Lec 4 Math 341 2/7/19

Let θ be a r.v. then this Bayes Rule makes sense:

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

likelihood; $p(\theta)$: prior: your thoughts about θ before you see any data

posterior: thoughts about θ after you see data

$p(x)$: prior predictive distr.

$$p(x) = \sum_{\theta\in\Theta} p(x|\theta)\,p(\theta) \quad \text{or} \quad \int_{\Theta} p(x|\theta)\,p(\theta)\,d\theta$$

$p(x|\theta) = p(x;\theta)$ — a r.v. with a value conditioned on; a constant in a formula

$= \mathcal{L}(\theta; x)$

example

F = iid bernoulli, $x = \langle 0,1,1\rangle$ $\quad p(x|\theta) = \theta^2(1-\theta)$

let $\Theta_0 = \{0.5, 0.75\}$ i.e. not the full parameter space.

$p(\theta = 0.75|x) \;(>)\; p(\theta = 0.5|x)$? Let's see how it works...

$$p(\theta=0.75|x) = \frac{p(x|\theta=0.75)\,p(\theta=0.75)}{p(x|\theta=0.75)\,p(\theta=0.75) + p(x|\theta=0.5)\,p(\theta=0.5)}$$

$p(x|\theta=0.75) = 0.75^2\,0.25 = .141$

$p(x|\theta=0.5) = 0.5^2\,0.5 = 0.125$

We need $p(\theta=0.75)$ & $p(\theta=0.5)$. That is $p(\theta)$! It is subjective! What do you think??

Principle of Indifference

all $\theta \in \Theta$ equally likely a priori i.e. $P(\theta) = \frac{1}{|\Theta|}$ $\forall \theta$ if θ is discrete

Here...

$$P(\theta) = \begin{cases} 0.5 \text{ if } \theta = 0.75 \\ 0.5 \text{ if } \theta = 0.5 \end{cases}$$

$$P(\theta = 0.75 | x) = \frac{(.141)(0.5)}{(.141)(0.5) + (.125)(.5)} = 0.53$$

$$P(\theta = 0.5 | x) = \frac{.125}{.141 + .125} = 0.47$$

After the data, the more likely parameter value is 0.75.

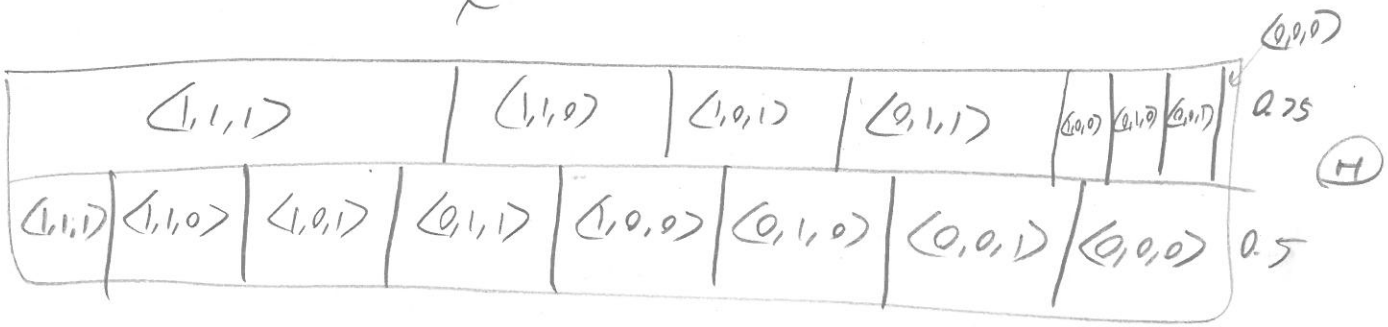
$$P(\theta = 0.75) = 0.5 \xrightarrow{X} P(\theta = 0.75 | x) = 0.53$$

Bayesian Confirmation

Let's look at the **full** space: all x and all Θ

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

$\mathcal{X}$



$P(X=\langle 1,1,1\rangle \mid \theta=.75) = .422$

$P(X=\langle 1,1,0\rangle \mid \theta=.75) = .141$

$P(X=\langle 1,0,0\rangle \mid \theta=.75) = .047$

$P(X=\langle 0,0,0\rangle \mid \theta=.75) = .016$

Is θ indep of X? NO

X tells you something about θ... that's the whole pt. of this course!

