Mr. Fisher wants

$\hat{V}_F(\hat{\theta}_{MLE}) \rightarrow$

STAT 206
3 Feb 21

wed AM
discussion
section /
lecture

$\hat{SE}_{IID}(\hat{\theta}_{MLE}) = \sqrt{\hat{V}_{IID}(\hat{\theta}_{h})}$

(1)

$$\hat{I}(\hat{\theta}_{MLE}) = \left[-\frac{\partial^2}{\partial \theta^2} \ell\ell(\theta \mid y\, B) \right]_{\theta = \hat{\theta}_{MLE}}$$

(Fisher) information provided by

$\hat{\theta}_{MLE}$ about θ

Mr. Fisher's recipe for (estimation-based) inference with parametric sampling models

(1) based on context $\mid \mathbb{P} = (\mathbb{Q}, \mathbb{C})$ (2)

commit to a single parametric sampling distribution/model $\mathbb{M}$

$$\boxed{\begin{gathered}(Y_i \mid \theta \, \mathcal{B} \, \mathbb{M}) \stackrel{IID}{\sim} p(y_i \mid \theta \, \mathcal{B} \, \mathbb{M}) \\ (i = 1, \dots, n) \qquad (\theta \in \Theta) \\ (\dim(\Theta) = 1 \text{ for now})\end{gathered}}$$

i.e., we (pretend that) we know the functional form of $p(y_i \mid \theta \, \mathcal{B} \, \mathbb{M})$

(&) the only unknown thing is θ

(Ex) (ex./ $\left(Y_i \mid \theta \, \mathcal{B} \overset{\mathbb{M}}{\left(\begin{matrix}\text{Bernoulli} \\ \text{sampling} \\ \text{model}\end{matrix}\right)}\right) \stackrel{IID}{\sim} \text{Bernoulli}(\theta)$ $\quad \Theta = (0, 1)$

i.e., $p_{Y_i}(y_i \mid \theta \, \mathcal{B} \, \mathbb{M}) = \theta^{y_i} (1 - \theta)^{1 - y_i} I(y_i = 1 \text{ or } 0)$

1 write down marginal sampling dist. of a single Y_i

2 $\underset{\sim}{Y} = (Y_1, \dots, Y_n)$

$\underset{\sim}{y} = (y_1, \dots, y_n)$

2 write down joint sampling dist. for $\underset{\sim}{Y}$:

$$p_{\underset{\sim}{Y}}(\underset{\sim}{y} \mid \theta\, \mathcal{B}\, M) = \prod_{i=1}^{n} p_{Y_i}(y_i \mid \theta\, \mathcal{B}\, M)$$

3 define likelihood function:

$$\ell(\theta \mid \underset{\sim}{y}\, \mathcal{B}\, M) = c\, p_{\underset{\sim}{Y}}(\underset{\sim}{y} \mid \theta\, \mathcal{B}\, M)$$

for arbitrary $(c > 0)$, & $\boxed{\text{plot it}}$

4 so $\ell(\theta \mid \underset{\sim}{y}\, \mathcal{B}\, M) = \prod_{i=1}^{n} p_{Y_i}(y_i \mid \theta\, \mathcal{B}\, M)$

for all $y_i = \{1 \text{ or } 0\}$

ex. $p_{\underset{\sim}{Y}}(\underset{\sim}{y} \mid \theta\, \mathcal{B}\, M) = \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i}$

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

$\boxed{4}$ define log likelihood function:

$$\ell\ell(\theta \mid y\, \mathcal{B}\, \mathbb{M}) = \log\left[\ell(\theta \mid y\, \mathcal{B}\, \mathbb{M})\right]$$

$$= \sum_{i=1}^{n} \log p_{Y_i}(y_i \mid \theta\, \mathcal{B}\, \mathbb{M})$$

& plot it

---

ex. $\ell\ell(\theta \mid y\, \mathcal{B}\, \mathbb{M}) = s \log \theta +$

(Bernoulli) $(n-s)\log(1-\theta)$

$\boxed{5}$ find the θ that maximizes the log likelihood function:

$$\hat{\theta}_{MLE} = \underset{\theta \in \Theta}{\operatorname{argmax}}\ \ell\ell(\theta \mid y\, \mathcal{B}\, \mathbb{M})$$

---

do this analytically, if you can; (set $\frac{\partial}{\partial\theta}\ell\ell(\theta \mid y\, \mathcal{B}\, \mathbb{M}) = 0$ & solve for θ) otherwise do it numerically ex. $\binom{\text{quiz}}{2}$

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

led immediately to

$$\hat{\theta}_{MLE} = \frac{s}{n} = \frac{s}{n}(\text{data})$$

random variable

6 $$\widehat{SE}(\hat{\theta}_{MLE}) = \sqrt{\hat{V}(\hat{\theta}_{MLE})}$$

$$= \sqrt{\left[I(\hat{\theta}_{MLE})\right]^{-1}}, \quad I(\hat{\theta}_{MLE}) = O(n)$$

where $$I(\hat{\theta}_{MLE}) = \left[-\frac{\partial^2}{\partial\theta^2}\, \ell\ell(\theta \mid y\, \mathcal{B}\, \mathcal{M})\right]_{\theta = \hat{\theta}_{MLE}}$$

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

$$= O\left(\frac{1}{\sqrt{n}}\right)$$

$\boxed{7}$ construct an approximate (large-n) $100(1-\alpha)\%$ likelihood-based confidence interval:

$$\hat{\theta}_{MLE} \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\left[I(\hat{\theta}_{MLE})\right]^{-1}}.$$

Bayesian: nasty integrals (harder)

likelihood story: differentiation (easier)
