Lec 5 2/14/18 Math 341

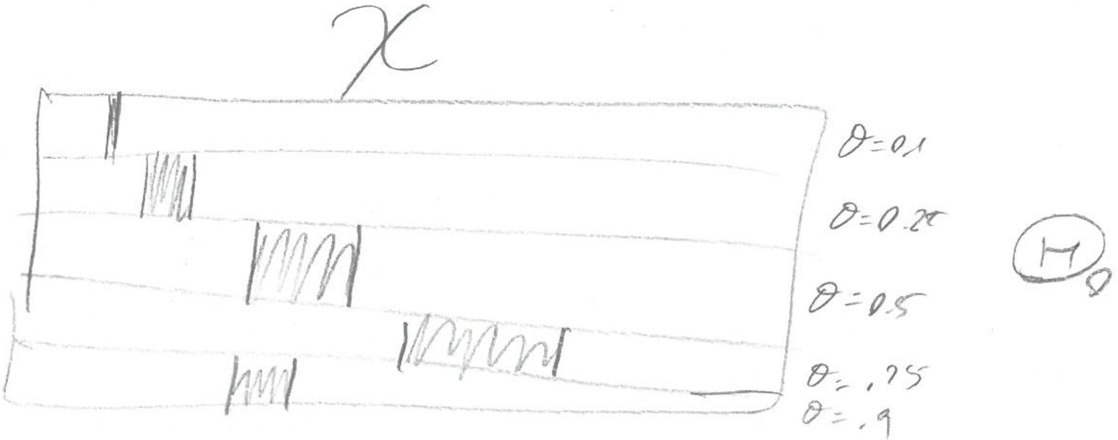
$\mathcal{F}$ = Bernoulli

$X = (0,1,1)$

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

$\theta \sim U(\Theta_0)$ i.e. discrete uniform on the elements of Θ_0 i.e. $P(\theta) = 0.2$ $\forall \theta \in \Omega$

We want to find $P(\theta \mid X)$. Before we do this... draw picture



$P(X \mid \theta)$ represents the prop. of area in any stripe.

$P(X \mid \theta = 0.1) = 0.009$

$P(X \mid \theta = 0.25) = 0.047$

$P(X \mid \theta = 0.5) = 0.125$

$P(X \mid \theta = .75) = 0.141$

$P(X \mid \theta = 0.9) = 0.061$

$P(\theta \mid X)$ is each of the slices div by the total sum of all slices

What is best model already? Biggest slice of the slices $\theta = .75$

Already you see its $P(\theta \mid X)$ is going to be the largest

Bigger size of the slices is a form of <u>pt. estimation</u> (best guess of Θ)! inference

$$\hat{\theta}_{MAP} = \underset{\theta \in \Theta_0}{\operatorname{argmax}} \left\{ P(\theta|x) \right\} = \underset{\theta \in \Theta_0}{\operatorname{argmax}} \left\{ \frac{P(x|\theta)P(\theta)}{P(x)} \right\}$$

Max. a posteriori Bayesian estimate (AKA "posterior mode")

$$= \underset{\theta \in \Theta_0}{\operatorname{argmax}} \left\{ P(x|\theta)P(\theta) \right\}$$ since $P(x)$ is a normalizing constant $\neq f(\theta)$

$$= \underset{\theta \in \Theta_0}{\operatorname{argmax}} \left\{ P(x|\theta) \right\}$$ since $P(\theta)$ is the same for all Θ_0 $\neq f(\theta)$

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

$$P(\theta|x) = P(x|\theta) \cdot \underbrace{P(\theta)}_{\text{scale based on prior belief (height)}} \cdot \underbrace{\frac{1}{P(x)}}_{\text{normalize so all } P(\theta|x)\text{'s add up to 1 (relative to other strips)}}$$

Under principle of indifference ..

$$P(\theta|x) = \frac{P(x|\theta)\cancel{P(\theta)}}{\sum_{\theta \in \Theta_0} P(x|\theta)\cancel{P(\theta)}} = \frac{P(x|\theta)}{P(x|\theta_1) + \dots P(x|\theta_k)}$$

use this here... $P(\theta = .75 | x = \langle 0,1 \rangle)$

$$= \frac{0.141}{0.009 + .047 + .125 + .141 + .061} = \frac{.141}{.383} \approx \boxed{37\%}$$

$\hat{\theta}_{MAP} = \hat{\theta}_{MLE}$ but $0.75 \neq 0.66$ Why? Our choice of prior did not cover all of the parameter space! $\Theta_0 \neq \Theta = (0,1)$ (for 7 = bundle)

Main reason to be skeptic of Bayesian stats: Prior could be wrong!

