this time: likelihood intervals.

next time: Bayes ↓

Hit Gelman et al. ch. 2 hard now

AMS 206
5 Feb 19

CS 2

$(Y_i \mid \theta\, \mathcal{B}) \overset{IID}{\sim} \text{Bernoulli}(\theta)$ $(i = 1, \dots, n)$

$n = 403$
$s = 72$

$\hat{\theta} = \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$

$y = (y_1, \dots, y_n)$

(1) assume $p(y_i \mid \theta\, \mathcal{B}) = \theta^{y_i} (1-\theta)^{1-y_i}$

(2) $p(Y_1 = y_1, \dots, Y_n = y_n \mid \theta\, \mathcal{B})$

$\overset{ID}{=} \prod_{i=1}^{n} p(y_i \mid \theta\, \mathcal{B})$

$\overset{IID}{=} \prod_{i=1}^{n} \theta^{y_i} (1-\theta)^{1-y_i}$

$= \underline{\underline{\theta}}^{s} (1-\theta)^{n-s}$ $\quad s = \sum_{i=1}^{n} y_i$

(3) $\boxed{\ell(\theta \mid y\, \mathcal{B})} = c \cdot \theta^{s} (1-\theta)^{n-s}$ $(c > 0)$

Neyman's approx 100(1 − α)% confidence interval (α = 0.05)

$\hat{\theta} \pm \underbrace{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}_{1.96} \underbrace{\hat{SE}(\hat{\theta})}_{\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}}$

(4) ~~(c=1)~~

$\ell\ell(\theta \mid y\, \mathcal{B}) = s \log \theta + (n-s) \log(1-\theta) = \ell\ell(\theta \mid s\, \mathcal{B})$

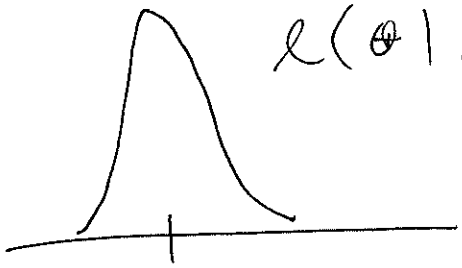
(lik, log lik) depend on y only through s

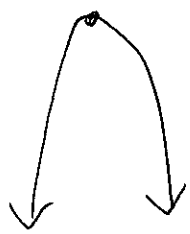
(5) sufficient statistic

convention: n is always secret element 0 in the suff. statistic vector

(n, s) (0, 1)

②





$$\frac{\partial}{\partial \theta} \ell\ell(\theta \mid s\,\mathcal{B}) = 0$$

$\rightarrow$ solve for θ $\rightarrow$ $\hat{\theta}_{MLE} = \bar{y}$

Def. A statistical procedure (inference, prediction, decision) is <u>well calibrated</u> if it gets the truth about as often as it claims to.

Mr. Neyman's large-sample claim

everything in math: if A then B

If the sampling dist assumption <u>You</u> make in step 1 is accurate for <u>Your</u> data & n is big enough for CLT to work well, then Your method for creating Your $100(1-\alpha)\%$ CI is <u>well-calibrated</u>

Huge advantage of frequentist methods: when properly constructed & when their assumptions are met, they are automatically well-calibrated

CS 2 Fisher $\ell(\theta | y\, \mathcal{B}) = \ell(\theta | s\, \mathcal{B})$

$= \theta^{s} (1-\theta)^{n-s}$

$\ell\ell(\theta | y\, \mathcal{B}) = \ell\ell(\theta | s\, \mathcal{B}) = s \log \theta + (n-s) \log(1-\theta)$

$\hat{\theta}_{MLE} = \bar{y}$ Fisher was a rabid frequentist when he invented MLE

next: $\hat{SE}_{RS}(\hat{\theta}_{MLE}) = ?$ $n = 403$, $s = ?$

random variable estimator

$\sqrt{V_{RS}(\hat{\theta}_{MLE})} = \underline{\underline{SD}}_{RS}(\hat{\theta}_{MLE}) = SE(\hat{\theta}_{MLE})$

let's devise a good estimate:

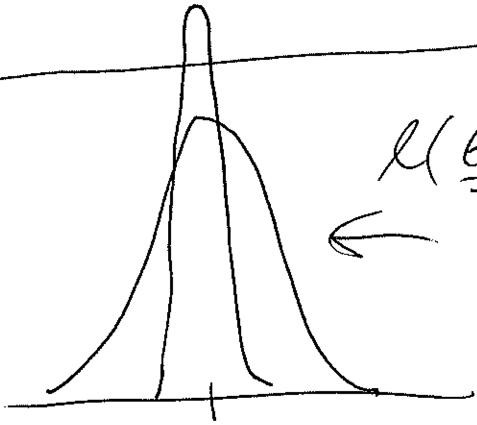
$\hat{V}_{RS}(\hat{\theta}_{MLE})$

as $n \uparrow$ $V_{RS}(\hat{\theta}_{MLE}) \downarrow$

we already know (Neyman) that $V_{RS}(\hat{\theta}_{MLE}) = \frac{\theta(1-\theta)}{n}$