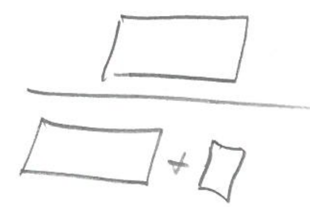
Cool questions:

$P(X=\langle 1,0,0\rangle, \theta=0.5) = .125 \cdot .5 = .0625$

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

$= .047 \cdot .5 + .125 \cdot .5 = .086$

$P(\theta=0.50 \mid X=\langle 1,0,0\rangle) = \frac{.0625}{.086} = 0.73$



$\sum_{\theta\in\Theta} P(\theta) = 1, \quad \sum_{\theta\in\Theta} P(\theta \mid x) = 1 \quad \sum_{\theta\in\Theta} P(x\mid\theta) = ?$ not guaranteed to be anything

---

Same data $X=\langle 0,1,1\rangle$ $\quad \Theta_0 = \{0.1, 0.25, 0.5, 0.75, 0.9\}$

$P(\theta) = \{0.2$ for $\forall \theta \in \Theta$

$P(\theta \mid x) = \frac{1}{P(x)} P(x\mid\theta) P(\theta) \propto P(x\mid\theta) P(\theta) \propto P(x\mid\theta)$

(always — over the first $\propto$; always a constant in θ under principle of indiff — over the second $\propto$)

View this as a function of θ (under $P(\theta \mid x)$); same for all θ (under $\frac{1}{P(x)}$)

$P(\theta \mid x) = P(x\mid\theta) \frac{1}{P(x)} P(\theta)$ — scale based on prior belief ($P(\theta)$); scale vs lots of data ($\frac{1}{P(x)}$)

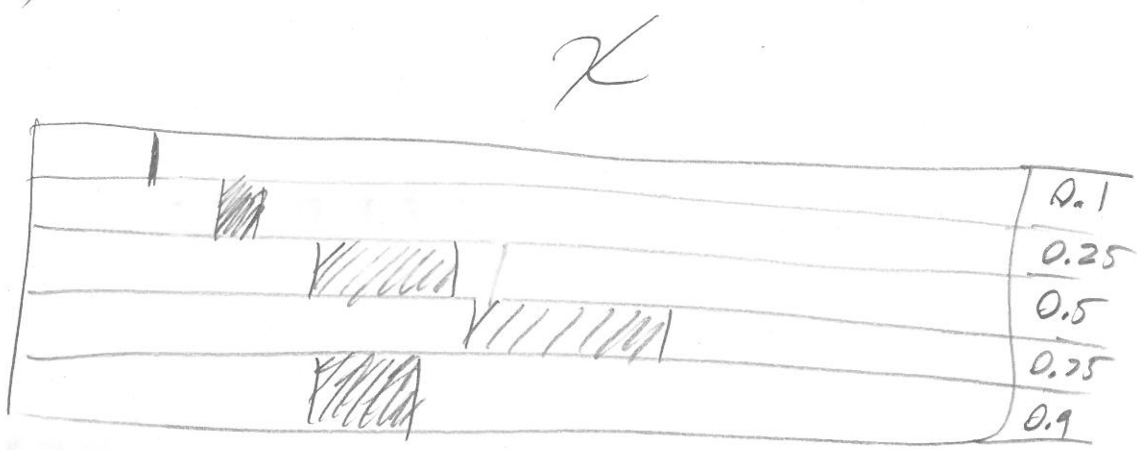
$P(X|\theta=0.1) = .1^2 \cdot .9 = .009$

$P(X|\theta=0.25) = .25^2 \cdot .75 = .047$

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

$P(X|\theta=.75) = \quad = .141$

$P(X|\theta=.9) = .9^2 \cdot .1 = .081$



If we want the most likely value of θ given the data, what is it? $\theta = 0.75$

always true since $P(X)$ const in θ

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

true if principled indifference

Maximum a posteriori (posterior mode) a Bayesian point estimate!

but $\hat{\theta}_{MLE} = 0.66 \neq 0.75$ What happened?

not true $\hat{\theta}_{MLE}$ requires search over full Θ. We're not there yet.