Let's look at data one at a time $\Theta = \{.25, .75\}$ (indifferent) $X_1 = 0$

What we know now ... prior and X_1

$$P(\theta = .25 \mid X_1 = 0) = \frac{P(X_1 = 0 \mid \theta = .25)}{P(X_1 = 0 \mid \theta = .25) + P(X_1 = 0 \mid \theta = .75)} = \frac{.75}{.75 + .25} = .75$$

let $P(\theta \mid X_1)$ be our new prior.

$\Rightarrow$ No longer indifferent!

$$\Rightarrow P(\theta = .75 \mid X_1 = 0) = .25 = 1 - P(\theta = .25 \mid X_1 = 0)$$

$X_2 = 1$

$$P(\theta = 0.25 \mid X_2 = 1) = \frac{P(X_2 = 1 \mid \theta = 0.25)\, P(\theta = 0.25 \mid X_1)}{P(X_2 = 1 \mid \theta = 0.25)\, P(\theta = 0.25 \mid X_1) + P(X_2 = 1 \mid \theta = 0.75)\, P(\theta = 0.75 \mid X_1)} = \frac{0.25 \cdot 0.75}{0.25 \cdot 0.75 + 0.75 \cdot 0.25} = 0.5$$

We're back to square 1. For this prior, no information learned. Make sense?

Now we know prior, X_1, X_2. Use this as prior

let $P(\theta \mid X_2, X_1)$ be new prior. Then $X_3 = 1$.

$$P(\theta = 0.25 \mid X_3 = 1) = \frac{P(X_3 = 1 \mid \theta = 0.25)\, P(\theta = 0.25 \mid X_2, X_1)}{P(X_3 = 1 \mid \theta = 0.25)\, P(\theta = 0.25 \mid X_2, X_1) + P(X_3 = 1 \mid \theta = 0.75)\, P(\theta = 0.75 \mid X_2, X_1)} = \frac{0.25 \cdot 0.5}{0.25 \cdot 0.5 + 0.75 \cdot 0.5} = 0.25$$

Same as $P(\theta = 0.25 \mid X = \langle 0, 1, 1 \rangle)$ from previously.

Is this true in general?

$$P(\theta \mid X_1, \ldots, X_n) = \frac{P(X_1, \ldots, X_n \mid \theta)\, P(\theta)}{P(X_1, \ldots, X_n)}$$

Why? iid

$$= \frac{P(X_n \mid \theta) \cdots P(X_2 \mid \theta) \boxed{P(X_1 \mid \theta)\, P(\theta)}}{P(X_n, \ldots, X_2 \mid X_1) \boxed{P(X_1)}} = P(\theta \mid X_1)$$

$$= \frac{P(X_n \mid \theta) \cdots P(X_3 \mid \theta) \boxed{P(X_1, X_2 \mid \theta)\, P(\theta)}}{P(X_n, \ldots, X_3 \mid X_1, X_2) \boxed{P(X_1, X_2)}} = P(\theta \mid X_1, X_2)$$

etc...

New question: we haven't seen X_4 yet. What is its distr?

Of course $P(X_4 \mid \theta) = \theta^{X_4}(1-\theta)^{1-X_4}$ but.. you don't know θ.

Previously ... what did you do?

$$P(X_4 \mid \theta) \approx P(X_4 \mid \theta = \hat{\theta}_{MLE} = 0.66) = \text{Bern}(0.66)$$

What's the problem? Uncertainty in $\hat{\theta}_{MLE}$ ...

Bayesian Solution:

Seek: $P(X_4 \mid X_1, X_2, X_3)$

posterior	$\theta \mid X_1, X_2, X_3$		X_4	$P(X_4, \theta \mid X_1, X_2, X_3)$
.25	0.25	0.25	1	.0625
		0.75	0	.1875
.75	0.75	0.75	1	.5625
		0.25	0	.1875

$.0625 + .5625 = .625$

$.1875 + .1875 = .375$

$$\Rightarrow P(X_4 \mid X_1 X_2 X_3) = \text{Bern}(.625)$$

this incorporates all uncertainty of θ assuming the prior & form.