AMS 206
1 Feb 18

this time: maximum likelihood

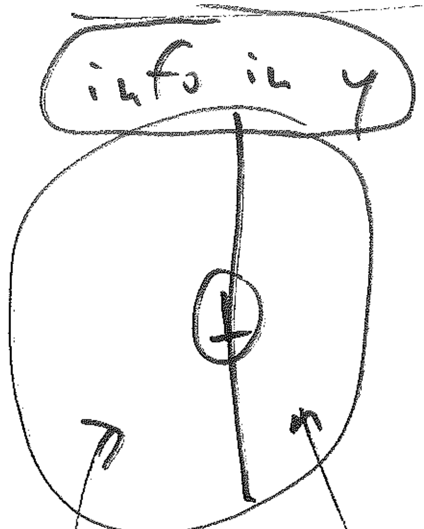
next time: & Bayes for $\theta \in (0,1)$

read: ① ch. 1-4, app. A, B
Ⓖ ch 1-2

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

$$= \ell(\theta \mid s\, \underline{\underline{B}})$$

Fisher: find your suff. stat., throw away data vector y



$n = 10$

	y	s
1	1 1 0 1 0 1 0 0 1 0	5
2	0 0 0 0 0 1 1 1 1 1	5
3	1 0 1 0 1 0 1 0 1 0	5

info in y above & beyond suff. stat useful for model checking/criticism

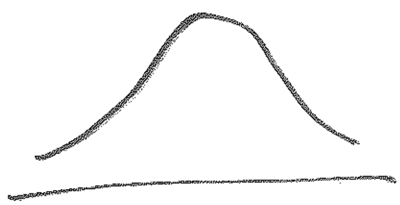
info in suff. stat. about unknown θ in your sampling model

$$\ell(\theta \mid y) \doteq \text{Normal} = c_1 e^{-c_2(\theta - c_3)^2}$$

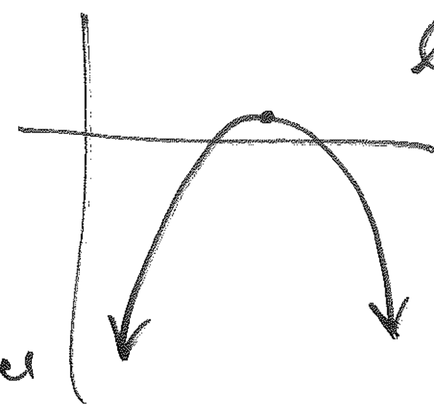
$(c_2 > 0)$

$$\ell\ell(\theta \mid y) \doteq \log c_1 - c_2\left(\theta - c_3\right)^2$$

log likelihood f'n



$\ell(\theta \mid y)$



$\ell\ell(\theta \mid y)$

IID Bernoulli sampling model

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

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

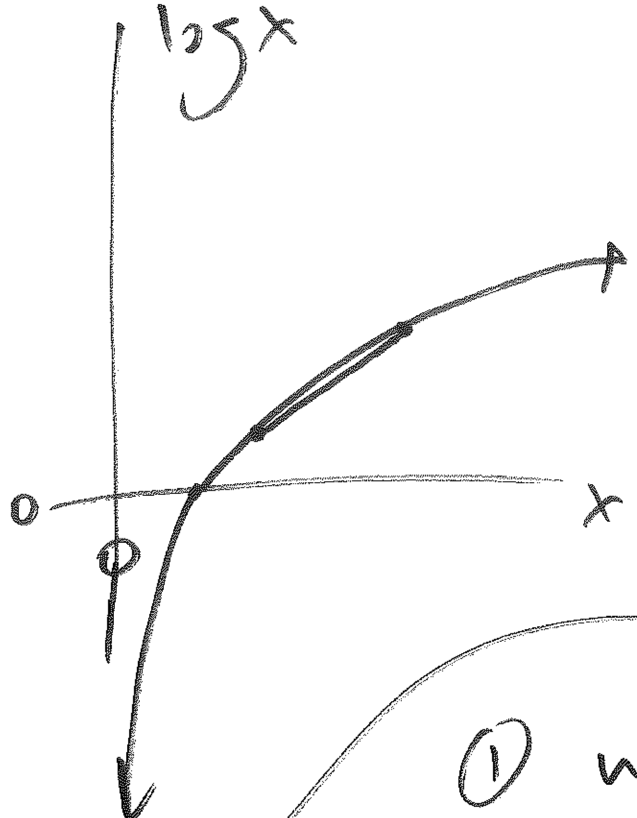
Fisher: trust me: $\hat{\theta}_{MLE}$ ← Fisher: this is a good estimate of θ

