Math 320

Lecture 4 2/10/16

"$X_1, X_2, X_3 \overset{iid}{\sim} \text{Bern}(\theta)$" correct? [1]

One at a time

$\langle 0, 1, 1 \rangle$

1st

$$P(\theta=0.25 \mid X_1=0) = \frac{P(X_1=0 \mid \theta=0.25)\, P(\theta=0.25)}{\underbrace{P(X_1=0)}_{P(X_1=0|\theta=0.25)P(\theta=0.25) + P(X_1=0|\theta=0.75)P(\theta=0.75)}}$$

$$= \frac{.75 \cdot .5}{.75 \cdot .5 + .25 \cdot \cancel{.5}} = .75$$

$$P(\theta=0.25) = 0.5 \longrightarrow P(\theta=0.25 \mid X_1) = 0.75$$

$\theta = 0.25$ is looking better!

Now pretend this is the prior...

$$P(\theta=0.25 \mid X_2=1) = \frac{P(X_2=1 \mid \theta=0.25)\, P(\theta=.25)}{P(X_2=1)}$$

$P(X_2=1) = P(X_2=1|\theta=0.25)P(\theta=.25) + P(X_2=1|\theta=.75)P(\theta=.75)$

$$= \frac{.25 \cdot .75}{.25 \cdot .75 + .75 \cdot .25} = .5$$

$$P(\theta=0.25) = 0.5 \longrightarrow P(\theta=0.25 \mid X_1, X_2) = .5$$

Our information didn't tell us anything!

Now pretend this is the prior

$$= .25 \cdot \cancel{.5}$$

$$P(\theta = 0.25 \mid X_3 = \phi) = \frac{P(X_3=\phi \mid \theta = 0.25)\, P(\theta = 0.25)}{P(X_3=\phi)} = .25$$

$$P(X_3=\phi) = P(X_3=\phi \mid \theta=.25)\, P(\theta=.25) + P(X_3=\phi \mid \theta=.75)\, P(\theta=.75)$$

$$= .75 \cdot \cancel{.5} + .25 \cdot \cancel{.5}$$

ALSO recall...

$$P(\theta = 0.25 \mid X_1 = 0, X_2 = 1, X_3 = 1) = .25$$

Is it possible that each piece of data comes in, you update your prior to the new posterior?? Etc.?

Let's see...

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

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

$$\cancel{= \frac{P(X_n \mid \theta) \cdots P(X_2 \mid \theta)}{P(X_n, \dots, X_2 \mid X_1)} \cdot \frac{P(X_1 \mid \theta)\, P(\theta)}{P(X_1)}}$$

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

prior of $X_n, \ldots, X_2$ now seeing X_1

this now takes the place of the prior

do one more...

$$\frac{P(X_3, \ldots X_n|\theta)}{P(X_3, \ldots, X_n|X_1, X_2)} \underbrace{\frac{P(X_2|\theta)}{P(X_2)} P(\theta|X_1)}$$

$$\underbrace{\frac{P(X_2|\theta)}{P(X_2)} \frac{P(X_1|\theta) P(\theta)}{P(X_1)}}$$

$$\underbrace{\frac{P(X_1, X_2|\theta) P(\theta)}{P(X_1, X_2)}}$$

$$P(\theta|X_1, X_2)$$ ← now the prior when considering $X_3, X_4, \ldots, X_n$

etc...

Explain $X \perp \theta \Rightarrow$ posterior = prior

explain why this is conceptually

explain why $X \not\perp \theta$ given F

Another concept... we have X_1, X_2, X_3. What does X_4 look like?

How would you do this before?

$$\hat{\theta}_{MLE} = \frac{0+1+1}{3} = 0.66 \Rightarrow X_4 \sim \text{Bern}(0.66)$$ ← You couldn't go further unluckily

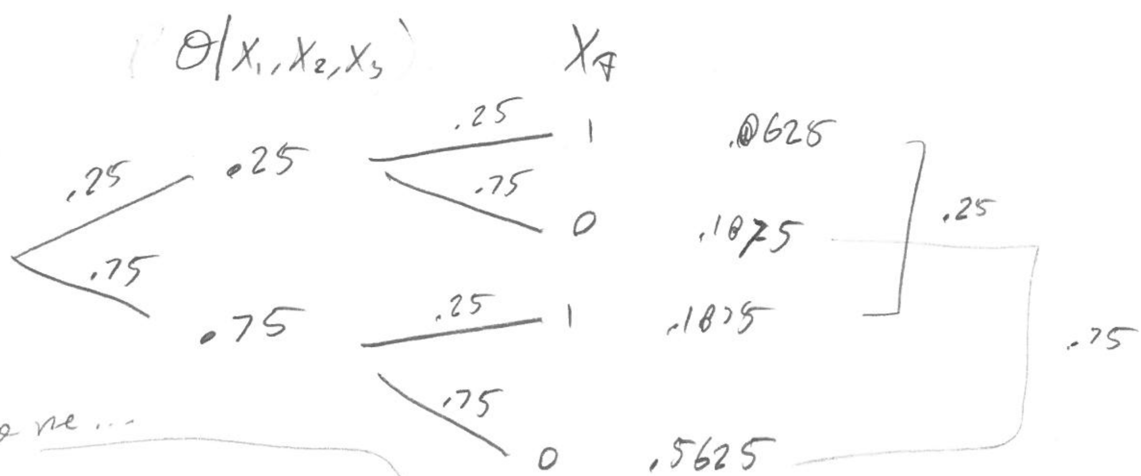
Or you can say $H_0: \theta = 0.5$ $H_a: \theta \neq 0.5$ $X = 0,1,1 \Rightarrow$ Retain H_0

