Lec 3 2/7/17 Math 341

Law of total prob $B_1, \dots, B_k$ mut. excl. coll. exh.

$$\Rightarrow P(A) = \sum_{i=1}^{k} P(AB_i) = \sum_{i=1}^{k} P(A|B_i)\,P(B_i)$$

likelihood / prob of data

prior on param of interest

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)} = \frac{P(A|B)\,P(B)}{\sum_{i=1}^{k} P(A|B_i)\,P(B_i)}$$

$P(A) \leftarrow ?$

Bayes Thm

posterior on param of interest

$$P(B) \xrightarrow[]{A} P(B|A)$$

data/evidence

Your answer gets better with data.

Another way to think about prob:

$$\text{Odds}(A) := \frac{P(A)}{P(A^c)} = \frac{P(A)}{1-P(A^c)}$$

assume $P(A) \in (0,1)$ i.e. $\neq 0, \neq 1$

Odds = 4 "4:1" i.e. 4 times out of 5 it will occur on avg.

Range? $(0, \infty)$

$$\text{OddsA}(A) := \frac{1}{\text{Odds}(A)} = \frac{P(A^c)}{P(A)} = \frac{1-P(A)}{P(A)}$$

"odds against"

4:1 odds against $\Rightarrow$ 4 times out of 5 it will <u>not</u> occur.

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

$$P(A|B) = \frac{.036}{.06} = 0.6, \quad P(A|B^c) = \frac{P(AB^c)}{P(B^c)} = \frac{.164}{.94} = .174$$

$P(A) = P(AB) + P(AB^c)$

$.2 = .036 + P(AB^c) \Rightarrow P(AB^c) = .164$

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

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

$$\Rightarrow \underbrace{\frac{P(B|A)}{P(B^c|A)}}_{\text{posterior odds}} = \underbrace{\frac{P(A|B)}{P(A|B^c)}}_{\text{likelihood ratio}} \overbrace{\frac{P(B)}{P(B^c)}}^{\text{prior odds}}$$

$= 0.22$ $\qquad = 3.44$ $\qquad = 0.064$

16:1 odds against

data

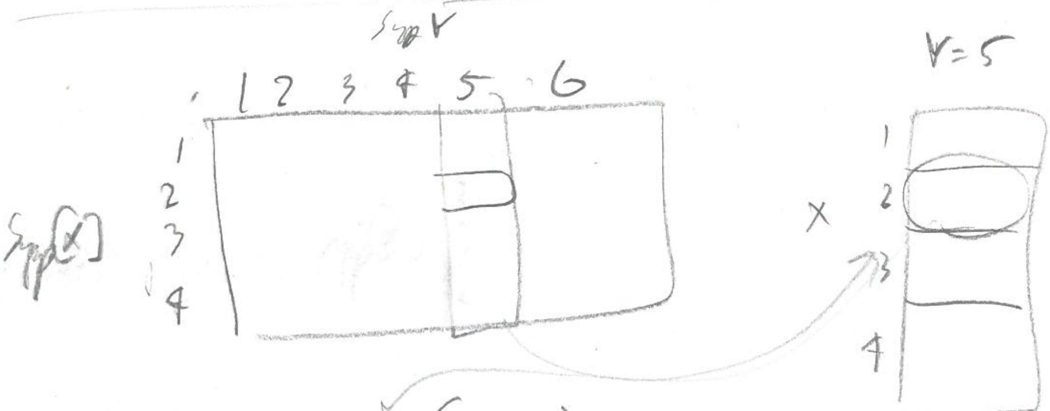
5:1 odds against

$P(A|B) = \frac{.036}{.06} = 0.6$

$P(A|B^c) = \frac{.164}{.94} = .174$

$P(A,B^c) = P(A) - P(A,B) = .2 - .036 = .164$

What is Bayesianism?



$$P(X=2|Y=5) = \frac{P(X=2, Y=5)}{\underbrace{P(Y=5)}_{P(X=1,Y=5) + P(X=2,Y=5) + P(X=3,Y=5) + P(X=4,Y=5)}}$$

Bayes Rule for r.v. usually denoted

in shorthand... $P(X|Y) = \frac{P(X,Y)}{P(Y)}$

or in longhand... $P(X=x|Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)}$

~~Bayesianism~~ Law of Total Prob

$\Rightarrow P(y) = \sum_{x \in Supp(X)} P(x,y)$ for PMFs or $p(y) = \sum_{x \in Supp(X)} p(y|x)\,p(x)$

$\Rightarrow f(y) = \int_{x \in Supp(X)} f(x,y)\,dx$ for PDFs or $f(y) = \int_{x \in Supp(X)} f(y|x)\,f(x)\,dx$

Back to the story...

MLE's have issues ... all frequentist stats has issues ...

Why don't we consider θ ~~our~~ one of interest? X our data.

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

What's wrong?

(1) $P(\theta)$?? θ is one immovable value!! $P(\theta)$ is degenerate! $\sim \text{Deg}(\theta)$ and you don't know it!!

