EM/DA & Swendsen-Wang

Note Title 5/21/2006

Missing Data:

$$p(\underline{\theta}, \underline{y}_{mis} | \underline{y}_{obs})$$

Estimate $\hat{\underline{\theta}} = \underset{\underline{\theta}}{\text{ARG MAX}}\ p(\underline{\theta} | \underline{y}_{obs})$

with $p(\underline{\theta} | \underline{y}_{obs}) = \sum_{\underline{y}_{mis}} p(\underline{\theta}, \underline{y}_{mis} | \underline{y}_{obs})$

EM. Introduce a new variable $q(\underline{y}_{mis})$

and the Kullback-Leibler divergence

$$K(q(\underline{y}_{mis}) \| p(\underline{y}_{mis} | \underline{y}_{obs}))$$

$K(q|p) \geq 0$
$K(q|p) = 0$ only if $q = p$

$$= \sum_{\underline{y}_{mis}} q(\underline{y}_{mis}) \log \frac{q(\underline{y}_{mis})}{p(\underline{y}_{mis} | \underline{y}_{obs}, \underline{\theta})}$$

Minimize:

$$F[\theta, q] = -\log P(\underline{\theta} | \underline{y}_{obs}) + \sum_{y_{mis}} q(\underline{y}_{mis}) \log \left\{ \frac{q(\underline{y}_{mis})}{p(\underline{y}_{mis} | \underline{y}_{obs}, \underline{\theta})} \right\}$$

w.r.t. θ & $q(.)$

$$-\log P(\underline{\theta} \mid \underline{y}_{obs}) + \sum_{\underline{y}_{mis}} q(\underline{y}_{mis}) \log \frac{q(\underline{y}_{mis})}{p(\underline{y}_{mis} \mid \underline{y}_{obs}, \underline{\theta})}$$

$$F[\theta, q] = \sum_{\underline{y}_{mis}} q(\underline{y}_{mis}) \log q(\underline{y}_{mis}) - \sum_{\underline{y}_{mis}} q(\underline{y}_{mis}) \left\{ \log P(\underline{\theta} \mid \underline{y}_{obs}) + \log p(\underline{y}_{mis} \mid \underline{y}_{obs}, \underline{\theta}) \right\}$$

$$F[\underline{\theta}, q] = \sum_{\underline{y}_{mis}} q(\underline{y}_{mis}) \log q(\underline{y}_{mis}) - \sum_{\underline{y}_{mis}} q(\underline{y}_{mis}) \log P(\underline{y}_{mis}, \underline{\theta} \mid \underline{y}_{obs})$$

EM Algorithm is equivalent to minimizing $F[\theta, q]$ w.r.t. θ & $q(.)$ alternatively.

Hence converges to a (local) minimum of $F[\underline{\theta}, q]$.

$$q^t(\underline{y}_{mis}) = p(\underline{y}_{mis} \mid \underline{\theta}^t, \underline{y}_{obs})$$

$$\underline{\theta}^{t+1} = \underset{\underline{\theta}}{\mathrm{ARG\,MIN}} \left\{ -\sum_{\underline{y}_{mis}} q(\underline{y}_{mis}) \log p(\underline{y}_{mis}, \underline{\theta} \mid \underline{y}_{obs}) \right\}$$

Example:

Back to the two-Gaussian example from Lecture 14.

Replace binary random variables V^i by probabilities

$$q(V^i=1) = q_i \qquad q(V^i=0) = 1-q_i$$

Compare with DA (on left).

