2/11 화 베이지안

$\mathcal{F}$ iid Bernoulli

$\Theta_0 = \{0.5, 0.75\} \subset (0,1)$

$x = \langle 0, 1, 1 \rangle$, $\; x_1 = 0,\ x_2 = 1,\ x_3 = 1$

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

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

$= 0.429$.

$$P(\theta \mid x_2 = 1) = \begin{cases} 0.429 & \text{if } \theta = 0.75 \\ 0.571 & \text{if } \theta = 0.5 \end{cases}$$

$$P(\theta = 0.75 \mid x_2 = 1) = \frac{\overset{0.75}{P(x_3 = 1 \mid \theta = 0.75)}\ \overset{0.429}{P(\theta = 0.75)}}{\underset{0.75}{P(x_3 = 1 \mid \theta = 0.75)}\ \underset{0.429}{P(\theta = 0.75)} + \underset{0.5}{P(x_3 = 1 \mid \theta = 0.5)}\ \underset{0.571}{P(\theta = 0.5)}} = 0.53$$

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

- $P(\theta \mid x) = \dfrac{P(x \mid \theta) P(\theta)}{P(x)} = \dfrac{P(x \mid \theta) P(\theta)}{\sum_{\theta \in \Theta} P(x \mid \theta) P(\theta)}$

Generally we want to show.

$$P(\theta \mid x_1 \cdots x_n) = \frac{P(x_n \mid \theta) \cdot P(\theta \mid x_1 \cdots x_{n-1})}{\sum_{\theta \in \Theta} P(x_n \mid \theta) P(\theta \mid x_1 \cdots x_{n-1})}$$

Start with full formula.

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

$P(A, B) = P(A \mid B) P(B)$

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

$= P(\theta \mid x_1 \cdots x_{n-1})$

: posterior when seeing the data $x_1 \cdots x_{n-1}$

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

$\boxed{P(X) = \sum_Y P(X, Y)}$

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

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

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

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

$$= P(x_n \mid \theta).$$

$\hat{\theta}_{MAP}$: maximum of posterior estimate

$$:= \underset{\theta \in \Theta_0}{\operatorname{argmax}} \{P(\theta \mid x)\} = \underset{\theta \in \Theta_0}{\operatorname{argmax}} \{P(x \mid \theta) \cdot P(\theta)\}$$

$$= \underset{\theta \in \Theta_0}{\operatorname{argmax}} \{P(x \mid \theta)\}$$

↑ if $P(\theta)$ is determined by the principle of indifference.

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

↑ if $\Theta_0 = \Theta = (0, 1)$ for the iid Bernoulli F.

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

↳ b/c $\neq (0,1)$

close the door: doesn't make sense.

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

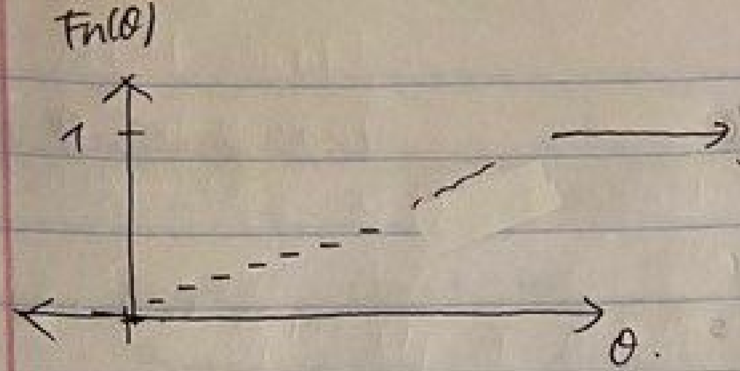
$\Theta_0 = \{0, 1/10, \dots 9/10, 1\}$, $P(\theta) = \{\frac{1}{11}$ $\forall \theta$.

⋮

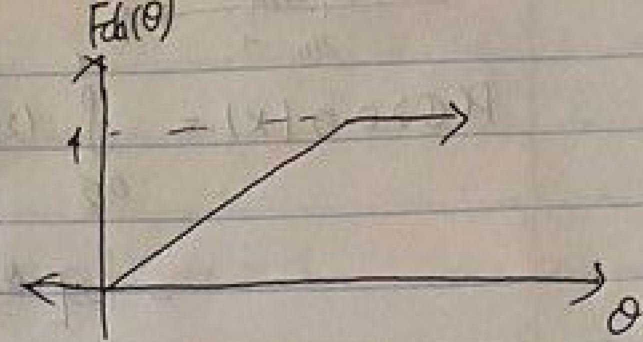
$\Theta_{0(n)} = \{0, \frac{1}{n} \dots \frac{n-1}{n}, 1\}$, $P(\theta) = \frac{1}{n-1}$ $\forall \theta$.

