02/05/2020

Bayes Rule for 2 rv's X, Y

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

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

Assumed: θ was fixed i.e $\theta \sim Deg(\theta_0)$

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

$$P(\theta) = \begin{cases} 1 & \text{if } \theta=\theta_0 \\ 0 & \text{o/t} \end{cases} = P(\theta|X) = \begin{cases} \frac{P(X|\theta=\theta_0)}{P(x)}^{=1} & \text{if } \theta=\theta_0 \\ 0 & \text{if } \theta\neq\theta_0 \end{cases}$$

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

No info about ~~theta~~ θ $= \int_{\Theta} P(x|\theta)P(\theta)\,d\theta$

Assume: θ is a non-degenerate r.v.

$$P(\theta|X) = \frac{\overbrace{P(X|\theta)}^{\text{likelihood (joint mass)}} \; P(\theta)}{\underbrace{P(X)}_{\text{prior predictive distribution}}} = \frac{P(X,\theta)}{\sum_{\theta\in\Theta} P(X,\theta)} = \frac{P(X,\theta)}{P(X)}$$

$P(\theta)$ → prior (your thoughts about θ before you see any data)

$P(\theta|X)$ ↑ posterior (your thoughts about θ after you see data; x)

For example ...

$\mathcal{F}$ = iid Bernoulli but $\Theta_0 = \{0.5, 0.75\} \subset \Theta = (0,1)$

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

$$P(\theta=0.75|X) \; > \; P(\theta=0.5|X)$$

$$\frac{\overbrace{P(x|\theta=0.75)}^{0.25\ 0.75^2} P(\theta=0.75)}{\underbrace{P(x|\theta=0.75)P(\theta=0.75)+P(x|\theta=0.5)P(\theta=0.5)}_{P(x)=\sum_{\theta\in\Theta} P(x|\theta)P(\theta)}} \qquad \frac{\overbrace{P(x|\theta=0.5)}^{0.5^3} \overbrace{P(\theta=0.5)}^{0.5}}{\underbrace{P(x|\theta=0.75)P(\theta=0.75)+P(x|\theta=0.5)P(\theta=0.5)}_{P(x) \text{ is the same.}}}$$

$\Rightarrow P(\theta=0.75|x)=0.53$, $P(\theta=0.5|x)=0.47$

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

Principle of Indifference

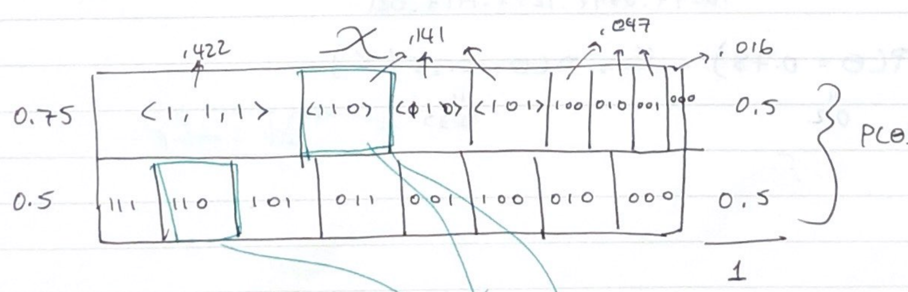
all $\theta\in\Theta$ are equally likely i.e. $P(\theta)=\frac{1}{|\Theta|}$

Here, $P(\theta) = \begin{cases} \frac{1}{2} & \text{if } \theta=0.5 \\ \frac{1}{2} & \text{if } \theta=0.75 \\ 0 & \text{if o/w} \end{cases}$

$P(\theta) \xrightarrow{x} P(\theta|X)$ Bayesian Conditional

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

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



$$P(\theta=0.75|x=\langle 1,1,0\rangle) = \frac{\boxed{\text{top}}}{\boxed{\text{bottom}}+\boxed{\text{top}}}$$

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

$\mathcal{F}$ = iid Bernoulli but $\Theta_0 = \{0.1, 0.25, 0.5, 0.75, 0.9\}$

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

Prior ? $P(\theta) = \begin{cases} 0.2 & \text{if } \theta \in \Theta_0 \\ 0 & \text{o/w} \end{cases}$ principle of indifference.

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

$$\hat{\theta} := \underset{\theta \in \Theta_0}{\operatorname{argmax}} \{ P(\theta | X) \} = \operatorname{argmax} \left\{ \frac{P(X|\theta)\, P(\theta)}{P(X)} \right\}$$

$$\hat{\theta}_{map} = \operatorname{argmax} \{ P(X|\theta)\, P(\theta) \} \underset{\text{under the principle of indifference}}{=} \operatorname{argmax} \{ \underbrace{P(X|\theta)}_{= \mathcal{L}(\theta; x)} \} = \hat{\theta}_{MLE} \quad \text{if } \Theta_0 = \Theta$$

Maximum a posterior

$P(\theta) = \frac{1}{|\Theta|} \neq f(\theta)$

$P(X=0|\theta=0.1)\, P(X=1|\theta=.1)\, P(X=1|\theta=.1)$

$\quad = P(X|\theta=0.1) = 0.1^2\, 0.9 = .009$

$P(X|\theta=.25) = 0.25^2\, 0.75 = .047$

$P(X|\theta=.5) = 0.5^2\, 0.5 = .125$

$P(X|\theta=.75) = 0.75^2\, .025 = .141$

$P(X|\theta=.9) = 0.9^2\, 0.1 = \underline{0.081}$

$\neq \Sigma 1$ ← not anything important

$\Rightarrow \hat{\theta}_{MAP} = 0.75$

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

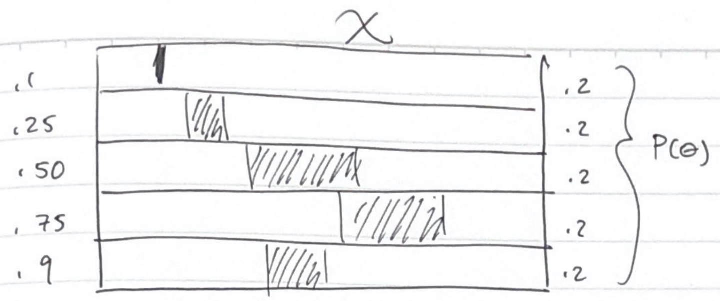
$\sum_{\theta \in \Theta_0} P(\theta) = 1 \qquad \sum_{\theta \in \Theta_0} P(\theta|X) = 1$

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

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

$$\underset{= 0.2}{P(\theta = 0.75)} \xrightarrow{X} \underset{= 0.35}{P(\theta = 0.75 | X)}$$

← best you can get with 3 data pts.



For example

$\mathcal{F}$ = iid Bernoulli but $\Theta_0 = \{0.5, 0.75\}$

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

Principle of indifference $\quad P(\theta = .75 \mid x_1, x_2, x_3)$

$P(\theta) = 0.5$

After seeing X_1...

$$P(\theta = 0.75 \mid X_1 = 0) = \frac{P(X_1 \overset{=0}{} \mid \theta = .75)\, \cancel{P(\theta = .75)}}{P(X_1 = 0 \mid \theta = .75)\, \cancel{P(\theta = .75)} + P(X_1 = 0 \mid \theta = .5)\, \cancel{P(\theta = .5)}} = \frac{1}{3}$$

$$P(\theta = .5 \mid X = 0) = \frac{2}{3}$$

Now my prior changes

$$P(\theta) = \begin{cases} \frac{1}{3} & \text{if } \theta = 0.75 \\ \frac{2}{3} & \text{if } \theta = 0.25 \end{cases}$$

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