = 1 - (1/n) \xrightarrow[n\to\infty]{} 1 = g(\theta^*),$$

but θ_n does not converge at all.

The problem here is not that g is discontinuous, but that the sequence θ_n is unbounded.

There are two ways to deal with this issue. One is to bound θ, which is artificial. The other is to guarantee that as $\|\theta_n\| \to \infty$ that $g(\theta_n)$ is bounded away from $g(\theta^*)$. □

The next lemma is a special case of the Berge Maximum Theorem [1, Theorem 12.1, p. 64].

S5.1.4 Lemma *Let Θ be a closed bounded subset of $\boldsymbol{R}^{\mathrm{n}}$, and let $g\colon \Theta \to \boldsymbol{R}$ be continuous. Assume that θ^* is the unique maximizer of g over Θ. If θ_n is a sequence in Θ satisfying $g(\theta_n) \to g(\theta^*)$, then $\theta_n \to \theta^*$.*

Proof: We wish to show that for any small $\varepsilon > 0$, there is an N such that for all $n \geqslant N$, we have $\|\theta_n - \theta^*\| < \varepsilon$. Let $\Theta' = \{\theta \in \Theta : \|\theta_n - \theta^*\| \geqslant \varepsilon\}$. (If Θ is a singleton $\{\theta^*\}$, the conclusion is trivial, so assume that Θ has at least one other point.) If ε is small enough, then Θ' is nonempty, and closed and bounded. Therefore by the well-known Weierstrass Theorem [2, pp. 89–90], g achieves a maximum value m on Θ'. By assumption, $g(\theta^*) > m$, and $g(\theta_n) \to g(\theta^*)$. Consequently there is some N such that for all $n \geqslant N$, we have $g(\theta_n) > m$, which implies $\theta_n \notin \Theta'$, so $\|\theta_n - \theta^*\| < \varepsilon$. ∎

S5.1.5 Corollary *Let Θ be a closed subset of $\boldsymbol{R}^{\mathrm{n}}$, and let $g\colon \Theta \to \boldsymbol{R}$ be continuous. Assume that θ^* is the unique maximizer of g over Θ. Assume that there is some $M > 0$ and some $m < g(\theta^*)$ such that*

$$\|\theta - \theta^*\| > M \implies g(\theta) < m.$$

If θ_n is a sequence in Θ satisfying $g(\theta_n) \to g(\theta^*)$, then $\theta_n \to \theta^*$.

Proof: The set $\Theta'' = \{\theta \in \Theta : \|\theta - \theta^*\| \leqslant M\}$ is a closed bounded set, and if $g(\theta_n) \to g(\theta^*) > m$ there is some N such that for all $n \geqslant N$, we have $\theta_n \in \Theta''$. Now apply the lemma to Θ''. ∎

S5.2 Assumptions for consistency

We now turn to the assumptions of Wald [3] and Wolfowitz [4]. I have renumbered some of them, and slightly strengthened a few others. Another difference is that Wald states some of his hypotheses in terms of the true parameter θ_0, but since we do not know what the true parameter is, we essentially to have to verify the hypotheses for every possible θ_0. I will make that more explicit.

S5.2.1 Assumptions Assume the following:

1. (Basic nature of likelihood) The parameter space Θ is a closed subset of $\boldsymbol{R}^{\mathrm{k}}$.

Either (i) for every $\theta \in \Theta$, $f(\cdot;\theta)$ is a density; or (ii) for every $\theta \in \Theta$, $f(\cdot;\theta)$ is a probability mass function.

Even in case (ii), I will write $\int f(x;\theta)\,dx$ instead of $\sum_x f(x;\theta)$. You should not get confused.

2. (Identification) For every $\theta \neq \theta'$,

$$P_\theta\left(f(X;\theta) \neq f(X;\theta')\right) > 0.$$

(That is, different parameters define different distributions.)

3. (Integrability of log-likelihood) For each θ,

$$\boldsymbol{E}_\theta|\ln f(X;\theta)| = \int |\ln f(x;\theta)| f(x;\theta)\,dx < \infty.$$

4. (Continuity in θ) For each x, if $\theta_n \to \theta$, then $f(x;\theta_n) \to f(x;\theta)$.

(This is nominally stronger than Wald's Assumption 3, which allows for some x to be exceptions, provided the set of such x has probability zero under the "true" parameter θ_0.)

5. For each x, if $\|\theta_n\| \to \infty$, then $f(x;\theta_n) \to 0$.

(This assumption is true for every distribution we have looked at, and serves to avoid the difficulties in Example S5.1.3.)

6. (Technical conditions to ensure that certain expectations are finite and continuous in θ)

[To be written down at some later date.]

(This is Wald's assumptions 2)

Bibliography

[1] K. C. Border. 1985. *Fixed point theorems with applications to economics and game theory*. New York: Cambridge University Press.

[2] W. Rudin. 1976. *Principles of mathematical analysis*, 3d. ed. International Series in Pure and Applied Mathematics. New York: McGraw Hill.

[3] A. Wald. 1949. Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics* 20(4):595–601. http://www.jstor.org/stable/2236315

[4] J. Wolfowitz. 1949. On Wald's proof of the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics* 20(4):601–602. http://www.jstor.org/stable/2236316