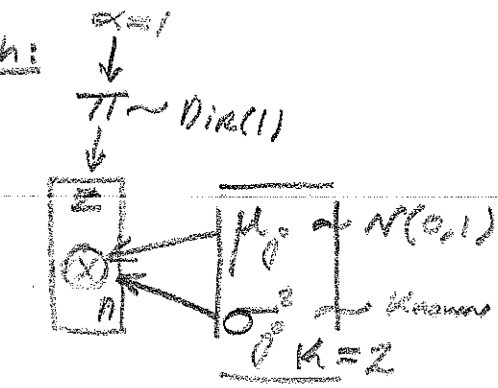
HW 5

Colbert Sesanker

1) Graph:



* $z \sim Mult(n, \pi)$

$\sigma_j^2 = \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \end{pmatrix}_j$

$\theta := \{\mu, \sigma^2, \pi\}$

$p(x_i|\theta) = \sum_j \pi_j \mathcal{N}(x_i|\mu_j, \sigma_j^2)$

$\prod_{i=1}^{n} p(x_i|\theta) = \prod_i \left(\sum_j \pi_j \mathcal{N}(x_i|\mu_j, \sigma_j^2)\right)$

if $\vec{z} = (z_1, \ldots, z_n)$ observed
$z_i \sim Cat(\pi)$

$\mathcal{L}(\theta|\vec{x}, \vec{z}) = \prod_i^n p(x_i, z_i|\theta) =$

$\prod_i \pi_{z_i} \mathcal{N}(x_i|\mu_{z_i}, \sigma_{z_i}^2)$

From the graph, conditional independencies can be simplified using the Markov Blanket:

$p(\alpha|\theta, \vec{x}, \vec{z}) = p(\alpha|\text{Blanket}(\alpha))$, $\alpha \in \theta \setminus \{\alpha\}$

$p(z_j|\theta, \vec{x}, \vec{z}) = p(z_j|\text{Blanket}(z_j) \cup z_{-j})$, $z_{-j} = \{z_1, \ldots z_{j-1}, z_{j+1}, \ldots z_n\}$
↳ Markov Blanket

Thus:

$$\begin{cases} p(\mu_j|\theta, \vec{x}, \vec{z}) = p(\mu_j|\sigma_j^2, \vec{x}, \vec{z}) \\ p(z_i = j|\theta, \vec{x}, \vec{z}) = p(z_i = j|\pi_j, \mu_j, \sigma_j^2, \vec{x}, z_{-j}) \\ p(\pi|\theta, \vec{x}, \vec{z}) = p(\pi|\vec{z}, \alpha = 1) \end{cases}$$

Using the posterior update formula for normal:

$$p(\mu_j|\sigma_j^2, \vec{x}, \vec{z}) = \mathcal{L}(\theta|\vec{x}, \vec{z}) \cdot \mathcal{N}(0,1) = \mathcal{N}\left(\frac{\sum_{i \in I}^{n} x_i}{\sigma_j^2 + n_j}, \frac{n_j}{\sigma_j^2} + 1\right)$$

$p(\pi|\vec{z}) = p(\vec{z}|\pi)p(\pi) =$
$\text{mult}(n, \pi_1, \pi_2) \cdot \text{Dir}(1,1) =$
$\text{Dir}(1+n_1, 1+n_2)$

$$\begin{cases} I = \{i : z_i = j\} \\ n_j = |\{z_i = j\}| \end{cases}$$

#1 cont...

$$p(z_i = j \mid \underbrace{\pi_j, \mu_j, \sigma_j^2, \vec{x}, \vec{z}_{-j}}_{\theta_{z_i}}) = \frac{p(\theta_{z_i} \mid z_i = 1)\, p(z_i = 1)}{p(\theta_{z_i} \mid z_i = 1)\, p(z_i = 1) + p(\theta_{z_2} \mid z_i = 2)\, p(z_2 = 1)}$$

ignore here?

$$= \frac{\pi_j \mathcal{N}(x_i \mid \mu_j, \sigma_j^2)}{\left[\pi_1 \mathcal{N}(x_i \mid \mu_1, \sigma_1^2) + (1-\pi) \mathcal{N}(x_i \mid \mu_2, \sigma_2^2)\right]}$$

Gibbs Sampler Algorithm:

1) initialize $\mu_1^0, \mu^0, \pi^0, \sigma^{-2}$

for $t = 1:n$

$p_i = p(z^t = 1 \mid \theta_{z_{i}}) \propto \mathcal{N}(x_i \mid \mu_1, \sigma_1^2)\, \pi^{t-1}$ # "=" means "←" in R here

$z_i^t \sim \text{Bernoulli}(p_i)$, $\vec{z}^t = [z_1, \dots z_n]$

$n_1^t = \text{sum}(\vec{z})$, $n_2^t = n - n_1$

$\pi^t \sim \text{Dir}(1 + n_1^t, 1 + n_2^t)$ # ~ means draw sample here

$$\mu_j^t \sim \mathcal{N}\left(\frac{\sum_{i=1}^n \mathbb{I}(z_i^t = j)\, x_i}{\sigma_j^{-2} + n_j^t}, \frac{n_j^t}{\sigma_j^{-2}} + 1 \right)$$

Known Known

End for

return ($\theta = \text{list}(\vec{z}, \pi, \mu, \sigma^{-2})$)

Colbert Sesanker HW 5

2) a) Sample $p(Z|X,P)$ with MCMC. $p(Z|X,P)$ is interpreted as a frequency.

b) Sample from $Dir(\lambda_1,\ldots,\lambda_j)$

c) Let $\vec{X} \sim Dir(\vec{X}|K=K_1)$, $\vec{Y} \sim Dir(\vec{X}, K=K_2)$ sample $|\vec{X}-\vec{Y}|$, indicies largest in magnitude are the significant Alleles.

d) same as c but take indicies closest to zero

e) Compute $p(Z_i=2|X,P)$

f) Can't, unless data from dachshund included

g) Consider P_{2lj}, l=locus, 2=poodle, j = allele

Compute: $$\frac{1}{n}\sum_i Sig\left(\sum_j |E[P_{2lj}] - P^i_{Klj}|\right)$$

- P^i_{Klj} is observed freq, at locus l, of allele j of the i^{th} mutt.
- $\sum_j$ is a sum over alleles on locus l
- $Sig: \mathbb{R} \to \{0,1\}$, is a significance measure of the difference between allele frequencies
- n = # of mutts