Lecture 5

02/11/20

$\mathcal{F}$: iid Bernoulli

$\Theta = \{0.5, 0.75\} \subset (0,1)$ $\quad X = \langle \underset{X_1}{0}, \underset{X_2}{1}, \underset{X_3}{1} \rangle$

$$P(\theta \mid X_1 = 0) = \begin{cases} \frac{1}{3} & \text{if } \theta = 0.75 \\ \frac{2}{3} & \text{if } \theta = 0.5 \end{cases}$$

$$P(\theta = 0.75 \mid X_2 = 1) = \frac{\overbrace{P(X_2 = 1 \mid \theta = 0.75)}^{.75}\ \overbrace{P(\theta = 0.75)}^{\frac{1}{3}}}{\underbrace{P(X_2 = 1 \mid \theta = 0.75)}_{.75}\underbrace{P(\theta = 0.75)}_{\frac{1}{3}} + \underbrace{P(X_2 = 1 \mid \theta = .5)}_{0.5}\underbrace{P(\theta = .5)}_{\frac{2}{3}}}$$

$= .429$

Check this...

$$* P(\theta = .75 \mid X_3 = 1) = \frac{P(X_3 = 1 \mid \theta = .75)}{P(X_3 = 1 \mid \theta = 0.75) P(\theta = .75) + P(X_3 = 1 \mid \theta = 0.5) P(\theta = 0.5)}$$

$= .53^{*}$

$$P(\theta \mid X_3 = 1) = \begin{cases} .53 & \text{if } \theta = 0.75 \\ .47 & \text{if } \theta = 0.5 \end{cases}$$

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

Generally, we want to show:

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

~~Start with full formula:~~

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

Start with full formula:

$$P(\theta|\underbrace{X_1,\dots,X_n}_{X}) = \frac{P(\overbrace{X_1,\dots,X_n}^{X}|\theta)P(\theta)}{\underbrace{P(X_1,\dots,X_{n-1},X_n)}_{P(X)}} = \rightarrow$$

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

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

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

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

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

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

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

(maximum a posterior estimate)

$$= \underset{\theta \in \Theta_0}{\operatorname{argmax}} \{ P(X|\theta) P(\theta) \}$$

$$= \underset{\theta \in \Theta_0}{\operatorname{argmax}} \{ P(X|\theta) \}$$

This is true if $P(\theta)$ is determined by principle of indifference

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

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

for the iid Bernoulli $\tilde{F}$

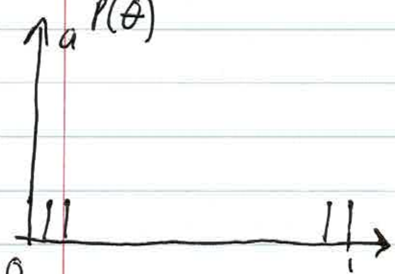
Why is $\Theta_0 = \{ \theta_1, \theta_2, \dots \}$ a bad idea?

$\neq (0,1)$

$$\Theta_0 = \{ 0, \tfrac{1}{4}, \tfrac{1}{2}, \tfrac{3}{4}, 1 \} \quad P(\theta) = \{ \tfrac{1}{5} \ \forall \theta$$

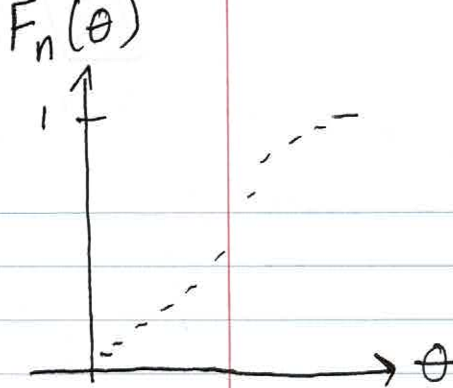
$$\Theta_0 = \{ 0, \tfrac{1}{10}, \dots, \tfrac{9}{10}, 1 \} \quad P(\theta) = \{ \tfrac{1}{11} \ \forall \theta$$

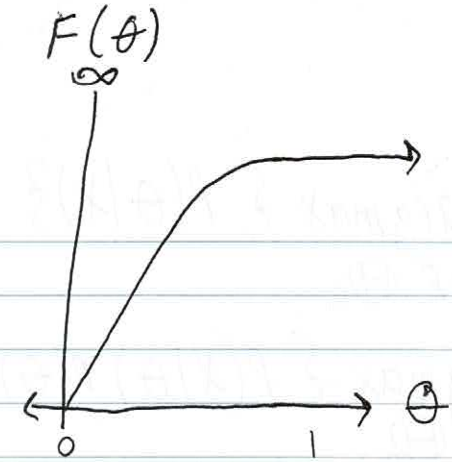
$$\Theta_{0(n)} = \{ 0, \tfrac{1}{n}, \dots, \tfrac{n-1}{n}, 1 \} \quad P(\theta) = \frac{1}{n-1} \ \forall \theta$$



$$\lim_{n \to \infty} P_n(\theta) = 0 \Rightarrow \Theta_{0(\infty)}$$ is not a discrete r.v. anymore

$$\lim_{n \to \infty} F_n(\theta) = \begin{cases} \theta & \text{if } \theta \in (0,1) \\ 0 & \text{if } \theta < 0 \\ 1 & \text{if } \theta > 1 \end{cases}$$