$$\tilde{\mu}_1 = \frac{\bar{\sigma}_m^2 \left(\sum_{i=1}^{M} V^i x^i \right) + \bar{\sigma}^2 \alpha_1}{\bar{\sigma}_m^2 \sum_{i=1}^{M} V^i + \bar{\sigma}^2} \longrightarrow \tilde{\mu}_1^{t+1} = \frac{\bar{\sigma}_m^2 \sum_{i=1}^{M} q_i^t x^i + \bar{\sigma}^2 \alpha_1}{\bar{\sigma}_m^2 \sum_{i=1}^{M} q_i^t x^i + \bar{\sigma}^2}$$

$$\tilde{\sigma}_1^2 = \frac{\bar{\sigma}^2 \bar{\sigma}_m^2}{\bar{\sigma}^2 + \bar{\sigma}_m^2 \sum_{i=1}^{M} V^i} \longrightarrow \tilde{\sigma}_1^{2^{t+1}} = \frac{\bar{\sigma}^2 \bar{\sigma}_m^2}{\bar{\sigma}^2 + \bar{\sigma}_m^2 \sum_{i=1}^{M} q_i^t}$$

$$\tilde{\mu}_2 = \frac{\bar{\sigma}_m^2 \sum_{i=1}^{M} (1-V^i) X^i + \bar{\sigma}^2 \alpha_2}{\bar{\sigma}_m^2 \sum_{i=1}^{M} (1-V^i) + \bar{\sigma}^2} \longrightarrow \tilde{\mu}_2^{t+1} = \frac{\bar{\sigma}_m^2 \sum_{i=1} (1-q_i^t) X^i}{\bar{\sigma}_m^2 \sum_{i=1}^{M} (1-q_i^t) + \bar{\sigma}^2}$$

$$\bar{\sigma}_2^2 = \frac{\bar{\sigma}^2 \bar{\sigma}_m^2}{\bar{\sigma}^2 + \bar{\sigma}_m^2 \sum_{i=1}^{M} (1-V^i)} \longrightarrow \bar{\sigma}_2^{2^{t+1}} = \frac{\bar{\sigma}^2 \bar{\sigma}_m^2}{\bar{\sigma}^2 + \bar{\sigma}_m^2 \sum_{i=1}^{M} (1-q_t^i)}$$

Ising $x_i \in \{-1, 1\}$

Edges E.

Swendsen-Wang.

Standard Metropolis is slow to converge when sampling from the Ising Model.

$$P(\underline{x}) = \frac{1}{Z} e^{\sum_{i,j \in E} x_i x_j / T}$$

particularly slow at low temperature.

Swendsen-Wang (SW) is a way to speed up the sampling by "dynamically grouping" sites into clusters.

Then the state of the entire cluster can be switched.

SW is also an example of Data Augmentation.

(*) Each subset V_i is connected.

Swendsen-Wang

First, Swendsen-Wang for the Potts model.

Potts Model.

Each node s, can take colors $\underline{c}_s \in \{1,2 \dots Q\}$

$$p(\underline{c}) = \frac{1}{Z} e^{\beta \sum_{s,t \in E} I(\underline{c}_s = \underline{c}_t)}$$

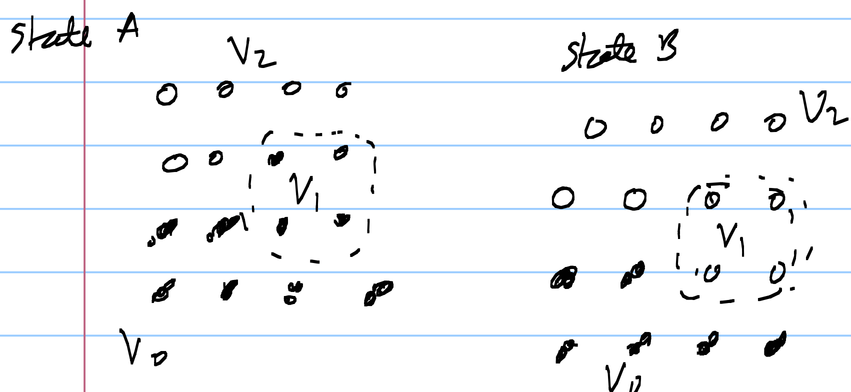
$\{ I(\underline{c}_s = \underline{c}_t) = 1$, if $\underline{c}_s = \underline{c}_t$
$= 0$, otherwise $\}$

Ising model is a special case of Potts with $Q = 2$.

