Lecture 5:

$\mathcal{F}$: iid Bernoulli

$\Theta = \{0.5, 0.75\} \subset (0,1)$

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

$$P(\theta \mid x_1 = 0) = \begin{cases} 1/3 & \text{if } \theta = .75 \\ \frac{2}{3} & \text{if } \theta = .5 \end{cases}$$

$$P(\theta = 0.75 \mid x_2 = 1) = \frac{\overbrace{P(x_2 = 1 \mid \theta = 0.75)}^{.75}\,\overbrace{P(\theta = 0.75)}^{1/3}}{\underbrace{P(x_2 = 1 \mid \theta = 0.75)}_{.75}\,P(\theta = 0.75) + \underbrace{P(x_2 = 1 \mid \theta = .5)}_{.5}\,\underbrace{P(\theta = .5)}_{2/3}}$$

Picture

$$= .429$$

$$P(\theta \mid x_2 = 1) = \begin{cases} .429 & \text{if } \theta = .75 \\ .571 & \text{if } \theta = .5 \end{cases}$$

$$P(\theta = .75 \mid X_3 = 1)$$

$$= \frac{\overbrace{P(X_3 = 1 \mid \theta = 0.75)}^{.75}\ \overbrace{P(\theta = .75)}^{.421}}{\underbrace{P(X_3 = 1 \mid \theta = .75)}_{.75}\ \underbrace{P(\theta = .75)}_{.429} + \underbrace{P(X_3 = 1 \mid \theta = 0.5)}_{.5}\ \underbrace{P(\theta = 0.5)}_{.571}}$$

$$P(\theta \mid X_3 = 1) = \begin{cases} .53 \text{ if } \theta = 0.75 \\ .47 \text{ if } \theta = .5 \end{cases}$$

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

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

Generally, We want to show;

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

Start with the full formula,

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

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

$P(AB) = P(A|B)\,P(B)$

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

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

posterior where seeing the data $X_1,\dots X_{n-1}$

$$P(X) = \sum_{Y} P(X, Y)$$

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

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

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

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

$$\Downarrow$$

$$P(X_n \mid \theta)$$

once you know θ nothing we need.

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

map → Maximum a posterior estimate

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

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

if $\Theta_0 = \Theta = (0,1)$ for iid bernoulli F

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

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

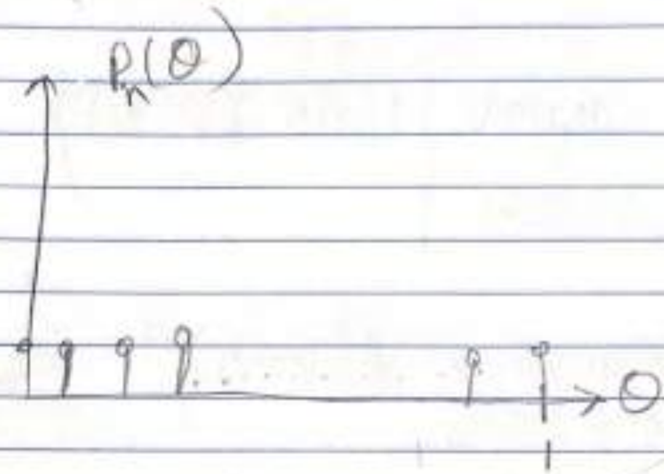
Why is $\Theta_0 = \{\theta_1, \theta_2, \ldots\}$ a bad idea? $\cancel{(0,1)}$

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

$$\Theta_0 = \left\{ 0, \tfrac{1}{10}, \ldots, \tfrac{9}{10}, 1 \right\} \quad P(\theta) = \left\{ \tfrac{1}{11} \; \forall \theta \right.$$

