Definition 1. Maximum Likelihood Estimator (MLE) For $X_i \overset{iid}{\sim} f(x|\theta)$, the MLE is

$$\hat{\theta} = \arg\max_{\theta} f(x^n|\theta) \qquad \text{for } \ \theta \in \Omega \subset \mathbb{R}^k$$

1 Consistency of the MLE

This is a summary of Wald [1949] which shows that $\hat{\theta}$ is consistent. It consists of a set of assumptions, some introductory lemmas, and the proof of the main theorem.

Consider the following 8 assumptions

1. The distribution function $F_\theta(x)$ is either discrete for all θ or absolutely continuous for all θ.
2. Let
$$\begin{aligned} f(x|\theta,p) &= \sup_{|\theta-\theta'|\leq p} f(x|\theta') \\ \phi(x|r) &= \sup_{|\theta|>r|} f(x|\theta) \\ f^*(x|\theta,p) &= \max(1, f(x|\theta,p)) \\ \phi^*(x|\theta,p) &= \max(1, \phi(x|r)) \end{aligned}$$
Assume that for sufficiently small p and sufficiently large r the expected values
$$\int_{-\infty}^{\infty} \ln f^*(x|\theta,p)\, dF_{\theta_0}(x) \quad \text{and} \quad \int_{-\infty}^{\infty} \ln \phi^*(x|r)\, dF_{\theta_0}(x)$$
are finite, $\theta_0 = \theta_{\text{true}}$.
3. If $\lim_{i\to\infty} \theta_i = \theta$ then $\lim_{i\to\infty} f(x|\theta_i) = f(x|\theta)$ for all x, except a set which may depend on θ (but not $\langle\theta_i\rangle$) and has P_{θ_0} probability 0.
4. If $\theta_0 \neq \theta_1$, there exists x such that $F_{\theta_0}(x) \neq F_{\theta_1}(x)$
5. If $\lim_{i\to\infty} |\theta_i| = \infty$ then $\lim_{i\to\infty} f(x|\theta_i) = 0$ except for a set of P_{θ_0} with probability 0 (and independent of $\langle\theta_i\rangle$)
6. $\int_{-\infty}^{\infty} |\log f(x|\theta_0)| dF_{\theta_0}(x) < \infty$
7. Ω is a closed subset of $\mathbb{R}^k$
8. $f(x|\theta,p)$ is measurable in x for any θ, p (not necessary for discrete θ)

Summary Wald's hypotheses can be succinctly stated as:

1. $\forall\, \theta \ \ \exists\, p = p(\theta) \ E_\theta \sup_{|\psi-\theta|>p} \ln \frac{f(X|\psi)}{f(X|\theta)} < 0$
2. $\forall \theta, \delta > 0$ small enough, $E_\theta \ln \frac{f(X|\theta)}{\sup_{(\theta'-\theta)<\delta} f(X|\theta)} < \infty$
3. $p(x|\theta) \to 0$ as $||\theta|| \to \infty$

Lemma 2. *For any $\theta \neq \theta_0$, we have $E \log f(X|\theta) < E \log f(X|\theta_0)$, where X is a chance variable with the distribution $F(X, \theta_0)$*

Proof:
Assumption 2 means that the expected values exist.
Assumption 6 means $E|\log f(X, \theta_0) < \infty$.
If $E \log f(X, \theta) = -\infty$, the lemma obviously holds.
If $E \log f(X, \theta) > -\infty$, then $E|\log f(X, \theta)| < \infty$.
Let $u = \log f(X, \theta) - \log f(X, \theta_0)$. $E|u| < \infty$. We know that for any RV u that is non-constant with probability 1 with finite expectation, $Eu < \log Ee^u$.
Since $Ee^u \leq 1$, $Eu < \log Ee^u \leq 0$ and thus $Eu < 0$.

Lemma 3. $\lim_{p=0} E \log f(X, \theta, p) = E \log f(X, \theta)$

Proof:

Let $f^*(x, \theta, p) = \begin{cases} f(x, \theta, p) & f(x, \theta, p) \geq 1 \\ 1 & \textit{otherwise} \end{cases}$ *and* $f^*(x, \theta) = \begin{cases} f(x, \theta) & f(x, \theta) \geq 1 \\ 1 & \textit{otherwise} \end{cases}$

$f^{**}(x, \theta, p) = \begin{cases} f(x, \theta, p) & f(x, \theta, p) \leq 1 \\ 1 & \textit{otherwise} \end{cases}$ *and* $f^{**}(x, \theta) = \begin{cases} f(x, \theta) & f(x, \theta) \leq 1 \\ 1 & \textit{otherwise} \end{cases}$

$\lim_{p \to 0} \log f^*(x, \theta, p) = \log f^*(x, \theta)$ *a.e.* *Assumption 3*

$\lim_{p \to 0} E \log f^*(X, \theta, p) = E \log f^*(X, \theta)$ *since* $\log f^*(x, \theta, p)$ *is incr in* $p +$ *Assumption 2*

$|\log f^{**}(x, \theta, p)| \leq |\log f^{**}(x, \theta)|$

$\lim_{p \to 0} \log f^{**}(x, \theta, p) = \log f^{**}(x, \theta)$ *for* x *a.e.*

$\lim_{p \to 0} E \log f^{**}(X, \theta, p) = E \log f^{**}(X, \theta)$ *follows from the two previous lines.*

Since $\lim_{p \to 0} E \log f^*(X, \theta, p) = E \log f^*(X, \theta)$
and $\lim_{p \to 0} E \log f^{**}(X, \theta, p) = E \log f^{**}(X, \theta)$, *we are done.*

Lemma 4.

$$\lim_{r \to \infty} E \log \psi(X|r) = -\infty$$

Proof:
Assumption 5 implies $\lim_{r \to \infty} \log \psi(X|r) = -\infty$ *a.e.*
Assumption 2 says $E \log \psi^*(X, r) < \infty$.
$\log \psi(x, r) - \log \psi^*(x, r)$ *and* $\log \psi^*(x, r)$ *are decreasing functions of* r, *so the lemma follows by the monotone convergence theorem.*

References

A. Wald. Note on the Consistency of the Maximum Likelihood Estimate. *The Annals of Mathematical Statistics*, 20(4):595–601, 1949. ISSN 0003-4851. URL https://www.jstor.org/stable/

2236315. Publisher: Institute of Mathematical Statistics.