Problem set 2

$$\mathcal{L}(\theta) = \sum_{i=1}^{N} \ell(y_i | x_i, \theta) = \sum_{i=1}^{N} l_i(\theta)$$

$$s_i(\theta) = \nabla \ell_i(\theta) \; ; \; H_i(\theta) = \nabla^2 \ell_i(\theta)$$

$$E(s_i(\theta_0)) = 0$$

$$V(s_i(\theta_0)) = -E(H_i(\theta_0))$$

$$\theta_0 = \underset{\theta}{\operatorname{argmax}} \, E(\mathcal{L}(\theta))$$

$\operatorname{plim} \hat{\theta} = \theta_0$ (under assumptions)

1c

FOC: $\nabla \mathcal{L}(\hat{\theta}) = 0$

$$\sum_{i=1}^{N} s_i(\hat{\theta}) = 0$$

Want to get: Expression for $(\hat{\theta} - \theta_0)$

Taylor expansion of $s(\theta)$ around θ_0

$$\sum_{i=1}^{N} s_i(\hat{\theta}) = s_i(\theta_0) + \nabla \sum s_i(\theta_0)(\hat{\theta} - \theta_0) + R(\hat{\theta})$$

$R(\tilde{\theta})$ converges to 0 faster than

$\hat{\theta} \to \theta_0$

$$\sum_{i=1}^{N} s_i(\hat{\theta}) \overset{a}{=} s_i(\theta_0) + \sum_{i=1}^{N} H_i(\theta_0)(\hat{\theta} - \theta_0)$$

$$\hat{\theta} - \theta_0 \overset{a}{=} \left[\sum_i H_i(\theta)\right]^{-1} \cdot \left(\sum_{i=1}^{N} \underbrace{s_i(\hat{\theta})}_{=0} - s_i(\theta_0)\right)$$

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{p} -\left[\frac{1}{N}\sum H_i(\theta_0)\right]^{-1} \frac{1}{\sqrt{N}} \sum s_i(\theta_0)$$

LLN + CMT

Variance is B_0

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{p} \underbrace{-E(H_i(\theta_0))}_{A_0}{}^{-1} \sqrt{N} \overbrace{\frac{1}{N}\sum s_i(\theta_0)}$$

CLT: $\sqrt{N}(\hat{\theta} - \theta_0) \overset{a}{\sim} N(0, A_0^{-1} B_0 A_0^{-1})$

$B_0 = \text{Avar}(s_i(\theta_0))$

But we know that

$$\text{Avar}(s_i(\theta)) = -E(H_i(\theta_0)) = A_0 = B_0$$

Then $\sqrt{N}(\hat{\theta} - \theta_0) \overset{a}{\sim} N(0, A_0^{-1})$

$$\text{Avar}(\hat{\theta}) = \frac{1}{N} A_0^{-1}$$

1d

Fisher information

$$I(\theta) = E(s_i(\theta) s_i(\theta)')$$

We have shown that

$$\underbrace{I(\theta_0) = -E(H_i(\theta_0))}_{\text{The information matrix equality}} = A_0$$

Question 2

a) The urn example. We have an urn filled with balls which are either red or blue. We draw N balls want to estimate p — the probability that a randomly drawn ball is red.

$$y_i = \begin{cases} 0 & \text{if the ball is blue} \\ 1 & \text{otherwise} \end{cases}$$

$\theta = p$ a scalar

No x's.

Bernoulli scheme

$\sum_{i=1}^{N} y_i = N_1 \sim$ Binomial distribution

$E y_i = p$

$\text{Var}\, y_i = p(1-p)$

$E N_1 = N \cdot p$

$\text{Var}\, N_1 = N p (1-p)$

$$L(p) = \prod_{i=1}^{N} p^{y_i} (1-p)^{1-y_i}$$

$$\mathcal{L}(p) = \sum_{i=1}^{N} \{ y_i \ln p + (1-y_i) \ln(1-p) \}$$

$$= N_1 \ln p + (N - N_1) \ln(1-p)$$

$$l_i(p) = y_i \ln p + (1-y_i) \ln(1-p)$$

$$S_i(p) = \frac{y_i}{p} - \frac{1-y_i}{1-p} = \frac{d\, l_i(p)}{dp}$$

$$H_i(p) = -\frac{y_i}{p^2} - \frac{1-y_i}{(1-p)^2} = \frac{d\, S_i(p)}{dp}$$

$$A_0 = -E(H_i(p))$$

$$-E(H_i(p)) = E\left(\frac{y_i}{p^2} + \frac{1-y_i}{(1-p)^2}\right)$$

$$= \frac{E(y_i)}{p^2} + \frac{1-E(y_i)}{(1-p)^2} = \frac{\cancel{p}}{p^{\cancel{2}}} + \frac{\cancel{1-p}}{(1-p)^{\cancel{2}}}$$

