Lec 5 2/14/19 Math 341

$$\hat{\theta}_{MAP} := \underset{\theta \in \Theta_0}{\text{argmax}} \{ P(\theta \mid X) \} \overset{\text{if?}}{=} \text{argmax} \{ P(X \mid \theta) P(\theta) \} \overset{\text{if?}}{=} \underset{\theta \in \Theta_0}{\text{argmax}} \{ P(X \mid \theta) \} \overset{\text{if?}}{=} \hat{\theta}_{MLE}$$

Let $\mathcal{F}$ = iid Bernoulli $\quad \Theta = (0,1)$

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

You are zeroing out the prob. of the rest of the values of θ unnecessarily!

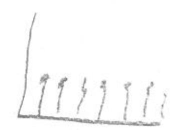
How do we include all values of θ in the prior

$\Theta_0 = \{0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, 1\} \qquad P(\theta) = U(\Theta_0) = \{\frac{1}{4} \; \forall \theta$ ← discrete uniform

$\Theta_0 = \{0, \frac{1}{10}, \frac{2}{10}, \dots, \frac{9}{10}, 1\} \qquad P(\theta) = U(\Theta_0) = \{\frac{1}{10} \; \forall \theta$

$\vdots$

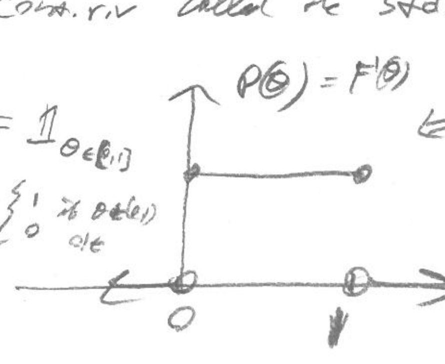
$\Theta_{0,n} = \{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\} \qquad P_n(\theta) = U(\Theta) = \{\frac{1}{n} \; \forall \theta$

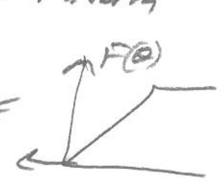


$\lim_{n \to \infty} \Theta_{0,n} = (0,1) \qquad \lim_{n \to \infty} P_n(\theta) = 0 \; \forall \theta$ if $P_n(\theta)$ is a PMF

$\lim_{n \to \infty} F_n(\theta) = F(\theta) \qquad$ But the CDF → cont. r.v. called the std. uniform

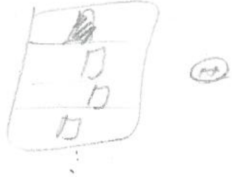
$\underbrace{P(\theta)}_{\text{density}} = U(0,1) = \mathbb{1}_{\theta \in [0,1]} = \begin{cases} 1 & \text{if } \theta \in [0,1] \\ 0 & \text{o/w} \end{cases}$





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

Let's see how this works

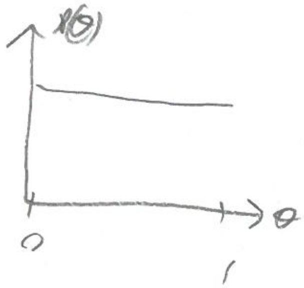


$X = \langle 0,1,1 \rangle$

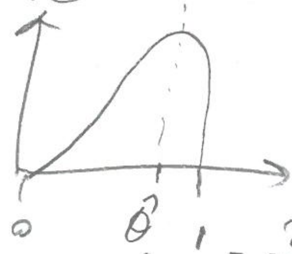
$$P(\theta|X) = \frac{P(X|\theta)\,P(\theta)}{P(X)} = \frac{P(X|\theta)}{P(X)} = \frac{P(X|\theta)}{\int_{\Theta} P(X|\theta)\,P(\theta)\,d\theta} = \frac{\theta^2(1-\theta)}{\int_0^1 \theta^2(1-\theta)\,d\theta}$$

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

$$P(\theta) \xrightarrow{X} P(\theta|X)$$



X



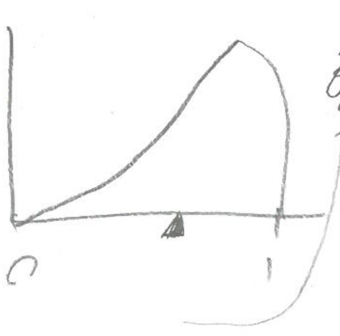
Mode = argmax

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

What's the prob the coin is unfairly biased to Heads (given: X and a prior of indifference)

$$P(\theta > 0.5|X) = \int_{0.5}^{1} 12\,\theta^2(1-\theta) = 12\left[\frac{\theta^3}{3} - \frac{\theta^4}{4}\right]_{0.5}^{1} = 12\left(\frac{1}{12} - \left(\frac{1}{24} - \frac{1}{64}\right)\right) = 0.688$$

$\hat{\theta}_{MAP}$: Bayesian point estimate. It is a measure of central tendency of the distr $P(\theta|X)$. Are there other measures of central tendency? the mean



$$\hat{\theta}_{MMSE} = E[\theta|X] = \int_{\Theta} \theta\, P(\theta|X)\, d\theta$$

posterior mean

$$= \int_0^1 \theta\,(12\,\theta^2(1-\theta))\,d\theta = 12\left[\frac{\theta^4}{4} - \frac{\theta^5}{5}\right]_0^1 = 12\left(\frac{1}{4} - \frac{1}{5}\right) = \frac{12}{20} = 0.6$$

$$\hat{\theta}_{MMSE} = \underset{\hat{\theta} \in \Theta}{\operatorname{argmin}} \left\{ E\left[(\theta - \hat{\theta})^2 \mid X\right] \right\}$$

this is the most common Bayesian pt. estimate

minimum mean squared error

Another pt. estimate?

