2/7/17

• Bayes Rule: $$P(B|A) = \frac{P(A|B)\overbrace{P(B)}^{\text{Prior or Model/Theory}}}{P(A)}$$

- Posterior on model theory
- Theory given data

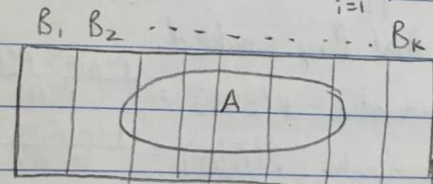
?

A = Data (Effect)
B = Model/Theory (Cause)

Bayes Theorem: $$P(B|A) = \frac{P(A|B)P(B)}{\sum_{i=1}^{K} P(A|B_i)P(B_i)}$$

• Law of Total Probability: Let $B_1, \dots, B_K$ are mutually exclusive and collectively exhaustive. Then

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



Bayesian Conditionalism: Take $P(B) \xrightarrow{\text{intro with data (A)}} P(B|A)$.

• Another way to think about the $P(A)$:

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

Ex: Assume $P(A) \in (0,1)$. Let's say an event has an odds of 4 or "4:1 odds" (4 to 1) then the event has a probability of occuring of 0.8 or 80%.

$\frac{0.8}{0.2} = 4$ or 5 chances. $100 \div 5 = 20$

$20 \times 4 = 80\%$ ✓

To get Odds Against $= \text{Odds}(A)^{-1} = \frac{P(\bar{A})}{P(A)} = \frac{1-P(A)}{P(A)}$

EX: At a racing track, they want you to lose, so they'll give you "odds againts" odds. If you bet $1 you'll get $4 if you win..."

lung cancer example : (A) = smoker, (B) = lung cancer

$P(A) = 0.2$

$P(B) = 0.06$

$P(A,B) = P(A \cap B) = 0.036$

* $P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{0.036}{0.06} = 0.6$.

* $P(A|\bar{B}) = \frac{P(A\cap\bar{B})}{P(\bar{B})} = \frac{P(A\cap\bar{B})}{1-P(B)} = \frac{P(A)-P(A\cap B)}{0.94} = \frac{0.2-0.036}{0.94} = 0.174$

↳ look at all the ppl who don't have linge cancer, probability that they smoked.

* $\frac{P(B|A)}{P(\bar{B}|A)} = ?$ Numerator : $P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}$

Denominator : $P(\bar{B}|A) = \frac{P(A|\bar{B})\,P(\bar{B})}{P(A)}$

So $$\underbrace{\frac{P(B|A)}{P(\bar{B}|A)}}_{\text{Posterior odds.}} = \overbrace{\frac{P(A|B)}{P(A|\bar{B})}}^{\text{likelihood ratio}} \underbrace{\left(\frac{P(B)}{P(\bar{B})}\right)}_{\text{Prior odds}}$$

Prior odds : Based on the general populatio[n]. Then, when you're given data everything Δ's.

$$0.22 = (3.44)(0.064)$$

$$0.22 = \left(\frac{0.6}{0.174}\right)\left(\frac{0.06}{0.94}\right)$$

The odds of getting lung cancer given that the person smokes is 0.22 22%.

* Posterior > Prior with introduction of data

$\frac{P(B)}{P(\bar{B})} \xrightarrow{\text{intro to A}} \frac{P(B|A)}{P(\bar{B}|A)}$

Marginilism:

Imagine 2 r.v.'s X, Y the joint pmf can be represented:

Supp(Y)
=

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

$$P(X|Y) = \frac{P(Y|X)\,P(X)}{P(Y)}$$ shorthand for $$P\{X=x \mid Y=y\} = \frac{P(Y=y|X=x)\,P(X=x)}{P(Y=y)}$$

same as

$$P(B|A) = \frac{P(A|B)(P(B))}{P(A)}$$

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

=

Law of Total Probability.

- Generally, $P\{Y=y\} = \sum_{x \in Supp\{X\}} P(Y=y, X=x\}$ aka Marginilization or Margining out x.

Discrete (Pmfs) $= \sum_{x \in Supp\{X\}} P(Y=y|X=x)P(X=x)$

continuous (pdf) $= \int_{x \in Supp\{X\}} f_{YX}(y,x)\,dx = \int_{x \in Supp\{X\}} f_{Y|X}(y|x)\,f_X(x)\,dx$

Pick $\mathcal{F}$ (model Family). θ = parameter (s) & X = data

$$\mathcal{L}(\theta ; x) = P(x ; \theta)$$

(look at all models, give me back the maximum likelihood estimate)

Consider the following as the Frequentist POV.

$$\underbrace{P(\theta | X)}_{P(\text{cause}|\text{effect})} = \frac{\overbrace{P(X|\theta)}^{p(\text{effect}|\text{cause})} P(\theta)}{P(X)} \rightarrow P(\theta) \text{ is either } 0 \text{ or } 1$$

Problems: 1. $P(\theta) = \text{Deg}(\theta_0) = \{0,1\}$ because we $\neq$ know θ exactly.

