Lecture 05   2-11-20

F ii Bernoulli

$\Theta_0 = \{0.5, 0.25\} \subset (0,1)$

$X = \langle \underset{x_1}{0}, \underset{x_2}{1}, \underset{x_3}{1} \rangle$

$$P(\Theta \mid x_1 = 0) = \begin{cases} 1/3 & \text{if } \theta = 0.25 \\ 2/3 & \text{if } \theta = 0.5 \end{cases}$$

$$P(\theta = 0.25 \mid x_2 = 1) = \frac{\overbrace{P(x_2 = 1 \mid \theta = 0.75)}^{0.25}\ \overbrace{P(\theta = 0.25)}^{1/3}}{\underbrace{P(x_2 = 1 \mid \theta = 0.75)}_{0.25}\ \underbrace{P(\theta = 0.75)}_{1/3} + \underbrace{P(x_2 = 1 \mid \theta = 0.5)}_{0.5}\ \underbrace{P(\theta = 0.5)}_{2/3}}$$

$= 0.429$

$$P(\Theta \mid x_2 = 1) = \begin{cases} .429 & \text{if } \theta = 0.75 \\ .571 & \text{if } \theta = 0.5 \end{cases}$$

$P(\theta = 0.25 \mid x_3 = 1)$

$$= \frac{\overbrace{P(x_3 = 1 \mid \theta = 0.25)}^{0.75}\ \overbrace{P(\theta = 0.25)}^{0.429}}{\underbrace{P(x_3 = 1 \mid \theta = 0.25)}_{0.75}\ \underbrace{P(\theta = 0.25)}_{0.429} + \underbrace{P(x_3 = 1 \mid \theta = 0.5)}_{0.5}\ \underbrace{P(\theta = 0.5)}_{0.571}}$$

$$P(\theta|x) = \frac{P(x|\theta)\,P(\theta)}{P(x)} = \frac{P(x|\theta)\,P(\theta)}{\sum_{\theta \in \Theta} P(x|\theta)\,P(\theta)}$$

$= 0.53$

$$P(\theta|x_3=1) = \begin{cases} 0.53 & \text{if } \theta = 0.75 \\ 0.47 & \text{if } \theta = 0.5 \end{cases}$$

(X) Generally we want to show:—

$$P(\theta|x_1,\ldots,x_n) = \frac{P(x_n|\theta)\,P(\theta|x_1,\ldots,x_{n-1})}{\sum_{\theta\in\Theta} P(x_n|\theta)\,P(\theta|x_1,\ldots x_{n-1})}$$

start with full formula

$$P(\theta|\underbrace{x_1,\ldots,x_n}_{X}) = \frac{P(\overbrace{x_1,\ldots,x_n}^{X}|\theta)\,P(\theta)}{\underbrace{P(x_1,\ldots x_{n-1},x_n)}_{P(X)}}$$

$$= \frac{P(x_1|\theta)\ldots P(x_{n-1}|\theta)\,P(x_n|\theta)\,P(\theta)}{P(x_n|x_1,\ldots,x_{n-1})\,P(x_1,\ldots x_{n-1})}$$

$$= \frac{P(x_n|\theta)\,P(x_1,\ldots,x_{n-1}|\theta)\,P(\theta)}{P(x_n|x_1,\ldots x_{n-1})\,\underbrace{P(x_1,\ldots x_{n-1})}_{}}$$

$P(\theta|x_1,\ldots,x_{n-1})$

$$P(x_n|x_1,\dots,x_{n-1}) = \sum_{\theta\in\Theta} P(x_n,\theta|x_1,\dots,x_{n-1})$$

$$P(x) = \sum_y P(x,y) = \sum_{\theta\in\Theta} P(x_n|\theta,x_1,\dots,x_{n-1})\, P(\theta|x_1,\dots,x_{n-1})$$

$$P(x_n|\theta,x_1,\dots,x_{n-1}) = \frac{P(x_1,\dots,x_{n-1},x_n,\theta)}{P(x_1,\dots,x_{n-1},\theta)}$$

$$\parallel$$

$$P(x_n \| \theta)$$

$$= \frac{\cancel{P(x_1|\theta)}\cdots\cancel{P(x_{n-1}|\theta)}\,P(x_n|\theta)}{\cancel{P(x_1|\theta)}\cdots\cancel{P(x_{n-1}|\theta)}}$$

$$= P(x_n|\theta)$$

x $\hat{\theta}_{map} := \underset{\theta\in\Theta_0}{\arg\max}\{P(\theta|x)\} = \underset{\theta\in\Theta_0}{\arg\max}\{P(x|\theta)\,P(\theta)\}$

$= \underset{\theta\in\Theta_0}{\arg\max}\{P(x|\theta)\}$

If $P(\theta)$ is determined by the principal of indifference

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

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

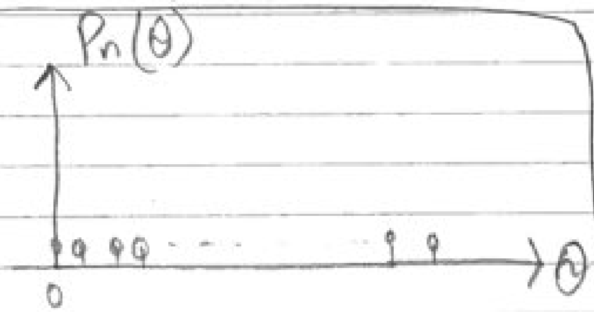
for the iid Bernoulli

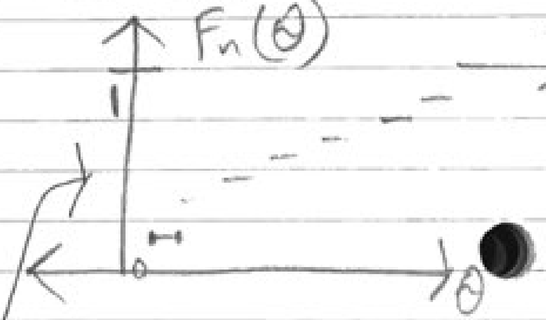
(Q) Why is $\Theta_0 = \{\theta_1, \theta_2, \ldots\} \neq (0,1)$ a bad idea?

