Lecture 4

02/06/20

$$\underbrace{P(Y|X)}_{P(X=x|Y=y)} = \frac{\overset{P(X=x|Y=y)}{P(X|Y)}\ \overset{P(Y=y)}{P(Y)}}{P(X) \rightarrow P(X=x)}$$

$$P(\theta|X) = \frac{\overbrace{P(X|\theta)}^{P(X=x|\theta=\theta)} P(\theta)}{P(X)} \quad P(\theta) = \begin{cases} 1 & \text{if } \theta = \theta_0 \\ 0 & \text{otherwise} \end{cases}$$

$P(X) \rightarrow$ No information about θ

Bayes Rule for a r.v. θ

Assume: θ was fixed ie. $\theta \sim Deg(\theta_0)$

$P(X=x|\theta=\theta)$ JMF, JDF, equal to likelihood

$$P(\theta) = \begin{cases} 1 & \text{if } \theta=\theta_0 \\ 0 & \text{otherwise} \end{cases} \Rightarrow P(\theta|X) = \begin{cases} \dfrac{P(X|\theta=\theta_0)}{P(X)} & \text{if } \theta=\theta_0 \\ 0 & \text{if } \theta \neq \theta_0 \end{cases}$$

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

$$\text{"} \int_{\Theta} P(X|\theta)P(\theta)\,d\theta$$

Assume θ is a non-degenerate r.v.

$$\underbrace{P(\theta|X)}_{\text{Posterior}} = \frac{\overbrace{P(X|\theta)}^{\text{likelihood}}\ \overbrace{P(\theta)}^{\text{Prior}}}{\underbrace{P(X)}_{\text{Prior Predictive Distribution}}}$$

Prior is your thoughts about θ before you see any data.

Posterior is your thoughts about θ after you see data.

Example

$\tilde{F}$ = iid Bernoulli ; $X = \langle 0,1,1 \rangle$

let $\underbrace{\Theta_0}_{\text{Subset of } \Theta = (0,1)} = \{0.5, 0.75\}$, $P(X|\theta) = \theta^2(1-\theta)$

$P(\theta = .75|X)$ $>$ $P(\theta = .5|X)$?

Why does this work?

$$P(\theta=0.75|X) = \frac{\overbrace{P(X|\theta=0.75)}^{.25(.75)^2}\,\overbrace{P(\theta=.75)}^{0.5}}{\underbrace{P(X|\theta=0.75)\,\underbrace{P(\theta=0.75)}_{0.5} + P(X|\theta=0.5)\,\underbrace{P(\theta=0.5)}_{0.5}}_{P(X) = \sum P(X|\theta)P(\theta)}}$$

$$P(\theta=0.5|X) = \frac{\overbrace{P(X|\theta=0.5)}^{(.5)^3}\,P(\theta=0.5)}{\underbrace{P(X|\theta=.75)P(\theta=0.75) + P(X|\theta=0.5)P(\theta=0.5)}_{P(X) \text{ is the same}}}$$

$P(X|\theta=.75) = .25(.75)^2 = .141$
$P(X|\theta=0.5) = (.5)^3 = .125$

we need $P(\theta=0.75)$ and $P(\theta=0.5)$

Principal of Indifference

All $\theta \in \Theta$ are equally likely i.e., $P(\theta) = \frac{1}{|\Theta|}$ $\forall\theta$ if θ is discrete

Here $P(\theta) = \begin{cases} \frac{1}{2} & \text{if } \theta = .5 \\ 0 & \text{otherwise} \end{cases}$

$$P(\theta=0.75|X) = \frac{(.141)(0.5)}{(.141)(0.5)+(.125)(0.5)} \doteq 0.53$$

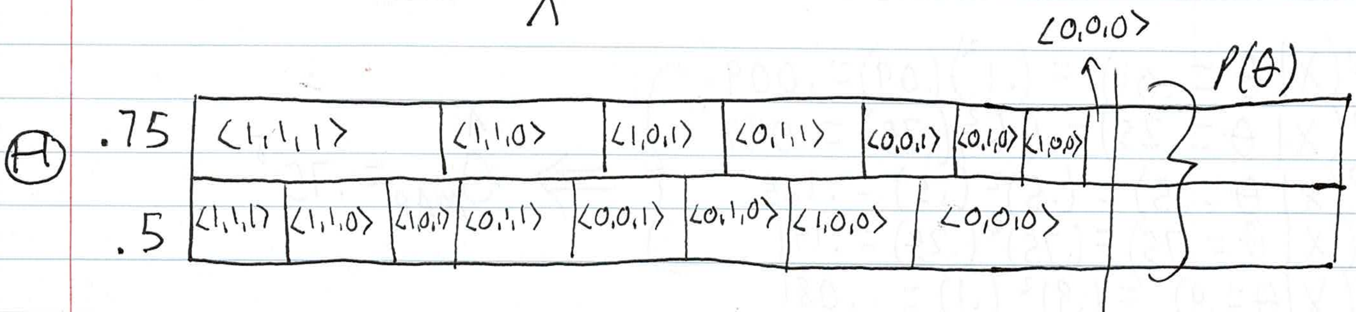
$$P(\theta = 0.5 \mid x) = \frac{.125}{.141 + .125} = 0.47$$

$$P(\theta) \xrightarrow[\text{(by knowing } x)]{x} P(\theta \mid x)$$

Bayesian Conditioning

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

$\mathcal{X}$



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

$$P(\theta \mid x) = \frac{P(x \mid \theta)\, P(\theta)}{P(x)} = \frac{P(x,\theta)}{\sum_{\theta \in \Theta} P(x,\theta)} = \frac{P(x,\theta)}{P(x)}$$