2. $P(X)$ makes no sense bc $\neq$ given a θ. If $\theta = \theta_0$ (given) the we would have

$$P(X) = \sum_{\theta \in \Theta} P(X|\theta_0)P(\theta_0) = P(X|\theta_0)$$

Then $P(\theta | X) = \dfrac{P(X|\theta_0)\, P(\theta)}{P(X|\theta_0)}$ and still can only be 0 or 1

The problems began with $P(\theta)$. There is one true value of θ (call it θ_0). In the Frequentist Approach $P(\theta)$ is degenerate. In the Bayesian Approach, we allow θ to represent ~~the~~ our prior uncertainty AKA Prior Information.

$$\underbrace{P(\theta|X)}_{\text{posterior}} = \frac{\overbrace{P(X|\theta)}^{\text{likelihood}} \overbrace{P(\theta)}^{\text{prior}}}{P(X)} = \frac{P(X|\theta)\, P(\theta)}{\sum_{\theta_i \in \Theta} P(X|\theta_i)P(\theta_i)} \quad (\theta = \text{Discrete})$$

$$= \frac{P(X|\theta)\, P(\theta)}{\int_{\Theta} P(X|\theta_i)\, P(\theta_i)\, d\theta_i} \quad (\theta = \text{cont.})$$

Ex: Let $\mathcal{F}$: Bernoulli iid $X = (0,1,1)$ (coin flip T,H,H)

estimating $\theta = 0.75$: $P(X|\theta=0.75) = (0.25)(0.75)(0.75) = 0.141$

estimating $\theta = 0.25$: $P(X|\theta=0.25) = (0.75)(0.25)(0.25) = 0.047$

So, we assume $\Theta = \{0.25, 0.75\}$ (only 2 values of θ)

What is $P(\theta = 0.75 \mid X = (0,1,1)) = ?$

$$P(\theta=0.75 \mid X=(0,1,1)) = \frac{\overbrace{P(X=0,1,1 \mid \theta=0.75)}^{0.141}\,P(\theta=0.75)}{P(X=0,1,1)}$$

represents $P(\theta)$ before any data is introduced. So $= 0.5$

use L.O.T.P.

(Principle of Indifference) all models equally likely.

$$P(\theta) = \begin{cases} 0.5 & \text{if } \theta = 0.25 \\ 0.5 & \text{if } \theta = 0.75 \end{cases}$$

since we only have 2 θ's to pick from

Max Rule.

$$= \frac{(0.141)(0.5)}{P(X=0,1,1, \theta=0.75)(\theta=0.75) + P(X=0,1,1, \theta=0.25)(\theta=0.25)}$$

$$= \frac{(0.141)(0.5)}{(0.141)(0.5) + (0.047)(0.5)}$$

$$= \frac{0.070}{0.070 + 0.023}$$

$$P(\theta=0.75 \mid X=0,1,1) = \boxed{0.75}$$

Gone from a 50% belief in the model to 75% with intro to data, and Bayesian Approach.

Now that I know $P(\theta=0.75 \mid X=(0,1,1))$, we know that

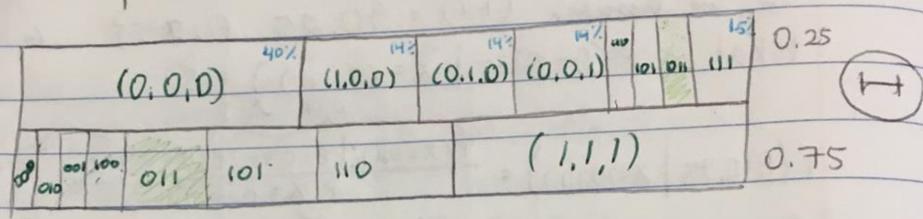
$P(\theta=0.25 \mid X=0,1,1) = 1 - P(\theta=0.75 \mid X=(0,1,1)) = 1-0.75 = 0.25$

- Let X and θ be r.v.'s. Thus there's a joint distribution.

