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

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

(effect) A: data → B: model/theory (cause) 2/7/17

$$P(B) \xrightarrow{A} P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$ ← prior on model/theory

$P(A)$ ← ?

theory | data

Bayes Rule

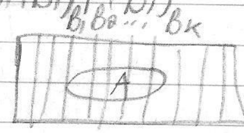
take whole space

Law of Total Prob.

Bayesian Conditionality

$B_1, \dots, B_K$ are mutually exclusive & collectively exhaustive

$$\underbrace{P(A) = \sum_{i=1}^{k} P(A, B_i) = \sum_{i=1}^{k} P(A|B_i)P(B_i)}_{\text{Bayes Thm.}}$$



Another way to think about prob. of A.

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

Assume $P(A) \in (0,1)$

Odds = 4 or "4 in 1 odds"

$$\underbrace{\text{Odds } \bar{A}}_{\text{Odds against}} := \text{Odd}(A)^{-1} = \frac{P(A^c)}{P(A)} = \frac{1-P(A)}{P(A)}$$

Ex: Lung Cancer

$P(A) = .2$ (smoking) $\quad P(A,B) = .036$

$P(B) = .06$ (lung cancer) $\quad P(A|B) = \frac{.036}{.06} = .6$

$P(A) - P(AB) = .164$

$$P(A|B^c) = \frac{\overbrace{P(A\,B^c)}}{0.94 = 1 - P(B) \ \{P(B^c)} = .174$$

$$\left.\begin{aligned} P(B|A) &= \frac{P(A|B)\,P(B)}{P(A)} \quad .036 \\ P(B^c|A) &= \frac{P(A|B^c)\,P(B^c)}{P(A)} \end{aligned}\right\} \Rightarrow \underbrace{\frac{\overset{\approx 20\%}{P(B|A)}}{P(B^c|A)}}_{\text{posterior odds}} = \overbrace{\frac{P(A|B)}{P(A|B^c)}}^{\text{likelihood ratio}} \underbrace{\left(\frac{P(B)}{P(B^c)}\right)}_{\text{prior odds}}$$

posterior odds $= .22$ 3.44 prior odds $= .064$

$$\frac{P(B)}{P(B^c)} \xrightarrow{A} \frac{P(B|A)}{P(B^c|A)}$$

5 to 1   16 to 1

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

Two r.v.'s X, Y

joint PMF

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

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

shorthand for

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

Law of Total Probability

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

Marginalization "marginality over X"

PMF's $\quad P(Y) = \sum_{x \in \text{Supp}(X)} P(Y=y, X=x) = \sum_{x \in \text{Supp}(X)} P(Y=y|X=x)\,P(X=x)$

PDF's $\quad f_Y(y) = \int_{x \in \text{Supp}(X)} f_{YX}(y,x)\,dx = \int_{x \in \text{Supp}(X)} f_{Y|X}(y|x)\,f_X(x)\,dx$

Pick F (odd family)

θ : parameter(s)

X : data $\quad \mathcal{L}(\theta; x) = P(X; \theta)$

Frequentist approach : one value of θ

$\Rightarrow P(\theta) = \underset{\text{degenerate}}{Deg}(\theta_0) = \{\theta_0 \text{ w.p. } 1\}$

Consider the following as a Frequentist:

P(Cause|Effect) $\rightarrow P(\theta|X) = \dfrac{P(X|\theta)P(\theta)}{P(X)}$ Deg 1 $\leftarrow$ P(Effect|Cause)

P(X) makes no sense

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

$P(\theta|X) = \dfrac{P(X|\overset{\theta_0}{\theta})P(\theta)}{P(X) = P(X|\theta)} = 1$ if $\theta = \theta_0$

- There is one true value of θ (call it θ_0)
- Frequentist Approach $P(\theta)$ is degenerate

Bayesian Approach: We use $P(\theta)$ to express our prior uncertainty a.k.a. prior information

$\underset{\text{posterior}}{P(\theta|X)} = \dfrac{\underset{\text{likelihood}}{P(X|\theta)}\,\underset{\text{prior}}{P(\theta)}}{P(X)} = \dfrac{P(X|\theta)P(\theta)}{\sum_{\theta \in \Theta} P(X|\theta_i)P(\theta_i)}$

