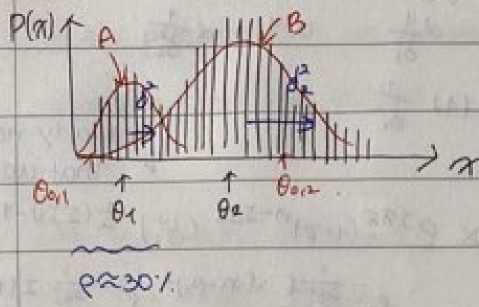
05/12/2020 [MATH341] Lecture 22

Consider the following data.



[theory]

$f(x) = \rho N(\theta_1, \sigma_1^2) + (1-\rho) N(\theta_2, \sigma_2^2)$

How many parameters are in this model?

↳ 5

[posterior]

$P(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho \mid X) \propto P(X \mid \theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho) \cdot P(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho)$

$P(X_1, \ldots, X_n \mid \theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho) = \prod_{i=1}^{n} \left(\rho \cdot \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x_i-\theta_1)^2} + (1-\rho) \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(x_i-\theta_2)^2} \right)$

→ we're in trouble because we can't simplify this .. product of $(a_i + b_i)$

introduce B

$P(A \mid X) = \int P(A, B \mid X)\, dB$

"Data Argumentation" or "parameter Argumentation" We introduce parameters:

$I_1 := \mathbb{1}_{x_1}$ comes from distribution A $= \begin{cases} 1, & \text{if } x_1 \text{ from distr A} \\ 0, & \text{if } x_1 \text{ from distr B} \end{cases}$

$I_2 := \mathbb{1}_{x_2}$

$\vdots$

$I_n := \mathbb{1}_{x_n}$

New Bayesian Setup:

$P(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho, I_1, \ldots, I_n \mid X) \propto P(X \mid \theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho, I_1, \ldots, I_n) \cdot P(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho, I_1, \ldots, I_n)$

$$P(X \mid \text{—}) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x_i-\theta_1)^2} \right)^{I_i} \left(\frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(x_i-\theta_2)^2} \right)^{1-I_i}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_1^2}} \right)^{\sum I_i} e^{-\frac{1}{2\sigma_1^2}\sum I_i (x_i-\theta_1)^2} \left(\frac{1}{\sqrt{2\pi\sigma_2^2}} \right)^{n-\sum I_i} e^{-\frac{1}{2\sigma_2^2}\sum (1-I_i)(x_i-\theta_2)^2}$$

$P(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho, I_1, \ldots, I_n) = P(I_1, \ldots, I_n \mid \rho, \theta_1, \sigma_1^2, \theta_2, \sigma_2^2) \cdot P(\rho, \theta_1, \sigma_1^2, \theta_2, \sigma_2^2)$

$= \underbrace{P(I_1, \ldots, I_n|\rho)} \; \underbrace{P(\rho)} \; \underbrace{P(\theta_1|\sigma_1^2)} \; \underbrace{P(\sigma_1^2)} \; \underbrace{P(\theta_2|\sigma_2^2)} \; \underbrace{P(\sigma_2^2)}$ • Jeffreys' priors

$= \prod_{i=1}^{n} P(I_i|\rho) \quad U(0,1) \quad \propto 1 \quad \propto \frac{1}{\sigma_1^2} \quad \propto 1 \quad \propto \frac{1}{\sigma_2^2}$

$= \rho^{\Sigma I_i}(1-\rho)^{n-\Sigma I_i}\,(1)\,(1)\,\frac{1}{\sigma_1^2}\,(1)\,\frac{1}{\sigma_2^2}$

clearly not a kernel that we know.

$P(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho, I_1, \ldots, I_n|x) \propto \rho^{\Sigma I_i}(1-\rho)^{n-\Sigma I_i}(\sigma_1^2)^{-\frac{1}{2}(\Sigma I_i)-1}(\sigma_2^2)^{-\frac{1}{2}(n-\Sigma I_i)-1}$

$\cdot\, e^{-\frac{1}{2\sigma_1^2}\Sigma I_i(x_i-\theta_1)^2} \cdot e^{-\frac{1}{2\sigma_2^2}\Sigma(1-I_i)(x_i-\theta_2)^2}$

Create a Gibbs sampler:

$P(\rho|-) \propto \rho^{\Sigma I_i+1-1}(1-\rho)^{n-\Sigma I_i+1-1} \propto Beta(\Sigma I_i+1,\; n-\Sigma I_i+1)$

$P(\theta_1|-) \propto e^{-\frac{1}{2\sigma_1^2}\Sigma I_i(x_i-\theta_1)^2}$

$\propto e^{-\frac{\Sigma I_i x_i^2}{2\sigma_1^2}} e^{+\frac{\Sigma I_i x_i\theta_1}{\sigma_1^2}} e^{-\frac{\Sigma I_i\theta_1^2}{2\sigma_1^2}}$

$\propto e^{\frac{\Sigma I_i x_i}{\sigma_1^2}\cdot\theta_1 - \frac{\Sigma I_i}{2\sigma_1^2}\theta_1^2}$ $\quad a = \frac{\Sigma I_i x_i}{\sigma_1^2}$, $b = \frac{\Sigma x_i}{2\sigma_1^2}$

$\propto N\left(\frac{\Sigma I_i x_i}{\Sigma I_i}, \frac{\sigma_1^2}{\Sigma I_i}\right)$ $\quad$ let $n_A = \Sigma I_i$: # of first distr.

