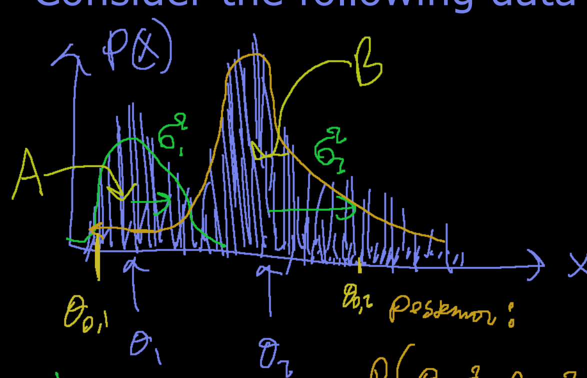
Consider the following data




$\rho \approx 30\%$

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?

posterior:

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

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

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 Augmentation" or "parameter augmentation". We introduce parameters:

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

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

$\vdots$

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

New Bayesian Setup:

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

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

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

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

$$= \underbrace{P(I_1, \ldots, I_n \mid \rho)}_{= \prod_{i=1}^{n} P(I_i \mid \rho)} \underbrace{P(\rho)}_{U(0,1)} \underbrace{P(\theta_1 \mid \sigma_1^2)}_{\propto 1} \underbrace{P(\sigma_1^2)}_{\propto \frac{1}{\sigma_1^2}} \underbrace{P(\theta_2 \mid \sigma_2^2)}_{\propto 1} \underbrace{P(\sigma_2^2)}_{\propto \frac{1}{\sigma_2^2}}$$

Jeffreys priors

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

clearly not a kernel that we know

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

Create a Gibbs Sampler:

$$P(\rho \mid \text{---}) \propto \rho^{(\sum I_i + 1) - 1} (1-\rho)^{(n - \sum I_i + 1) - 1} \propto \text{Beta}\left( \sum I_i + 1, n - \sum I_i + 1 \right)$$

$$P(\theta_1 \mid \text{---}) \propto e^{-\frac{1}{2\sigma_1^2} \sum I_i (x_i - \theta_1)^2} \propto e^{-\frac{\sum I_i x_i^2}{2\sigma_1^2}} e^{+\frac{\sum I_i x_i \theta_1}{\sigma_1^2}} e^{-\frac{\sum I_i \theta_1^2}{2\sigma_1^2}} \propto e^{\frac{\sum I_i x_i}{\sigma_1^2} \theta_1 - \frac{\sum I_i}{2\sigma_1^2} \theta_1^2}$$

$$\propto N\left( \frac{\sum I_i x_i}{\sum I_i}, \frac{\sigma_1^2}{\sum I_i} \right) = N\left( \bar{x}_A, \frac{\sigma_1^2}{n_A} \right) \quad \text{let } n_A = \sum I_i$$

$$P(\theta_2 \mid \text{---}) \propto N\left( \frac{\sum (1 - I_i) x_i}{n - \sum I_i}, \frac{\sigma_2^2}{n - \sum I_i} \right) = N\left( \bar{x}_B, \frac{\sigma_2^2}{n_B} \right) \quad \text{let } n_B = n - \sum I_i$$

$$P(\sigma_1^2 \mid \text{---}) \propto (\sigma_1^2)^{-\overset{\alpha}{\frac{1}{2}\sum I_i} - 1} e^{-\frac{\overset{\beta}{\sum I_i (x_i - \theta_1)^2 / 2}}{\sigma_1^2}} \propto \text{InvGamma}\left( \frac{\sum I_i}{2}, \frac{\sum I_i (x_i - \theta_1)^2}{2} \right)$$

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

$$P(\sigma_2^2 \mid \text{---}) \propto \text{InvGamma}\left( \frac{n - \sum I_i}{2}, \frac{\sum (1 - I_i)(x_i - \theta_2)^2}{2} \right) = \text{InvGamma}\left( \frac{n_B}{2}, \frac{n_B \hat{\sigma}_2^2}{2} \right)$$

$$P(I_i \mid \text{---}) \propto \Big( \overbrace{\rho \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x_i - \theta_1)^2}}^{A} \Big)^{I_i} \Big( \overbrace{(1-\rho) \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(x_i - \theta_2)^2}}^{B} \Big)^{1 - I_i}$$

$$\propto A^{I_i} B^{1 - I_i} \propto \underbrace{\left( \frac{A}{A+B} \right)}_{p}^{I_i} \underbrace{\left( \frac{B}{A+B} \right)}_{1-p}^{1 - I_i} \propto \text{Bern}(p) = \text{Bern}\left( \frac{A}{A+B} \right)$$

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

Set $\theta_{0,1}$ = 25%ile of data $\quad \sigma_{0,1}^2, \sigma_{0,2}^2 = 1$ (doesn't matter)

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

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

This will force the lower group to be centered at theta1 and the upper group to be centered at theta2.

---

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

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

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

if $> 1 \Rightarrow M_1$ better

if $< 1 \Rightarrow M_2$ better

marginal prob. avg'd over all $\theta$ for $P_j(X \mid \theta)$ and $P_j(\theta)$ for both models.

Often Bayes Factors compare $H_0$ to $H_a$. $H_0$ and $H_a$ differ on $\Theta$.

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

$$\text{Retainment Region} = \left[ \theta_0 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\theta_0(1-\theta_0)}{n}} \right] = \left[ 0.5 \pm 2\sqrt{\frac{.5(1-.5)}{100}} \right] = [0.40, 0.60]$$

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

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

$$CR_{\theta, 95\%} = \left[ \text{qbeta}(0.025, 61+1, 39+1), \text{qbeta}(0.975, 61+1, 39+1) \right] = [.511, .700]$$

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