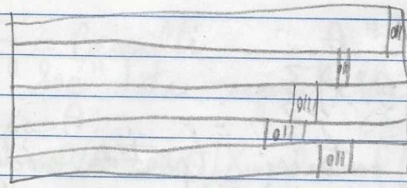


.1
.25
.5
.75 *
.9

these .75 is most likely model

$P(X|\theta = .1) = .009$ " $0.5 = .125$

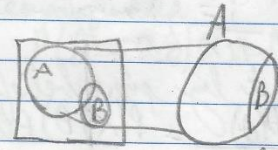
" $0.25 = .047$ " $.75 = .141$

" $.9 = .061$

$P(\theta = .75 | X = 011) = \ldots$

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

$P(\theta) = \frac{1}{k}$ df principle of indifference



$0 \propto 0$

My "best guess" in Bayesian framework is $\theta = .75$

$\hat{\theta}_{MLE} = 0.68 \neq 0.75$ Why?

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

2/14/17 ♡

F = Bernoulli

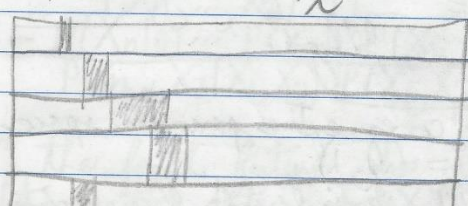
$X = \langle 0, 1, 1 \rangle \Rightarrow n = 3$

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

$\theta \sim U(\Theta_0)$

We want $P(\theta|x)$

x



$\theta = 0.1$
$\theta = .25$ Θ_0
$\theta = .5$
$\theta = .75$
$\theta = .9$

Prob. / likelihood $P(X|\theta)$

$P(X|\theta = .1) = .009$ $P(X|\theta = .5) = .125$ $P(X|\theta = .9) = .061$

$P(X|\theta = .25) = .047$ $P(X|\theta = .75) = .141$

Idea to find "best" θ.

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

maximum a posterior (posterior mode)

$$= \text{argmax} \left\{ \frac{P(X|\theta) P(\theta)}{P(X)} \right\} \quad \text{Bayes Rule}$$

$$= \text{argmax} \{ P(X|\theta) P(\theta) \} \quad \text{Because } P(X) \text{ is a constant} \neq f(\theta)$$

$\text{argmax}\{-cx^2\}$
$= \text{argmax}\{-x^2\}$

$$= \text{argmax} \{ P(X|\theta) \} \quad \text{due to the principle of indifference } P(\theta) \neq f(\theta)$$

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

$$P(\theta|X) = P(X|\theta) \cdot \underbrace{P(\theta)}_{\text{scale it by prior belief}} \cdot \frac{1}{P(X)}$$ normalization constant so all [illegible] sum to 1

$$P(\theta|X) = \frac{P(X|\theta) P(\theta)}{P(X)} = \frac{P(X|\theta) P(\theta)}{\sum_{\theta_0 \in \Theta_0} P(X, \theta_0)} = \frac{P(X|\theta) P(\theta)}{\sum_{\theta_0 \in \Theta_0} P(X|\theta_0) P(\theta_0)} =$$

Under principle of indifference....

$$= \frac{P(X|\theta)}{P(X|\theta_1) + \ldots + P(X|\theta_n)}$$

$$P(\theta = .75 | X) = \frac{0.141}{.009 + .047 + .125 + .141 + .061} = \frac{.141}{.383} \approx 37\%$$

$$\hat{\theta}_{MLE} = \hat{\theta}_{MLE}$$
$$.75 \neq .67$$

Why?

Prior did not cover entire parameter space

$$\Theta_0 = \Theta = (0, 1)$$

Main reason to be skeptical of Bayesianism: prior could be wrong

$$\Theta_0 = \{.25, .75\}$$

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

$F = $ Bernoulli $\qquad \theta \sim U(\Theta_0)$ points of indifference

"See" data sequentially

$X_1 = 0$

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

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

$$\Rightarrow P(\theta=.75 \mid x_1=0) = .25$$

Now... $X_2 = 1$

Let's let our prior here be the posterior from previous data

$$P(\theta=.25 \mid X_2=1) = \frac{P(X_2=1 \mid \theta=.25)\,P(\theta=.25 \mid X_1=0)}{\underbrace{P(X_2=1 \mid \theta=.25)\,P(\theta=.25 \mid X_1=0) + P(X_2=1 \mid \theta=.25)\,P(\theta=.25 \mid X_1=0)}_{P(X_2)}}$$

$$= \frac{(.25)(.5)}{\underbrace{.25+.5+.75+.5}_{P(X)}} = .25 \Rightarrow P(\theta=.75 \mid X=\langle 0,1 \rangle)$$

$$P(\theta \mid X_1,\dots,X_n) = \frac{P(X_1,\dots,X_n \mid \theta)\,P(\theta)}{P(X_1,\dots,X_n)} \qquad \text{Bayes Rule}$$

independence

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

Bayes Rule

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

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

X^* the fourth future observation

What is the dist. of X^4?

$$X^* \sim Bern(?)$$

the next unseen observation

Exam q about expectation

$P(X^*|\theta)$ Frequentist approach:

Previously...

