Lec 23 5/9/19 Math 381 (Last lecture)

Consider data: $X_1, \dots, X_n$

Looks like $\rho N(\theta_1, \sigma_1^2) + (1-\rho) N(\theta_2, \sigma_2^2)$ i.e. a mixture model

each of these pts
unknown which
distr. it
is realized from

Likelihood?

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

This can't be worked out! $(a_1+b_1)(a_2+b_2)\cdots(a_n+b_n)$ doesn't simplify!

Introduce more parameters $I_1 = \mathbb{1}_{X_1 \text{ comes from } N(\theta_1,\sigma_1^2)}$

$I_2 = \mathbb{1}_{X_2 \,\dots}$

$\vdots$

$I_n = \mathbb{1}_{X_n \,\dots}$

"Data Augmentation"

$$p(X_1,\dots,X_n \mid \theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\rho, I_1,\dots,I_n) = \prod_{i=1}^{n} \left(\rho \frac{1}{\sqrt{2\pi\sigma_1^2}} \cdots\right)^{I_i} \left((1-\rho)\frac{1}{\sqrt{2\pi\sigma_2^2}}\cdots\right)^{1-I_i}$$

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

$\dim(\Theta) = n+5$ params!

$$p(\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\rho,I_1,\ldots,I_n \mid X_1,\ldots,X_n) \propto p(X_1,\ldots,X_n \mid \text{—})\, p(\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\rho,I_1,\ldots,I_n)$$

prior structure

$$p(\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\rho,I_1,\ldots,I_n) = \underbrace{p(I_1,\ldots,I_n \mid \theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\rho)}_{=\prod_{i=1}^n p(I_i\mid\rho)\ \text{Bernoulli}}\, p(\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\rho)$$

$$= p(I_1,\ldots,I_n \mid \rho)\, \underbrace{p(\theta_1)}_{\propto 1}\, \underbrace{p(\theta_2)}_{\propto 1}\, \underbrace{p(\sigma_1^2)}_{\propto \frac{1}{\sigma_1^2}}\, \underbrace{p(\sigma_2^2)}_{\propto \frac{1}{\sigma_2^2}}\, \underbrace{p(\rho)}_{\propto 1}$$

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

$$\Rightarrow p(\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\rho,I_1,\ldots,I_n \mid X_1,\ldots,X_n) \propto (\quad)(\quad)$$

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

$$\propto N\left(\frac{\Sigma I_i x_i}{\Sigma I_i}, \frac{\sigma_1^2}{\Sigma I_i}\right)$$

$$p(\theta_2 \mid \text{———}) \propto e^{-\frac{1}{2\sigma_2^2}\Sigma(1-I_i)(x_i^2 - 2x_i\theta_2 + \theta_2^2)} \propto e^{\frac{\Sigma(1-I_i)x_i}{\sigma_2^2}\theta_1 - \frac{n-\Sigma I_i}{2\sigma_2^2}\theta_2^2}$$

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

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

$$p(\sigma_2^2 \mid \text{———}) \propto (\sigma_2^2)^{\frac{1}{2}(n-\Sigma I_i) - 1}\, e^{-\frac{\Sigma(1-I_i)(x_i-\theta_2)^2/2}{\sigma_2^2}} \propto \text{InvGamma}\left(\frac{n-\Sigma I_i}{2}, \frac{\Sigma(1-I_i)(x_i-\theta_2)^2}{2}\right)$$

$$p(\rho \mid \text{—}) \propto \rho^{\Sigma I_i + 1 - 1}(1-\rho)^{n-\Sigma I_i + 1 - 1} \propto \text{Beta}(\Sigma I_i + 1, n - \Sigma I_i + 1)$$

$$p(I_i \mid \text{—}) \propto \Big(\underbrace{\rho \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x_i-\theta_1)^2}}_{a}\Big)^{I_i} \Big(\underbrace{(1-\rho)\frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(x_i-\theta_2)^2}}_{b}\Big)^{1-I_i}$$

$$\propto \text{Bernoulli}\left(\frac{a}{a+b}\right)$$

COURSE OVER

Metropolis-Hastings Algorithm

During a Gibbs Sampler... You need...

$p(\theta_1 | \theta_2, \dots, \theta_p, X)$

$p(\theta_2 | \theta_1, \theta_3, \dots, \theta_p, X)$

$p(\theta_3 | \theta_1, \theta_2, \theta_4, \dots \theta_p, X)$

$\vdots$

$p(\theta_p | \theta_1, \dots, \theta_{p-1}, X)$

What if you don't have $p(\theta_2 | \theta_1, \theta_3, \dots, \theta_p, X)$? But only $k(\theta_2 | \theta_1, \theta_3, \dots \theta_p, X)$?

And...

I) You can't solve for the constant analytically

II) Grid sampling is ~~too~~ slow or unrealistic because you don't know $\theta_{2,min}$, $\theta_{2,max}$ because it is a function of $\theta_1, \theta_3, \dots, \theta_p$ and the changes at every iteration of the Gibbs sampler?

Then you can do the following...

