Lecture 4:

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

Bayes rule for 2 r.v's x, y

$$P(y \mid X) = \frac{\overset{P(X=x,\,Y=y)}{P(X \mid Y)}\;P(Y)}{P(X)} \qquad P(Y) = P(Y=y), \quad P(X) = P(X=x)$$

Assume θ was fixed i.e. $\theta \sim \mathrm{Deg}(\theta_0)$

$\rightarrow P(X=x \mid \theta=\theta)$ Jmf, JDF. equal to likelihood

$$P(\theta) = \begin{cases} 1 & \text{if } \theta = \theta_0 \\ 0 & \text{otherwise} (\theta \notin t) \end{cases}$$

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

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

$$= \int_{\Theta} P(x \mid \theta)\,P(\theta)\,d\theta$$

no information about θ. Whole thing. $\boxed{\text{USELESS}}$

Assume θ in a non degenerate r.v:

$P(\theta|X)$.

$$P(\theta|X) = \frac{\overbrace{P(X|\theta)}^{\text{Likelihood}}\ \overbrace{P(\theta)}^{\text{Priore}}}{\underbrace{P(X)}_{\text{Priore predictive distribution}}}$$

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

Priore:- Youre thoughts about θ before you see any data.

Posteriore:- Youre thought about θ after You see the data, X.

For example; F = iid Bernoulli, But $\Theta = (0,1)$

Fore example

$$\mathcal{F} = \text{iid Bernoulli}, \quad \text{But}$$

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

$$\Theta_0 = \{0.5, 0.75\}$$

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

$$P(\theta = .75 \mid X) > P(\theta = 0.5 \mid X)$$

$$= \frac{\overbrace{P(X \mid \theta = .75)}^{.25 \times .75^2}\ \overbrace{P(\theta = .75)}^{.5}}{\underbrace{P(X \mid \theta = .75)P(\theta = .75) + P(X \mid \theta = .5)P(\theta = .5)}}$$

$$\hookrightarrow P(X) = \sum_{\theta \in \Theta} P(X \mid \theta) P(\theta)$$

$$\frac{\overbrace{P(X \mid \theta = .5)}^{.5^3}\ \overbrace{P(\theta = .5)}^{.5}}{P(X \mid \theta = .75)P(\theta = .75) + P(X \mid \theta = .5)P(\theta = .5)}$$

$P(X)$ is the same.

$\Rightarrow P(\theta = .75 \mid X) = .53$

$\Rightarrow P(\theta = .5 \mid X) = .47$

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

Assume Principle of indifference:-

all $\theta \in \Theta$ are equally likely i.e

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

$$P(\theta) = \begin{cases} \frac{1}{2} & \text{if } \theta = .5 \\ \frac{1}{2} & \text{if } \theta = .75 \\ 0 & \text{o/t} \end{cases}$$

$P(\theta) \xrightarrow{X} P(\theta \mid X)$ Baysian Conditionalism

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

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

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

$P(x = \langle 1,1,1\rangle \mid \theta = .75)$ $\mathcal{X}$ .141 0.047 0.016

.422

.75	$\langle 1,1,1\rangle$	$\langle 1,1,0\rangle$	$\langle 1,0,1\rangle$	$\langle 0,1,1\rangle$	$\langle 0,0,1\rangle$	$\langle 0,1,0\rangle$	$\langle 1,0,0\rangle$	$\langle 0,0,0\rangle$	.5	$P(\theta)$
.5	$\langle 1,1,1\rangle$	$\langle 1,1,0\rangle$	$\langle 0,1\rangle$	$\langle 0,1,1\rangle$	$\langle 0,0,1\rangle$	$\langle 0,1,0\rangle$	$\langle 0,1,0\rangle$	$\langle 1,0,0\rangle$	.5	

1

$$P(\theta = 0.75 \mid x = \langle 1,1,0\rangle) = \frac{\square}{\square + \square}$$

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

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

Another example:

F = iid Bernoulli, But

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

$$\underset{\downarrow \text{subset}}{}$$

$$\Theta_0 = \{ \underset{1}{0.1}, \underset{2}{0.25}, \underset{3}{0.5}, \underset{4}{0.75}, \underset{5}{0.9} \}$$

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

$$P(\theta) = \frac{1}{|\Theta_0|}$$

Prior? $P(\theta) = \begin{cases} \overset{1/5}{.2} & \text{if } \theta \in \Theta_0 \\ 0 & \text{o/t} \end{cases}$

Principle of indifference.