[Uniform distribution]

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



$\xRightarrow{n \to \infty}$



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

$\theta \sim U(0,1)$. *Principle of Indifference

default prior: standard uniform (continuous)

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

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

$\theta \sim U(0,1)$
$$f(\theta) = \begin{cases} 1, & \text{if } \theta \in (0,1) \\ 0, & \text{o.w} \end{cases}$$

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

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

← $\theta \sim U(0,1)$

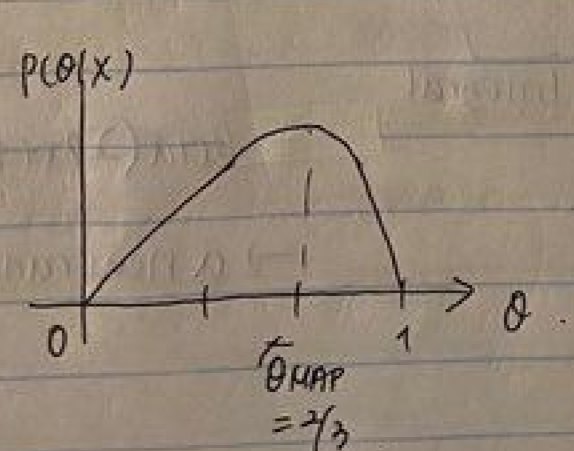
$$= \frac{\theta^2(1-\theta)}{\left.\frac{\theta^3}{3} - \frac{\theta^4}{4}\right]_0^1}$$

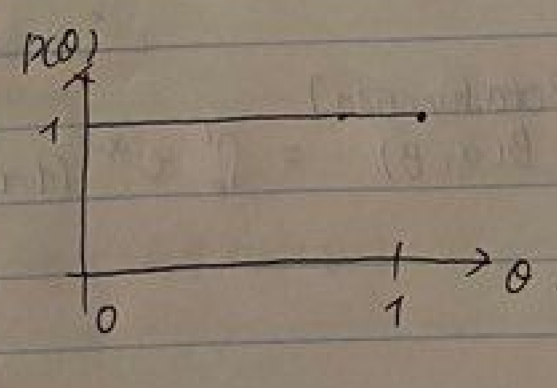
$$= 12\theta^2(1-\theta)$$

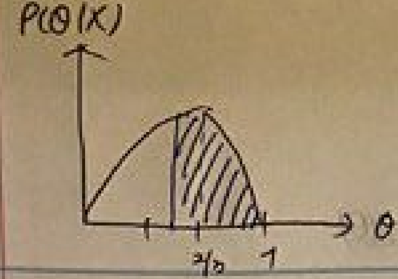
$$\hat{\theta}_{MAP} = \underset{\theta\in(0,1)}{\operatorname{argmax}} \{ \underbrace{12}_{\text{does nothing!}}\theta^2(1-\theta) \} = \underset{\theta\in(0,1)}{\operatorname{argmax}} \{ \overbrace{\theta^2(1-\theta)}^{\theta^2-\theta^3} \}.$$

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

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



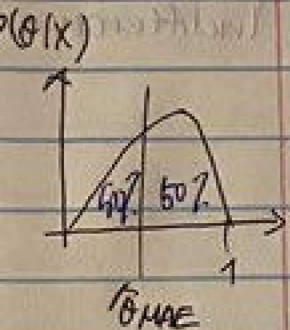




$$P(\theta>0.5|X) = \int_{0.5}^{1} 12\theta^2(1-\theta)\,d\theta$$

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

$$= 0.688$$



posterior medium

$\hat{\theta}_{MMAE} := Med[\theta|X] = a$ such that (s.t) $\int_{-\infty}^{a} p(\theta|x)\,d\theta = \frac{1}{2}$. $= \underset{\theta\in\Theta}{\arg\min}\{|\hat{\theta}-\theta|\}$

minimum medium absolute estimate

posterior expectation

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

minimum mean square error.

$X_1, \dots X_n$ (general case)

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

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

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

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

$$= \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}{\underbrace{\int_0^1 \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\,d\theta}_{F\text{— integral}}} = \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}{B(\sum x_i+1,\ n-\sum x_i+1)}$$

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

→ a new random variable

[Beta function]

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

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

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

Parameter Space $\alpha, B > 0$

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

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

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

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

factorial function for all positive r — . not defined!

Facts

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

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

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

$\cdot$ Med[Y] = not ____ form

qbeta(0.5, α, B)

(quantile)

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

$= \underset{y\in(0,1)}{\operatorname{argmax}} \left\{ y^{\alpha-1}(1-y)^{B-1} \right\}$

log ↓

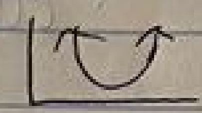
$= \operatorname{argmax} \underbrace{\left\{ (\alpha-1)\cdot\ln(y) + (B-1)\ln(1-y) \right\}}_{f(y)}.$

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

$(\alpha-1)(1-y) - (\beta-1)y = 0.$

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

if $\alpha,\beta > 1$ x



$\Rightarrow$ x mode.

· If you're careful, $f'' y_{mode} < 0$ if $\alpha,\beta > 1$.

** expectation & mode shouldn't be different

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

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

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

standard uniform is special case of the Beta distribution