CS2: AMI mortality

STAT 206

28 Jan 20

$n = 403$ patients

$s = 72$ died

in 1^{st} 30 days

Neyman: $\hat{\theta} = \frac{s}{n}$ is "reasonable" estimate of θ

sample mean

population mean

pop

sample

$\begin{bmatrix} 1s \\ \& \\ 0s \end{bmatrix}$ IID

1 = Y

0 = N

died?

$\begin{array}{c} Y_1 \\ \vdots \\ Y_n \end{array} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ $n = 403$

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

$= \hat{\theta} = \frac{s}{n} \doteq 0.18$

mean $\theta = ?$

unknown constant

The entire diagram can be expressed in math as follows:

(RS = repeated sampling)

$(Y_i \mid \theta) \overset{IID}{\underset{RS}{\sim}} \text{Bernoulli}(\theta)$

$(i = 1, \ldots, n)$ $\circledast$

random variables (sampling dist.)

frequentist model in CS2

Lecture notes part 8

Fisher (1925) ②

① given θ, write down marginal dist. (PMF or PDF) for Y_i: $p(y_i \mid \theta) = P(Y_i = y_i \mid \theta)$

for $y_i = 0$ or 1

$$p(y_i \mid \theta) = \theta^{y_i}(1-\theta)^{1-y_i} = \begin{cases} \theta & \text{for } y_i = 1 \\ 1-\theta & y_i = 0 \\ 0 & \text{else} \end{cases}$$

computer science

② write down joint dist. of $\underset{\sim}{Y} = (Y_1, \dots, Y_n)$ & simplify

joint = product of marginals

$$p_{\underset{\sim}{Y}}(\underset{\sim}{y} \mid \theta) = P(Y_1 = y_1 \text{ and } Y_2 = y_2, \dots, Y_n = y_n)$$

$$= \prod_{i=1}^{n} P(Y_i = y_i \mid \theta)$$

comment: unrealistic to pretend ③

θ constant for all pts., but

$(I_i \mid \theta_i) \overset{I}{\sim} \text{Bernoulli}(\theta_i)$ is
$(i = 1, \dots, n)$ impossible to fit

(# data values = n, # unknowns = n) $\left(\begin{array}{l} y_6 = 1 \\ \hat{\theta}_6 = 1 \end{array} \right)$

$\begin{array}{l} y_{21} = 0 \\ \hat{\theta}_{21} = 0 \end{array}$

$$p_{\underset{\sim}{I}}(y \mid \theta) = \prod_{i=1}^{n} p_{I_i}(y_i \mid \theta)$$

$$= \prod_{i=1}^{n} \left[\theta^{y_i} (1-\theta)^{1-y_i} \right] =$$

$$= \left[\theta^{y_1} (1-\theta)^{1-y_1} \right] \cdot \left[\theta^{y_2} (1-\theta)^{1-y_2} \right] \cdots \left[\theta^{y_n} (1-\theta)^{1-y_n} \right]$$

$$\cdot\ \theta^{y_1 + y_2 + \dots + y_n} (1-\theta)^{(1-y_1)+(1-y_2)+\dots+(1-y_n)}$$

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

③ Laplace / Fisher

$p_{\underset{\sim}{Y}}(y \mid \theta)$ looks like a function of y for fixed θ; but let's just reverse our thinking: regard it instead as fn of θ for fixed y:

$$\ell(\theta \mid y) \triangleq c\, p_{\underset{\sim}{Y}}(y \mid \theta)$$

(the) <u>likelihood</u> function for θ $\qquad$ (c real > 0)

here

$$\ell(\theta \mid \underset{\sim}{y}) = c\, \theta^{s} (1-\theta)^{n-s}$$

$$= \ell(\theta \mid s) \qquad s(\underset{\sim}{y}) = \sum_{i=1}^{n} y_i$$

Def:

If $\ell(\theta \mid \underset{\sim}{y})$ depends on $\underset{\sim}{y}$ only through some function $s(\underset{\sim}{y})$, $s(\underset{\sim}{y})$ is said to be sufficient for θ

"s is a suff. statistic for θ"

superb dimensionality-reduction in data space from n to 1

$\dim(\underset{\sim}{y}) = n$

$\dim(\theta) = 1$

$\dim(s) = 1$

usually if a suff. stat. s exists, $\dim(\underset{\sim}{s}) = \dim(\underset{\sim}{\theta})$

uncertainty about sampling list. ↔ <u>model</u> uncertainty

Fisher claimed: no model unc.

in practice: model "walks in door with client": usually FALSE

---

total <u>info</u> in y

info in y to check whether your sampling distn is good

info in y about θ

model diagnostic info

$\underline{y}_1 = 1\ 0\ 1\ 1\ 0\ 0\ 0\ 1\ 0\ 1 \to s = 5$

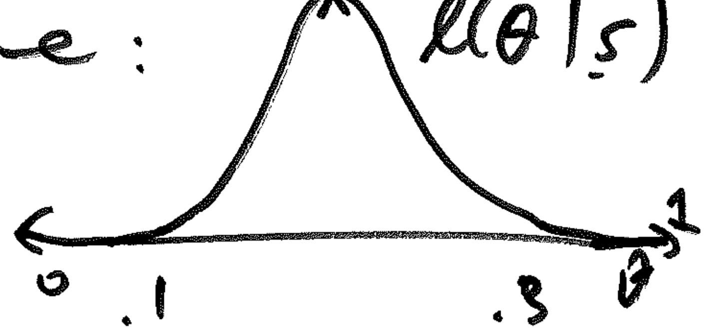
$\underline{y}_2 = 1\ 1\ 1\ 1\ 1\ 0\ 0\ 0\ 0\ 0 \to s = 5$

