Stat 543 4-11-05

"Recall" ... Asymptotics (large sample Thy) of "Maximum Likelihood"

Folklore: MLE's are ① consistent
② asymptotically normal (with variance $I^{-1}(\theta)$)

Handout (iid case)

Thm 1 The $\theta_0$ probability that the score function has a 0 within $\epsilon$ of $\theta_0$ goes to 1

Corollary 2 Under additional conditions sufficient to make sure that the likelihood equation

$$\ell_n'(\theta) = 0 \qquad (*)$$

makes sense for all $\theta \in \Theta$, define

$$\delta_n(X) = \begin{cases} \text{the root of the likelihood equation when there is exactly 1} \\ \text{anything convenient when there is not exactly 1 root of } (*) \end{cases}$$

Then if for $\theta_0 \in \Theta$

$$P_{\theta_0}\left[\, (*) \text{ has single root} \,\right] \longrightarrow 1$$

then $\delta_n(X)$ is consistent for $\theta$ at $\theta_0$

Example $\quad X_1, X_2, \ldots, X_n$ iid Bernoulli $p \quad p \in (0,1)$

except when $\bar{x}_n = \hat{p}_n = 0$ or $1$ $\quad \bar{x}_n$ is the only root of the likelihood equation $l_n'(p) = 0$

– but

$$P_p\left[\bar{X}_n = 0 \text{ or } \bar{X}_n = 1\right] = (1-p)^n + p^n$$

$$\longrightarrow 0 \quad \left(\text{provided } p \in (0,1)\right)$$

corollary says $\bar{X}_n$ is consistent for $p$

– this follows more directly from WLLN

$$\bar{X}_n \xrightarrow{P_p} EX_1 = p$$

Corollary 3   Under the hypotheses of Thm 1 if $\{T_n(X)\}$ is a sequence of estimators of $\theta$ consistent at $\theta_0$ and

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

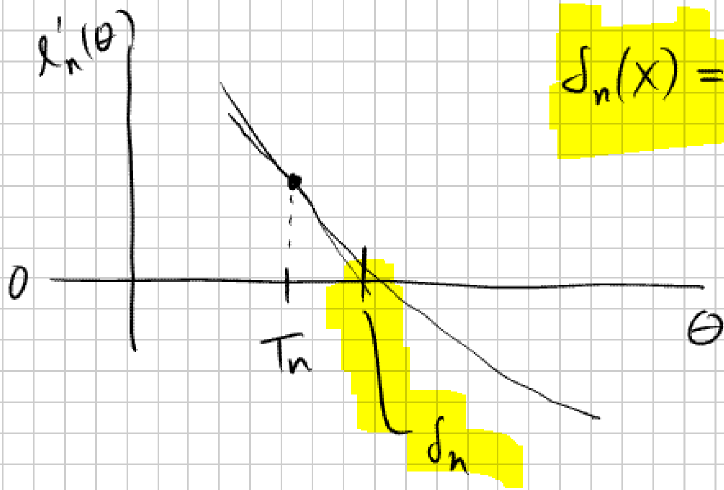
then $\{\delta_n(X)\}$ is consistent for $\theta$ at $\theta_0$

---

this has practical problems (I'd like to use it because roots of $\ell_n'(\theta)=0$ have other nice properties) unless I can do calculus to implement

?? . . .

As an alternative, it is common to use "1-step Newton improvements on consistent estimators" —




$$\delta_n(X) = T_n(X) - \frac{\ell_n'(T_n(X))}{\ell_n''(T_n(X))}$$

this turns out to have limiting properties that are as good as those of roots of the likelihood equation

The multi-parameter version of this is $\Theta \subset \mathbb{R}^k$

for

$$\underset{k \times k}{H}(\theta) \equiv \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} l_n(\theta) \right)$$

use

$$\underset{k \times 1}{\hat{\theta}_n} = \underset{k \times 1}{T_n} - \underbrace{H^{-1}(T_n)}_{k \times k} \underbrace{\nabla l_n(T_n)}_{k \times 1}$$

---

Now what about limiting normality of "MLE's" –

Theorem 6  If $\Theta \subset \mathbb{R}^1$ under appropriate regularity conditions, if with $\theta_0$ probability approaching 1,

$$\ell_n'(\delta_n(X)) = 0$$

and $\delta_n(X)$ is consistent for $\theta$ at $\theta_0$, then

$$\sqrt{n}\,(\delta_n(X) - \theta_0) \xrightarrow{\mathcal{L}_{\theta_0}} N\left(0, \frac{1}{I_1(\theta_0)}\right)$$

(The $I_1$ above is the FI in a single observation, $X_1$, about $\theta$ at $\theta_0$)

roughly "MLE's are asymptotically normal with variance of the limit dsn $\frac{1}{n I_1(\theta_0)}$"

A multivariate version of this would say that under appropriate conditions if $\hat{\theta}_n$ is an "MLE" of $\theta \in \mathbb{R}^k$ then under the $\theta_0$ dsn for $X_i$'s

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{L}} N_k\left(0, I_1^{-1}(\theta_0)\right)$$

<u>Example</u> (obvious) iid Bernoulli$(p)$

$$\bar{X}_n = \frac{1}{n}\sum X_i = \hat{p}_n$$ is a root of the likelihood equation with $p$ probability approaching 1 for any $p \in (0,1)$ — $\bar{X}_n$ is also consistent for $p \in (0,1)$ — so the theorem promises

$$\sqrt{n}(\bar{X}_n - p) \xrightarrow{\mathcal{L}_p} N\left(0, \frac{1}{I_1(p)}\right)$$

$$I_1(p) = \text{FI in } X_1 \text{ about } p = \frac{1}{p(1-p)}$$

So the theorem promises that

$$\sqrt{n}(\bar{X}_n - p) \xrightarrow{\mathcal{L}_p} N(0, p(1-p))$$

which comes far more simply from CLT