AMS 206
31 Jan 19

today: Fisher

next time: Bayes

$(Y_i \mid \theta\, \mathcal{B}) \overset{IID}{\sim} N(\mu, \sigma^2)$
$(i = 1, \dots, n)$

$\theta = (\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}^+$

CS2 Fisher's algorithm for frequentist inference

① (same as Neyman) $\mathbb{C} \longrightarrow p(y_i \mid \theta\, \mathcal{B})$

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

ex. ((CS2) $(Y_i \mid \theta\, \mathcal{B}) \overset{IID}{\sim}$ Bernoulli(θ) marginal sampling dist
$(i = 1, \dots, n)$

② use ① to write down the joint sampling dist. for $(Y_1, \dots, Y_n) = Y$ | CS2 Y_i discrete

$p(Y_1 = y_1$ and $Y_2 = y_2, \dots, Y_n = y_n \mid \theta\, \mathcal{B})$

$$\underbrace{P(\mathcal{I}_i = y_i \mid \theta\,\mathcal{B})}_{\text{Bernoulli}} = \begin{cases} \theta & \text{if } y_i = 1 \\ 1-\theta & y_i = 0 \end{cases}$$

$$= \theta^{y_i}(1-\theta)^{1-y_i}$$

$$= p(y_i \mid \theta\,\mathcal{B})$$

$$P(\mathcal{I}_1 = y_1, \dots, \mathcal{I}_n = y_n \mid \theta\,\mathcal{B}) \overset{\text{IID}}{=} \prod_{i=1}^{n} p(y_i \mid \theta\,\mathcal{B})$$

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

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

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

$$p(\mathcal{I}_1 = 1, \dots, \mathcal{I}_n = y_n \mid \theta\,\mathcal{B}) = \theta^{s}(1-\theta)^{n-s}$$

$y = (y_1, \dots, y_n)$ (Laplace) (1785)

③ Define the likelihood function:

f'n of θ for fixed y (any # > 0) — f'n of y for fixed θ

$$\ell(\theta \mid y\,\mathcal{B}) \triangleq c \cdot P(\mathcal{I}_1 = y_1, \dots, \mathcal{I}_n = y_n \mid \theta\,\mathcal{B})$$

ex. 4.2) $P(Y_1 = y_1, \ldots, Y_n = y_n \mid \theta \mathcal{B}) = \theta^s (1-\theta)^{n-s}$

$$\ell(\theta \mid \underline{y}\, \mathcal{B}) \triangleq c\, \theta^s (1-\theta)^{n-s}$$

$s = \sum_{i=1}^{n} y_i$

data-compression

dimensionality reduction from

$y = (y_1, \ldots, y_n)$ (n-dimensional)

to s (1-dimensional) (1925)

terabyte 10^{12} bytes

850

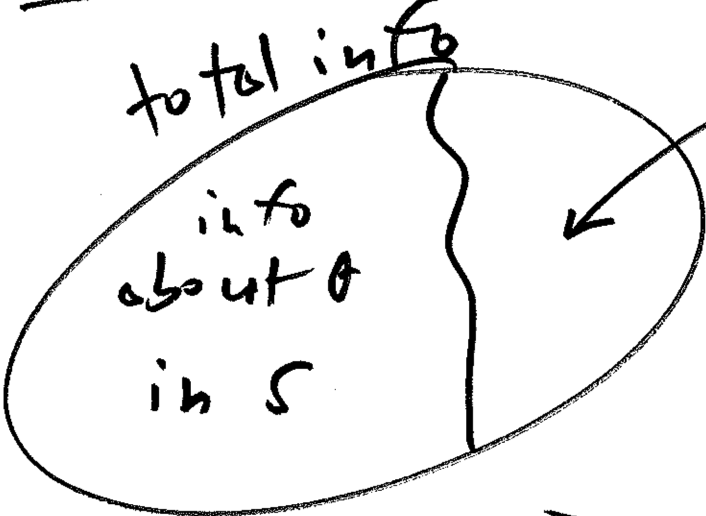
1,000,000,000,000

m G

Def. if $\ell(\theta \mid y\, \mathcal{B})$ depends on y only through a (lower-dimensional) (function / summary of y) s,

s is said to be (a) sufficient (statistic) for θ in this sampling model

Fisher: "throw y away & keep s" (when suff. stats exists) | wrong



info allowing you to check/criticize your sampling dist.

(uncertainty about sampling dist.) $\longleftrightarrow$ (model uncertainty)

Fisher — horseshit

$$\mathbb{P} = (\mathcal{Q}, \mathcal{C}) \xrightarrow{\text{unique}} p(y \mid \theta, \mathcal{B})$$

④ find θ to maximize $\ell(\theta \mid y\, \mathcal{B})$:

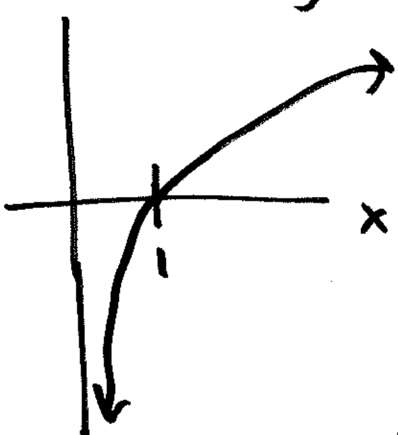
result is $\hat{\theta}_{MLE}$ — maximum likelihood estimate

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \ell(\theta \mid y\, \mathcal{B})$$

4' maximize log likelihood fn:

$$\hat{\theta}_{MLE} = \arg\max \ell\ell(\theta \mid y\,\mathcal{B})$$

$f(x) = \ln x$ is monotonic increasing



Def. $\ell\ell(\theta \mid y\,\mathcal{B}) = \log \ell(\theta \mid y\,\mathcal{B})$

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

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

$$\frac{d}{d\theta}\ell\ell(\theta \mid y\,\mathcal{B}) = \frac{s}{\theta} + \frac{n-s}{1-\theta}(-1) = 0$$

$$= \frac{s(1-\theta) - (n-s)\theta}{\theta(1-\theta)} = 0 \text{ iff}$$

$$s - \cancel{s\theta} - n\theta + \cancel{s\theta} = 0$$

$$\theta = \hat{\theta}_{MLE} = \frac{s}{n} = \bar{y}$$

5 Fisher (frequentist) wants

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

he wants a formula relating

$\hat{V}_{RS}(\hat{\theta}_{MLE})$ to $\ell\ell(\theta \mid y\,\mathcal{B})$