I) Draw $\theta_{2,t}^{proposal}$ from $q(\theta_{2,t-1}, \phi)$ where q is called a "proposal dist". It is not the cond. dist!

other params

e.g. $q = N(\theta_{2,t-1}, 1^2)$

(II) Calc.

$$r := \frac{\dfrac{P(\theta_{1,t}, \theta_{2,t}^{\text{proposal}}, \theta_{3,t-1}, \ldots \theta_{p,t-1} \mid x)}{P(\theta_{2,t}^{prop} \mid \theta_{2,t-1}, \phi)}}{\dfrac{P(\theta_{1,t}, \theta_{2,t-1}, \theta_{3,t-1}, \ldots \theta_{p,t-1} \mid x)}{P(\theta_{2,t-1} \mid \theta_{2,t}^{prop}, \phi)}}$$

← posterior with proposal

← transition prob. for $\theta_{2,t-1} \Rightarrow \theta_{2,t}^{prop}$

← posterior without proposal

← transition prob backwards

Metropolis-Hastings Ratio

(III) Accept $\theta_{2,t}^{prop}$ w.p. r. Draw u from $U(0,1)$. If $u \leq r \Rightarrow$ accept $\theta_{2,t} = \theta_{2,t}^{prop}$

Otherwise reject $\theta_{2,t} = \theta_{2,t-1}$. Done move!

AKA "rejection-sampling"

Hastings (1974), which generalized Metropolis et al. (1953) which demands ← Teller!

$$P(\theta_t^{prop} \mid \theta_{t-1}, \phi) = P(\theta_{t-1} \mid \theta_t^{prop}, \phi)$$

symmetrical transitions

$$\Rightarrow r = \frac{P(\theta_{1,t}, \theta_{2,t}^{prop}, \theta_{3,t-1}, \ldots \theta_{p,t-1} \mid x)}{P(\theta_{1,t}, \theta_{2,t-1} \ldots \mid x)}$$

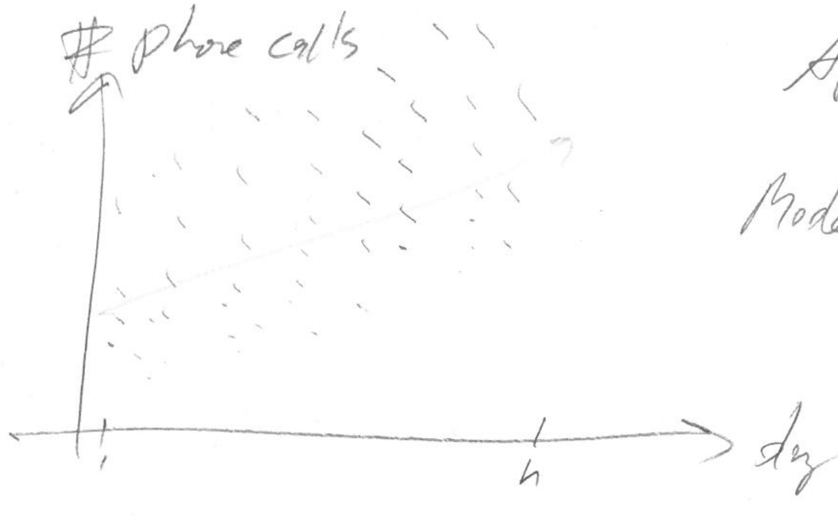
Metropolis Alg.

If you know the conditionals, you let the transition distr. = cond.

$$r = \frac{\dfrac{P(\theta_1, \ldots, \overset{\theta_j = \theta^*}{\ldots}, \theta_p \mid x)}{P(\theta_j = \theta_j^* \mid \theta_{-j}, x)}}{\dfrac{P(\theta_1, \ldots \theta_j = \theta_j^t, \ldots \theta_p \mid x)}{P(\theta_j = \theta_{j,t+1} \mid \theta_{-j}, x)}} = \frac{\cancel{P(\theta_j = \theta_j^* \mid \theta_{-j}, x)}\, P(\theta_{-j} \mid x)}{\cancel{P(\theta_j = \theta_{j,t+1} \mid \theta_{-j}, x)}\, P(\theta_{-j} \mid x)} = 1$$

Gibbs sampling is a very special case of Metropolis-Hastings Alg.

Let's see an example



Appears to be increasing linearly?

Model: $X_t \overset{ind}{\sim} \text{Poisson}(\underbrace{a+bt}_{\lambda})$

let $P(a,b) = \overset{\propto 1}{P(a)}\,\overset{\propto 1}{P(b)}$

$$P(a,b \mid X_1,\ldots,X_n) \propto P(X_1,\ldots,X_n \mid a,b)\, P(a,b)$$

$$= \prod_{t=1}^{n} \frac{e^{-(a+bt)}(a+bt)^{X_t}}{X_t!}$$

$$\propto e^{-n(a+b\bar{t})} \prod_{t=1}^{n} (a+bt)^{X_t}$$

$$P(a \mid \text{——}) \propto e^{-na} \prod_{t=1}^{n} (a+bt)^{X_t} = k(a \mid \text{——})$$

$$P(b \mid \text{——}) \propto e^{-n\bar{t}b} \prod_{t=1}^{n} (a+bt)^{X_t} = k(b \mid \text{——})$$

Use $q_a = N(a_{t-1}, 1^2)$ and $q_b = N(b_{t-1}, 1^2)$ as cand. distr.

Since transition prob's same forward-backward for both, this is technically a Metropolis sampler.