Then

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

(I_1 the $k \times k$ FI matrix for a single X_1)

Example ("Obvious") iid Bernoulli(p) problem

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

The theorem promises that

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

for any $p \in (0,1)$ — But

$$I_1(p) = \frac{1}{p(1-p)}$$

so the theorem promises that for $p \in (0,1)$

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

which is exactly what the CLT already promises

So, why does Thm 6 work in general? The following is not a tight proof, but an outline for the $\Theta \subset \mathbb{R}^1$ case —

day 36

Still: $\ell_n(\theta) = \sum_{i=1}^{n} \log f(X_i|\theta)$

$\delta_n(X) = \hat{\theta}_n$ a consistent root of the likelihood equation

Expand $l_n'(\cdot)$ around θ_0 in a Taylor series

$$0 = l'(\hat{\theta}_n) = l_n'(\theta_0) + (\hat{\theta}_n - \theta_0) l_n''(\theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)^2 l'''(\theta_1)$$

for some θ_1 between $\hat{\theta}_n$ and θ_0

So $$-l_n'(\theta_0) = (\hat{\theta}_n - \theta_0)\left[l_n''(\theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0) l'''(\theta_1)\right]$$

and thus

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{-\sqrt{n}\, l_n'(\theta_0)}{l_n''(\theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0) l_n''(\theta_1)}$$

$$= -\frac{\overbrace{\frac{1}{\sqrt{n}} l_n'(\theta_0)}^{A_n}}{\underbrace{\frac{1}{n} l_n''(\theta_0)}_{B_n} + \underbrace{\frac{1}{n}\frac{1}{2}(\hat{\theta}_n - \theta_0) l_n'''(\theta_1)}_{C_n}}$$

$$A_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{d}{d\theta} \log f(X_i|\theta)\Big|_{\theta=\theta_0}$$

But under $\theta = \theta_0$ the variables

$$\frac{d}{d\theta} \log f(X_i|\theta)\Big|_{\theta=\theta_0}$$

are iid with mean 0 (θ_0 mean of the score function at θ_0 is 0) and variance

$$E_{\theta_0}\left(\frac{d}{d\theta} \log f(X_i|\theta)\Big|_{\theta=\theta_0}\right)^2 = I_1(\theta_0)$$

So $A_n \xrightarrow{\mathcal{L}_{\theta_0}} N(0, I_1(\theta_0))$

$$B_n = \frac{1}{n}\sum_{i=1}^{n} \frac{d^2}{d\theta^2}\log f(X_i|\theta)\Big|_{\theta=\theta_0}$$

But under $\theta=\theta_0$ the variables $\frac{d^2}{d\theta^2}\log f(X_i|\theta)\Big|_{\theta=\theta_0}$

are iid with mean

$$E_{\theta_0}\left(\frac{d^2}{d\theta^2}\log f(X_1|\theta)\Big|_{\theta=\theta_0}\right) = -I_1(\theta_0)$$

2nd form of FI

and WLLN says $B_n \xrightarrow{P_{\theta_0}} -I_1(\theta_0)$

To handle C_n, hypothesis of consistency gets

$$\hat{\theta}_n - \theta_0 \xrightarrow{P_{\theta_0}} 0$$

and there is enough built into the hypotheses of the theorem to show that

$$\frac{1}{n}\ell_n'''(\theta_1)$$

is "bounded in P_{θ_0} probability" so that ultimately

$$C_n = \frac{1}{2}(\hat{\theta}_n - \theta_0)\left(\frac{1}{n}\ell'''(\theta_1)\right) \xrightarrow{P_{\theta_0}} 0$$

So we then consider

$$g(u,v,w) = \frac{u}{-v-w}$$

and conclude that since this function is cont^s except where $v+w=0$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = g(A_n, B_n, C_n)$$

$$\xrightarrow{\mathcal{L}_{\theta_0}} \frac{N(0, I_1(\theta_0))}{I_1(\theta_0) - 0}, \text{ i.e. } N(0, I_1^{-1}(\theta_0))$$

Theorem 6 doesn't really provide all we need for inference purposes -

$$\sqrt{n}\,(\delta_n(X)-\theta_0) \underset{\theta_0}{\dot\sim} N\left(0, \frac{1}{I_1(\theta_0)}\right)$$

says

$$\frac{\delta_n(X)-\theta_0}{\frac{1}{\sqrt{nI_1(\theta_0)}}} \dot\sim N(0,1)$$

gets me the unusable confidence limits for θ

$$\delta_n(X) \pm z \frac{1}{\sqrt{nI_1(\theta)}}$$

This depends on θ (that is to be estimated)

What is needed is some rationale for replacing $nI_1(\theta)$ by something that is data-based - there are two standard possibilities - the 1st is to replace θ with its estimator $\delta_n(X)$ - this is supported by

