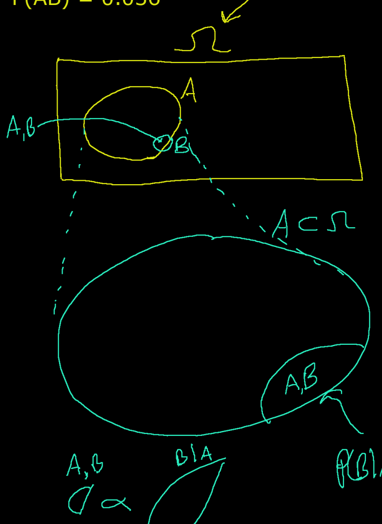
Let A: smoking and B: lung cancer
P(A) = 0.200
P(B) = 0.060
P(AB) = 0.036



P(lung cancer | smoking)

$= P(B|A)$

given, conditional universe

$$P(B|A) \propto P(AB)$$

$$P(B|A) = c\,P(AB) = \frac{\overbrace{P(\Omega)}^{1}}{P(A)} P(AB) = \frac{P(AB)}{P(A)}$$

definition of conditional probability

zoom factor $= \frac{1"}{0.5"} = 2$

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

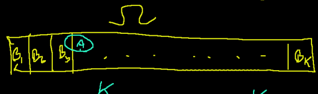
Bayes Rule

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

If B_1, B_2, ..., B_K are mutually exclusive and collectively exhaustive

$\forall i \neq j \quad B_i \cap B_j = \phi$

$\Omega = \bigcup_{i=1}^{K} B_i$



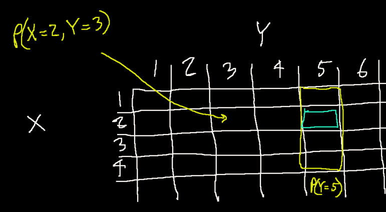
$$P(A) = \sum_{k=1}^{K} P(A, B_k) = \sum_{k=1}^{K} P(A|B_k)P(B_k)$$

"margining out the B_k's or integrating out the B_k's"

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

Bayes Theorem

Bayes Rule and Bayes Thm for rv's. Imagine two rv's X, Y and the Supp[X] = {1,2,3,4} and Supp[Y] = {1,2,3,4,5,6}.



$$P(Y=5) = P(Y=5, X=1) + P(Y=5, X=2) + P(Y=5, X=3) + P(Y=5, X=4) = \sum_{x \in Supp[X]} P(Y=5, X=x)$$

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

$$p(y) := P(Y=y) = \sum_{x \in Supp[X]} \underbrace{P(Y=y, X=x)}_{\text{JMF}} = \sum_{x} p(y,x)$$

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

conditional PMF

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

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

Back to the story... Can we use Bayes Rule to tell use anything about inference for parameter θ given data x (x = <x_1,...,x_n>). Consider:

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

$P(x_1, \ldots, x_n; \theta) = \mathcal{L}(\theta; x_1, \ldots, x_n)$ JMF but called "likelihood"

What is wrong with this equation? Previously, θ, the unknown parameter was assumed to be a fixed real value. Thus, θ ~ Deg(θ_0). Then, this equation is trivial. If you plugin the actual value of $\theta = \theta_0$ on the r.h.s. then you get:

$$P(\theta = \theta_0|x) = \frac{P(x|\theta_0)(1)}{\sum_{\theta \in \Theta} P(x|\theta)P(\theta)} = \frac{P(x|\theta_0)}{P(x|\theta_0)} = 1$$

$$P(\theta \neq \theta_0|x) = \frac{P(x|\theta_0)(0)}{\sum P(x|\theta)P(\theta)} = \frac{0}{P(x|\theta_0)} = 0$$

This was a mean exam problem but not super interesting since you don't know θ_0 and even if you did, this doesn't help with the three goals of inference.

The big leap: let θ be a rv! Then P(θ) has a distribution (either discrete or continuous). But it's a constant! This is the big philosophical problem in Bayesian Statistics / Bayesian Inference. Some authors say it's still a constant but P(θ) represents uncertainty in its value. Purists say that's nonsense.

$$\underbrace{P(\theta|x)}_{\text{posterior}} = \frac{\overbrace{P(x|\theta)}^{\text{likelihood}}\ \overbrace{P(\theta)}^{\text{prior}}}{P(x)}$$

$P(x)$: prior predictive distribution

$$P(x) = \sum_{\theta \in \Theta} p(x|\theta)p(\theta) \text{ if } \theta \text{ discrete}$$

$$= \int_{\Theta} p(x|\theta)p(\theta)\,d\theta \text{ if } \theta \text{ continuous}$$

prior: thoughts summed up in a distribution over Θ the parameter space **before** to seeing any data. There is no x within in. Frequentists say this is "subjective" and not real!

posterior: thoughts summed up in a distribution over Θ the parameter space **after** seeing the data x which is why it's conditional on x!

Notation for the rest of class: "p" now denotes discrete PMF / conditional mass function **or** continuous PDF / conditional density function. I won't use "f" anymore.

$\mathcal{F}$ = iid Bernoulli, $x = \langle 0, 1, 1 \rangle$ $\quad P(x|\theta) = \theta^2(1-\theta)$

let $\Theta_0 = \{0.5, 0.75\} \neq (0,1)$

$$P(\theta = 0.75|x) \overset{?}{>} P(\theta = 0.5|x)$$

$$P(\theta = 0.75|x) \overset{\text{Bayes Thm}}{=} \frac{P(x|\theta = 0.75)P(\theta = 0.75)}{P(x|\theta = 0.5)P(\theta = 0.5) + P(x|\theta = 0.75)P(\theta = 0.75)}$$

$P(x|\theta = 0.75) = 0.75^2 \cdot 0.25 = .141$, $P(x|\theta = 0.5) = .5^3 = .125$

We need P(θ = 0.75) and P(θ = 0.5) to complete the calculation. That's the prior, P(θ). It's subjective. What do you think it should be? Amir says P(θ = 0.75) = 0.2 and P(θ = 0.5) = 0.8 because he feels that way.

An automatic rule is called the "principle of indifference" (Laplace's idea so it's sometimes called the "Laplace prior"). This principle says that all values of θ in the parameter space are equally likely. In our case,

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

In general, $P(\theta) = \frac{1}{|\Theta|}$ this formula only works for finite parameter spaces

$$P(\theta = 0.75|x) = \frac{.141 \cdot \cancel{.5}}{.125 \cdot \cancel{.5} + .141 \cdot \cancel{.5}} = 0.53^*$$

$\theta = 0.75$ is your pt. estimate

$$P(\theta = 0.5|x) = \frac{.125 \cdot \cancel{.5}}{.125 \cdot \cancel{.5} + .141 \cdot \cancel{.5}} = 0.47$$

$$P(\theta = 0.75) = 0.5 \xrightarrow{x} P(\theta = 0.75|x) = 0.53$$

This is called Bayesian Conditionalism