2. Asymptotics for likelihood methods

2.1. Estimation

Let $X_1, X_2, \ldots$ be iid with density $f(x;\theta)$. Write $L_n(\theta)$ for the likelihood, and $L_n^*(A) = \sup_{\theta \in A} L_n(\theta)$.

Theorem 2.1.1 (Consistency of mle using Wald's technique)
Assume

(i) $f(x;\theta) \not\equiv f(x;\theta_0)$ for all $\theta \neq \theta_0$.
(ii) For all x, f is continuous in θ
(iii) Let H be a compact subset of $\Theta \subset \mathbf{R}^d$ with $\theta_0 \in H$, and for some r $\mathbf{E}_{\theta_0} \log(L_r(\theta_0)/L_r^*(H)) > -\infty$.

Then there exists a sequence $\hat{\theta}_n \in H$ such that $L_n(\hat{\theta}_n) = L_n^*(H)$ and $\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta_0$ as $n \to \infty$. If in addition

(iv) $\mathbf{E}_{\theta_0} \log(L_r(\theta_0)/L_r^*(H^c)) > 0$, then a.s. $L_n(\hat{\theta}_n) = L_n^*(\Theta)$ for n large enough.

Let $S_n(\theta) = \dot{l}_n(\theta)$ be the **score function**. The **likelihood equation(s)** is $S_n(\theta) = 0$.

Theorem 2.1.2 (Cramér's approach to asymptotic normality of likelihood equation solutions)
Assume

(1) $\theta \neq \theta^*$ implies that $f(x;\theta) \not\equiv f(x;\theta^*)$
(2) The support of $f(x;\theta)$ does not depend on θ
(3) Θ contains an open neighborhood N_0 of θ_0 in which
(i) $\log f(x;\theta)$ is twice continuously differentiable as a function of θ for a.e. x
(ii) $\dfrac{\partial^3}{\partial\theta_i \partial\theta_j \partial\theta_k} \log f(x;\theta)$ exists, bounded in absolute value by $M_{ijk}(x)$ with $m_{ijk} = \mathbf{E}(M_{ijk}(X)) < \infty$ for all $\theta \in N_0$, $1 \leq i, j, k \leq d$.
(4) (i) $\mathbf{E}_{\theta_0} S_1(\theta_0) = 0$
(ii) $\mathbf{E}_{\theta_0} S_1(\theta_0)^T S_1(\theta_0) < \infty$
(iii) $I_1(\theta_0) = -\mathbf{E}_{\theta_0}(\dot{S}_1(\theta_0))$ is positive definite.

Then with probability convergin to one as $n \to \infty$ there is a sequence $\hat{\theta}_n = \hat{\theta}_n(\mathbf{X}_n)$ of likelihood equation solutions such that

(a) $\hat{\theta}_n$ is consistent for θ_0
(b) $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathrm{N}(0, I_1(\theta_0)^{-1})$.