Lecture 5

$\mathcal{F}$: Bernoulli

let $\Theta = \{0.1, 0.25, 0.50, 0.75, 0.9\}$ $X = \langle 0,1,1 \rangle$

then $P(\theta)$ $\{0.1, 0.25, 0.5, 0.75, 0.9\}$ all w/ prob. $1/5$ by "principle of indifference"

$$P(X \mid \theta_1 = 0.1) = (0.9)(0.1)^2 = 0.009$$
$$P(X \mid \theta_2 = 0.25) = (0.75)(0.25)^2 = 0.047$$
$$P(X \mid \theta_3 = 0.50) = (.5)(.5)^2 = 0.125$$
$$P(X \mid \theta_4 = 0.75) = (0.25)(0.75)^2 = 0.141$$
$$P(X \mid \theta_5 = 0.90) = (0.1)(0.9)^2 = 0.081$$

each is in proportion to $1/5$ of each respective θ

$P(X) = \sum_{i=1}^{5} P(X \mid \theta_i) P(\theta_i)$ * we can solve $P(\theta_i \mid X)$ similar to previous.

- point estimation : best guess of θ
  $X = \langle 0,1,1 \rangle$

Note $\hat{\theta}_{MLE} = \frac{2}{3}$ (frequentism)

for Bayesian perspective we use largest $P(\theta \mid X)$

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

$P(\theta)$ is constant for all θ
$P(X)$ is also the same for all θ } so $P(\theta \mid X) \propto P(X \mid \theta)$

- so $\hat{\theta}_{MAP} := \underset{\theta \in \Theta}{\arg\max} \{P(\theta \mid X)\} = \underset{\theta \in \Theta}{\arg\max} \left\{ \frac{P(X \mid \theta) P(\theta)}{P(X)} \right\} = \underset{\theta \in \Theta}{\arg\max} \{P(X \mid \theta)\} = \hat{\theta}_{MLE}$

(maximum A posteriori)

uninformative prior

$f\{\theta ; X\}$ (from earlier)

$$\left[P(\theta_i \mid X) = \frac{P(X \mid \theta_i) P(\theta)}{P(X)} = \frac{P(X \mid \theta_i) P(\theta)}{\sum_{\theta \in \Theta} P(X \mid \theta) P(\theta)} = \frac{P(X \mid \theta_i)}{\sum_{\theta \in \Theta} P(X \mid \theta)} \right]$$

under principle of indifference $P(\theta_i) = P(\theta_k)$

* each $\theta \in \Theta$ has a (%) of occurring, the largest is the $\hat{\theta}_{MAP}$

$P(\theta = 0.75 \mid X = \langle 0,1,1 \rangle) = 37\%$

$\hat{\theta}_{MAP} = 0.75$; $\hat{\theta}_{MLE} = 0.66$

$\hat{\theta}_{MAP} \neq \hat{\theta}_{MLE}$ b/c $\Theta_0 \neq \Theta$

The prior didn't give probability of entire parameter space, we want a uniform prior to range over all θ

- Utilizing bayesian

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

$\Theta_0 = \{0.25, 0.75\}$ and $\theta \sim \text{Uniform}(\Theta_0)$

- start with x_1 or $\boxed{X_1 = 0}$

$$P(\theta_1 = 0.25 \mid x_1 = 0) = \frac{P(x_1 = 0 \mid \theta = 0.25)\ \overbrace{P(\theta_1)}^{=1/2}}{P(x_1=0 \mid \theta=0.25)\underbrace{}_{P(\theta_1)=1/2} + P(x_1 = 0 \mid \theta = 0.75)\underbrace{}_{P(\theta_2)=1/2}} = \frac{0.75}{0.75+0.25} = 0.75$$

$$P(\theta_2 = 0.75 \mid x_1 = 0) = 1 - \theta_1 = 1 - 0.75 = 0.25$$

value of $P(\theta)$ changes based on the new info.