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

Principle of indifference

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

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} = \frac{P(X|\theta)P(\theta)}{\int_{\Theta_0} P(X|\theta)P(\theta)d\theta}$$

not discrete anymore

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

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

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

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

$$\hat{\theta}_{MAP} = \underset{\theta \in (0,1)}{\operatorname{argmax}} \{12\theta^2(1-\theta)\}$$

$$= \underset{\theta \in (0,1)}{\operatorname{argmax}} \{\theta^2(1-\theta)\}$$

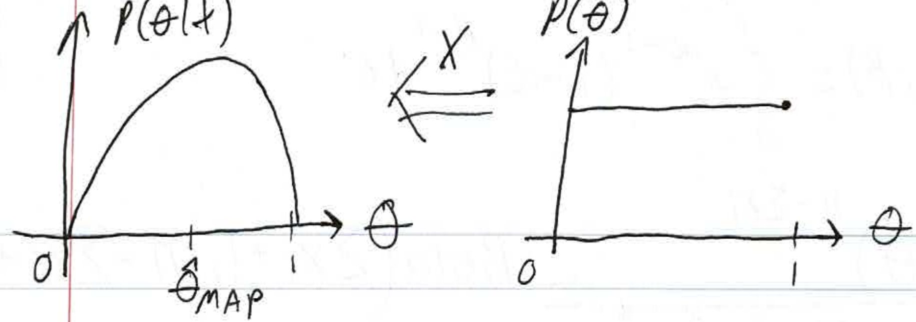
$f(\theta) = \theta^2 - \theta^3$

$f'(\theta) = 2\theta - 3\theta^2 \xRightarrow{set=0} \theta(2-3\theta) = 0$

$2 = 3\theta$

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

$= \bar{X} = \hat{\theta}_{MLE} F(\theta)$



$$P(\theta > 0.5|X) = \int_{0.5}^{1} 12\theta^2(1-\theta)\,d\theta = 12\left[\frac{\theta^3}{3} - \frac{\theta^4}{4}\right]_{.5}^{1} = .688$$

Posterior Median

$$\hat{\theta}_{MMAE} = Med[\theta|X] = a \quad s.t.$$

(Minimum mean absolute error)

$$\int_{-\infty}^{a} P(\theta|X)\,d\theta = \frac{1}{2}$$

Posterior Expectation

$$\hat{\theta}_{MMSE} = E[\theta|X] = \underset{\theta \in \Theta}{\arg\max} \{ \hat{\theta} - \theta^2 \}$$

(Minimum mean squared error)

$$\hat{\theta}_{MMAE} = Med[\theta|X] = \underset{\theta \in \Theta}{\arg\max} \{ |\hat{\theta} - \theta| \}$$

$X_1, \dots X_n$ (general case)

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

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} = \frac{P(X|\theta)P(\theta)}{\int_{\Theta} P(X|\theta)P(\theta)\,d\theta}$$

$$= \frac{P(X|\theta)}{\int_0^1 P(X|\theta)\,d\theta} = \frac{\theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}}{\underbrace{\int_0^1 \theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}}_{\text{Famous Integral}}} =$$

Continued on next page

<u>Beta Function</u> : $B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt$

From previous page $= \dfrac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}{B(\sum x_i + 1, n-\sum x_i + 1)} = \text{Beta}(\sum x_i + 1, n - \sum x_i + 1)$

$\rightarrow$ A new r.v.

$$Y \sim \text{Beta}(\alpha,\beta) = \frac{1}{B(\alpha,\beta)} = \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1}$$

$$\text{Supp}[Y] = (0,1) \Rightarrow \int_0^1 \frac{1}{B(a,\beta)} y^{\alpha-1}(1-y)^{\beta-1}\,dy = 1$$

parameter space $\alpha, \beta > 0$

$$E[Y] = \int_0^1 y \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1}\,dy$$

$$= \frac{1}{B(\alpha,\beta)} \int_0^1 y^{(\alpha+1)-1}(1-y)^{\beta-1}\,dy$$

$$= \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)}$$

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}\,dt \quad \text{for } \alpha > 0$$

factorial function for all positive reals

<u>Facts</u>

1) $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$
2) $B(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+B)}$

$$= \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+1)\,\Gamma(\beta)}{\Gamma(\underbrace{\alpha+1+\beta}_{\text{Check this ?}})} \Big/ B(a,\beta) =$$

by Fact 1:

$$\frac{\dfrac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)}}{\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}} = \frac{\alpha}{\alpha+\beta}$$

Median [Y] = no closed form
MMSE

To calculate in R: qbeta(0.5, α, β)

$$\text{Mode}[Y] = \underset{y\in(0,1)}{\operatorname{argmax}} \left\{ \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1} \right\}$$
MAP

$$= \operatorname{argmax} \left\{ y^{\alpha-1}(1-y)^{\beta-1} \right\}$$

$$= \operatorname{argmax} \left\{ \underbrace{(\alpha-1)\ln(y) + (\beta-1)\ln(1-y)}_{\Rightarrow f(y)} \right\}$$

$$f'(y) = \frac{\alpha-1}{y} - \frac{\beta-1}{1-y} \overset{set}{=} 0$$

$$\Rightarrow Y_{mode} = \frac{\alpha-1}{\alpha+\beta-2}$$

If being careful: $f''(y_{mode}) < 0$ if $\alpha, \beta > 1$

$$Y \sim Beta(\underset{\alpha}{1}, \underset{\beta}{1}) = \frac{1}{B(1,1)} y^{(1)-1}(1-y)^{(1)-1}$$

$$= \frac{1}{\int_0^1 y^{(1)-1}(1-y)^{(1)-1}} (1)(1)$$

$$= \frac{1}{\int_0^1 (1)(1)\,dy} = 1 \Rightarrow Y \sim U(0,1) = Beta(1,1)$$

$$\downarrow$$

Standard Uniform
is a special case
of the Beta distribution