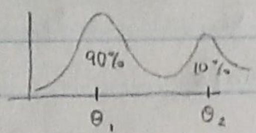
Lecture 22

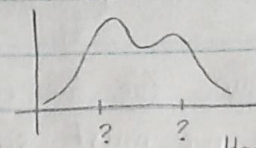
Suppose:

$$X_1,\dots X_n \mid \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \rho \overset{iid}{\sim} \rho N(\theta_1, \sigma_1^2) + (1-\rho) N(\theta_2, \sigma_2^2)$$

ex:



OR



Harder

$\boxed{\text{Note: } \rho \sim U(0,1)}$

- $P(\theta_1 \theta_2 \sigma_1^2 \sigma_2^2 \rho \mid X) \propto P(X \mid \theta_1 \theta_2 \sigma_1^2 \sigma_2^2 \rho)\; P(\theta_1 \theta_2 \sigma_1^2 \sigma_2^2 \rho)$

$P(\theta_1)P(\theta_2)P(\sigma_1^2)P(\sigma_2^2)P(\rho)$

are all independent.

assume indifference

$$\propto \prod_{i=1}^{n}\left(\rho \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{\frac{-1}{2\sigma_1^2}(x_i-\theta_1)^2} \underset{\substack{\uparrow \\ \text{creates issues w/ computations}}}{+} (1-\rho)\frac{1}{\sqrt{2\pi\sigma_2^2}} e^{\frac{-1}{2\sigma_2^2}(x_i-\theta_2)^2}\right)(1)(1)\underbrace{\left(\frac{1}{\sigma_1^2}\right)\left(\frac{1}{\sigma_2^2}\right)}_{\text{Jeffrey's}}(1)$$

- Data Augmentation : let $I_1 := \begin{cases} 1 & \text{if } X_1 \text{ comes from } N(\theta_1, \sigma_1^2) \\ 0 & \text{if } X_1 \text{ comes from } N(\theta_2, \sigma_2^2) \end{cases}$

$\vdots$

Assuming we can find I_i $\quad I_n := \begin{cases} 1 & \text{if } X_n \;\dots\; N(\theta_1, \sigma_1^2) \\ 0 & \text{if } X_n \;\dots\; N(\theta_2, \sigma_2^2) \end{cases}$

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

more parameters than pts.

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

$$\propto \prod_{i=1}^{n}\left(\rho \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{\frac{-1}{2\sigma_1^2}(x_i-\theta_1)^2}\right)^{I_i}\left((1-\rho)\frac{1}{\sqrt{2\pi\sigma_2^2}} e^{\frac{-1}{2\sigma_2^2}(x_i-\theta_2)^2}\right)^{1-I_i} \cdot (\text{Priors})$$

*Assume $I_i \sim \text{Bern}(0.5)\ \forall i$

$$\propto (\rho)^{\Sigma I_i}\left(\frac{1}{\sqrt{2\pi\sigma_1^2}}\right)^{\Sigma I_i}\left(e^{-\Sigma I_i \frac{1}{2\sigma_1^2}(x_i-\theta_1)^2}\right)(1-\rho)^{\Sigma 1-I_i}\left(\frac{1}{\sqrt{2\pi\sigma_2^2}}\right)^{\Sigma 1-I_i}\left(e^{(\Sigma 1-I_i)\left(\frac{1}{2\sigma_2^2}\right)(x_i-\theta_2)^2}\right) \cdot \text{Priors}$$

$$\propto \rho^{\Sigma I_i}(1-\rho)^{1-\Sigma I_i}(\sigma_1^2)^{-\frac{\Sigma I_i}{2}}(\sigma_2^2)^{\frac{-(\Sigma 1-I_i)}{2}}\; e^{-\Sigma I_i \frac{1}{2\sigma_1^2}(x_i-\theta_1)^2}\; e^{-(\Sigma 1-I_i)\frac{1}{2\sigma_2^2}(x_i-\theta_2)^2}$$

$$\cdot\ \left.\underset{\theta_1}{(1)}\underset{\theta_2}{(1)}\underset{\sigma_1^2}{\left(\frac{1}{\sigma_1^2}\right)}\underset{\sigma_2^2}{\left(\frac{1}{\sigma_2^2}\right)}\underset{\rho}{(1)}\underset{I_1 \dots I_n}{(0.5)^n}\right\}\ \text{priors}$$

$$= \rho^{\Sigma I_i}(1-\rho)^{\Sigma 1-I_i}(\sigma_1^2)^{-\frac{1}{2}\Sigma I_i -1}(\sigma_2^2)^{-\frac{1}{2}(\Sigma(1-I_i))-1}\underbrace{\underbrace{e^{-(\Sigma I_i)\frac{1}{2\sigma_1^2}X_i^2}}_{X_i}\underbrace{e^{(\Sigma I_i)\frac{X_i\theta_1}{\sigma_1^2}}e^{-(\Sigma I_i)\frac{1}{2\sigma_1^2}\theta_1^2}}_{\theta_1}}_{\sigma_1^2}\underbrace{\underbrace{e^{-(\Sigma 1-I_i)\frac{1}{2\sigma_2^2}X_i^2}}_{X_i}\underbrace{e^{(\Sigma 1-I_i)\frac{X_i\theta_2}{\sigma_2^2}}e^{-(\Sigma 1-I_i)\frac{1}{2\sigma_2^2}\theta_2^2}}_{\theta_2}}_{\sigma_2^2}$$