$\bar{y}$

$\hat{V}_{RS}(\hat{\theta}_{MLE}) = \frac{\hat{\theta}(1-\hat{\theta})}{n}$



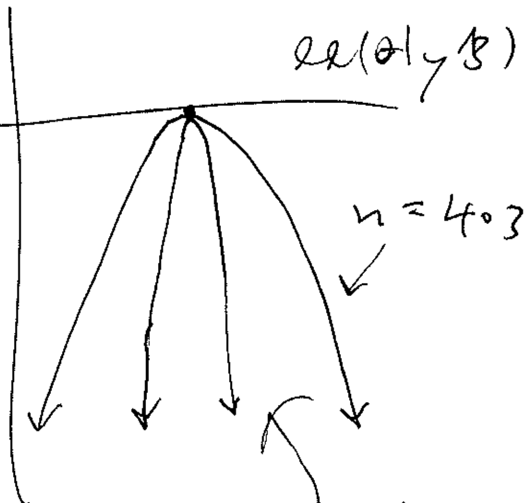
$\ell(\underline{\theta} \mid y\mathcal{B})$ ← $n = 403$, $s = 72$

$n = 4030$, $s = 720$

$\hat{\theta}_{MLE}$

$\ell(\theta \mid y\mathcal{B}) = \underline{c} \cdot$ joint sampling dist $(c > 0)$

$\ell\ell(\theta \mid y\mathcal{B}) = \underline{c'} + \log(\quad)$



as $n \uparrow$,

this $\rightarrow$ $\left. -\frac{\partial^2}{\partial\theta^2} \ell\ell(\theta \mid y\mathcal{B}) \right|_{\theta = \hat{\theta}_{MLE}}$ $\uparrow$

as $n \uparrow$, our information about $\theta \uparrow$

~~estimate (expected Fisher info.)~~

Fisher: def.

$\hat{I}(\hat{\theta}_{MLE}) = \left. -\frac{\partial^2}{\partial\theta^2} \ell\ell(\theta \mid y\mathcal{B}) \right|_{\theta = \hat{\theta}_{MLE}}$

(1925) observed Fisher information

information about θ

Fisher showed (under regularity conditions) ⑤

that $\boxed{\hat{V}_{RS}(\hat{\theta}_{MLE}) \doteq \frac{1}{\hat{I}(\hat{\theta})}}$

$$\therefore \quad \hat{SE}_{RS}(\hat{\theta}_{MLE}) = \sqrt{\frac{1}{\hat{I}(\hat{\theta})}}$$

$$ll(\theta | y B) = s \log \theta + (n-s) \log(1-\theta)$$

$$\frac{\partial}{\partial\theta} ll(\theta | y B) = \frac{s}{\theta} - \frac{(n-s)}{(1-\theta)} \qquad -(n-s)(1-\theta)^{-1}$$

$$s\theta^{-1}$$

$$\frac{\partial^2}{\partial\theta^2} ll(\theta | y B) = \frac{-s}{\theta^2} - \frac{(n-s)}{(1-\theta)^2} = \frac{-s(1-\theta)^2 - \theta^2(n-s)}{\theta^2(1-\theta)^2}$$

$$-\frac{\partial^2}{\partial\theta^2} ll(\theta | y B) = \frac{s(1-\theta)^2 + (n-s)\theta^2}{\theta^2(1-\theta)^2}$$

$$-\frac{d^2}{d\theta^2}\, \ell\ell(\theta \mid y^B)\Big|_{\theta=\hat{\theta}_{MLE}} = \hat{I}(\hat{\theta})$$

$$= \frac{s\left(1-\frac{s}{n}\right)^2 + (n-s)\frac{s^2}{n^2}}{\left(\frac{s}{n}\right)^2\left(1-\frac{s}{n}\right)^2}$$

after pain

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

$$\hat{V}_{RS}(\hat{\theta}_{MLE}) = \frac{1}{\hat{I}(\hat{\theta})} = \frac{\hat{\theta}(1-\hat{\theta})}{n}$$

$$\hat{SE}_{RS}(\hat{\theta}_{MLE}) = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

---

by CLT $\quad \hat{\theta}_{MLE} \pm \Phi\left(1-\frac{\alpha}{2}\right)\frac{1}{\sqrt{\hat{I}(\hat{\theta})}}$ is an approx $100(1-\alpha)\%$ CI for θ

Decision theory

CS1

$\theta = \begin{cases} 1 & \text{if Bob is HIV+} \\ 0 & \text{is not} \end{cases}$

$\theta \in \Theta$, $\Theta = \{0, 1\}$

$Y_1 \in \mathbb{D} = \{0, 1\}$

$Y_1 = \begin{cases} 1 & \text{if ELISA blood test says HIV+} \\ 0 & \text{HIV-} \end{cases}$

$= D$

ingredients for decision

① ~~A~~ Specify action space $(\mathcal{a} \mid \mathcal{B})$ given based on problem context $C + \mathcal{B}$

ex. possible $(\mathcal{a} \mid \mathcal{B}) = \{ a_1, a_2 \}$

a_1 = (run ELISA once; if ⊕ say Bob ⊕, if ⊖ say Bob ⊖)

a_2 = (~~run~~ ELISA twice on 2 different samples of Bob's blood; declare ⊕ iff ELISA = ⊕ both times) utility function

$$u(a, \theta \mid \mathcal{B}) \in \mathbb{R}$$

$a \in (\mathcal{A} \mid \mathcal{B})$, $\theta \in \Theta$

8

without loss of generality, convention: large u better than small

one idea

choose the a^* that maximizes $u(a, \theta \mid \mathcal{B})$

this doesn't work because θ is unknown