$$\Rightarrow X_4 \sim \text{Bern}(0.5)$$

You guessed H_0!

In Bayesian paradise, we seek $P(X_4 \mid X_1, X_2, X_3)$

AKA the "posterior predictive distribution". I.e. seeing X_1, X_2, X_3, what does X_4 look like?



Unexpected to me...

$$X_4 \mid X_1, X_2, X_3 \sim \text{Bern}(0.25)$$

How to find this? Recall $P(Y \mid A,B) = \sum_{X \in Supp(X)} P(X, Y \mid A,B) = \sum P(X \mid Y, A,B)\, P(Y \mid A,B)$

I can condition on whatever I want everywhere

$$P(X_4 \mid X_1, X_2, X_3) = \sum_{\theta \in \Theta_0} P(X_4, \theta \mid X_1, X_2, X_3)$$

$$= \sum P(X_4 \mid \theta, X_1, X_2, X_3)\, P(\theta \mid X_1, X_2, X_3) = \sum_{\theta \in \Theta_0} P(X_4 \mid \theta)\, P(\theta \mid X_1, X_2, X_3)$$

$P(X_4 \mid \theta, X_1, X_2, X_3) = P(X_4 \mid \theta)$

Why?

genlly $P(X^* \mid X) = \sum_{\theta \in \Theta_0} P(X^* \mid \theta)\, P(\theta \mid X)$ -OR- $\int_{\Theta_0} P(X^* \mid \theta)\, P(\theta \mid X)\, d\theta$

$$P(X_4 \mid \theta, X_1, X_2, X_3) = \frac{P(\theta, X_1, X_2, X_3, X_4)}{P(\theta, X_1, X_2, X_3)}$$

$$= \frac{P(X_1, X_2, X_3, X_4 \mid \theta) \cancel{P(\theta)}}{P(X_1, X_2, X_3 \mid \theta) \cancel{P(\theta)}}$$

$$= \frac{\cancel{P(X_1, X_2, X_3 \mid \theta)} P(X_4 \mid \theta)}{\cancel{P(X_1, X_2, X_3 \mid \theta)}}$$

Why?? Once you know θ, the previous info is not relevant... you don't get anything out of it.

But $P(X_4 \mid X_1, X_2, X_3) \overset{?}{=} P(X_4)$???? Aren't the iid? $P(X \mid Y) = P(X)$ if $X \perp Y$?

Previously, $X_1, X_2, X_3 \overset{iid}{\sim} \text{Bern}(\theta)$? In Frequentist world this is true since the underlying assumption is θ is fixed

Now X_1, X_2, X_3 Not iid $\text{Bern}(\theta)$

Since if I know X_1, my θ has changed!

$$P(\theta) \rightarrow P(\theta \mid X_1)$$

So $\underbrace{X_1 \mid \theta,\ X_2 \mid \theta,\ X_3 \mid \theta}$ $\overset{iid}{\sim} \text{Bern}(\theta)$

You need the Cond'l r.v.'s now!

Let's return to the beginning

with I marble chosen before #15

$$\cancel{P(\theta)} \quad \theta \sim \begin{cases} 0.25 & \text{wp } \frac{1}{2} \\ 0.75 & \text{wp } \frac{1}{2} \end{cases} \Rightarrow \theta \mid X \sim \begin{cases} 0.25 & \text{wp } 0.25 \\ 0.75 & \text{wp } 0.25 \end{cases}$$

We said our $\hat{\theta}_{Bayes} = \hat{\theta}_{MAP} = \text{argmax}\{P(\theta|X)\} = 0.75$

What about $\hat{\theta}_{Bayes} = E[\theta|X] = \sum_{\theta \in \Theta_0} \theta \, P(\theta|X)$

AKA the "posterior mean"