Kernal of posterior for two combined Normals

- to find values we must <u>sample</u>

- $$P\left(\theta_1 \,\middle|\, \begin{matrix}\text{All} \\ \text{variables} \\ \text{Known}\end{matrix}\right) \propto e^{\left((\Sigma I_i)\frac{X_i}{\sigma_1^2}\right)\theta_1} e^{\left(\frac{-\Sigma I_i}{2\sigma_1^2}\right)\theta_1^2} \propto N\left(\underbrace{\frac{\Sigma(I_i X_i)}{\Sigma I_i}}_{=\bar{X}}, \underbrace{\frac{\sigma_1^2}{\Sigma I_i}}_{=\frac{\sigma^2}{n}}\right)$$

- $$P(\theta_2 \mid \text{—}) \propto e^{\left(\frac{(\Sigma 1-I_i)X_i}{\sigma_2^2}\right)\theta_2} e^{\left(\frac{-\Sigma(1-I_i)}{2\sigma_2^2}\right)\theta_2^2} \propto N\left(\frac{\Sigma(1-I_i)X_i}{\Sigma(1-I_i)}, \frac{\sigma_2^2}{\Sigma(1-I_i)}\right)$$

- $$P(\sigma_1^2 \mid \text{—}) \propto \sigma_1^{2\,(-\frac{1}{2}\Sigma I_i - 1)} e^{-\frac{\Sigma I_i (X_i-\theta_1)^2/2}{\sigma_1^2}} \propto \text{InvGam}\left(\frac{\Sigma I_i}{2}, \frac{\Sigma I_i (X_i-\theta_1)^2}{2}\right)$$

- $$P(\sigma_2^2 \mid \text{—}) \propto \sigma_2^2 \ \text{\&} \ \Sigma 1-I_i \propto \text{InvGam}\left(\frac{\Sigma(1-I_i)}{2}, \frac{\Sigma(1-I_i)(X_i-\theta_2)^2}{2}\right)$$

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

- $$P(I_n \mid \text{—}) \propto \left(\rho\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{1}{2\sigma_1^2}(X_i-\theta_1)^2}\right)^{I_n}\left((1-\rho)\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2\sigma_2^2}(X_i-\theta_2)^2}\right)^{1-I_n}$$

for all n $\{1,\ldots,n\}$

since we have ρ and $1-\rho$

$$\left(\frac{a}{a+b}\right)^{I_n}\left(\frac{b}{a+b}\right)^{1-I_n} \propto \text{Bern}\left(\frac{(\rho)(N_1)}{(\rho)(N_1) + (1-\rho)(N_2)}\right)$$

To sample choose start values and sample from each conditional

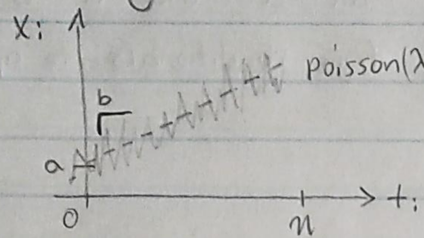
ex: $\theta_{1_0} = 0 \quad \theta_{2_0} = 0 \quad \sigma^2_{1_0} = 1 \quad \sigma^2_{2_0} = 1 \quad \rho = 0.5 \quad I_{1,\ldots,n_0} = \begin{cases} 0 \\ 1 \end{cases}$

- run for many iterations
- choose burn in
- Thin out based on latest point of autocorrelation.

M-H

- Metropolis - Hastings Algorithm

ex. Assuming a constant change point model



$\lambda = a + bt$ "poisson regression"

- a, b are our parameters now •

* Note: each t_i is paired with an X_i

$$P\left(\begin{matrix} X_1 \dots X_n \\ t_1 \dots t_n \end{matrix} \middle| a,b \right) = \left(\prod_{i=1}^{n} \underbrace{\frac{e^{-(a+bt_i)} (a+bt_i)^{X_i}}{(X_i)!}}_{\text{poisson }(a+bt)} \right) \propto \left(e^{-\Sigma(a+bt_i)} \prod_{i=1}^{n} (a+bt_i)^{X_i} \right)$$

non-conjugate

$$P(a,b|x) \propto P(X|a,b) \quad \underbrace{P(a,b)}_{P(a)\,P(b)}$$

$$\left(\propto e^{-\Sigma(a+bt_i)} \prod_{i=1}^{n} (a+bt_i)^{X_i}\right) \underbrace{(\propto 1)(\propto 1)}_{\text{uninformative}}$$

* Note
$a \in \mathbb{R}$
$b \in \mathbb{R}$
independent

Posterior

- $P(a,b|X) \propto e^{-\Sigma(a+bt_i)} \prod_{i=1}^{n} (a+bt_i)^{X_i}$

conditionals

- $P(a|X,b) \propto e^{-\Sigma(a+bt_i)} \prod_{i=1}^{n} (a+bt_i)^{X_i}$
- $P(b|X,a) \propto e^{-\Sigma(a+bt_i)} \prod_{i=1}^{n} (a+bt_i)^{X_i}$

conditionals are not useful for sampling since they are kernels (Grid sampling) too slow not accurate.

* Gibbs sampling moves around our conditionals to approximate our posterior

BUT

* our conditionals = Posterior so Gibbs won't approximate our posterior well