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

$\Theta_0 = \{0, \frac{1}{10}, \ldots, \frac{9}{10}, 1\}$ $P(\theta) = \{\frac{1}{11} \ \forall \theta$

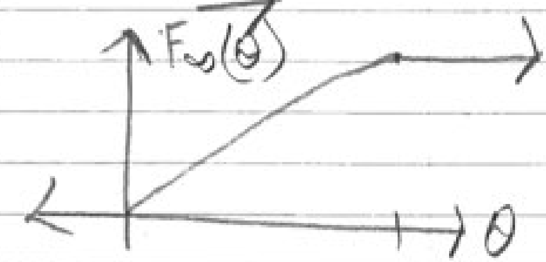
$\Theta_0 = \{0, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1\}$ $P(\theta) = \frac{1}{n+1} \ \forall \theta$





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

$\xRightarrow{n \to \infty}$



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

$P(\theta) = U(0,1)$

default prior — Standard uniform (Cont.)

Principles of Indifference

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

$\mathcal{F}$ = iid Bernoulli

fun one x

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

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

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

$$= \frac{P(x|\theta)}{\int_0^1 P(x|\theta)\,d\theta}$$

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

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

density func

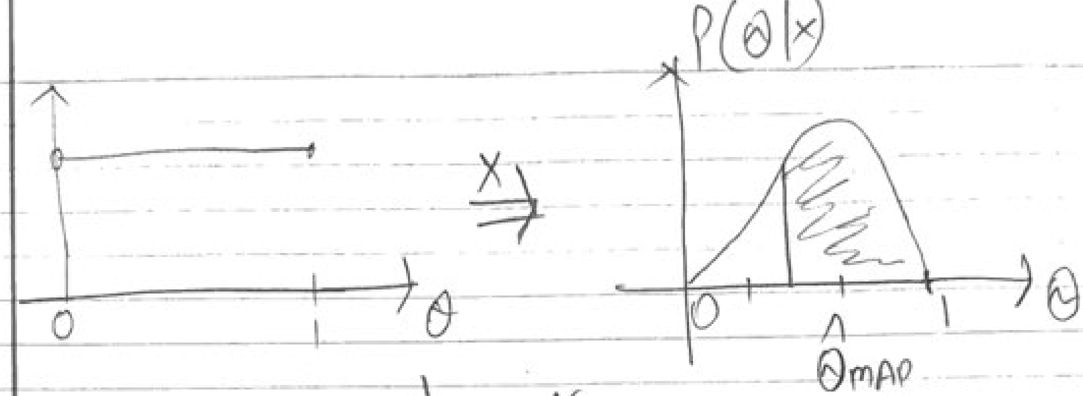
$$\hat{\theta}_{MAP} = \underset{\theta \in (0,1)}{\arg\max} \left\{12\,\theta^2(1-\theta)\right\} = \underset{\theta \in (0,1)}{\arg\max} \left\{\theta^2(1-\theta)\right\}$$

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

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

$\Rightarrow 2 - 3\theta = 0 \quad \Rightarrow 3\theta = 2 \quad \Rightarrow \hat{\theta}_{MAP} = \frac{2}{3}$

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



Posterior mode

$$P(\theta > 0.5|x) = \int_{0.5}^{1} 12\theta^2(1-\theta)\,d\theta$$

$$= 12\left[\frac{\theta^3}{3} - \frac{\theta^4}{4}\right]_{0.5}^{1}$$

Posterior Median

$$= 0.688$$

$$\hat{\theta}_{MMAE} := Med[\theta|x] = \underset{\theta \in \Theta}{\arg\min}\{|\hat{\theta} - \theta|\}$$

minimum mean absolute error

Posterior expectation/mean

$$\hat{\theta}_{MMSE} := E[\theta|x] = \underset{\theta \in \Theta}{\arg\min}\{(\hat{\theta} - \theta)^2\}$$

Minimum mean squared error

(X) $x_1, \ldots, x_n$ (general case)

$$P(\theta|x) = \frac{P(x|\theta)\,P(\theta)}{P(x)} = \frac{P(x|\theta)\,P(\theta)}{\int P(x|\theta)\,P(\theta)\,d\theta}$$

$$B(\alpha,\beta) := \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt$$

Beta function

(H)

$$= \frac{P(x|\theta)}{\int_0^1 P(x|\theta)\,d\theta}$$

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

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

$$= \frac{\theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}}{B(\Sigma x_i+1,\, n-\Sigma x_i+1)}$$

a new R.V.

$$= \mathrm{Beta}(\Sigma x_i+1,\, n-\Sigma x_i+1)$$

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

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

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 needs.

$(3.7)!$ not defined

Facts:

① $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$

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

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

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

$\text{Med}[Y] =$ no closed form
$q\text{beta}(0.5, \alpha, \beta)$

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

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

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

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

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

If you're carefull
$f''(y_{mode}) < 0$ if $\alpha, \beta > 1$

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

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

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