(2) $P(X)$ makes no sense ... you can't calc. prob of data without knowing θ, so using $P(X) = \sum_{\theta_0 \in \Theta} P(X|\theta_0)P(\theta_0)$ has $P(\theta_0)$ which will be 0 except for when $\theta_0 = \theta \Rightarrow P(\theta_0) = 1$
$\Rightarrow P(X) = P(X|\theta)$

$\Rightarrow$ (3) $P(\theta|X) = P(\theta) = 1$ if θ is its true value. Clearly ~~not~~ useful!

Frequentism: θ is one value

Bayesian: ditto ... but we can use $P(\theta)$ to represent uncertainty in this value a priori. $\Rightarrow \theta$ is a r.v.

$$\Rightarrow P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$ Now is coherent

$P(X) \leftarrow$ How is this calc?

and by Bayes Thm $P(\theta|X) = \frac{P(X|\theta)P(\theta)}{\sum_{\theta_0 \in \Theta} P(X|\theta_0)P(\theta_0)}$ or $\frac{P(X|\theta)P(\theta)}{\int_{\theta_0 \in \Theta} P(X|\theta_0)P(\theta_0)\,d\theta_0}$

X: Data (the effect)

θ: Model (the cause)

$P(X|\theta)$ effect | cause

$P(\theta|X)$ cause | effect $\Rightarrow$ the "inverse problem"

What does this mean. Let X be Bernoulli, $X = \langle 0, 1, 1 \rangle$
and there are two models $\theta = 0.75$ $\theta = 0.5$ ... absurd ... but let's go with it ...

$P(X|\theta = 0.75) = .25 \cdot .75 \cdot .75 = .141$

$P(X|\theta = 0.25) = .75 \cdot .25 \cdot .25 = .047$

Model #2 is more likely but what is explicitly $P(\theta = 0.75 | X)$?

$$= \frac{P(X|\theta=0.75)\ P(\theta=0.75)}{P(X)} = \frac{P(X|\theta=0.75)\ P(\theta=0.75)}{\underbrace{P(X|\theta=0.25)\ P(\theta=0.25) + P(X|\theta=0.75)\ P(\theta=0.75)}_{\text{Result of Bayes thm}}}$$

Need $P(\theta=0.75)$, $P(\theta=0.25)$. Remember... we are allowed to consider our prior uncertainty in the model param. What should we choose?

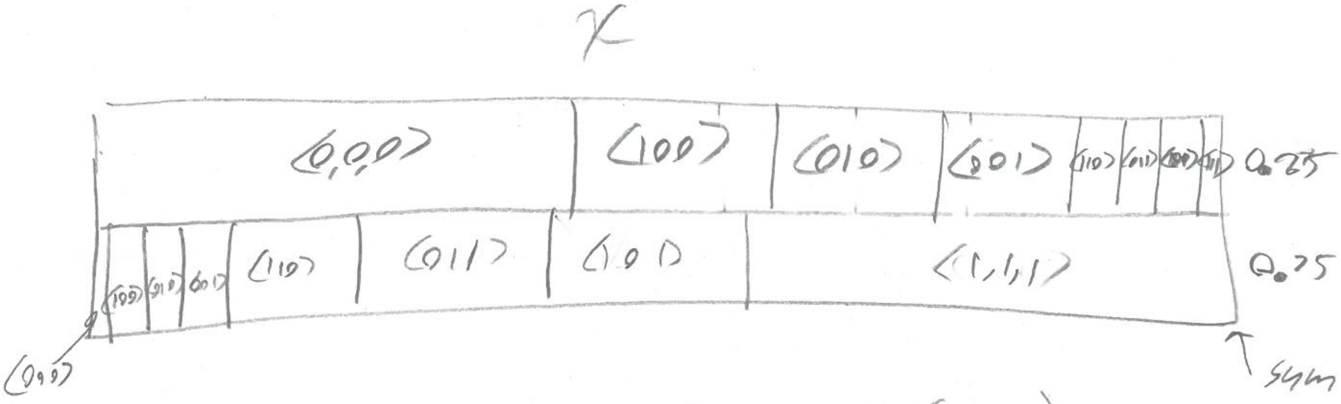
$$P(\theta) = \begin{cases} 0.75 & wp\ \frac{1}{2} \\ 0.25 & wp\ \frac{1}{2} \end{cases}$$

"Principle of indifference": All models equally likely a priori

X and θ are both r.v.'s. Let's visualize them

$\theta \in \Theta = \{0.75, 0.25\}$

$X \in \mathcal{X} = \text{Supp}[X]^3 = \{0,1\} \times \{0,1\} \times \{0,1\}$
"data space"
$= \{\langle 000\rangle, \langle 001\rangle, \langle 011\rangle, \langle 111\rangle, \langle 110\rangle, \langle 100\rangle, \langle 010\rangle, \langle 101\rangle\}$



$$P(X=\langle 000\rangle \,\&\, \theta=0.25) = P(\langle 000\rangle \mid \theta=0.25)\,P(\theta=0.25)$$

$$= \underbrace{0.75^3}_{0.422}\cdot 0.5 = .211$$

$$P(X=\langle 100\rangle \,\&\, \theta=0.25) = \underbrace{.25\cdot .75^2}_{0.141}\cdot 0.5 = .070$$

$$P(X=\langle 110\rangle \,\&\, \theta=0.25) = \underbrace{.25^2\cdot 0.75}_{.047}\cdot 0.5 = .023$$

$$P(X=\langle 111\rangle \,\&\, \theta=0.25) = \underbrace{.25^3}_{.016}\cdot 0.5 = .008$$

Is θ indep of X?