$\underline{y}_3 = 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0 \to s = 5$

---

(4) Fisher: trust me:

find θ that maximize $\ell(\theta \mid \underline{y})$:



$$\hat{\theta}_{MLE} = \underset{\theta \in (0,1)}{\arg\max}\ \ell(\theta \mid \underline{y})$$

maximum likelihood estimate

$$\frac{d}{d\theta}\ell(\theta \mid \underline{y}) = \frac{d}{d\theta}\left[\theta^s (1-\theta)^{n-s}\right]$$

Def. $\ell\ell(\theta \mid \mathbf{y}) \triangleq \log \ell(\theta \mid \mathbf{y})$

log likelihood

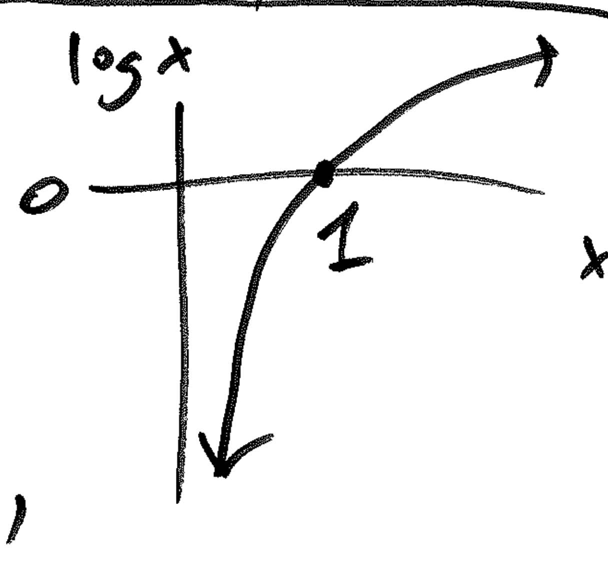
here

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

---

to find $\hat{\theta}_{MLE}$, can diff. either $\ell(\theta \mid \mathbf{y})$ or $\ell\ell(\theta \mid \mathbf{y})$ & set to 0

log is strictly increasing (inequality preserving)



$$\frac{d}{d\theta}\, \ell\ell(\theta|s) = \frac{s}{\theta} + \frac{n-s}{1-\theta}(-1)$$

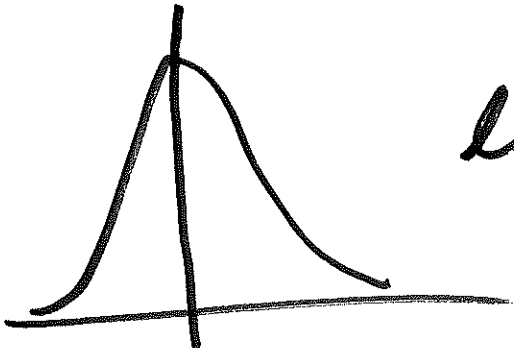
$$= \frac{s}{\theta} - \frac{n-s}{1-\theta}$$

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

$$= \frac{s - \cancel{s\theta} - n\theta + \cancel{s\theta}}{\theta(1-\theta)} = 0$$

iff $\theta = \hat{\theta}_{MLE} = \dfrac{s}{n} = \bar{y}$ (Neyman)

$$= \frac{72}{403} \doteq .179$$

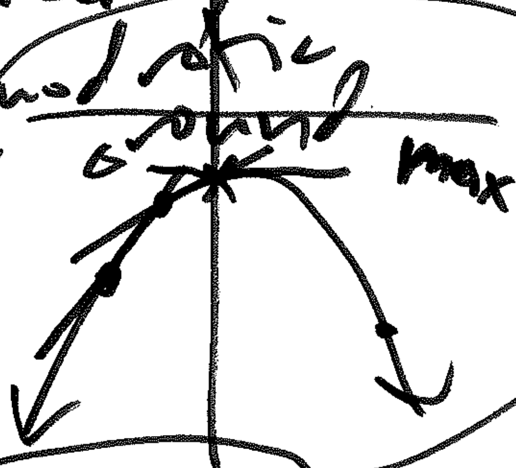


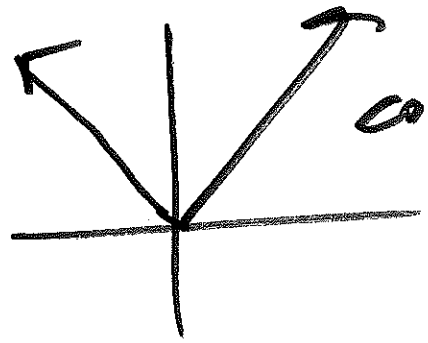
$\ell(\theta \mid x)$

$c_1 e^{-c_2(\theta - c_3)^2}$

$\ell\ell(\theta \mid x)$

locally quadratic around





convex

Newton/Raphson

concave

convex optimization

↓

$$\arg\max_{\theta} \ell\ell = \arg\min_{\theta} (-\ell\ell)$$

(gradient descent)