$X = \{(0,0,0), (0,0,1), (0,1,0), (1,0,0), (0,1,1), (1,0,1), (1,1,0), (1,1,1)\}$

= data space = all possible realizations of X.

opposites



- $P(000 \mid \theta = 0.25) = (0.75)^3 = 0.422 \cong 40\%$

$P(100 \mid \theta = 0.25) = (0.25)(0.75)(0.75) = 0.141 \cong 14\%$

$P(110 \mid \theta = 0.25) = (0.25)(0.25)(0.75) = 0.047 \cong 5\%$

$P(111 \mid \theta = 0.25) = (0.25)(0.25)(0.25) = 0.015 \cong 1.5\%$

Is X independent of θ? **NO**

Knowing $\theta \rightsquigarrow$ tells you about X

knowing X $\rightsquigarrow$ tells you about θ

- $P(X = (0,0,0), \theta = 0.25) = P\{X = 0,0,0 \mid \theta = 0.25\} P\{\theta = 0.25\}$

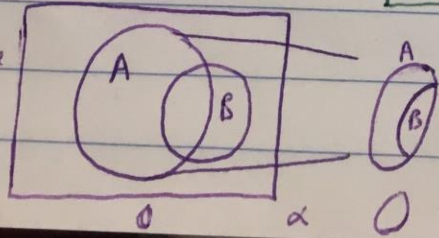
$= (0.422)(0.5)$

$= 0.211$

- $P(\theta = 0.75 \mid X = (0,1,1)) = \boxed{011}$ $\boxed{011}$ } Pieces of Picture we only care about

$= \dfrac{\boxed{011}}{\boxed{011} + \boxed{011}}$

Bug Example



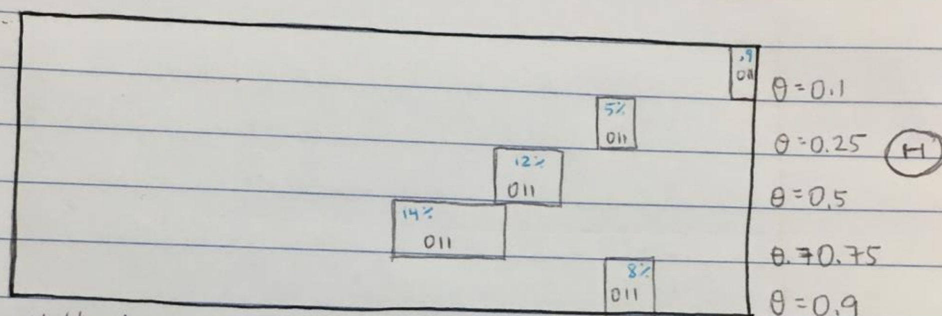
- Need to know A to normalize the universe
- Scale up by $\frac{1}{P(A)}$ like similar Δ's (same $\measuredangle$)

If $\Theta = \{0.1, 0.25, 0.5, 0.75, 0.9\}$. (same $X = (0,1,1)$)

$$P(\theta) = \begin{cases} 0.2 & \text{if } \theta = 0.1 \\ 0.2 & \text{if } \theta = 0.25 \\ 0.2 & \text{if } \theta = 0.5 \\ 0.2 & \text{if } \theta = 0.75 \\ 0.2 & \text{if } \theta = 0.9 \end{cases}$$

#MaxRule

(Apriori Rule before data)



$\underbrace{}_{}$ likelihoods

$P(X \mid \theta = 0.1) = 0.009 \cong 0.9\%$

$P(X \mid \theta = 0.25) = 0.047 \cong 5\%$

$P(X \mid \theta = 0.5) = 0.125 \cong 12\%$

$P(X \mid \theta = 0.75) = 0.141 \cong 14\%$

$P(X \mid \theta = 0.9) = 0.081 \cong 8\%$

if Principal of Indifference

if wanted $P(\theta = 0.75 \mid X = 011) = P(\theta \mid X) = \dfrac{P(X|\theta)P(\theta)}{P(X)} \propto P(X|\theta)P(\theta) \propto \underset{\text{likelihood}}{P(X|\theta)}$

$P(X)$ ↓ Σ boxes

- scale ↑ when ÷ by $P(X)$
- scale ↓ when ÷ by $P(\theta)$

$o \propto O \propto \delta$

$\theta = 0.75$ most likely model bc ↑est % (14%) ~~state~~

$\hat{\theta}_{MLE}$ for $X = (0,1,1)$ = Best guess in Bayesian framework $= \bar{X} = 2/3 = 0.66 \neq 0.75$

Not equal bc of Θ. You're only supposed to use a θ from there. None else