Lecture 3 Math 390.03-02

2/8/16

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

~~Another Bayes Factor:~~

~~Odds for A~~ ~~$\frac{P(B|A)}{P(B|A^c)} = 6$ very strong evidence~~

~~Seen i.e. decide between two scenarios~~

~~HW: $\frac{P(B|A)}{P(B|A^c)} = \frac{.18}{.03} = 6$ technically, a Bayes Factor in favor of A~~

In general $A_1, \ldots, A_n$ mut. excl. coll. exh.

$$\Rightarrow P(B) = P(B, A_1) + P(B, A_2) + \ldots + P(B, A_n)$$

$$P(A_\ell | B) = \frac{P(B|A_\ell)\, P(A_\ell)}{\sum_{i=1}^{n} P(B|A_i)\, P(A_i)}$$ Bayes Thm.

Likelihood given jmf

$p(X, Y)$,

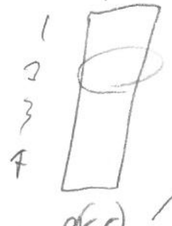
Supp(Y)

Supp(X)	1	2	3	4	5	6
1						
2						
3						
4						

$P(X=2 | Y=5)$ just like B, A

$= \dfrac{P(X=2, Y=5)}{P(Y=5)}$

Y=5



for $P(Y=5)$

$P(Y=5) = P(X=1, Y=5) + P(X=2, Y=5) + P(X=3, Y=5) + P(X=4, Y=5)$

disjoint

sum of joint prob

different!! → prob

So we can allow B, A to be random vars

PMF

$p(x|y) = \dfrac{p(x,y)}{p(y)} \quad \dfrac{p(y|x)\,p(x)}{p(y)}$

$p(y) = \sum_{x \in Supp(X)} p(x,y)$ marginalization

Note $X|Y$ is a new r.v.

$f(x|y) = \dfrac{f(x,y)}{f(y)} = \dfrac{f(y|x)\,f(x)}{f(y)}$

$f(y) = \int_{x \in Supp(X)} f(x,y)\,dx$

$Supp(X|Y) = Supp(X)$?

$E(X|Y) = E(X)$?

$Var(X|Y) = Var(X)$?

the parameter for F

Consider $P(\theta | X) = \dfrac{P(X|\theta)\,P(\theta)}{P(X)}$ ← prior for θ

the data

likelihood of X under a certain model

"prior for X"

"marginal likelihood"

makes no sense since θ is fixed! either 0 or 1

What is $P(\theta ; X) = \dfrac{P(X;\theta)\,P(\theta)}{P(X)}$

this also makes no sense since you cannot calculate $P(X)$ without knowing what θ is

Incoherent in frequentist statistics

likelihood function $\mathcal{L}(\theta; X) = P(X;\theta)$ jdf

We have to stare at this again ~~the~~ whole semester

Staring at this formula:

inverse inference about θ — effect — cause

$$P(\theta|x) = \frac{P(x|\theta)\,P(\theta)}{P(x)} \leftarrow \text{prior on } \theta$$

inverse prob question

P(cause|effect) vs. P(effect|cause)

What does this mean? Recall coin flips: $\underbrace{\langle 0,1,1 \rangle}_{X}$

$$P(X|\theta = 0.75) = .25 \cdot .75 \cdot .75 = .141$$
$$P(X|\theta = 0.5) = .5 \cdot .5 \cdot .5 = .125$$

presumably we calculated these...

$P(\theta = 0.25)$?? $P(\theta = 0.75)$??

this is our idea of what θ is a priori to continuing coin flips

Imagine for now these are the only two theories for the param θ

Ridiculous, yes, but go with it.

Also imagine $\underbrace{P(\theta = 0.25)}_{\text{model 1}} = \underbrace{P(\theta = 0.75)}_{\text{model 2}} = \frac{1}{2}$

principle of indifference.

All prior hypothesis about θ are equally likely.

