The Truth (More Than You Probably Wish to Know) About the Asymptotics of "Maximum Likelihood" and the LRT Statistic
Stat 543 Spring 2005

Theorem 1 *Suppose that* $X_1, X_2, ..., X_n$ *are iid random vectors and* $f(x|\theta)$ *is either the marginal probability mass function or the marginal probability density. Suppose that* $\Theta \subset \mathcal{R}^1$ *and there exists an open neighborhood of* θ_0*, say* $\mathcal{O}$*, such that*

i) $f(x|\theta) > 0\, \forall x$ *and* $\forall \theta \in \mathcal{O}$*,*

ii) $\forall x$*,* $f(x|\theta)$ *is differentiable at every point* $\theta \in \mathcal{O}$*, and*

iii) $E_{\theta_0} \ln f(X|\theta)$ *exists* $\forall \theta \in \mathcal{O}$ *and* $E_{\theta_0} \ln f(X|\theta_0) < \infty$*.*

Then for any $\epsilon > 0$*, the* θ_0 *probability that the likelihood equation*

$$\ell_n'(\theta) = \frac{d}{d\theta}\ell_n(\theta) = \frac{d}{d\theta}\sum_{i=1}^{n} \ln f(X_i|\theta) = 0 \qquad (1)$$

has a root within ϵ *of* θ_0 *converges to 1.*

Corollary 2 *Under the hypotheses of Theorem 1, suppose further that* Θ *is open and that* $\forall x$*,* $f(x|\theta)$ *is positive and differentiable at every point* $\theta \in \Theta$*, so that the likelihood equation* (1) *makes sense at all* $\theta \in \Theta$*. Define* $\delta_n(X)$ *to be the root of the likelihood equation when there is exactly one (and make any convenient definition for other cases). If with* θ_0 *probability approaching 1 the likelihood equation has a single root, then* $\delta_n(X)$ *is consistent in probability for* θ *at* θ_0*.*

Corollary 3 *Under the hypotheses of Theorem 1, if* $\{T_n(X)\}$ *is a sequence of estimators consistent for* θ *in probability at* θ_0*, and*

$$\delta_n(X) = \begin{cases} T_n(X) & \text{if the likelihood equation has no roots} \\ \text{the root closest to } T_n(X) & \text{if the likelihood equation has at least one root} \end{cases} ,$$

then $\delta_n(X)$ *is consistent in probability for* θ *at* θ_0*.*

Notation 4 *In an iid model, i.e. where* $X = (X_1, ..., X_n)$*, let* $I_1(\theta_0)$ *be the Fisher Information Matrix at* θ_0 *for a single observation. That is, with marginal distribution specified by* $f(x|\theta)$*,*

$$I_1(\theta_0) = \left(E_{\theta_0} \frac{\partial}{\partial\theta_i} \ln f(X_1|\theta)\bigg|_{\theta=\theta_0} \frac{\partial}{\partial\theta_j} \ln f(X_1|\theta)\bigg|_{\theta=\theta_0} \right) .$$

Theorem 5 *Suppose that* $X_1, X_2, ..., X_n$ *are iid random vectors and* $f(x|\theta)$ *is either the marginal probability mass function or the marginal probability density. If the model for* X_1 *is FI regular at* θ_0 *and with* $X = (X_1, ..., X_n)$*,* $I(\theta_0)$ *is the Fisher Information Matrix for* X *at* θ_0*,*

$$I(\theta_0) = nI_1(\theta_0) .$$

Theorem 6 *Suppose that* $X_1, X_2, ..., X_n$ *are iid random vectors and* $f(x|\theta)$ *is either the marginal probability mass function or the marginal probability density. Suppose that* $\Theta \subset \mathcal{R}^1$ *and there exists an open neighborhood of* θ_0*, say* $\mathcal{O}$*, such that*

i) $f(x|\theta) > 0\, \forall x$ *and* $\forall \theta \in \mathcal{O}$*,*

ii) $\forall x$*,* $f(x|\theta)$ *is three times differentiable at every point* $\theta \in \mathcal{O}$*,* *iii)* $\exists\, M(x) \geq 0$ *with* $E_{\theta_0} M(X_1) < \infty$ *and* $\left|\frac{d^3}{d\theta^3} \ln f(x|\theta)\right| \leq M(x)\ \ \forall x$ *and* $\forall \theta \in \mathcal{O}$,

iv)

$$0 = \sum_x \frac{d}{d\theta} f(x|\theta)\Big|_{\theta=\theta_0} \text{ and } 0 = \sum_x \frac{d^2}{d\theta^2} f(x|\theta)\Big|_{\theta=\theta_0} \text{ in the discrete case}$$

or

$$0 = \int \frac{d}{d\theta} f(x|\theta)\Big|_{\theta=\theta_0} dx \quad \text{and} \quad 0 = \int \frac{d^2}{d\theta^2} f(x|\theta)\Big|_{\theta=\theta_0} dx \quad \text{in the continuous case, and}$$

v) $I_1(\theta_0) \in (0, \infty)\ \forall \theta \in \mathcal{O}$.

Then if with θ_0 *probability approaching 1,* $\delta_n(X)$ *is a root of the likelihood equation and is consistent for* θ *in probability at* θ_0*, under the* θ_0 *distribution for* X

$$\sqrt{n}\,(\delta_n(X) - \theta_0) \ \xrightarrow{\mathcal{L}}\ N(0, \frac{1}{I_1(\theta)}) \quad .$$

Corollary 7 *Under the hypotheses of Theorem 6, if* $I_1(\theta)$ *is continuous at* θ_0*, then under the* θ_0 *distribution for* X

$$\sqrt{nI_1(\delta_n(X))} \cdot (\delta_n(X) - \theta_0) \ \xrightarrow{\mathcal{L}}\ N(0,1) \quad .$$

Corollary 8 *Under the hypotheses of Theorem 6, and the* θ_0 *distribution for* X

$$\sqrt{-\ell_n''(\delta_n(X))} \cdot (\delta_n(X) - \theta_0) \ \xrightarrow{\mathcal{L}}\ N(0,1) \quad .$$

Useful Fact If $Y \sim \mathrm{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $(Y - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(Y - \boldsymbol{\mu}) \sim \chi^2_k$.

Theorem 9 *Under the hypotheses of Theorem 6 if*

$$\lambda_n^*(X) = 2 \ln \left(\frac{L_n(\delta_n(X))}{L_n(\theta_0)} \right) = 2 \left(\ell_n(\delta_n(X)) - \ell_n(\theta_0) \right) \quad ,$$

then under the θ_0 *distribution for* X*,*

$$\lambda_n^*(X) \ \xrightarrow{\mathcal{L}}\ \chi^2_1 \quad .$$