Lec 5. Feb. 17

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

argmax → most likely for y value

maximum a posteriori

after data

side Note: 1 is heads

0 → tails

"is" one of three Bayesian point estimate we will study in this class.

F = iid Bern(θ), $n=3$, $x = \langle 0,1,1 \rangle$, $\Theta_0 = \{0.5, 0.75\}$

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

cartesian product (set product)

$\mathcal{X}$

$\theta = .75$	<1,1,1>	<1,1,0>	<1,0,1>	<0,1,1>	<1,0,0>	<0,1,0>	<0,0,1>	<0,0,0>	0.5
$\theta = .5$	<1,1,1>	<1,1,0>	<1,0,1>	<0,1,1>	<1,0,0>	<0,1,0>	<0,0,1>	<0,0,0>	0.5

Θ principal of indifference

$P(X = \langle 1,1,1 \rangle \mid \theta = 0.75)\ (0.75)^3 = .422$

$P(X = \langle 1,1,0 \rangle \mid \theta = 0.75) = (0.75)^2 \cdot (0.25) = 0.141$

$P(X = \langle 1,0,0 \rangle \mid \theta = 0.75) = 0.75 \cdot (0.25)^2 = 0.47$

$P(X = \langle 0,0,0 \rangle \mid \theta = 0.75) = (0.25)^3 = 0.016$

$P(X = \langle 1,0,0 \rangle) = P(X = \langle 1,0,0 \rangle, \theta = .75) + P(X = \langle 1,0,0 \rangle, \theta = .5) = .47 + .125$

$$P(\theta = 0.5 \mid X = \langle 1,0,0 \rangle) = \frac{P(\theta = 0.5,\ X = \langle 1,0,0 \rangle)}{P(X = \langle 1,0,0 \rangle)} = \frac{(.125).5}{(.047 + .125).5}$$

$\sum_{\theta \in \Theta} P(\theta) = 1$, $\sum_{\theta \in \Theta} P(\theta \mid x) = 1$, $\sum_{\theta \in \Theta} P(x \mid \theta)$ = could be anything

$P(\theta)$ is constant

Laplace's idea

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

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

$$\boxed{\overset{?}{=} \hat{\theta}_{MLE}}$$

if the MLE is in the parameter set you specify

Let $\Theta_0 = \{0.1, 0.25, 0.5, 0.75, 0.9\}$

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

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

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

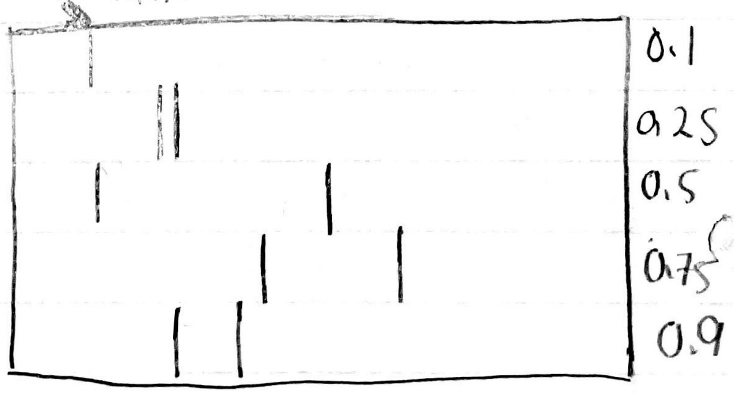
$P(X|\theta = 0.5) = .125$

$P(X|\theta = 0.75) = .141$

$P(X|\theta = 0.9) = 0.81$

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

$\hat{\theta}_{MLE} \in \Theta_0$



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

Let's examine Laplace's prior under many different parameter, Space approaching the full space.

$$\Theta_{0,3} = \{ 1/4, 1/2, 3/4 \} \Rightarrow P(\theta) = U(\Theta_0) = \{ 1/3 \ \forall \theta$$

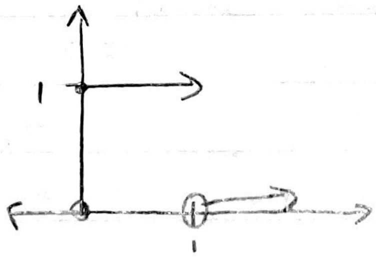
principal of indifference

$$\Theta_{0,2} = \{ 1/10, 2/10, \ldots 9/10 \} \Rightarrow P(\theta) = U(\theta) = \{ 1/2 \ \forall \theta.$$

$$\Theta_{0,n} = \{ \tfrac{1}{n+1}, \tfrac{2}{n+1}, \ldots, \tfrac{n}{n+1} \} \ P(\theta) = U(\theta) = \tfrac{1}{n} \ \forall \theta$$