maximum likelihood estimate

is that θ value that maximizes the lik. fn.

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}}\ \ell(\theta \mid sBB)$$

$$\boxed{= \underset{\theta}{\operatorname{argmax}}\ \ell\ell(\theta \mid sBB)}$$



concave
↓
strictly increasing

Convex ∀

Fisher's MLE algorithm

① write down joint sampling dist. of data random variables $p(Y \mid \theta)$

② think of ① as f'n of θ for fixed y, & mult. it by $(c > 0)$, this yields lik. f'n; take logs to get $\ell\ell(\theta \mid y)$

$\ell(\theta \mid y)$ ←

③ find θ to max. $\ell\ell(\theta|y)$; ④

this can often (but not always) be done as follows:

$$\boxed{\frac{d}{d\theta}\,\ell\ell(\theta|y) = 0 \quad \text{solve for } \theta}$$

$\ell\ell(\theta|y) = \ell\ell(\theta|s) =$

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

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

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

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

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

$$\therefore \quad \hat{\theta}_{MLE}: \quad s - n\theta = 0$$

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

---

method of moments

Karl Pearson (1880s)

equate sample moments to population moments

$$E(Y_i \mid \theta) = \theta \longrightarrow \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i^{1}$$

$$\bar{Y} = \hat{\theta}_{MLE} = \hat{\theta}_{MoM} = \hat{\theta}_{\text{common sense (Neyman)}}$$

AmII

all pts similar to sample in all relevant ways

Fisher

sample of

Re-observed AmI pts

repeated sampling

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

die?

1s

&

0s

mean θ = ?

(actual)

like

IID

die?

y_1

y_n

1s

&

0s

n

MLE $\hat{\theta}_{MLE} = \bar{y} = 0.180$

hyp.