NO

Knowing θ tells you something about X; when x tells you something about θ.

$$P(\langle 000\rangle \,\&\, \theta=0.75) = \underbrace{.25^3}_{.016}\cdot 0.5 = .008$$

Opposite!

easily understand the denominator

$$P(X=\langle 011\rangle) = P(X=\langle 011\rangle \,\&\, \theta=0.25) + P(X=011 \,\&\, \theta=0.75)$$

$$= .023 + .070 = .094$$

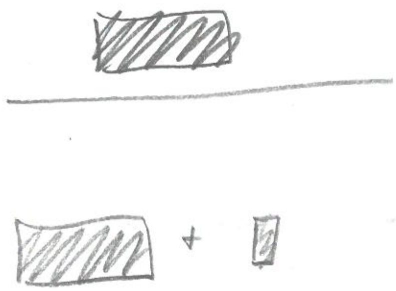
Is and a different piece of universe? YES

Now what's the prob of $P(\theta = 0.75 \mid X = \langle 0 1 1 \rangle)$? $\neq P(\langle 0,1,1 \rangle \mid \theta = 0.75)$

$= .0141$

$$= \frac{P(X = 011 \ \& \ \theta = 0.75)}{P(X = 011)}$$

$=$



$$= \frac{.020}{.023 + .020} = \underbrace{0.75}_{\text{(coincidence that } \theta = 0.75)}$$

and of course $P(\theta = 0.25 \mid X = \langle 0 1 1 \rangle) = 1 - P(\theta = 0.75 \mid X = \langle 0 1 1 \rangle) = 0.25$

$$\underbrace{P(\theta = 0.75)}_{0.5} \xrightarrow{X} \underbrace{P(\theta = 0.75 \mid X = \langle 0,1,1 \rangle)}_{0.75}$$ Bayesian conditionalism

$$0.75 = \underbrace{\frac{0.141}{0.09}}_{1.504} 0.5$$

$\frac{0.5}{0.5} = 1$ prior odds $= 1$

$$\underbrace{\frac{P(\theta = 0.75 \mid X = \langle 0,1,1 \rangle)}{P(\theta = 0.25 \mid X = \langle 0,1,1 \rangle)}}_{\frac{0.75}{0.25}} = \frac{P(X = \langle 0,1,1 \rangle \mid \theta = 0.75)}{P(X = \langle 0,1,1 \rangle \mid \theta = 0.25)} \frac{P(\theta = 0.75)}{P(\theta = 0.25)}$$

Note: odds doesn't require computing $P(X)$

$$\frac{.141}{.047} = \frac{.25 \cdot .75 \cdot .75}{.75 \cdot .25 \cdot .25} = 3$$

posterior odds $= 3$

$P(X = \langle 0,1,1 \rangle)$ is the prior on the data... what does the prob of this data look like averaged over prior on θ. "prior pred. distr"

Let $\Theta = \{ 0.1, 0.25, 0.5, 0.75, 0.9 \}$

$P(\theta) = \begin{cases} .1 \\ .25 \\ .5 \\ .75 \\ .9 \end{cases}$ w.p. $\frac{1}{5}$

equally likely

Saw data $x = \langle 0, 1, 1 \rangle$

$$P(\theta | x) = \frac{P(x|\theta)\, P(\theta)}{P(x)} = \left(\frac{1}{P(x)}\right) P(x|\theta)\, P(\theta) \propto P(x|\theta) \underbrace{P(\theta)}^{= \frac{1}{5}}$$

$$\propto P(x|\theta)$$

same for all values of θ

$\theta = 0.1$ $P(X|\theta) = 0.009$

$\theta = 0.25$ $P(X|\theta) = 0.047$

$\theta = 0.5$ $P(X|\theta) = 0.125$

$\theta = 0.75$ $P(X|\theta) = 0.141$

$\theta = 0.9$ $P(X|\theta) = 0.081$

relative sizes

$P(\theta = 0.75 | X) \propto P(X | \theta = 0.75)$ But how to get constant?

mult by $\dfrac{P(\theta)}{P(X)}$

$=$ Σ of all boxes

How likely is this data over all possible theories?

Recall $\hat{\theta}_{MLE} = 0.66$ was "best guess" of θ i.e. the "pt. estimate".

What's our best guess of θ now?

$\theta = 0.75$. Why? Most likely given data!

Kind of like MLE!