- add next data point. x_2 or $\boxed{X_2 = 1}$

$$P(\theta_1 = 0.25 \mid x_1 = 0\ x_2 = 1) = \frac{P(X_2 \mid \theta_1)\ P(\theta_1 \mid X_1)}{P(X_1, X_2)} = \frac{(0.25)\cdot(0.75)}{P(X_2 \mid \theta_1)\,P(\theta_1 \mid X_1) + P(X_2 \mid \theta_2)\,P(\theta_2 \mid X_1)} = \frac{0.25 \cdot 0.75}{(0.25)(0.75) + (0.75)(0.25)} = 0.5$$

$$P(\theta_2 = 0.75 \mid x_1 = 0\ x_2 = 1) = 1 - 0.5 = 0.5$$

- Next data point x_3 or $\boxed{X_3 = 1}$

keeps adjusting to previous data

$$P(\theta_1 = 0.25 \mid x_1 = 0\ x_2 = 1\ x_3 = 1) = \frac{P(X_3 \mid \theta_1)\ P(\theta_1 \mid X_1 X_2)}{P(X_1\, X_2\, X_3)} = \frac{(0.25)(0.5)}{(0.25)(.5) + (0.75)(.5)} = 0.25$$

$$P(X_3 \mid \theta_1)\, P(\theta_1 \mid X_1 X_2) + P(X_3 \mid \theta_2)\, P(\theta_2 \mid X_1 X_2)$$

$$P(\theta_2 = 0.75 \mid \langle 0,1,1 \rangle) = 0.75$$

* these values are almost the same as find $P(\theta \mid X = \langle 0,1,1 \rangle)$ directly where instead we'd need to calculate $P(X \mid \theta)$ each time here we simply update the prior.

Proof/ $P(\theta | X) = \dfrac{P(X|\theta)\, P(\theta)}{P(X)}$ $\mathcal{X}$ = Bernoulli iid

$$P(\theta | x_1 \ldots x_n) = \frac{P(x_1 \ldots x_n | \theta)\, P(\theta)}{P(x_1 \ldots x_n)} = \frac{P(x_1|\theta)\, P(x_2 \ldots x_n | \theta)\; P(\theta)}{P(x_1)\;\; P(x_2 \ldots x_n | x_1)}$$

definition of conditional prob.

1st event

$$= \frac{P(x_2 \ldots x_n | \theta)}{P(x_2 \ldots x_n | x_1)} \cdot \left[\frac{P(x_1|\theta)\, P(\theta)}{P(x_1)} \right] \rightarrow P(\theta | x_1)$$

2nd event

$$= \frac{P(x_3 \ldots x_n | \theta)}{P(x_3 \ldots x_n | x_1 x_2)} \cdot \left[\frac{P(x_2|\theta) P(x_1|\theta) P(\theta)}{P(x_2|x_1)\; P(x_1)} \right] \rightarrow \frac{P(x_2 x_1 | \theta)\; P(\theta)}{P(x_1 x_2)}$$

$$\rightarrow P(\theta | x_1 x_2)$$

... and so on

- What about x_4? X_4 distribution $P(X_4|\theta) = \theta^{x_4}(1-\theta)^{1-x_4}$

  with $x = <0,1,1>$ $X_4 \sim$ Bern(?)

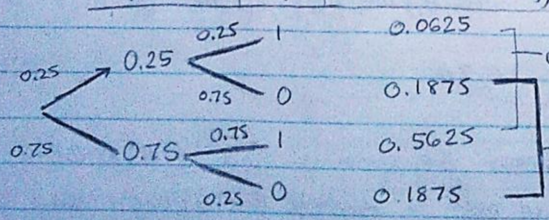
  $\hat{\theta}_{MLE} = 0.66$ but θ is unknown

  But too certain

  Needs uncertainty b/c X_4 is unknown

by Bayesian selection we can find $P(X_4 | x_1 x_2 x_3)$

$\theta / x_1 x_2 x_3$	X_4	$P(X_4\, \theta \mid x_1 x_2 x_3)$



0.625 $= P(X_4 = 1 | x_1 x_2 x_3)$

0.375 $= P(X_4 = 0 | x_1 x_2 x_3)$

So $X_4 | x_1 x_2 x_3 \sim$ Bern(0.625)

[possibility of getting $x_4 = 1$]

so

$$P(X_4 | x_1 x_2 x_3) = \sum_{\theta \in \Theta} P(X_4\, \theta | x_1 x_2 x_3)$$

$$= \sum_{\theta \in \Theta} P(X_4 | \theta, x_1 x_2 x_3)\, P(\theta | x_1 x_2 x_3)$$

$$= \sum_{\theta \in \Theta} P(X_4 | \theta)\, P(\theta | x_1 x_2 x_3)$$