$= \dfrac{P(X|\theta)P(\theta)}{\int_{\Theta} P(X|\theta_i)P(\theta_i)d\theta_i}$

Distributions for Probability

F Bernoulli, iid $x = \langle 0, 1, 1 \rangle$

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

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

Assume $\Theta = \{.25, .75\}$

$P(\theta = .75 | X = (0,1,1)) = \dfrac{\overbrace{P(X = \langle 0,1,1\rangle | \theta = .75)}^{.141}\overbrace{P(\theta = .75)}^{.5}}{P(X = \langle 0,1,1\rangle)}$ ← must use law of total prob

$P(\theta) = \begin{cases} .5 & \text{if } \theta = .25 \\ .5 & \text{if } \theta = .75 \end{cases}$

Principle of indifference: all models equally likely

$= \dfrac{\overset{.141}{P(X = \langle 0,1,1\rangle | \theta = .75)}\, P(\theta = .75)\}.5}{\underbrace{P(X = \langle 0,1,1\rangle, \theta = .75)}_{(.141)(.5)\ P(X = \langle 0,1,1\rangle|\theta = .75)P(\theta = .75)} + \underbrace{P(X = \langle 0,1,1\rangle, \theta = .25)}_{(.047)\ P(X = \langle 0,1,1\rangle|\theta = .25)P(\theta = .25)(.5)}} = \dfrac{.070}{.070 + .023} = \boxed{.75}$

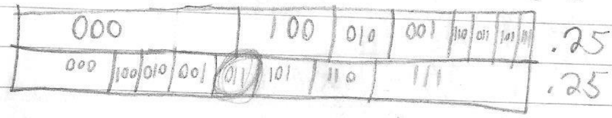
Logical vs. Sound Argument

$P(\theta = .25 \mid X = \langle 0,1,1 \rangle) = .25$

X, θ are r.v.'s thus there's a joint distribution

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

"data space"



$P(\langle 000 \rangle \mid \theta = .25) = (.75)^3 = .422$

$P(\langle 000 \rangle, \theta = .25) = .211$

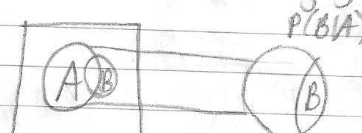
↖ Do Bayes rule in head

$P(\langle 100 \rangle \mid \theta = .25) = (.75)^2 (.25) = .141$

$P(\langle 110 \rangle \mid \theta = .25) = (.25)^2 (.75) = .047$

Is X indep. of θ? NO

$P(\theta = .75 \mid X = \langle 0,1,1 \rangle)$



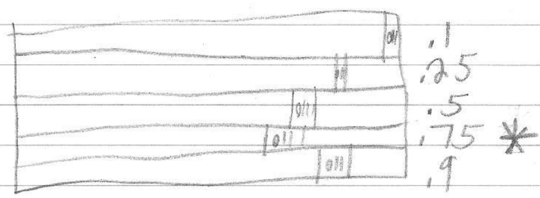
$P(B|A)$

$$\frac{\boxed{011}\ \theta = .75 \; + \; \boxed{011}\ \theta = .25}{P(X = \langle 0,1,1 \rangle)}$$

$$P(\theta = .75 \mid X = \langle 0,1,1 \rangle) = \frac{\boxed{}}{\boxed{} + \boxed{}}$$

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

$$P(\theta) = \begin{cases} .2 & \text{if } \theta = .1 \\ .2 & \text{if } \theta = .25 \\ .2 & \text{if } \theta = .5 \\ .2 & \text{if } \theta = .75 \\ .2 & \text{if } \theta = .9 \end{cases}$$



show .75 is most likely model

$P(X|\theta=.1)=.009$

" $0.25=.047$

" $0.5=.125$

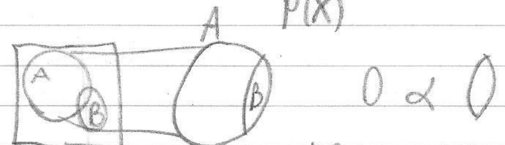
" $.75=.141$

" $.9=.061$

$P(\theta=.75|X=011)=\ldots$

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

$P(\theta) = \frac{1}{k}$ of principle of indifference



My "best guess" in Bayesian framework is $\theta=.75$

$\hat{\theta}_{MLE}=0.68\neq 0.75$ Why? for $X=\langle 0,1,1\rangle$