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

$$P(\theta|X) := \underset{\theta \in \Theta_0}{\arg\max} \{ P(\theta|X) \}$$

$$= \arg\max \left\{ \frac{P(X|\theta)\, P(\theta)}{P(X)} \right\}$$

MAP = Maximum of Posterior.

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

under the principle of indifference.

$$\overset{\uparrow}{=} \underset{\theta \in \Theta}{\operatorname{argmax}} \{ P(X|\theta) \}$$

$$\Downarrow$$

$$\mathcal{L}(\theta, X) = \hat{\theta}_{MLE}.$$

Pmf of Bernoulli

$= p^x (1-p)^x$

$$\boxed{\begin{matrix} p & \text{if } x=1 \\ 1-p & \text{if } x=0 \end{matrix}}$$

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

Standard Pmf of Bernoulli

$$P(X|\theta = 0.1) = 0.1^2 \times 0.9 = 0.009.$$

$$P(X|\theta = 0.25) = .25^2 \times .75 = 0.047$$

$$P(X|\theta = .5) = .5^2 \times .5 = 0.125$$

$$P(X|\theta = 0.75) = .75^2 \times .25 = .141$$

$$P(X|\theta = .9) = 0.9^2 \times 0.01 = 0.081$$

$$\hat{\theta}_{map} = 0.75$$

$\rightarrow$ Pmf $\qquad$ likelihood $\sum_{x \in \mathcal{X}} P(x|\theta) = 1$

$$\sum_{x \in \mathcal{X}} P(X) = 1, \quad \sum_{\theta \in \Theta_0} P(x|\theta) \overset{?}{=}$$

Pmf. $\sum_{\theta \in \Theta} P(\theta) = 1, \quad \sum_{\theta \in \Theta_0} P(\theta|x) = 1$

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

under principle of indifference

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

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

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

$$= \frac{.141}{.403} = 0.35$$

$$P(\theta = .75) \longrightarrow P(\theta = .75|x)$$

$\Downarrow$

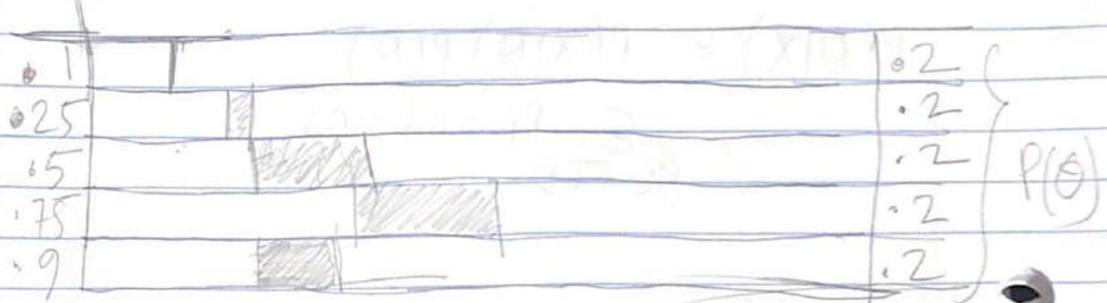
$.35$

$0.2 \xrightarrow{\quad X \quad} 0.35$

$\Downarrow$ $\qquad\qquad\qquad\qquad$ $\Downarrow$

$P(\theta = .75)$ $\qquad\qquad$ $p(\theta = .75 \mid X)$

without data $\qquad\qquad$ given data.



For example,

F = iid Bernoulli $\qquad$ But

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

$\Theta_0 = \{0.5, 0.75\}$

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

Principle of indifference

$P(\Theta) = 0.5$

$P(\theta = .75 \mid x_1, x_2, x_3)$

After seeing X,

$$P(\theta = 0.75 \mid X_1 = 0)$$

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

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

$\Rightarrow$ ~~$P(0.5 \mid$~~

$$P(\theta = 0.5 \mid X = 0) = 2/3.$$

Now my prior changes;

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

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