Lecture - 05 02/11/2020

T: iid Bernoulli

$\Theta_o = [0.5, 0.75], \subset (0,1)$

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

$$P(\theta \mid x_1 = 0) = \begin{cases} 1/3 & \text{if } \theta = 0.75 \\ 2/3 & \text{if } \theta = 0.5 \end{cases}$$

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

$$= 0.429$$

$$P(\theta \mid x_3 = 1) = \begin{cases} 0.429 & \text{if } \theta = 0.75 \\ 0.571 & \text{if } \theta = 0.5 \end{cases}$$

$$P(\theta = 0.75 \mid x_3 = 1) = \frac{\overset{0.75}{P(x_3 = 1 \mid \theta = 0.75)}\ \overset{0.429}{P(\theta = 0.75)}}{\underbrace{P(x_3 = 1 \mid \theta = 0.75)}_{0.75}\ \underbrace{P(\theta = 0.75)}_{0.429} + \underbrace{P(x_3 = 1 \mid 0.05)}_{0.5}\ \underbrace{P(\theta = 0.5)}_{0.571}}$$

$$= 0.53$$

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

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

Generally we want to show:

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

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

Start with full formula

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

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

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

Posterior where seeing the data $x_1,\dots,x_n$

$P(x) = \sum_y P(x,y) \Rightarrow$ to get rid of y

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

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

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

↳ once you know θ, you can get rid of $x_1,\dots,x_{n-1}$

$$\parallel$$

$$P(x_n|\theta)$$

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

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

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

$\hat{\theta}_{MAP}$ = Maximum a posterior estimate.

$$\hat{\theta}_{MAP} = \underset{\theta \in \Theta_0}{\text{argmax}} \left[P(\theta|x) \right]$$

$$= \underset{\theta \in \Theta_0}{\text{argmax}} \left[P(x|\theta)\, P(\theta) \right]$$

$$= \underset{\theta \in \Theta_0}{\text{argmax}} \left[P(x|\theta) \right]$$

It's true if $P(\theta)$ is determined by the principle of indifference.

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

It's true if $\Theta_0 = \Theta = (0,1)$

for the iid Bernoulli $\mathbb{F} \neq (0,1)$

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

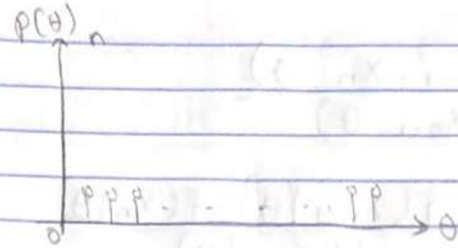
$\Theta_0 = \{0, 1/4, 1/6, 3/4, 1\}$

$P(\theta) = \{1/5 \ \forall \theta$

$\Theta_0 = \{0, 1/10, \ldots, 3/10, 1\}$ $P(\theta) = \{1/11 \ \forall \theta'$

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

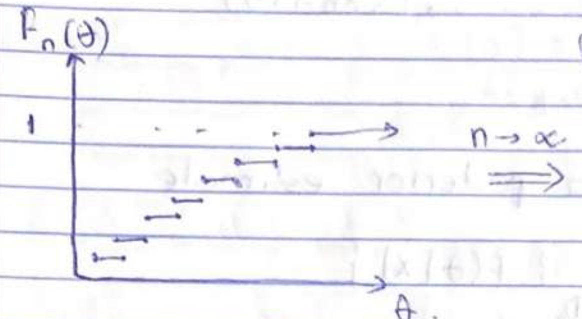
uniform distribution r.v

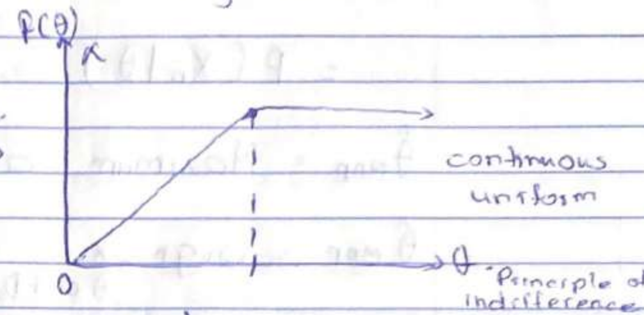


$$\lim_{n \to \infty} P_n(\theta) = 0.$$

$\Rightarrow \Theta_{(m)}$ is not a discrete r.v any more.

CDF





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

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

default prior. standard uniform (continuous)

F = iid Bernoulli

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

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

$f_\theta(\theta) = \begin{cases} 1 & \text{if } \theta \in (0,1) \\ 0 & \text{O/T} \end{cases}$

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

for our X

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

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

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

$$= \underset{\theta \in (0,1)}{\arg\max} [\theta^2(1-\theta)]$$

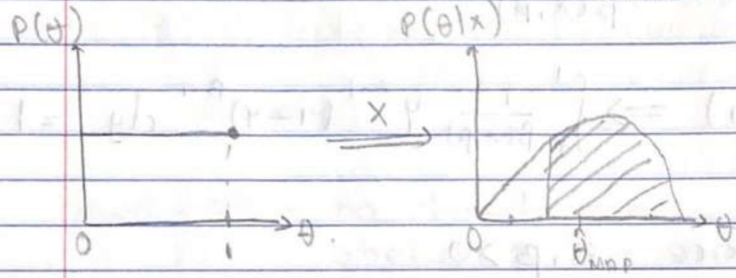
$f(\theta) = \theta^2 - \theta^3$.

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

$2 - 3\theta = 0$.

$3\theta = 2$.

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



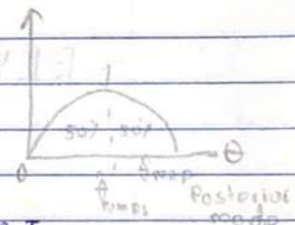
$P(\theta > 0.5 | x) = \int_{0.5}^{1} 12\theta^2(1-\theta)\,d\theta$

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

$= 0.688.$

$\hat{\theta}_{MMAE} := med[\theta | x] = a$ s.t. $\int_{-\infty}^{a} P(\theta|x)\,d\theta = 1/2$.

Minimum mem absolute error



Posterior expectation / mean

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

Minimum mem squard error

$X_1, \dots, X_n$ (General Case)

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

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

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

$$= \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}{\int_0^1 \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\,d\theta}$$

famous integral

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

Beta function

continuous from previous

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

$= \text{Beta r.v}(\Sigma x_i + 1,\, n - \Sigma x_i + 1) \to$ a new r.V

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

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

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

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

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

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

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

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

Facts

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

II $B(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ (proof is math 368.)

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

$\text{med}[Y]$ = no closed form

qbeta$(0.5, \alpha, \beta)$

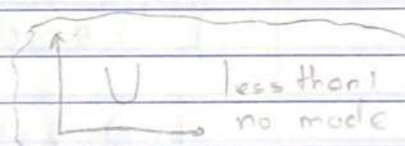
$$\text{Mode}[Y] = \underset{y\in(0,1)}{\text{avgmax}} \left\{ \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1} \right\}$$

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

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

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

$$\Rightarrow y \text{ mode} = \frac{\alpha-1}{\alpha+\beta-2}.$$



If you're careful

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

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

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

$$\Longrightarrow Y \sim U(0,1) = \text{Beta}(1,1)$$

standard uniform is a special case of the Beta distribution.