b) for C ... $P(C_i = k)$

$g(C) = 1$ ... $g(+) = 1[C_i = k]$

$\hat{g}(m_1, \ldots, m_M) = \frac{1}{M} \sum_{m=1}^{M} g(C[m])$

$P(C_i = k \mid v_1, \ldots, v_K)$ for $k = 1, \ldots, K$

Let $g(C_i) = 1[C_i = k]$

$\hat{g}(C_i[1], \ldots, C[M]) = \frac{1}{M} \sum_{m=1}^{M} g(C[m])$

c) For Gibbs Sampling to work, it must be true that

$$A(c' \mid c) = 1 = \min(1, 1) =$$

$$\rightarrow \frac{P(v_1 \mid C_1 = 1) \cdot P(v_2 \mid C_2 = 2)}{P(v_1 \mid C_1 = 2) \cdot P(v_2 \mid C_2 = 1)} = 1$$

This will not work because there is no guarantee this quantity $== 1$

$\frac{\partial}{\partial x} f(x, y)$, $f(x, y) = xy$, $\hookrightarrow = y$

2 b)

$$\frac{\partial}{\partial \theta_i} g(\theta ; D) = \frac{1}{|D|} \sum_{x,y \in D} f_i(x, y) - \frac{1}{|D|} \sum_{x \in D} \overset{(1-\alpha)}{E_\theta[f_i(x, y)]} - \frac{1}{|D|} \sum_{y \in D} \overset{(\alpha)}{E_\theta[f_i(x, y)]}$$

if Q = empirical dist:

if $Q = P_\theta(W \mid Z = z)$:

2) Multi-conditional Parameter Learning, Markov Networks

$D = \{(x^1, y^1), \ldots, (x^n, y^n)\}$

model has parameters $\Theta = [\theta_1, \ldots, \theta_n]$, estimated w/:

$$g(\Theta ; D) = (1-\alpha) \ell_{Y|X}(\Theta ; D) + \alpha \ell_{X|Y}(\Theta ; D)$$

$\ell_{X|Y}(\Theta ; D)$ is conditional log-likelihood of D using $P_\Theta(X \mid Y)$ defined by Markov Network w/ parameters Θ. Same w/ $\ell_{Y|X}$

We have n features $f_i(X_i, Y_i)$ where X_i, Y_i are possibly empty subsets of X, Y respectively

a) Write full objective function $g(\theta ; D)$ in terms of all f_i and θ_i:

$$\ell_{Y|X}(\Theta ; D) = \frac{1}{|D|} \log p(D, \theta) = \frac{1}{|D|} \sum_{x,y \in D} \Theta^T f(x, y) - \frac{1}{|D|} \sum_{x \in D} \log Z(x, \Theta)$$

$$\ell_{X|Y}(\Theta ; D) = \frac{1}{|D|} \sum_{x,y \in D} \Theta^T f(x, y) - \frac{1}{|D|} \sum_{y \in D} \log Z(y, \Theta)$$

$$g(\Theta ; D) = (1-\alpha) \left[\frac{1}{|D|} \sum_{x,y \in D} \Theta^T f(x, y) - \frac{1}{|D|} \sum_{x \in D} \log Z(x, \Theta)\right]$$

$$+ \alpha \left[\frac{1}{|D|} \sum_{x,y \in D} \Theta^T f(x, y) - \frac{1}{|D|} \sum_{y \in D} \log Z(y, \Theta)\right]$$

$$= (1-\alpha) \left[\frac{1}{|D|} \sum_{x,y \in D} \Theta^T f(x, y)\right] - (1-\alpha) \left[\frac{1}{|D|} \sum_{x \in D} \log Z(x, \Theta)\right] + //$$

$$= 1 \cdot \left[\frac{1}{|D|} \sum_{x,y \in D} \Theta^T f(x, y)\right] - (1-\alpha) \left[\frac{1}{|D|} \sum_{x \in D} \log Z(x, \Theta)\right] - \alpha \left[\frac{1}{|D|} \sum_{y \in D} \log Z(y, \Theta)\right]$$

where $Z(X, \Theta) = \sum_{y \in D} \Theta^T f(x, y)$, $Z(y, \Theta) = \sum_{x \in D} \Theta^T f(x, y)$