$= N\left(\bar{x}_A, \frac{\sigma_1^2}{n_A}\right)$

$P(\theta_2|-) \propto N\left(\frac{\Sigma(1-I_i)x_i}{n-\Sigma I_i}, \frac{\sigma_2^2}{n-\Sigma I_i}\right)$ $\quad$ let $n_B = n - \Sigma I_i$

$= N\left(\bar{x}_B, \frac{\sigma_2^2}{n_B}\right)$

$P(\sigma_1^2|-) \propto (\sigma_1^2)^{-\frac{1}{2}(\Sigma I_i)-1} e^{-\frac{\Sigma I_i(x_i-\theta_1)^2/2}{\sigma_1^2}}$

$\propto InvGamma\left(\frac{\Sigma I_i}{2}, \frac{\Sigma I_i(x_i-\theta_1)^2}{2}\right)$

$= InvGamma\left(\frac{n_A}{2}, \frac{n_A\hat{\sigma}_1^2}{2}\right)$

$P(\sigma_2^2|-) \propto InvGamma\left(\frac{n-\Sigma I_i}{2}, \frac{\Sigma(1-I_i)(x_i-\theta_2)^2}{2}\right)$

$= InvGamma\left(\frac{n_B}{2}, \frac{n_B\hat{\sigma}_2^2}{2}\right)$

$P(I_i|-) \propto \Big(\underbrace{\rho\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{1}{2\sigma_1^2}(x_i-\theta_1)^2}}_{A}\Big)^{I_i}\Big(\underbrace{(1-\rho)\cdot\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2\sigma_2^2}(x_i-\theta_2)^2}}_{B}\Big)^{1-I_i}$

$\propto A^{I_i}B^{1-I_i}$

$\propto \Big(\underbrace{\frac{A}{A+B}}_{\rho}\Big)^{I_i}\Big(\underbrace{\frac{B}{A+B}}_{1-\rho}\Big)^{1-I_i}$

$\propto Bern(\rho) = Bern\left(\frac{A}{A+B}\right)$

We can now build a Gibbs sampler to have inference for all 5th parameters. We do need to be careful to specify a good starting location though.

Set $\theta_{0,1}$ = 25% of data $\quad \delta^2_{0,1}, \delta^2_{0,2} = 1$ (don't matter)

$\theta_{0,2}$ = 75% of data

$I_1, \ldots, I_{n/2} = 1$ and $I_{n/2+1}, \ldots, I_n = 0$.

This will force the first group to be θ_1 and the upper group to be centered at θ_2

Bayes Factors (B) AKA "ratio of evidences". we want to compare two models

$M_1 = \langle \mathcal{F}_1, P_1(\theta) \rangle$ model 1 is likelihood 1 and prior 1

$M_2 = \langle \mathcal{F}_2, P_2(\theta) \rangle$: model 2 " 2 " 2.

$$B = \frac{P_{M_1}(x)}{P_{M_2}(x)} = \frac{\int_{\Theta_1} P_1(x|\theta) \cdot P_1(\theta)\, d\theta}{\int_{\Theta_2} P_2(x|\theta)\, P_2(\theta)\, d\theta}$$

If $> 1 \Rightarrow M_1$ better

If $< 1 \Rightarrow M_2$ better.

marginal probability averaged over all θ for $P_J(x|\theta)$ and $P_J(\theta)$ for both models.

Often Bayes Factors compare H_0 to H_a. H_0 and H_a differ on Θ. Imagine $n=100$ coin flips and $x=61$ heads. Want to test if coin is unfair.

$H_0: \theta = 0.5$

$H_a: \theta \neq 0.5$

$\alpha = 5\%$

$\hat{\theta} = 61/100 = 0.61$

- Frequentist Test

Retainment Region $= [\theta_0 \pm z_{\alpha/2} \sqrt{\frac{\theta_0(1-\theta_0)}{n}}]$

$= [0.5 \pm 2 \cdot \sqrt{\frac{0.5(1-0.5)}{100}}]$

$= [0.40, 0.60]$

$\hat{\theta} \notin$ Ret. Region $\Rightarrow$ Reject H_0.

- Bayesian Test using CR method. let $\theta \sim U(0,1)$

$CR_{\theta, 95\%} = [\, qbeta(0.025, 61+1, 39+1),\ qbeta(0.975, 61+1, 39+1) \,]$

$= [\, 0.511,\ 0.700 \,]$

$\theta_0 = 0.5 \notin CR_{\theta, 95\%} \Rightarrow$ Reject H_0.