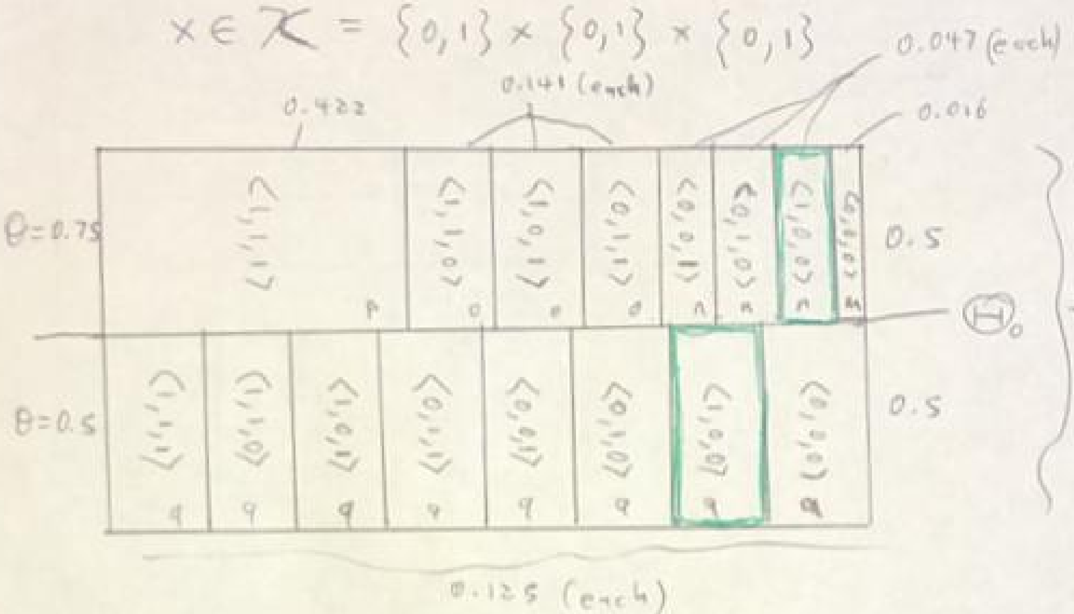
Lecture 5 2/17/21

$\hat{\theta}_{MAP} := \underset{\theta \in \Theta_0}{\text{argmax}} \{P(\theta \mid X)\}$ — is one of the three Bayesian point estimates we will study in this class.

maximum a posteriori → After data

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

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



— Principle of indifference

$p: P(X = \langle 1,1,1 \rangle \mid \theta = 0.75) = (0.75)^3 = 0.422$

$o: P(X = \langle 1,1,0 \rangle \mid \theta = 0.75) = (0.75)^2(.25) = 0.141$

"

"

$n: P(X = \langle 1,0,0 \rangle \mid \theta = 0.75) = (0.75)(.25)^2 = 0.047$

"

"

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

$q: P(X = \langle 1,1,1 \rangle \mid \theta = 0.5) = (0.5)^3 = 0.125$

$q: P(X = \langle 1,1,0 \rangle \mid \theta = 0.5) = (0.5)^3 = 0.125$

"

"

"

"

"

$q: P(X = \langle 0,0,0 \rangle \mid \theta = 0.5) = (0.5)^3 = 0.125$

$P(X = \langle 1,0,0 \rangle) = P(X = \langle 1,0,0 \rangle, \theta = 0.75) + \underbrace{P(X = \langle 1,0,0 \rangle, \theta = 0.5)}_{P(X = \langle 1,0,0 \rangle \mid \theta = 0.5) P(\theta = 0.75)} = (0.047)(.5) + (0.125)(.5)$

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

Application of law of total probability

$P(\theta = 0.5 \mid X = \langle 1,0,0 \rangle) = \dfrac{P(\theta = 0.5,\ X = \langle 1,0,0 \rangle)}{P(X = \langle 1,0,0 \rangle)} = \dfrac{(.125)\cancel{(.5)}}{(.047 + .0125)\cancel{(.5)}} \quad \dfrac{\square}{\square + \square}$

For discrete parameter spaces,

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

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

$P(\theta)$ is constant (Laplace's idea)

$$\Downarrow$$

$$\hat{\theta}_{MAP} = \underset{\theta \in \Theta_0}{\operatorname{argmax}} \{P(\theta|x)\} = \underset{\theta \in \Theta_0}{\operatorname{argmax}} \{P(x|\theta)P(\theta)\} = \underset{\theta \in \Theta_0}{\operatorname{argmax}} \{P(x|\theta)\} \overset{?}{=} \hat{\theta}_{MLE}$$

If the MLE is in the parameter set you specify.