IID

1s

&

0s

n = 400

MLE $\hat{\theta}_{MLE} = \bar{y} = 0.173$

⋮

0.18

0.173

⋮

m → ∞

long run mean $E(\hat{\theta}_{MLE}) = \theta + O(\frac{1}{n})$

long run SD $\hat{SE}(\hat{\theta}_{MLE}) = \sqrt{\hat{I}^{-1}}$

long run dist

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

repeated sampling

one "good" estimator criterion for frequentists: <u>unbiasedness</u>

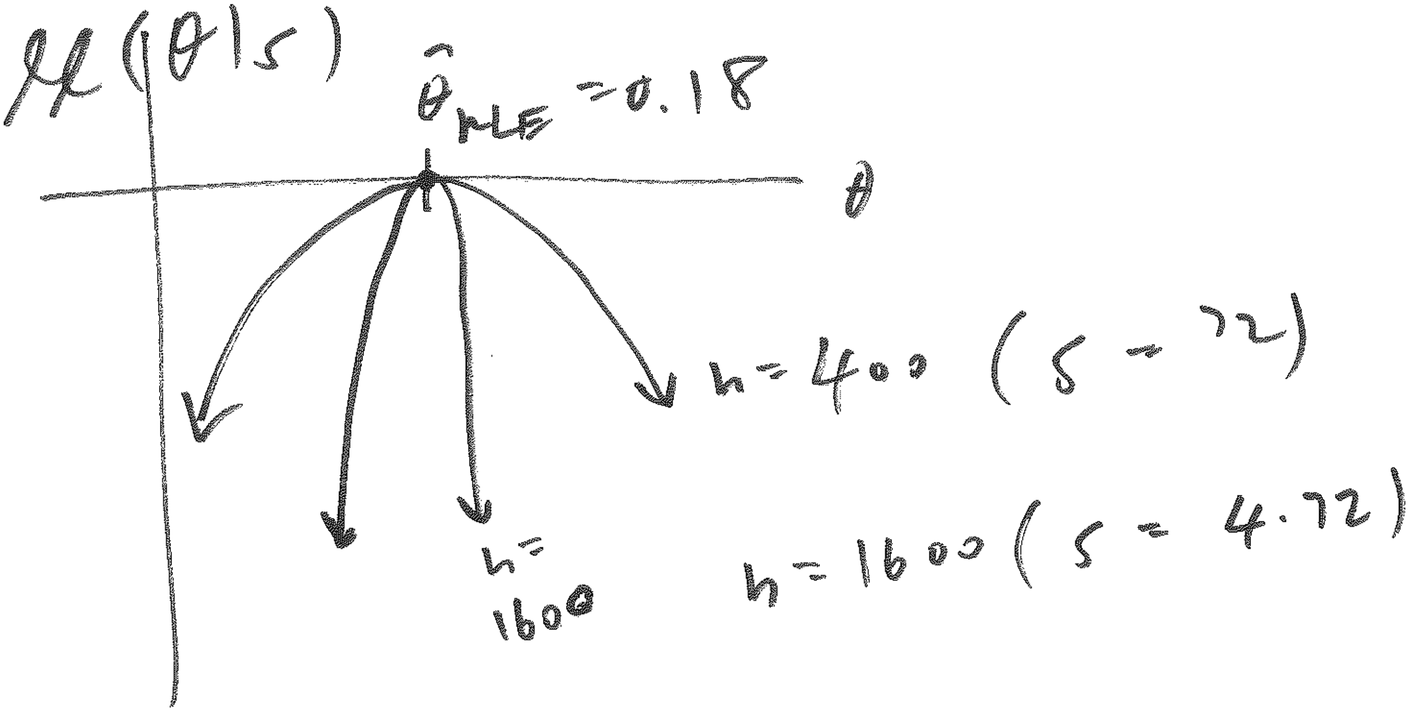
<u>def.</u> $\hat{\theta}$ is <u>unbiased</u> for θ iff $\boxed{E_{RS}(\hat{\theta}) = \theta}$