Corollary 7 (ML Handout) Under appropriate regularity conditions and the θ_0 dsn ($\delta_n(X)$ as before)

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

Why?

$$\sqrt{nI_1(\delta_n(X))}\,(\delta_n(X)-\theta_0) = \sqrt{\frac{I_1(\delta_n(X))}{I_1(\theta_0)}}\;\underbrace{\sqrt{nI_1(\theta_0)}\,(\delta_n(X)-\theta_0)}_{\xrightarrow{\mathcal{L}_{\theta_0}} N(0,1)}$$

$$\xrightarrow{P_{\theta_0}} 1$$ if, e.g., $I_1(\cdot)$ is cont$^{\underline{s}}$ at θ_0

Using this fact to do inference for θ is sometimes known as using "Expected" Fisher Information – it obviously requires computing the function $I_1(\theta)$ –

Example $X_1, X_2, \ldots, X_n$ iid Ber(p)

$I_1(p) = \dfrac{1}{p(1-p)}$ so we can replace the unusable confidence limits for p

$$\hat{p}_n \pm z\sqrt{\frac{1}{nI_1(p)}} \quad \text{i.e. } \hat{p}_n \pm z\sqrt{\frac{p(1-p)}{n}}$$

with the realizable ones

$$\hat{p}_n \pm z\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}$$

Another possibility is to note that

$$I(\theta_0) = nI_1(\theta_0) = -n\,E_{\theta_0}\left.\frac{d^2}{d\theta^2}\log f(X_1|\theta)\right|_{\theta=\theta_0}$$

and that

$$\frac{1}{n}\ell_n''(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}\left.\frac{d^2}{d\theta^2}\log f(X_i|\theta)\right|_{\theta=\theta_0}$$

so that by the LLN

$$-\frac{1}{n} l_n''(\theta_0) \xrightarrow{P_{\theta_0}} I_1(\theta_0)$$

and so it is plausible that under appropriate regularity conditions

$$-\frac{1}{n} l_n''(\delta_n(X)) \xrightarrow{P_{\theta_0}} I_1(\theta_0)$$

which in turn makes plausible

Corollary 8 (ML Handout) Under appropriate regularity conditions

$$\sqrt{-l_n''(\delta_n(X))}\,\left(\delta_n(X) - \theta_0\right) \xrightarrow{\mathcal{L}\,\theta_0} N(0,1)$$

The corresponding approximate confidence limits for θ are

$$\delta_n(X) \pm z \frac{1}{\sqrt{-l_n''(\delta_n(X))}}$$

This kind of logic is sometimes called use of the "Observed Fisher Information" (i.e. the negative 2nd derivative of the loglikelihood at the "MLE")

day 37

Example $X_1, X_2, \ldots, X_n$ iid Ber(p)

$$l_n(p) = \sum X_i (\ln p) + (n - \sum X_i) \ln(1-p)$$

$$l_n'(p) = \frac{\sum X_i}{p} - \frac{n - \sum X_i}{1-p}$$

$$l_n''(p) = -\frac{\sum X_i}{p^2} - \frac{n - \sum X_i}{(1-p)^2}$$

$$l_n''(\hat{p}_n) = -\frac{\sum X_i}{\left(\frac{\sum X_i}{n}\right)^2} - \frac{n-\sum X_i}{\left(\frac{n-\sum X_i}{n}\right)^2}$$

$$= -n\left(\frac{1}{\hat{p}_n} + \frac{1}{1-\hat{p}_n}\right)$$

$$= -\frac{n}{\hat{p}_n(1-\hat{p}_n)}$$

So we get approximate confidence limits for p

$$\hat{p}_n \pm z\frac{1}{\sqrt{-l_n''(\hat{p}_n)}}$$

i.e.

$$\hat{p}_n \pm z\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}$$

Here (as in all exponential family cases) the "Expected FI" and "Observed FI" analyses turn out to be equivalent

---

Use of the MVN_k limit dsn of an "MLE" of a k-dimensional parameter vector θ provides inferences for θ

<u>MVN Fact</u> : $Y \sim MVN_k(\mu, \Sigma) \Rightarrow (Y-\mu)'\Sigma^{-1}(Y-\mu) \sim \chi^2_k$

So if, e.g., c is the upper α pt. of the χ^2_k dsn

$$P\left[(Y-\mu)'\Sigma^{-1}(Y-\mu) < c\right] = 1-\alpha$$

and $\therefore$ for $Y=y$, the set

$$\left\{\mu \in \mathbb{R}^k \mid (y-\mu)'\Sigma^{-1}(y-\mu) < c\right\}$$

serves as a $(1-\alpha)\times 100\%$ confidence set for μ