Notation: set of nodes V, partition into subsets $\{V_i : i = 1 \text{ to } n\}$ *

such that $\bigcup_{i=1}^{n} V_i = V$, $V_i \cap V_j = \phi$, $\forall i \neq j$.

All nodes in each subset have the same color.



Partitions differ only by the state of the regions.

$\Pi_A = (V_0 \cup V_1, V_2)$ $\qquad$ $\Pi_B = (V_0, V_1 \cup V_2)$

SW Algorithm.

1. Define a new variable $V_{s,t}$ (binary-valued).

Set. $V_{s,t}=0$ if $c_s \neq c_t$.

If $c_s = c_t$, set $V_{s,t}=1$ with prob $q_0 = 1-e^{-\beta}$, (otherwise $V_{s,t}=0$).

This yields a number of connected components (each is a subset of vertices of the same color).

2. Randomly select a connected component (e.g. the region on previous page).

3. Randomly select a color for this connected component (uniform prob).

Claim: this algorithm is a Markov Chain Monte Carlo. – satisfies detailed balance.

It encourages large groups at small temperature (large β)

Detailed Balance of SW.

Define: $C_A = C(V_0, V_1) = \{(s,t) : s \in V_0, t \in V_1\}$

$C_B = C(V_0, V_1) = \{(s,t) : s \in V_0, t \in V_2\}$

These are the Swendsen-Wang cuts at π_A, π_B

Then, it follows that:

$$\frac{q(\pi_A \to \pi_B)}{q(\pi_B \to \pi_A)} = \frac{(1-q_0)^{|C_A|}}{(1-q_0)^{|C_B|}}$$

where $|C_A|, |C_B|$ are the sizes of C_A & C_B.

For the Potts model.

$$\frac{p(\pi_A)}{p(\pi_B)} = \frac{e^{-\beta |C_B|}}{e^{-\beta |C_A|}}.$$

Hence: if we set $q_0 = 1 - e^{-\beta}$,

then

$$\frac{q(\pi_B \to \pi_A)}{q(\pi_A \to \pi_B)} \frac{p(\pi_B)}{p(\pi_A)} = 1$$

which is detailed balance.

Alternative View of Swendsen-Wang.
on Potts model.

$$\pi(\underline{c}) \qquad \pi(\underline{V} \mid \underline{c}) = \prod_{s,t \in E} \pi(V_{s,t} \mid \underline{c}_s, \underline{c}_t)$$

with $\pi(V_{s,t}=1 \mid \underline{c}_s, \underline{c}_t) = 0$, if $\underline{c}_s \neq \underline{c}_t$

$\pi(V_{s,t}=1 \mid \underline{c}_s, \underline{c}_t) = 1 - e^{-\beta}$, if $\underline{c}_s = \underline{c}_t$

Then do Data Augmentation:

Fix $\underline{c}$,

sample from $\pi(\underline{V} \mid \underline{c})$ to get connected components.

Then sample from $\pi(\underline{c} \mid \underline{V})$ to get the colour of the connected components.

It can be shown that $\pi(\underline{c} \mid \underline{V})$ is a uniform distribution on each color for each connected component.

(By definition, each connected component must have the same colour).

Extending Swendsen-Wang (Barbu & Zhu)

Standard SW is restricted to a few special distributions — e.g. Ising model & Potts model.

The problem is that detailed balance only occurs for these models.

But, we can extend Swendsen-Wang by Metropolis-Hastings.

Use SW to make proposals $q(\pi_A \to \pi_B)$, then accept proposals with probability

$$\alpha(\pi_A \to \pi_B) = \min\left(1, \frac{q(\pi_B \to \pi_A)}{q(\pi_A \to \pi_B)} \frac{p(\pi_B)}{p(\pi_A)}\right)$$

(Note: original SW is a special case where $\alpha(\pi_A \to \pi_B) = 1$, the proposal is automatically accepted because of detailed balance.)