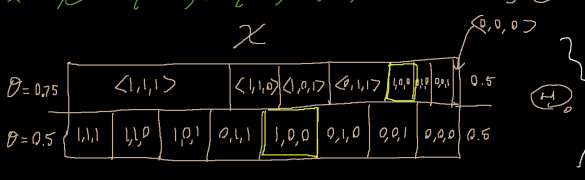
$\hat{\theta}_{MAP} := \underset{\theta \in \Theta_0}{\text{argmax}} \left\{ P(\theta | X) \right\}$ is one of three Bayesian point estimates we will study in this class

maximum a posteriori

after data

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

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




principle of indiff-erence

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

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

$P(X = \langle 1,0,0 \rangle | \theta = 0.75) = 0.75 \cdot 0.25^2 = 0.047$

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

$P(X = \langle 1,1,1 \rangle) = 0.5^3 = 0.125$

$P(X = \langle 1,0,0 \rangle | \theta = 0.5)\, P(\theta = 0.5)$

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

$P(X = \langle 1,0,0 \rangle) = P(X = \langle 1,0,0 \rangle, \theta = 0.75) + P(X = \langle 1,0,0 \rangle, \theta = 0.5) = \begin{matrix} .047 \cdot 0.5 + \\ .125 \cdot 0.5 \end{matrix}$

$P(\theta = 0.5 | X = \langle 1,0,0 \rangle) = \dfrac{P(\theta = 0.5, X = \langle 1,0,0 \rangle)}{P(X = \langle 1,0,0 \rangle)} = \dfrac{.125 \cdot \cancel{.5}}{(.047 + .125) \cdot \cancel{.5}}$

For discrete parameter spaces,

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

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

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

$\Downarrow$

$\hat{\theta}_{MAP} = \underset{\theta \in \Theta_0}{\text{argmax}} \left\{ P(\theta | X) \right\} = \underset{\theta \in \Theta_0}{\text{argmax}} \left\{ P(X | \theta) P(\theta) \right\} = \underset{\theta \in \Theta_0}{\text{argmax}} \left\{ P(X | \theta) \right\} = \hat{\theta}_{MLE}$

If the MLE is in the parameter set you specify

let $\Theta_0 = \{0.1, 0.25, 0.5, 0.75, 0.9\}$, Laplace Prior

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

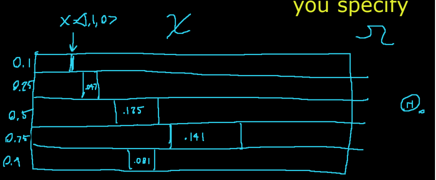
$P(X | \theta = 0.1) = 0.1^2\ 0.9 = .009$

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

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

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

$P(X | \theta = 0.9) = .081$




$\Downarrow$

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

$P(\theta = 0.75 | X = \langle 1,1,0 \rangle) = \dfrac{\square}{\square + \square + \square + \square + \square}$

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

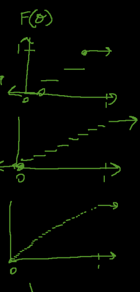
Let's examine Laplace's prior under many different parameter spaces approaching the full space.

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

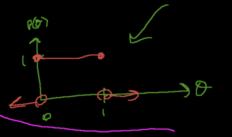
$\Theta_{0,9} = \left\{ \frac{1}{10}, \frac{2}{10}, \ldots, \frac{9}{10} \right\} \Rightarrow P(\theta) = U(\Theta) = \left\{ \frac{1}{9} \quad \forall \theta \right.$

$\Theta_{0,n} = \left\{ \frac{1}{n+1}, \frac{2}{n+1}, \ldots, \frac{n}{n+1} \right\} \Rightarrow P(\theta) = U(\Theta) = \left\{ \frac{1}{n} \quad \forall \theta \right.$

$\Theta = (0,1) = \lim_{n \to \infty} \Theta_{0,n} \Rightarrow P(\theta) = 0\ \forall \theta$ 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



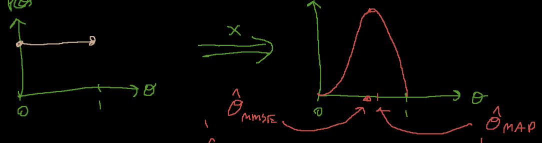

$\mathcal{F} =_{iid} \text{Bern}(\theta),\ X = \langle 1,1,0 \rangle,\ P(\theta) = U(0,1)$

$P(\theta | X) = \dfrac{P(X|\theta) \cancel{P(\theta)}^{1}}{P(X)} = \dfrac{P(X|\theta)}{P(X)} = \dfrac{P(X|\theta)}{\int_\Theta P(X, \theta)\, d\theta} = \dfrac{P(X|\theta)}{\int_\Theta P(X|\theta) \cancel{P(\theta)}^{1}\, d\theta} = \dfrac{\theta^2(1-\theta)}{\int_0^1 \theta^2(1-\theta)\, d\theta}$

$= \dfrac{\theta^2(1-\theta)}{\left[ \frac{\theta^3}{3} - \frac{\theta^4}{4} \right]_0^1} = 12\,\theta^2(1-\theta)$, $\quad \hat{\theta}_{MAP} = \text{argmax} \left\{ 12\,\theta^2(1-\theta) \right\} = \text{argmax} \left\{ \theta^2(1-\theta) \right\}$

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

$P(\theta) \overset{x}{\Longrightarrow} P(\theta | X)$




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

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

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

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

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