not model selection

Standard Frequentist flavor — throughout most of course, essentially

$x = \langle 0, 0, 1, 0, 1, 0 \rangle$ and assume i.i.d. then you pick an F, but θ unknown.

Do not know θ (mother nature)

Figuring out θ is the goal of statistical inference.

There are three main objectives:

(1) point estimation (best guess of θ)

(2) confidence set... a set of "likely" θ's.

(3) Theory testing (θ value testing)

(Hypothesis testing)

$$p(001010; \theta) = \prod_{i=1}^{6} p(x_i; \theta) = \prod_{i=1}^{6} \theta^{x_i}(1-\theta)^{1-x_i}$$

$p(001010; \theta = \frac{1}{2}) = 0.5^6 = 0.0156$

$p(001010; \theta = \frac{1}{4}) = (\frac{1}{4})^2(\frac{3}{4})^4 = 0.0198$

more likely b/c pdf matches better

$p(x_1, \dots, x_n; \theta) = \mathcal{L}(\theta; x_1, \dots, x_n)$

changes — joint density — locked — changes — likelihood — locked

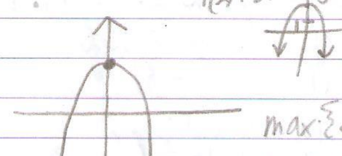
Mathematically same, conceptually different (perspective)

$$\underset{\theta \in \Theta}{\operatorname{argmax}} \{\mathcal{L}(\theta; x_1, \dots, x_n)\} = \hat{\theta}_{MLE}$$

maximum likelihood (#) estimate / estimator (r.v.)

$f(x)+2$

$f(x) = 1 - x^2$



$\max\{f(x)\} = 1$

$\operatorname{argmax}\{f(x)\} = 0$

If you take a 1:1 increasing function of $\mathcal{L}$, $\hat{\theta}_{MLE}$ won't change.

$$\ell(\theta; x_1, \dots, x_n) := \ln(\mathcal{L}(\theta; x_1, \dots, x_n)) \text{ "log-likelihood"}$$

2/2/17

$$P(X_1, \dots, X_n; \theta) = \mathcal{L}(\theta; X_1, \dots, X_n)$$

$$\hat{\theta}_{MLE} := \underset{\theta \in \Theta}{\operatorname{argmax}} \{\mathcal{L}(\theta; X_1, \dots, X_n)\}$$

$$= \underset{\theta \in \Theta}{\operatorname{argmax}} \{\ell(\theta; x_1, \dots, x_n)\}$$

Frequentist approach

$X_1, \dots, X_6 \overset{i.i.d.}{\sim} Bern(\theta)$

$0, 0, 0, 0, 1, 0$ $\quad \ell(\theta; x) = \ln\left(\prod_{i=1}^{6} \theta^{x_i}(1-\theta)^{1-x_i}\right)$

$= \sum_{i=1}^{6} \ln(\theta^{x_i}(1-\theta)^{1-x_i}) =$

$\sum_{i=1}^{6} x_i \ln(\theta) + (1-x_i)\ln(1-\theta) = \ln\theta \sum x_i + (6 - \sum x_i)\ln(1-\theta)$

$= \ln(\theta) \cdot 6\bar{x} + 6 \cdot 6\bar{x} \ln(1-\theta) = 6(\ln(\theta)\bar{x} + (1-\bar{x})\ln(1-\theta))$

$\bar{x} := \frac{1}{n}\sum x_i \Rightarrow \sum x_i = n\bar{x}$

$\frac{d}{d\theta}\left[\quad\right] = 6\underbrace{\left(\frac{\bar{x}}{\theta} - \frac{1-\bar{x}}{1-\theta}\right)}_{\text{independent of } n} \overset{set}{=} 0$

$(1-\theta)\bar{x} - \theta(1-\bar{x}) = 0$

$\bar{x} - \cancel{\theta\bar{x}} - \theta + \cancel{\theta\bar{x}} = 0 \Rightarrow \boxed{\hat{\theta}_{MLE} = \bar{x}}$

R.V. $\hat{\theta}_{MLE} = \bar{X}$ max lik. estimator

$\hat{\theta}_{MLE} = \bar{x}$ max lik. estimate

number

$\prod \Rightarrow \sum$

(easier)

$X_1, \dots, X_n \overset{i.i.d.}{\sim} Geom(\theta) := (1-\theta)^x \theta$

$X :=$ # of failure before the stopping success

$Supp(X) = \{0, 1, \dots\} = \mathbb{N}_0$

Parameter space of θ $\rightarrow \Theta = (0, 1)$

$P(X_1, \dots, X_n; \theta) = \mathcal{L}(\theta; x_1, \dots, x_n) = \prod_{i=1}^{n} (1-\theta)^{x_i}\theta$

$\ell(\theta; x) = \sum \ln(1-\theta)^{x_i}\theta$

$= \ln(1-\theta)\sum x_i + n\ln(\theta)$

$\ell'(\theta; x) = \frac{n}{\theta} - \frac{n\bar{x}}{1-\theta} \overset{set}{=} 0 \Rightarrow \frac{1}{\theta} = \frac{\bar{x}}{1-\theta} \Rightarrow$