Let's Listen to $\Theta_0 = \{0.5, 0.75\}$, $X = (0, 1, 1)$

And let's look at our $P(\theta = 0.75)$ iteratively..

$$P(\theta = 0.75 \mid X_1 = 0) = \frac{P(X_1 = 0 \mid \theta = 0.75) \cdot P(\theta = 0.75)}{P(X_1 = 0 \mid \theta = 0.75) P(\theta = 0.75) + P(X_1 = 0 \mid \theta = 0.5) P(\theta = 0.5)}$$

Use this as prior

$$= \frac{0.25}{0.25 + 0.5} = \frac{1}{3}$$

0.75 $\quad$ $\frac{1}{3}$

$$P(\theta = 0.75 \mid X_2 = 1) = \frac{P(X_2 = 1 \mid \theta = 0.75) \, P(\theta = 0.75 \mid X_1 = 0)}{P(X_2 = 1 \mid \theta = 0.75) \, P(\theta = 0.75 \mid X_1 = 0) + P(X_2 = 1 \mid \theta = 0.5) \, P(\theta = 0.5 \mid X_1 = 0)}$$

$P(X_2 \mid X_1) = \sum_{\theta \in \Theta} P(X_2 \mid \theta, X_1) \, P(\theta \mid X_1)$

0.75 $\quad$ $\frac{1}{3}$ $\quad$ 0.5 $\quad$ $\frac{2}{3}$

$$= 0.429$$

0.75 $\quad$ $.429$

$$P(\theta = 0.75 \mid X_3 = 1) = \frac{P(X_3 = 1 \mid \theta = 0.75) \, P(\theta = 0.75 \mid X_1 = 0, X_2 = 1)}{P(X_3 = 1 \mid \theta = 0.75) P(\theta = 0.75 \mid X_1 = 0, X_2 = 1) + P(X_3 = 1 \mid \theta = 0.5) P(\theta = 0.5 \mid X_1 = 0, X_2 = 1)}$$

0.75 $\quad$ $.429$ $\quad$ 0.5 $\quad$ $.571$

$$= 0.53$$

$$P(\theta = 0.75) = 0.5 \xrightarrow{X_1} 0.33 \xrightarrow{X_2} 0.429 \xrightarrow{X_3} 0.53$$

Constant refinement

WTS $P(\theta | X_1, \dots X_n) = \dfrac{P(X_n|\theta)\, P(\theta | X_1, \dots, X_{n-1})}{\sum_{\theta \in \Theta} P(X_n|\theta)\, P(\theta | X_1, \dots, X_{n-1})}$ $\forall n$

Start with $P(\theta | X_1, \dots X_n) = \dfrac{P(X_1, \dots, X_{n-1}, X_n | \theta)\, P(\theta)}{P(X_1, \dots, X_{n-1}, X_n)}$

iid r.v's

$= \dfrac{P(X_1|\theta) \cdot \dots \cdot P(X_{n-1}|\theta)\, P(X_n|\theta)\, P(\theta)}{P(X_n | X_1, \dots, X_{n-1})\, P(X_1, \dots, X_n)}$

def of cond prob

$= \dfrac{P(X_n|\theta)\, P(X_1, \dots, X_{n-1}|\theta)\, P(\theta)}{P(X_n | X_1, \dots X_{n-1})\, P(X_1, \dots X_n)}$ $P(\theta | X_1, \dots X_{n-1})$

$P(X_n | X_1, \dots X_{n-1}) = \sum_{\theta \in \Theta} P(X_n, \theta | X_1, \dots, X_{n-1})$

Why? I can introduce another r.v. and marginalize over it.

$P(x) = \sum_{y \in \text{Supp}(Y)} P(x, y)$

Then I can condition on a third r.v. Z

$P(x|z) = \sum_{y \in \text{Supp}(Y)} P(x, y | z)$

$= \sum_{\theta \in \Theta} P(X_n | \theta, X_1, \dots, X_{n-1})\, P(\theta | X_1, \dots X_{n-1})$

Now $P(X_n \mid \theta, X_1, \ldots, X_{n-1}) = \dfrac{P(X_1, \ldots, X_{n-1}, X_n, \theta)}{P(X_1, \ldots, X_{n-1}, \theta)}$

$$= \frac{P(X_1, \ldots, X_{n-1}, X_n \mid \theta) P(\theta)}{P(X_1, \ldots, X_{n-1} \mid \theta) P(\theta)}$$

$$= \frac{P(X_1 \mid \theta) \cdot \ldots \cdot P(X_{n-1} \mid \theta) P(X_n \mid \theta)}{P(X_1 \mid \theta) \cdot \ldots \cdot P(X_{n-1} \mid \theta)} = P(X_n \mid \theta)$$

Anyone may you can see this?

If θ is known, then $X_1, \ldots, X_{n-1}$ do not give any more information to the distr. of X_n. θ completely specifies the distr.