-OR-

$$= \int_{\theta \in \Theta_0} \theta \, f(\theta|x) \, d\theta$$

In this case $= 0.25 \cdot 0.25 + 0.75 \cdot 0.75 = .625$

Note $.625 \notin \Theta_0$

Not in scope of course

Without getting carried away, $E(\theta|X) = \hat{\theta}_{MMSE}$

min mean sq err

$l(\hat{\theta}, \theta)$

$$= E_{P(\theta)}\left[(\hat{\theta}_{\cancel{MMSE}} - \theta)^2\right] = \int_{\Theta_0} (\hat{\theta} - \theta)^2 P(\theta) \, d\theta$$

minimizes this is i.e. a "Bayes Estimator"

So... this is a good pt. estimator if you care about sq'd err loss

$\hat{\theta}_{MAP}$ is not a ~~Bayes Estimator~~ except for a weird loss function and still iso unless θ is discrete. Most params are not $p, \mu, \sigma^2, \lambda$, etc

So... $.625 \notin \Theta_0$. Maybe its fine to use $\Theta_0 = \Theta = (0,1)$

How about $\theta \sim U(0,1)$. This is now continuous... every value in $(0,1)$ is represented and given equal prob. It's principle of indifference on steroids!

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

$= 1$ since $f(X;\, a=0, b=1) = 1\, \mathbb{1}_{x \in (0,1)}$

Why him????

What are these?? "params of prior := hyperparameters"

$$\int_\Theta P(X_1, X_2, X_3 \mid \theta)\, P(\theta) = \int_\Theta P(X_1 = 0 \mid \theta)\, P(X_2 = 1 \mid \theta)\, P(X_3 = 1 \mid \theta)\, P(\theta)\, d\theta$$

$$= \frac{P(X_1 = 0 \mid \theta)\, P(X_2 = 1 \mid \theta)\, P(X_3 = 1 \mid \theta)}{\int_0^1 \cdots \underbrace{P(\theta)}_{1}\, d\theta}$$

$$= \frac{(1-\theta)\cdot\theta\cdot\theta}{\int_0^1 \underbrace{(1-\theta)\,\theta^2}_{\theta^2 - \theta^3}\, d\theta}$$

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

$$\hat{\theta}_{MAP} = \underset{\theta \in \Theta}{\operatorname{argmax}} \{\ast\} = \underset{\theta \in (0,1)}{\operatorname{argmax}} \{\theta^2 - \theta^3\}$$

$2\theta - 3\theta^2 = 0$

$\Rightarrow 2 - 3\theta = 0$

$\Rightarrow \theta = \frac{2}{3} = \hat{\theta}_{MLE}$ cool!!

$$\hat{\theta}_{MMSE} = E[\theta \mid X] = \int_0^1 \theta\left(12(1-\theta)\theta^2\right) d\theta = 12\int_0^1 (\theta^3 - \theta^4)\, d\theta = 12\left[\frac{\theta^4}{4} - \frac{\theta^5}{5}\right]_0^1 = 12\left(\frac{1}{4} - \frac{1}{5}\right) = \frac{12}{20} = 0.6$$

One more Bayes Estimator... if $\ell(\theta, \hat{\theta}) = a|\theta - \hat{\theta}|$ i.e. absolute loss,

$$\Rightarrow \hat{\theta}_{MAE} = \text{Median}[\theta|x]$$

$$= \left\{\theta_0 : \int_0^{\theta_0} p(\theta|x)\, d\theta = 0.5\right\}$$

$$= \int_0^{y} 12(1-\theta)\theta^2\, d\theta = 0.5$$

$$\left[\frac{\theta^3}{3} - \frac{\theta^4}{4}\right]_0^{y} = \frac{1}{24}$$

$$\Rightarrow \frac{y^3}{3} - \frac{y^4}{4} = \frac{1}{24}$$

$$\Rightarrow 4y^3 - 3y^4 = \frac{1}{2}$$

⋮

Compute ~~4 results, 2 imaginary, 1 > 1~~

⋮

$$y \in \{0.614272, 1.24748, -0.26-0.36i, -0.26+0.36i\}$$

We have $\hat{\theta}_{MAP}$, $\hat{\theta}_{MMSE}$, $\hat{\theta}_{MAE}$ ... three estimates...

$$p(X_4|X_1, X_2, X_3) = \int_{\theta \in \Theta} p(X_4|\theta)\, p(\theta|X_1, X_2, X_3)\, d\theta$$

$$= \int_0^1 \theta^{X_4}(1-\theta)^{1-X_4}\, 12(1-\theta)\theta^2\, d\theta$$

$$= 12 \underbrace{\int_0^1 \theta^{X_4+2}(1-\theta)^{2-X_4}\, d\theta}_{B(X_4+3,\ 3-X_4)}$$

$$\Gamma(x) = \int_0^\infty x^{t-1} e^{-x}\, dx$$

$$= (x-1)! \text{ if } x \in \mathbb{N}$$

extending! to all $\mathbb{R}$

Beta Function

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$