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

$$Avar(\hat{p}) = \frac{1}{N}\hat{A}_0^{-1} = \frac{p(1-p)}{N}$$

b) Cramér-Rao lower bound

For any consistent estimator $\hat{\theta}$,

$$Avar(\hat{\theta}) = I_N(\theta_0)^{-1} + \Delta$$

where Δ is positive semi-definite

For MLE,

$$I(\theta_0) = Var(S_i(\theta_0))$$

$$I_N(\theta_0) = E(\nabla \mathcal{L}(\theta_0)' \nabla \mathcal{L}(\theta_0))$$

In our example,

$$\mathcal{L}(p) = N_1 \ln p + (N - N_1)\ln(1-p)$$

$$\frac{d\mathcal{L}(p)}{dp} = \frac{N_1}{p} - \frac{N - N_1}{1-p}$$

$$E\left[\left(\frac{d\mathcal{L}(p)}{dp}\right)^2\right] = E\left[\left(\frac{N_1}{p} - \frac{N-N_1}{1-p}\right)^2\right]$$

$$= E\left(\frac{N_1^2}{p^2} + \frac{(N-N_1)^2}{(1-p)^2} - 2\frac{N_1(N-N_1)}{p(1-p)}\right)$$

$$= \frac{E(N_1^2)}{p^2} + \frac{N + E(N_1)^2 - 2NE(N_1)}{(1-p)^2}$$

$$- 2\frac{NE(N_1) - E(N_1^2)}{p(1-p)}$$

$$E(N_1) = N \cdot p$$

$$Var(N_1) = Np(1-p)$$

$$Var(N_1) = E(N_1^2) - E(N_1)^2$$

$$E(N_1^2) = N \cdot p(1 - p + Np)$$

$= \ldots$ (left as exercise)

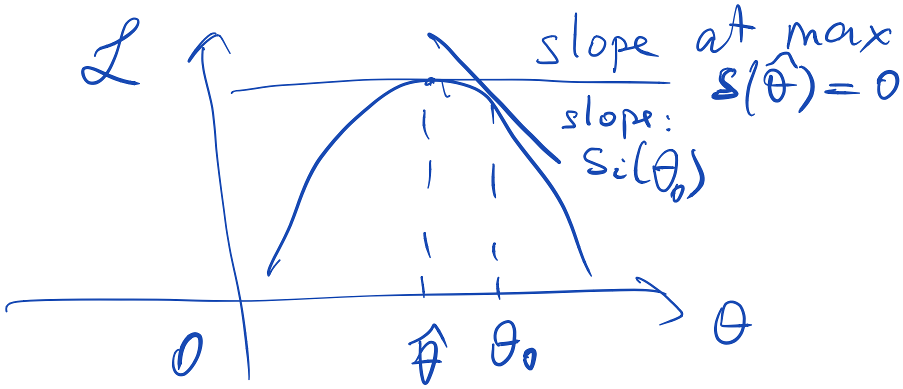
$$= \frac{N}{p(1-p)} = A_0$$

$\hat{p} = \frac{N_1}{N}$ attains Cramér-Rao l.b.

Summary

$$\hat{\theta} \xrightarrow{P} \theta_0$$

Used a linear approximation of $S_i(\theta_0)$ to get an expression for $\hat{\theta}$



$$\hat{\theta} - \theta = -\sum H_i(\theta_0)\left(\sum S_i(\theta)\right)$$

$$\sqrt{N}\left(\hat{\theta} - \theta\right) \overset{a}{\approx} N(0, A_0^{-1})$$

$A_0 = B_0 \qquad -E(\overset{A_0}{H_i}(\theta_0)) = \underbrace{Var(\overset{B_0}{S_i}(\theta_0))}_{I(\theta_0)}$

information matrix equality

$I_N(\theta_0)^{-1}$ the Cramér–Rao l.b.