$$\hat{SE}_{RS}(\hat{\theta}_{MLE}) = ?$$

here: $\hat{\theta}_{MLE} = \bar{y}$

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

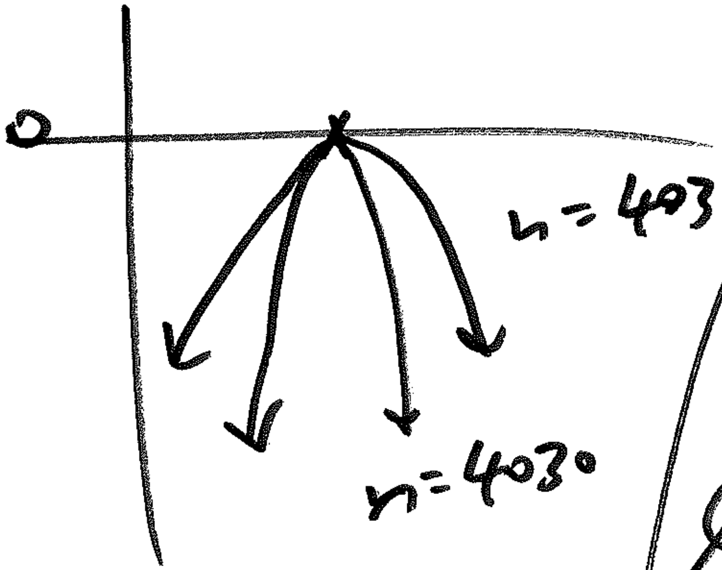
$$\therefore \hat{SE}_{RS}(\hat{\theta}_{MLE}) = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

$$\doteq 0.019$$

high uncertainty = low <u>information</u> about θ from $\hat{\theta}_{MLE}$

=

high RS variance of $\hat{\theta}_{MLE}$



$$\ell(\theta \mid s) = c_1\, \theta^{s} (1-\theta)^{n-s}$$

$$\ell\ell(\theta \mid s) = \underline{\underline{c_2}} + s \log \theta + (n-s)\log(1-\theta)$$

as $n \uparrow$

$$\left[\frac{d^2}{d\theta^2}\, \ell\ell(\theta \mid y)\right]_{\theta = \hat{\theta}_{MLE}} \uparrow$$

$$\parallel$$

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

as $n \uparrow$

info $\uparrow$

$\boxed{\text{Def.}}$ Fisher information of the $\hat{\theta}_{MLE}$:

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

here $\hat{I}(\hat{\theta}_{MLE}) = \dfrac{n}{\hat{\theta}(1-\hat{\theta})}$

Fisher showed $= O(n)$

$$\hat{V}_{RS}(\hat{\theta}_{MLE}) \doteq \left[\hat{I}(\hat{\theta}_{MLE}) \right]^{-1}$$

so

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

$$\hat{SE}_{RS}(\hat{\theta}_{MLE}) \doteq \sqrt{\left[\hat{I}(\hat{\theta}_{MLE})\right]^{-1}}$$

Fisher showed 3 things: in nice problems

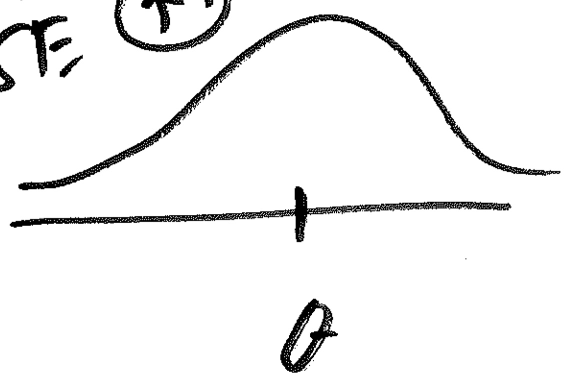
① $E_{RS}(\hat{\theta}_{MLE}) = \theta + \underline{\underline{O}}\left(\frac{1}{n}\right)$

i.e. MLE may be biased but bias $\to 0$ at $\frac{1}{n}$ rate (fast)

② $\hat{SE}_{RS}(\hat{\theta}_{MLE}) \doteq \sqrt{\left(\hat{I}(\hat{\theta}_{MLE})\right)^{-1}}$ = ⊛

③ for large n $\hat{\theta}_{MLE} \overset{RS}{\underset{\cdot}{\sim}} N(\theta, \hat{I}(\hat{\theta})^{-1})$

$\widehat{SE}$ **



RS dist of $\hat{\theta}_{MLE}$ (large n)

→ 95% likelihood-based interval for θ:

$$\hat{\theta}_{MLE} \pm 1.96\, \widehat{SE}_{RS}(\hat{\theta}_{MLE})$$