$P(X^*|X_1, X_2, X_3) \approx P(X^*|\theta = \hat{\theta}_{MLE}) = \text{Bern}(.67)$

posterior predictive distribution

$\leftarrow$ inaccurate

Does not account for uncertainty in $\hat{\theta}_{MLE}$

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

Θ_0

$P(X^* = 1 | X_1, X_2, X_3) = .625 \Rightarrow X^*|X_1, X_2, X_3 \sim \text{Bern}(.625)$

$P(X^*|X_1, X_2, X_3) = \sum_{\theta \in \Theta_0} P(X^*, \theta | X_1, X_2, X_3)$

posterior predictive distribution

$= \sum_{\theta \in \Theta_0} P(X^*|\theta, X_1, X_2, X_3) P(\theta|X_1, X_2, X_3)$

Averaging out possible θs from tree

$= \sum_{\theta \in \Theta_0} P(X^*|\theta) P(\theta|X_1, X_2, X_3)$

$= \sum_{\theta \in \Theta_0} P(X^*|\theta) \frac{P(X_1, X_2, X_3|\theta) P(\theta)}{P(X_1, X_2, X_3)}$

Procedure for posterior pred. dist.:

① draw θ from posterior

② examine $X^*|\theta$

③ repeat for all θ's and avg.

$P(X^*|\theta) \stackrel{?}{=} P(X^*|\theta, X_1, X_2, X_3)$

$= \frac{P(X^*, X_1, X_2, X_3, \theta)}{P(X_1, X_2, X_3, \theta)}$

$= \frac{P(X^*, X_1, X_2, X_3|\theta)}{P(X_1, X_2, X_3|\theta)} \cdot \frac{\cancel{P(\theta)}}{\cancel{P(\theta)}}$

$= \frac{P(X^*|\theta) \cancel{P(X_1|\theta)} \cancel{P(X_2|\theta)} \cancel{P(X_3|\theta)}}{\cancel{P(X_1|\theta)} \cancel{P(X_2|\theta)} \cancel{P(X_3|\theta)}}$

$P(X^* | X_1, \ldots, X_n)$ posterior predictive distribution

$$= \sum_{\theta \in \Theta_0} P(X^* | \theta) P(\theta | X_1, \ldots, X_n)$$

$$\text{or } = \int_{\Theta_0} P(X^* | \theta) P(\theta | X_1, \ldots, X_n)\, d\theta \neq P(X^* | \hat{\theta})$$

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

if principle of idifference

$0.75 \neq 0.67$

$\Theta_0 \neq \Theta = (0, 1)$ if F = Bernoulli

What prior should we use?

$Supp[\theta] =$ Parameter Space of F

$= (0, 1)$

Has to be all possible values to ensure the answer is not excluded

An idea:

$\theta \sim U(0, 1)$ const. R.V.

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

$$P(\theta | X) = \frac{P(X|\theta) P(\theta)}{P(X)} \propto P(X|\theta) = \frac{\theta^2 - \theta^3}{\int_0^1 P(X|\theta) P(\theta)\, d\theta}$$

$P(\theta) := 1$ if $\theta \in [0, 1]$

If I care about the $\hat{\theta}_{MAP}$

$\hat{\theta}_{MAP} = \text{argmax} \{P(\theta | X)\}$

$= \text{argmax} \{P(X|\theta)\} = \text{argmax} \{\theta^2 - \theta^3\}$

if principle of indifference

How? Calc. 101

Take derivative set $= 0$

$$\frac{d}{d\theta}[\theta^2 - \theta^3] \overset{set}{=} 0$$

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

$2 - 3\theta = 0$

$\Rightarrow \hat{\theta}_{MAP} = .67 = \hat{\theta}_{MLE}$

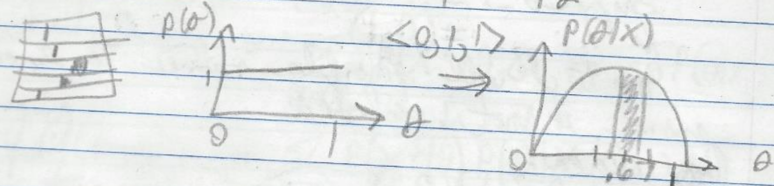
$P(\theta \in [0.6\ 0.7] \mid X)$ $= \boxed{.1765}$

frequentists can't do this $\quad 12(.0543 - .0396)$

$$= \int_{.6}^{.7} P(\theta|x)\,d\theta = \int_{.6}^{.7} 12(\theta^2-\theta^3) = 12\left[\frac{\theta^3}{3} - \frac{\theta^4}{4}\right]_{.6}^{.7}$$

$$\frac{\theta^2-\theta^3}{\int_0^1 P(X|\theta)P(\theta)\,d\theta} = \frac{\theta^2-\theta^3}{\int_0^1 (\theta^2-\theta^3)\,d\theta} = 12(\theta^2-\theta^3)$$

$$\frac{\theta^3}{3} - \frac{\theta^4}{4}\Big|_0^1 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$



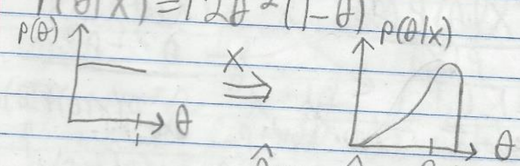
2/16/17

Maximal Apex Theory

F = Bernoulli $\quad X = \langle 0,1,1 \rangle$

$\theta \sim U(0,1)$

$P(\theta|X) = 12\theta^2(1-\theta)$



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

F = Bernoulli

$X_1, \ldots, X_n$ $\quad P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} = \frac{P(X|\theta)P(\theta)}{\int_{\Theta} P(X|\theta)P(\theta)\,d\theta}$ $\quad P(\theta) = 1$

$\theta \sim U(0,1)$ $\quad$ all data

$$P(X|\theta) = \prod_{i=1}^{n} P(X_i|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

$$= \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

$= \theta^{X}(1-\theta)^{n-X}$