$\frac{1}{\theta} - 1 = \bar{x} \Rightarrow \hat{\theta}_{MLE} = \frac{1}{\bar{x}+1}$

MLE's have nice properties

(1) Consistency $\hat{\theta}_{MLE} \xrightarrow{P} \theta$

$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P(|\hat{\theta}_{MLE} - \theta| \geq \varepsilon) = 0$

0.001

Confidence Interval, Hypothesis Interval
$\hat{\theta} \pm 2SE$ (-ish)

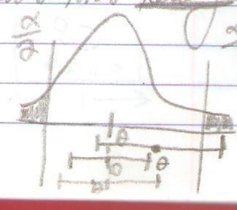
② Asymptotic Normality

$$\hat{\theta}_{MLE} \xrightarrow{d} N(\theta, SE[\hat{\theta}_{MLE}]^2)$$

③ Efficiency
$\hat{\theta}_{MLE}$ has the lowest std. error theoretically possible.

0.6 ± 0.01
0.6 ± 17

Inference w/ MLE's
① Pt. est. $\hat{\theta}_{MLE}$
② Conf. set $\left(I_{\theta, 1-\alpha} := [\hat{\theta}_{MLE} \pm Z_{\frac{\alpha}{2}} SE[\hat{\theta}_{MLE}]]\right.$
③ Hypothesis Testing
$H_0: \theta = \theta_0$
$H_1: \theta \neq \theta_0$
Retainment Region $= [\theta_0 \pm Z_{\frac{\alpha}{2}} SE[\hat{\theta}_{MLE}]]$
Observe data $\Rightarrow$ Pick $F \Rightarrow$ do inference w/ your MLE
assume have right model
Problems w/ the MLE approach
Problems
F = Bernoulli
① $X = 0, 0, 0$
$\hat{\theta}_{MLE} = \bar{x} = 0 \quad SE[\hat{\theta}_{MLE}] = \sqrt{\hat{\theta}_{MLE}(1-\hat{\theta}_{MLE})} = 0$
Tells us nothing; no inference
Needs at least one "1" in Bernoulli case
② What if I have prior knowledge about Θ?
I can't use it.
Can only use available data, not prior knowledge
③ Frequentist Confidence Interval Interpretation
$(I_{\theta, 95}) = [0.42, 0.47]$. What does this really mean?
(a) If experiment is repeated "many" times, then you cover (θ is itself interval) $1-\alpha$ proportion of the time.



KNOW BAYES RULE

b) **Before** I do my experiment $P(\theta \in CI) = 1-\alpha$
Largely Unsatisfactory

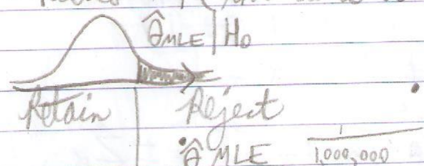
$CI_{\theta,\, 1-\alpha}$

confidence interval, param of interest, coverage level

③ 3 dec ⇒ (a) H_0 Retain
(b) H_0 Reject

p-values := P(this data or "more extreme" | H_0 true)

$\hat{\theta}_{MLE} | H_0$



p-value is 1%
Care about extreme data b/c indicates strength of rejection

Want to know: $P(H_0 | x)$?

$x = 001010$ $\hat{\theta} = \frac{1}{3}$

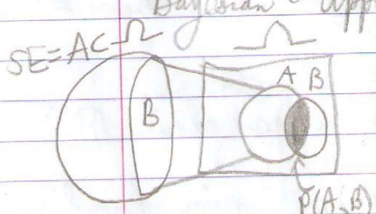
$CI_{\theta,\, 95\%} = \left[\frac{1}{3} \pm 2\sqrt{\frac{\frac{1}{3}\cdot\frac{2}{3}}{6}}\right] = [-0.60, 1.26]$

Frequentist MLE cannot have #s < 0 or > 1

MLE Std. Error; "2" is for normal distribution; 95% interval
it has not converged

Normal asymptotics didn't kick in yet by $n=6$!

Bayesian Approach to Estimation



A: smoking
B: lung cancer

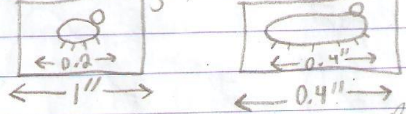
$P(A) = 0.2$
$P(B) = 0.06$
$P(A,B) = 0.036$

Cond. Prob.
$P(B|A)$
given

$P(A,B) \propto P(B|A)$

$\Rightarrow P(A,B) = c\,P(B|A)$

Zoom = $\frac{1''}{0.5''}$ = previous area / covered area



new size: $P(B|A) = \frac{P(\Omega)}{P(A)} P(A,B)$ old size $= \frac{P(AB)}{P(A)}$

Bayes Rule

$P(A) = P(A,B) + P(A,B^c) = P(A|B)P(B) + P(A|B^c)P(B^c)$

Law of Total Prob.

$$\Rightarrow P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

Form of Bayes Thm.

Posterior Prob. $\rightarrow P(B|A) = \dfrac{P(A|B)P(B)}{P(A)}$ ← prior prob.

$P(A)$ ← complicated

target of estimation

evidence/data

likelihood prob. of data

$$P(B) \xrightarrow{A\ (\text{data})} P(B|A)$$

Bayesian Conditionalism