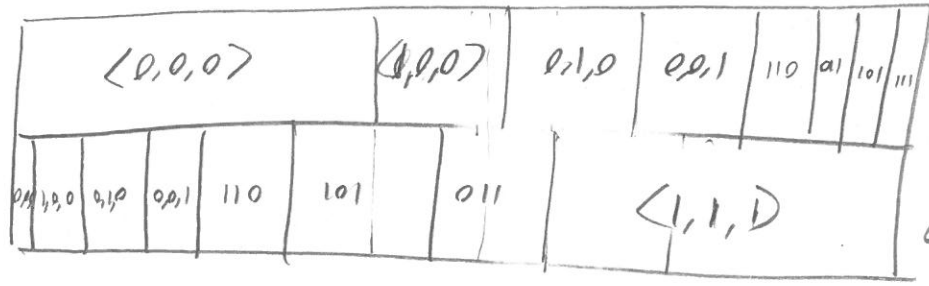
$\Theta_0 \subseteq \Theta$
prm

A

$\Omega = \mathcal{X} \times \Theta_0$

$\text{Supp}[X]^n$

$\text{Supp}[X]^3$ = sample space



$\mathcal{X} = \{\langle 0,0,0\rangle, \langle 0,0,1\rangle, \langle 0,1,0\rangle, \langle 1,0,0\rangle, \langle 0,1,1\rangle, \langle 1,0,1\rangle, \langle 1,1,0\rangle, \langle 1,1,1\rangle\}$

$P(\theta=.5)\ P(0,0,0 \mid \theta=0.25) = .75^3 = .421 \Rightarrow P(\langle 0,0,0\rangle \,\&\, \theta=.25)$

$= .211$

is $\langle 0,0,0\rangle, \theta=0.25 \cap \langle 0,0,0\rangle, \theta=.75$ $=\phi$? YES,

different pieces of the universe!

$P(\langle 0,0,1\rangle \mid \theta=0.25) = .75^2 \cdot .25 = .141$

$= P(\langle 0,1,0\rangle \mid \quad) \quad =$

$P(\langle 1,0,1\rangle \mid \quad) \quad =$

$P(\langle 0,1,1\rangle \mid \theta=0.25) = .75 \cdot .25^2 = .047$

$P(\langle 1,1,1\rangle \mid \quad) = .25^3 = .016$

$P(\langle 0,0,0\rangle \mid \theta=.75) = .25^3 = .016$

$P(\langle 1,0,0\rangle \mid \quad) = .75 \cdot .25^2 = .047$

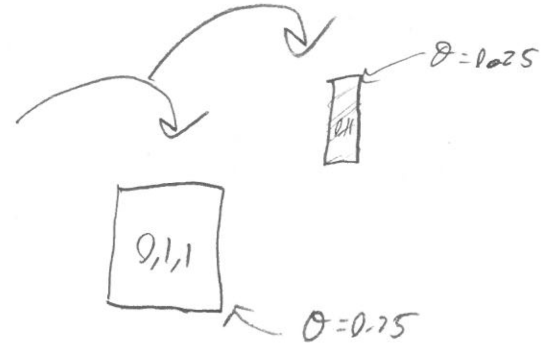
Now let's ask the inverse question. I've seen the effect, who's the cause. $X = \langle 0,1,1\rangle$. I live in that world now!

What is prob $\theta = 0.25$?

$\cancel{P(\langle 0,1,1\rangle \mid \theta = 0.25)}$ $\quad P(\theta = 0.25 \mid \langle 0,1,1 \rangle)$

Not this!!!! But this →

$\langle 0,1,1 \rangle = \Omega_0$



Obviously...
$$\frac{P(0,1,1,\ \theta=0.25)}{\underbrace{P(0,1,1,\ \theta=0.25) + P(0,1,1,\ \theta=0.75)}_{P(0,1,1)}} = \frac{P(0,1,1 \mid \theta=0.25)\,P(\theta=0.25)}{P(0,1,1 \mid \theta=0.25)\,P(\theta=0.25) + P(0,1,1,\ \theta=0.75)\,P(\theta=0.75)}$$

$$= \frac{.047 \cdot .5}{\underbrace{.047 \cdot .5 + .141 \cdot .5}_{.094}} = \boxed{.25}$$

$$\Rightarrow P(\theta = 0.75 \mid \langle 0,1,1\rangle) = P((\theta=0.25)^c \mid \langle 0,1,1\rangle) = 1 - 0.25 = 0.75$$

What is $P(0,1,1)$. This is the prior on the data.