$\tilde{F}$ = iid Bernoulli $\quad \Theta_0 = \{.1, .25, .5, .75\}$ Subset of $\Theta = (0,1)$

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

Prior? $\rightarrow$ $P(\theta) = \begin{cases} .2 & \text{if } \theta \in \Theta \\ 0 & \text{otherwise} \end{cases}$

Principle of Indifference

What if I want the most likely value of θ, given X?

$$\hat{\theta}_{MLE} = \underset{\theta \in \Theta_0}{\operatorname{argmax}} \{P(\theta|x\} = \operatorname{argmax} \left\{ \frac{P(x|\theta)P(\theta)}{P(+)} \right\}$$

$$\hat{\theta}_{MAP} = \operatorname{argmax}\{P(X|\theta)P(\theta)\} = \underset{\theta \in \Theta_0}{\operatorname{argmax}} \{P(X|\theta)\}$$

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

Maximum a Posterior

Under the principle of indifference

$$= \hat{\theta}_{MLE} \text{ if } \Theta_0 = \Theta$$

$$\left.\begin{array}{l} P(X|\theta=.01) = (.1^2)(.09) = .009 \\ P(X|\theta=.25) = (.5^2)(.75) = .047 \\ P(x|\theta=.5) = (.5)^2(.5) = .125 \\ P(X|\theta=.75) = (.75)^2(.25) = .141 \\ P(X|\theta=.9) = (.9)^2(.1) = .081 \end{array}\right\} \Rightarrow \hat{\theta}_{MAP} = .75$$

$P \neq 1$ Not anything important

$$\sum_{x \in X} P(x) = 1 \qquad \sum_{\theta \in \Theta_0} P(X|\theta) = ?$$

$$\sum_{\theta \in \Theta_0} P(\theta) = 1 \qquad \sum_{\theta \in \Theta} P(\theta|X) = 1$$

$$P(\theta|x) = \frac{P(X|\theta)P(\theta)}{\sum_{\theta \in \Theta} P(X|\theta)P(\theta)} = \frac{P(X|\theta)P(\theta)}{P(\theta)\sum P(X|\theta)}$$

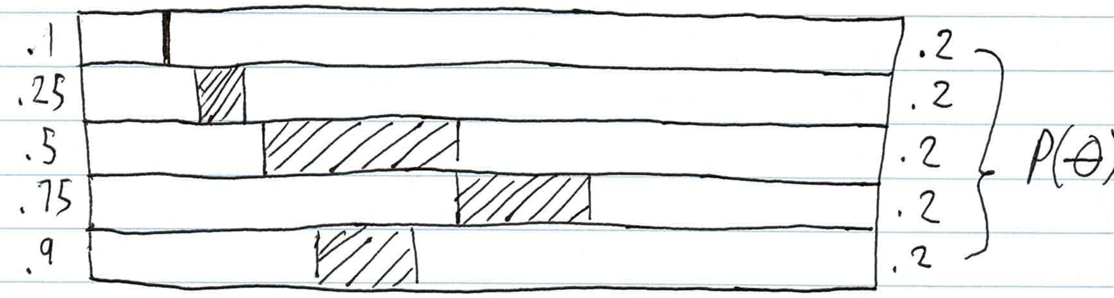
Under the Principle of indifference

$$P(\theta=.75|X) = \frac{.141}{.009+.047+.125+.141+.081} = \frac{.141}{.403} = .35$$

$$0.2 \xrightarrow{X} 0.35$$

Thought Before $P(\theta=.75)$

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



$\tilde{F}$ = iid Bernoulli ; $X = \langle 0, 1, 1 \rangle$

$\Theta_0 = \{0.5, .75\}$

$C\ \Theta = (0,1)$

$P(\theta = .75 \mid X_1, X_2, X_3)$

$P(\theta = .5)$

$$P(\theta = .75 \mid X_1 = 0) = \frac{P(X_1 = 0 \mid \theta = .75)\, P(\theta = .75)}{P(X_1 = 0 \mid \theta = .75)\, P(\theta = .75) + P(X_1 = 0 \mid \theta = .5)\, P(\theta = .5)}$$

$$= \frac{.25}{.25 + .5} = \frac{1}{3}$$

$$P(\theta = .75 \mid X_2 = 1) = \frac{2}{3}$$

After ~~X_T~~ seeing X_1 ...

Now my prior changes...

$$P(\theta) = \begin{cases} \frac{1}{3} & \text{if } \theta = .75 \\ \frac{2}{3} & \text{if } \theta = .25 \end{cases}$$

$$P(\theta \mid X_2) = \frac{P(X_2 \mid \theta)\, P(\theta)}{P(X_2)}$$