Math 341 Lec 4 2/7/18

$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}$ Bayes Rule for events

$P(Y=y|X=x) = \frac{P(X=x|Y=y)\,P(Y=y)}{P(X=x)}$ " for r.v.'s.

1

Back to the story...

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

Why don't we consider θ our r.v. of interest? X our data.

What is this? $$P(\theta|X) = \frac{P(X|\theta)\,P(\theta)}{P(X)}$$ incoherent!

What's wrong?

(1) $P(\theta)$?? θ is one immutable value!! $P(\theta)$ is degenerate! $\sim Deg(\theta)$ and you don't know it!! $P(\theta) = 0$ or 1

(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.

THE BIG LEAP!

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

$P(X)$ ← How is this calc? Bayes Thm,

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

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 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 $P(\theta = 0.75 | x)$?

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

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?

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

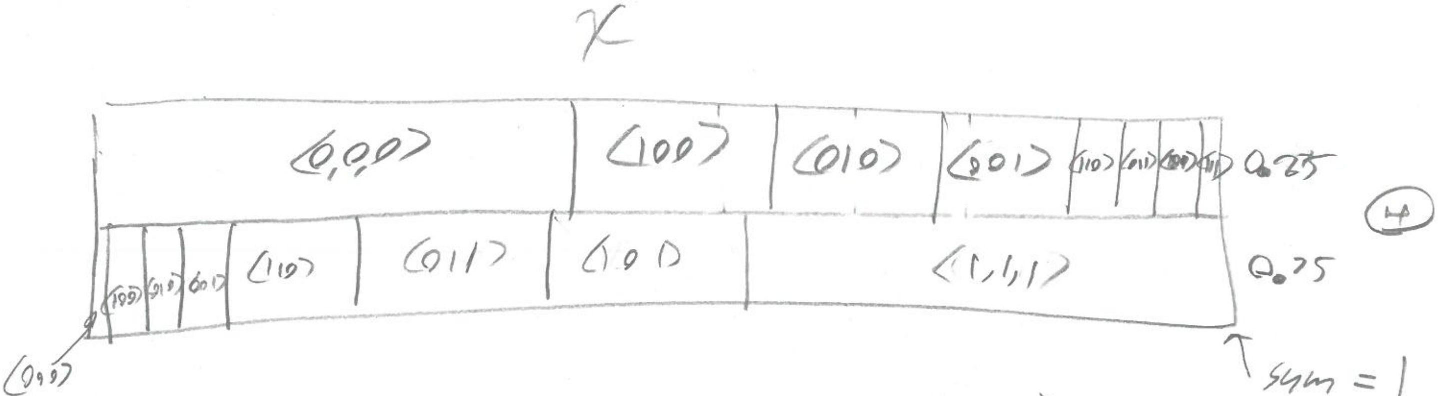
"principle of indifference": All models equally likely a priori

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"
$= \{\langle 000\rangle, \langle 001\rangle, \langle 011\rangle, \langle 111\rangle, \langle 110\rangle, \langle 100\rangle, \langle 010\rangle, \langle 101\rangle\}$

B



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

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

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

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

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

Is θ indep of X?

NO

Knowing θ tells you something about X's; likewise 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

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

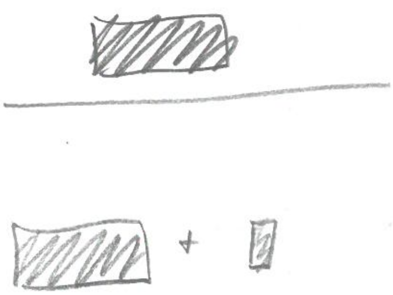
$= .023 + .070 = .094$

Is and a different piece of universe? YES

Now what's the prob of $P(\theta=0.75 \mid X=\langle 011\rangle)$? $\neq P(\langle 0,1,1\rangle \mid \theta=0.75)$ $= .141$

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

=



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

and of course $P(\theta=0.25 \mid X=\langle 011\rangle) = 1 - P(\theta=0.75 \mid X=\langle 011\rangle) = 0.25$

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

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

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

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

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

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

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

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

equally likely

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

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

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

Same for all values of θ

$\mathcal{X}$

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

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

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

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

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

relative sizes

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

mult by $\frac{P(\theta)}{P(x)}$

$P(x) = \Sigma$ 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!