$$(0,1) = \lim_{n \to \infty} \Theta_{0,n} \Rightarrow P(\theta) = 0 \ \forall \theta \quad \text{not a } \underline{\underline{PMF}}!$$

$$\lim_{n \to \infty} F(\theta) = \theta \Rightarrow P(\theta) = F'(\theta) = 1 \Rightarrow P(\theta) = U(0,1) \ \text{ex. continuous}$$



$$\mathcal{F} = \text{iid Bern}(\theta); \ X = \langle 1,1,0 \rangle, \ P(\theta = U(0,1)$$

$$P(\theta \mid X) = \frac{P(X \mid \theta) \cancel{P(\theta)}^{1}}{P(X)} = \frac{P(X \mid \theta)}{P(X)} = \frac{P(X \mid \theta)}{\int_{\Theta} P(X, \theta)\, d\theta} \rightarrow$$

$$\frac{P(X|\theta)}{\int_{\Theta} P(X|\theta)\, \cancel{P(\theta)}^{1}\, d\theta}$$

$$= \frac{\theta^2(1-\theta)}{\int_0^1 \theta^2(1-\theta)\, d\theta} = \frac{\theta^2(1-\theta)}{\left[\frac{\theta^3}{3} - \frac{\theta^4}{4}\right]_0^1}$$

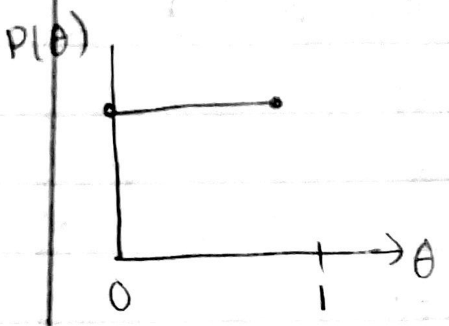
✓

$$= 12\,\theta^2(1-\theta), \quad \hat{\theta}_{MAP} = \operatorname{argmax}\{12\,\theta^2(1-\theta)\}$$

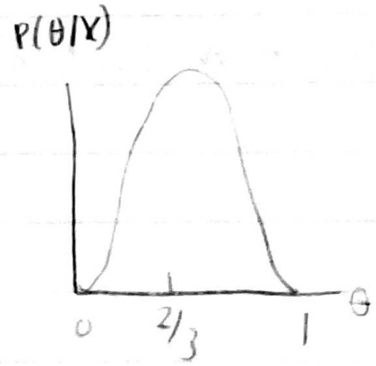
$$= \operatorname{argmax}\{\theta^2(1-\theta)\}$$

$$= \operatorname{argmax}\{\mathcal{L}(\theta; x)\} = 2/3$$

$$P(\theta) \xRightarrow{X} P(\theta|X)$$



$\Rightarrow$



$$P(\theta > 0.5|X) = \int_{0.5}^{1} 12\,\theta^2(1-\theta)\, d\theta = 12\left[\frac{\theta^3}{3} - \frac{\theta^4}{4}\right]_{0.5}^{1}$$

$$= 12\left(\frac{1}{12} - \right.$$

We talked about the MAP Bayesian point estimate. Are there other measurement of "best guess of θ" if you have the posterior distribution $P(\theta|X)$?

$\hat{\theta} := E[\theta|X]$

The minimum mean squared error Bayesian point estimate is the posterio mean (expectation). In our case:

Prior

$$\hat{\theta}_{MMSE} = \int_{\Theta} P(\theta) P(\theta|X) d\theta = \int_0^1 \theta \, 12 \, \theta^2 (1-\theta) d\theta = 12\left[\frac{\theta^4}{4} - \frac{\theta^5}{5}\right] = \frac{12}{20} = 0.6$$

Feb 22: Lec 6

$$\hat{\theta}_{MMAE} = \underset{\hat{\theta} \in \Theta}{\text{argmax}} \{F$$

minimum mean absolute error

$$\hat{\theta}_{MMAE} = \text{med}[\theta|X] = a \quad \text{such that} \quad \int_{-\infty}^{a} P(\theta|X) d\theta = \frac{1}{2}$$

Using our model: iid bern(θ) and data $X = \langle 0,1,1 \rangle$, we can compute the MMAE Bayesian point estimate:

$$\int_0^a 12\theta^2(1-\theta)d\theta = 12\left[\frac{\theta^2}{2} - \frac{\theta^3}{3}\right] = 12\left[\frac{\theta^3}{3} - \frac{\theta^4}{4}\right]$$

$$= 12\left[\frac{a^3}{3} - \frac{a^4}{4}\right] \overset{\text{set}}{=} \frac{1}{2} \rightarrow$$