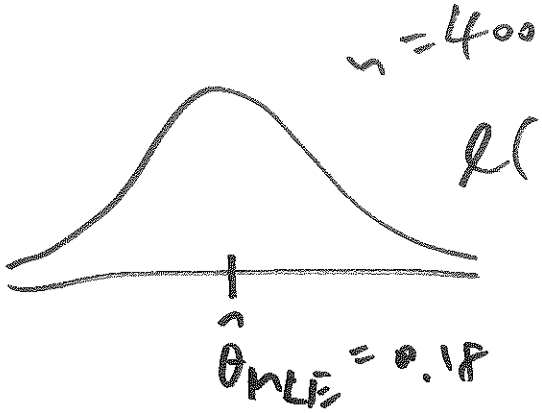
Fisher facts

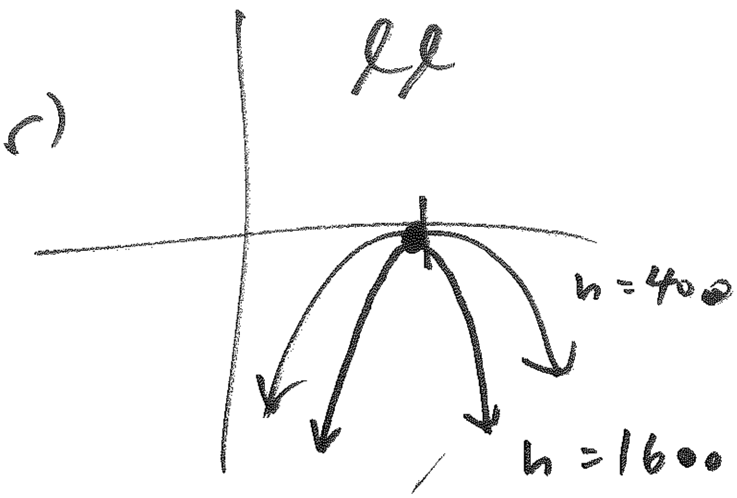
1. $E_{RS}(\hat{\theta}_{MLE,n}) = \theta + \underline{O}\left(\frac{1}{n}\right)$

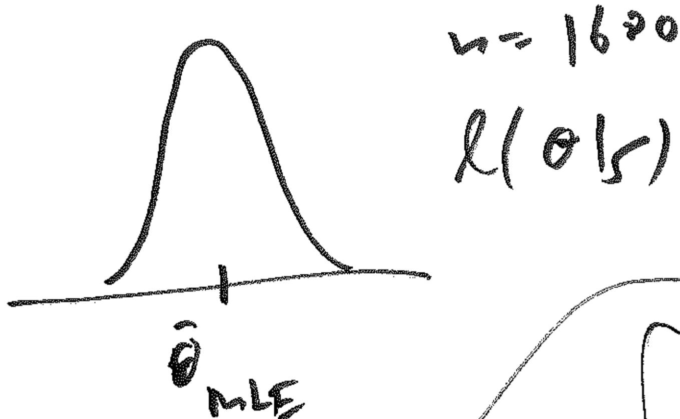
MLE based on n obs.

2. $\hat{V}_{RS}(\hat{\theta}_{MLE,n}) = ?$









as $n \uparrow$

$$\hat{I}(\hat{\theta}_{MLE}) = \left[-\frac{d^2}{d\theta^2} ll(\theta|s)\right]_{\theta=\hat{\theta}_{MLE}} \uparrow$$

observed (Fisher) information

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

$$\frac{d^2}{d\theta^2} \ell\ell(\theta \mid s) = -\frac{s}{\theta^2} + \frac{(n-s)}{(1-\theta)^2}\cancel{(-1)}$$

$$-\left[\quad \right] = \frac{s}{\theta^2} - \frac{n-s}{(1-\theta)^2}$$

$$-\left[\quad \right]_{\theta = \frac{s}{n}} = \frac{s}{\left(\frac{s}{n}\right)^2} - \frac{(n-s)}{\left(1-\frac{s}{n}\right)^2}$$

$$\hat{I}(\hat{\theta}_{MLE}) = \frac{n^3}{(n-s)s}$$

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

Fisher facts, ② $\hat{I}(\hat{\theta}_{MLE}, n) = \underline{\underline{O}}(n)$

for large n

③ $\hat{V}_{RS}(\hat{\theta}_{MLE}) \doteq \dfrac{1}{\hat{I}(\hat{\theta}_{MLE})} = \hat{I}^{-1}(\hat{\theta}_{MLE})$

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

---

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

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

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

(same as Neyman)

(4) for large n

$$\hat{\theta}_{MLE} \underset{RS}{\dot{\sim}} N\left[\theta, \hat{I}^{-1}(\hat{\theta})\right]$$

(is approx. distributed or, in repeated sampling)

(5) for large n an approx. 95% CI for θ looks like

$$\hat{\theta}_{MLE} \pm 1.96 \, \underbrace{\hat{SE}(\hat{\theta}_{MLE})}_{\sqrt{\hat{I}^{-1}(\hat{\theta}_{MLE})}}$$

caveat: only works accurately for large n

Q: how large?

A: simulation study