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

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

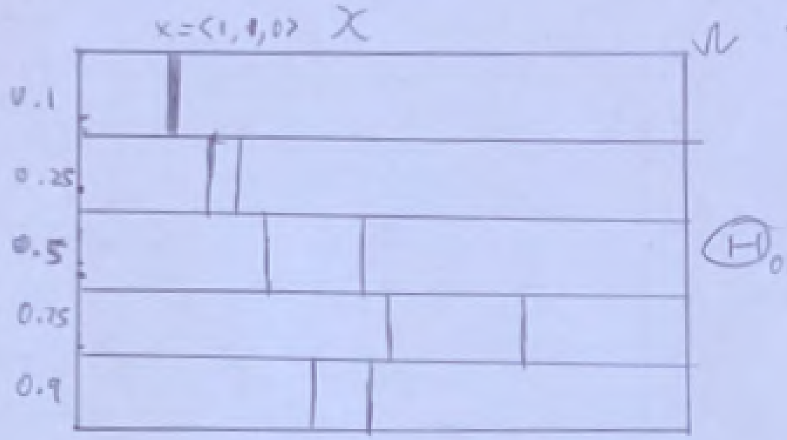
$P(x|\theta=0.1) = (0.1)^2(0.9) = 0.009$

$P(x|\theta=0.25) = (0.25)^2(0.75) = 0.047$

$P(x|\theta=0.5) = (0.5)^2(.5) = 0.125$

$P(x|\theta=0.75) = (.75)^2(.25) = 0.141$

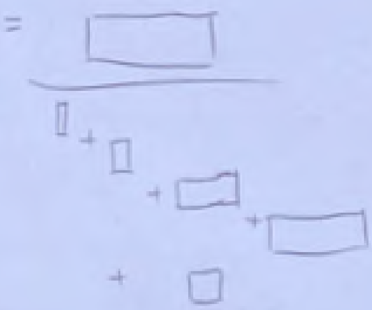
$P(x|\theta=0.9) = (0.9)^2(.1) = 0.081$



$$\Downarrow$$

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

$P(\theta=0.75 | x=\langle 1,1,0\rangle) =$



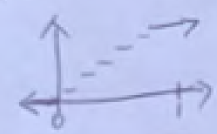
$\hat{\theta}_{MLE} = 0.67$

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

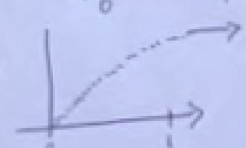
Let's examine Laplace's prior under many different parameter spaces approaching the full space.

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

$\Theta_{0,9} = \{\frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \cdots \frac{9}{10}\} \Rightarrow P(\theta) = U(\Theta_0) = \{\frac{1}{9} \ \forall\theta \longrightarrow$



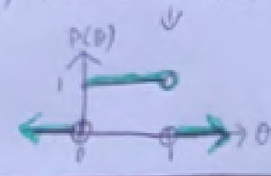
$\Theta_{0,n} = \{\frac{1}{n+1}, \frac{2}{n+1}, \cdots \frac{n}{n+1}\} \Rightarrow P(\theta) = U(\Theta_0) = \{\frac{1}{n} \ \forall\theta \longrightarrow$



Convergence in distribution

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

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



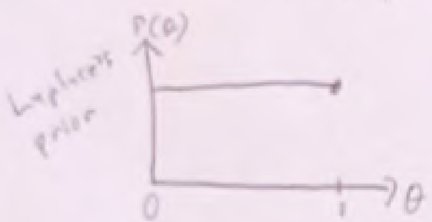
$\mathcal{F} = \text{iid Bern}(\theta)$, $x = \langle 1,1,0 \rangle$, $P(\theta) = U(0,1)$

$$P(\theta|x) = \frac{P(x|\theta)\overset{1}{P(\theta)}}{P(x)} = \frac{P(x|\theta)}{P(x)} = \frac{P(x|\theta)}{\int_{\Theta} P(x,\theta)\,d\theta} = \frac{P(x|\theta)}{\int_{\Theta} P(x|\theta)\cancel{P(\theta)}\,d\theta}$$

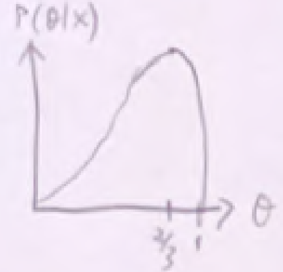
$$= \frac{\theta^2(1-\theta)}{\int_0^1 \theta^2(1-\theta)\,d\theta} = \frac{\theta^2(1-\theta)}{\left[\frac{\theta^3}{3} - \frac{\theta^4}{4}\right]_0^1} = \frac{\theta^2(1-\theta)}{\frac{1}{12}} = 12\theta^2(1-\theta),$$

$$\hat{\theta}_{MAP} = \text{argmax}\{12\theta^2(1-\theta)\} = \text{argmax}\{\theta^2(1-\theta)\}$$
$$= \text{argmax}\{\mathcal{L}(\theta; x)\} = \frac{2}{3}$$

$$P(\theta) \overset{x}{\Rightarrow} P(\theta|x)$$



$\overset{x}{\Rightarrow}$



$$P(\theta > 0.5|x) = \int_{0.5}^1 12\theta^2(1-\theta)\,d\theta = 12\left[\frac{\theta^3}{3} - \frac{\theta^4}{4}\right]_{0.5}^1 = 12\left(\frac{1}{12} - \left(\frac{1}{24} - \frac{1}{64}\right)\right) = 0.688$$

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

$$\hat{\theta}_{MMSE} := E[\theta|x] = \underset{\theta \in \Theta}{\text{argmin}}\left\{E[(\theta - \hat{\theta})^2|X]\right\}$$

The "minimum mean squared error" Bayesian point estimate is the posterior mean (expectation). This is the default estimator. In our case:

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