The chance of just seeing 0,1,1 without regards to θ.

s.t. $\Theta = \{\theta_1, \ldots, \theta_n\}$, n could be ∞

$$P(X) = \sum_{i=1}^{n} P(X, \theta_i) \quad \text{or} = \int_{\Theta} P(X,\theta)\, d\theta$$

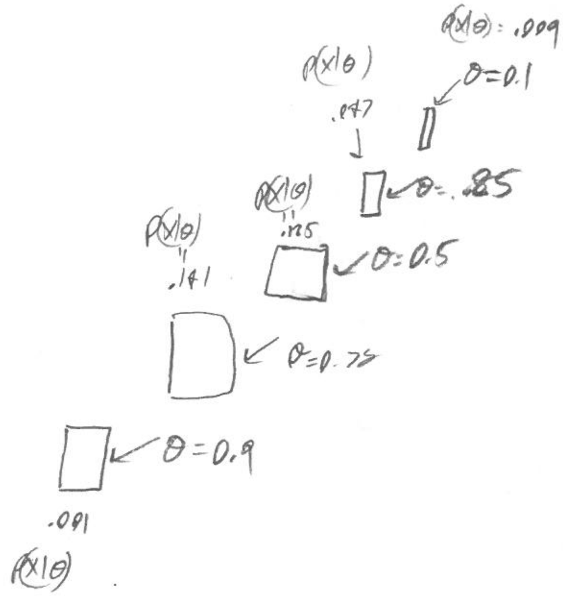
$$= \sum_{i=1}^{n} P(X \mid \theta_i)\, P(\theta_i) \qquad = \int_{\Theta} P(X \mid \theta)\, P(\theta)\, d\theta$$

Is X, θ independent? $P(X,\theta) = P(X)\,P(\theta)$? NO WAY

$$\overset{.047 \cdot .5}{P(0,1,1,\ \theta = .25)} \overset{?}{=} \overset{.094}{P(X)}\ \overset{.5}{P(\theta = 0.25)}$$

Imagine $\Theta = \{0.1, 0.25, 0.5, 0.75, 0.9\}$ s.t. $P(\theta_i) = \frac{1}{5} \; \forall i$

Observe $X = \langle 0,1,1 \rangle$. Want to ask $P(\theta \mid X)$ again.



Note that $$P(\theta \mid X) = \frac{P(X \mid \theta)\, P(\theta)}{P(X)} \propto P(X \mid \theta)\, P(\theta)$$

not a $f(\theta)$

hence is a constant

$\forall \theta \in \Theta$.

So if you just want to compare

$$\frac{P(\theta = 0.75 \mid X)}{P(\theta = 0.9 \mid X)} = \frac{P(X \mid \theta = 0.75)\; P(\theta = 0.75)}{P(X \mid \theta = 0.9)\; P(\theta = 0.9)}$$

I know this is a lot... but we're going to do it 100 times.

What is the "best" estimate of θ?

Given the data, isn't it the most likely θ?

$$\hat{\theta}^{MAP}_{Bayes} = \underset{\theta \in \Theta_0}{\arg\max} \{P(\theta|X)\} = \underset{\theta \in \Theta_0}{\arg\max} \left\{ \frac{P(X|\theta)\,P(\theta)}{P(X)} \right\} = \underset{\theta \in \Theta_0}{\arg\max} \{ P(X|\theta)\,P(\theta) \}$$

maximum a posteriori (MAP)

if $P(\theta)$ are the same for all $\theta \in \Theta_0$ $= \underset{\theta \in \Theta_0}{\arg\max} \{ P(X|\theta) \}$

if $\Theta_0 = \Theta$ $= \hat{\theta}_{MLE}$ (only)

In our case, $\hat{\theta}_{Bayes} = 0.75$

$\hat{\theta}_{MLE} = 0.66$ why? $\Theta_0 \neq \Theta = (0,1)$

Our choice of Θ_0 was weak. Unless you have good reason, $\Theta_0 = \Theta$.

Weakness #1: Your prior could be bad (mainly controversy).

$$P(\theta = 0.25) = 0.5 \longrightarrow P(\theta = 0.25 \mid \langle 0,1,1 \rangle) = 0.25$$

"Bayesian Conditionalism"

Data comes in ... update prior beliefs.

Let's see how this works more closely ... One at a time...

$X_1 = 0,\ X_2 = 1,\ X_3 = 1$ s.t. $X_1, X_2, X_3 \overset{iid}{\sim} \text{Bern}(\theta)$ little bit wrong