$$\hat{\theta}_{MAE} = Med\left[\theta \mid X\right] = a \text{ s.t. } \int_{-\infty}^{a} p(\theta \mid x)\, d\theta = \frac{1}{2} \quad \text{or} \quad \underset{\theta \text{ s.t.}}{F(\theta \mid x)} = \frac{1}{2}$$

median

(the 50%ile)

Let's solve...

$$\int_0^a 12\, \theta^2 (1-\theta)\, d\theta = 12\left[\frac{\theta^3}{3} - \frac{\theta^4}{4}\right]_0^a = 12\left(\frac{a^3}{3} - \frac{a^4}{4}\right) \stackrel{set}{=} \frac{1}{2}$$

$$\Rightarrow \frac{a^3}{3} - \frac{a^4}{4} = \frac{1}{24} \Rightarrow -\frac{1}{4} a^4 + \frac{1}{3} a^3 + 0a^2 + 0a - \frac{1}{24} = 0$$

quartic eq.

there's a really long formula for this!

the answer is $a = 0.614$

$$\hat{\theta}_{MMAE} = \underset{\hat{\theta} \in \Theta}{\operatorname{argmin}} \left\{ E\left[\,|\theta - \hat{\theta}| \mid X\right] \right\}$$

minimum mean absolute error

$x = \langle 0,1 \rangle$ specific case, Let's now solve for the general case of $x_1, \ldots x_n$ (★

$$P(\theta|x) = \frac{P(x|\theta)\,P(\theta)}{P(x)} = \frac{\theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}(1)}{\underbrace{\int_0^1 \theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}(1)\,d\theta}} = \frac{1}{B(\Sigma x_i+1, n-\Sigma x_i-1)}\theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}$$

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

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

Famous Integral, a special function known as the "beta function"

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

dummy variable

Back to probability class

let $Y \sim \text{Beta}(\alpha,\beta) := \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1} = p(y)$ the PDF, valid?

$\text{Supp}[Y] = (0,1)$

$$\int_{\text{Supp}[Y]} p(y)\,dy \stackrel{?}{=} 1 \qquad \int_0^1 \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1}\,dy = \frac{1}{B(\alpha,\beta)}\underbrace{\int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\,dy}_{B(\alpha,\beta)} = \frac{B(\alpha,\beta)}{B(\alpha,\beta)} = 1$$

Param space?

let $\alpha = 0$, $\beta = 1$ $\Rightarrow \int_0^1 \frac{1}{y}\,dy = \infty$ for $\alpha < 0 \Rightarrow$ worse divergence

$\alpha = 1$, $\beta = 0$ $\Rightarrow \int_0^1 \frac{1}{1-y}\,dy = \infty$ for $\beta < 0 \Rightarrow$ " " " " "

You can show that $\alpha > 0$ & $\beta > 0$ is the param space.

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

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

this can be simplified, but first we need to introduce a new special function:

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

Facts:

① $\Gamma(\alpha+1) = \alpha\,\Gamma(\alpha)$ Pf: integration by parts

② $B(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ using mult. calc. change of variables

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

$B(\alpha,\beta)$

$Var(Y) = \ldots =$ on homework

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

take log

$$\operatorname{argmax}\left\{(\alpha-1)\ln(y) + (\beta-1)\ln(1-y)\right\}$$

Take derivative, set = 0

$$\frac{d}{dy}\left[(\alpha-1)\ln(y)+(\beta-1)\ln(1-y)\right] \overset{set}{=} 0$$

Shouldn't negative

$$\Rightarrow \frac{\alpha-1}{y} - \frac{\beta-1}{y-1} = 0 \Rightarrow \frac{\alpha-1}{y} = \frac{\beta-1}{1-y} \Rightarrow \frac{1-y}{y} = \frac{\beta-1}{\alpha-1} \Rightarrow \frac{1}{y}-1 = \frac{\beta-1}{\alpha-1}$$

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

Need to check 2^{nd} deriv to ensure regime. See HW

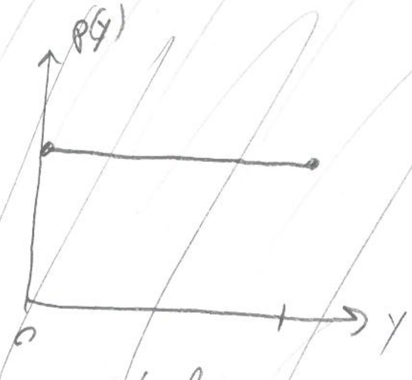
This only applies if $\alpha > 1$ & $\beta > 1$.

Med[θ] has no known closed form expression

To do the numeric regression, use the "qbeta" function in R.

qbeta(0.5, α, β) ← 50%ile

Shapes of distribution



$\alpha=\beta=1$

U(0,1): special case of beta

$\alpha=\beta=100$

concentrated about 0.5