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

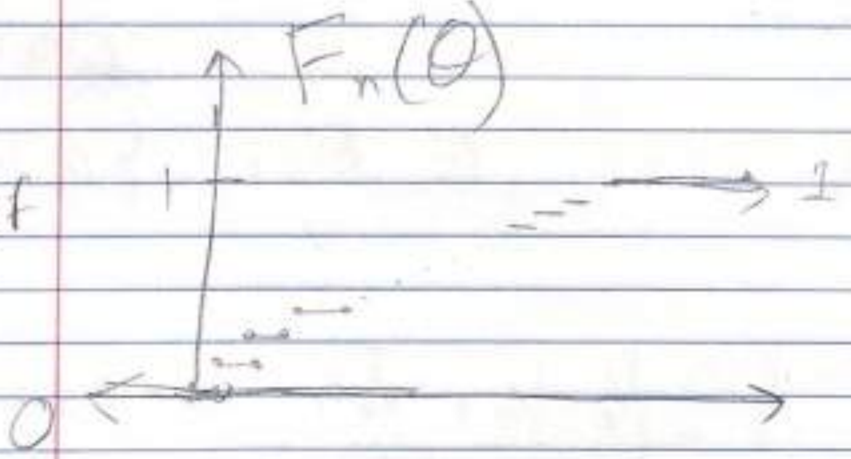
pdf



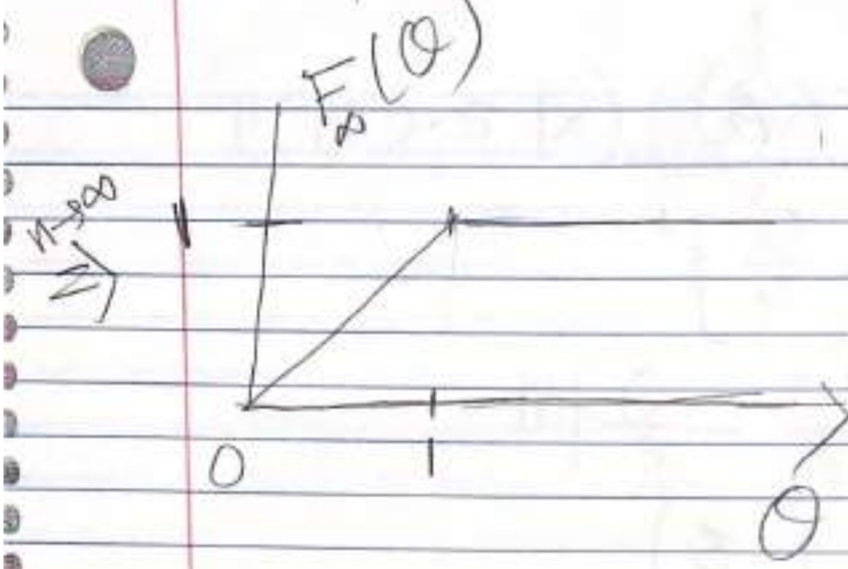
$$\lim_{n \to \infty} P_n(\theta) = 0 \Rightarrow \Theta(\infty) \text{ is}$$

not a discrete rv any more.

cdf



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



$\theta \sim U(0,1)$
Standard uniform
(cont)

$P(\theta) = U(0,1)$ Principle of Indifference

default prior

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

iid bernoulli

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

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

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

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

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

$f_\theta(\theta) = \begin{cases} 1 & \theta \in [0,1] \\ 0 & o/t \end{cases}$

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

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

$$= 12\theta^2(1-\theta)$$

density function

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

$f'(\theta) = 2\theta - 3\theta^2 = 0$

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

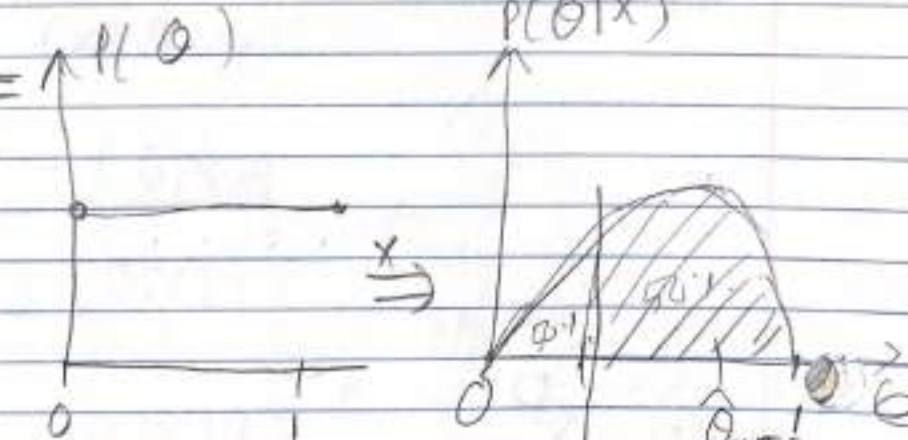
$2 - 3\theta = 0$

$\hat{\theta}_{MAP} = 2/3$

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

$= \bar{x}$

$= \theta_{MLE}$

$=$ 

$$P(\theta > .5 \mid X) = \int_{.5}^{1} 12\,\theta^2(1-\theta)\,d\theta$$

$$= 12\left[\frac{\theta^3}{3} - \frac{\theta^4}{4}\right]_{.5}^{1} = .688$$

Posterior Medium

$$\hat{\theta}_{MMAE} := \mathrm{Med}[\theta \mid X] = a \text{ s.t. } \int_{-\infty}^{a} P(\theta \mid X)\,d\theta = \frac{1}{2}$$

must

Mean absolute error.

$$= \arg\min_{\theta \in \Theta} \{ |\hat{\theta} - \theta| \}$$

MMAE → Minimum mean absolute error

Posterior expectation/mean

$$\hat{\theta}_{MMSE} := E[\theta \mid X]$$

$$= \arg\min_{\theta \in \Theta} \{ (\hat{\theta} - \theta)^2 \}$$

$x_1, \ldots x_n$ (general case)

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

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

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

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

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

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

next page after that

def

Beta function

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

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

$$= \text{Beta}(\sum x_i + 1, n - \sum x_i + 1)$$

→ a new r.v.

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

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

Parameter space: $\alpha, \beta > 0$

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

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

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

multiple

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

factorial function for all positive reals (3.7)! not defined!

Facts

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

② $\beta(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ proof 368

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

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

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

Med = no closed form

$$\mathcal{E}^{beta}(0.5,\alpha,\beta)$$

$$\mathrm{Mode}(Y) = \underset{y\in(0,1)}{\arg\max}\left\{\frac{1}{B(\alpha,\beta)}y^{\alpha-1}(1-y)^{\beta-1}\right\}$$

$$= \arg\max \left\{ y^{\alpha-1}(1-y)^{\beta-1} \right\}$$

$$= \arg\max \left\{ (\alpha-1)\cdot\ln(y) + (\beta-1)\ln(1-y) \right\}$$

$$f(y) = \frac{\alpha-1}{y} - \frac{\beta-1}{1-y}$$

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

if you are careful.

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

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

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

21

$Y \sim U(0,1) = Beta(1,1)$

Standard uniform is a

Special case of the Beta distri