Solutions to Homework 4

March 8, 2006

1. Imagine that you are in charge of a political survey where each subject is classified according to four category spaces—a) conventional (C): {liberal (0), conservative (1)}, b) nonlinear (N): {moderate (0), radical (1)}, c) orthogonal (O): {authoritarian (0), libertarian (1)} and d) moral (M): {exclusivist (0), inclusivist (1)}. In each space, please note that we have assigned binary values to the labels. For example, liberal is assigned '0' and conservative '1' etc. We are interested in studying the co-occurrences between these spaces.

- Given a pool of candidates who have been classified according to the four category structures above, how would you estimate the joint probability $\Pr(C, N, O, M)$ between the four spaces from the data? [Explain qualitatively how you'd build up the four-way probability distribution.]

- You are given the following: $\Pr(C = 0) = x$, $\Pr(N = 0) = y$, $\Pr(O = 0) = z$, and $\Pr(M = 0) = w$. Also, $\Pr(C = 0, N = 0) = a$, $\Pr(N = 0, O = 0) = b$, $\Pr(O = 0, M = 0) = c$, and $\Pr(M = 0, C = 0) = d$. Evaluate the pairwise joint probabilities $\Pr(C, N)$, $\Pr(N, O)$, $\Pr(O, M)$ and $\Pr(M, C)$ given this information. [You'll need to use basic rules relating two variable probability distributions to single variable probability distributions. Pretend that a, b, c, d and x, y, z, w are numbers. Now, write all the probabilities in terms of these 8 numbers. You'll need to know that $\sum_c \Pr(C = c, N) = \Pr(N)$.]

- Given the above pairwise probabilities, estimate the full joint probability $\Pr(C, N, O, M)$ two ways. In case 1, remove $\Pr(M, C)$ to get a tree. In case 2, remove $\Pr(C, N)$ to get a tree. List the conditional probability approximations in both cases. Write down all 16 possibilities for both cases. [Warning: This will take some time. However, since the Fall 2002 class completely botched this question, I'm making sure that if I asked you ten years from now to answer this question, you'll do it like a zombie. We're going for permanent memory etching here. If you think this is horse%$#@, please realize that it builds character.]

- For both cases above, evaluate $\Pr(C, O)$ which was not given to you. [Establish expressions such that both approaches give you the same answer for all four possibilities. You cannot assume that you have the co-occurrences from the data from which to estimate $\Pr(C, O)$.]

- a) Since there are 16 possibilities in total, the ideal way of estimating the joint probability $\Pr(C, N, O, M)$ is to get statistics for all 16 possibilities.
- b) This is a straightforward probability question. We'll write down a few of these probabilities.

Probability	Value
$\Pr(C = 1, N = 1)$	$a + 1 - x - y$
$\Pr(N = 1, O = 1)$	$b + 1 - y - z$
$\Pr(O = 1, M = 1)$	$c + 1 + z - w$
$\Pr(M = 1, C = 1)$	$d + 1 - w - x$

- c) You have to use the tree probability formula two ways to get the full expansion of the joint probability.
- d) When all four probabilities of $\Pr(C, O)$ are equated, we get $a = d$, $b = c$, and $y = w$. Going back to the question, we see that this implies that the probability of being a liberal and a moderate must equal the probability of being a liberal and a moral exclusivist. And, the probability of being an authoritarian and a moderate must equal the probability of being an authoritarian and a moral exclusivist. Finally, the probability of being a moderate must equal the probability of being a moral exclusivist. I think you'll agree that these are very serious constraints on the space of possibilities and this example serves to demonstrate exactly what happens when simplifying independence assumptions are made.

2. Show the equivalence of the following free energy to the Bethe free energy

$$
\begin{aligned}
F_{\text{equiv}}(\{p_{ij}, p_i, \sigma_{ij}, \rho_i, \gamma_{ij}, \lambda_{ij}\}) &= \sum_{ij:i>j} \sum_{x_i, x_j} p_{ij}(x_i, x_j) \log \frac{p_{ij}(x_i, x_j)}{\sigma_{ji}^{\delta}(x_i)\sigma_{ij}^{\delta}(x_j)} \\
&+ \delta \sum_{ij:i\neq j} \sum_{x_i} \sigma_{ij}(x_j) + \sum_i \sum_{x_i} p_i(x_i) \log \frac{p_i(x_i)}{\rho_i^{q_i}(x_i)} + \sum_i q_i \sum_{x_i} \rho_i(x_i) \\
&- \sum_{ij:i>j} \sum_{x_i, x_j} p_{ij}(x_i, x_j) \log \psi_{ij}(x_i, x_j) \psi_i^{\xi_i}(x_i) \psi_j^{\xi_j}(x_j) - \sum_i r_i \sum_{x_i} p_i(x_i) \log \psi_i(x_i) \\
&+ \sum_{ij:i>j} \sum_{x_j} \lambda_{ij}(x_j)[\sum_{x_i} p_{ij}(x_i, x_j) - p_j(x_j)] + \sum_{ij:i>j} \sum_{x_i} \lambda_{ji}(x_i)[\sum_{x_j} p_{ij}(x_i, x_j) - p_i(x_i)] \\
&+ \sum_{ij:i>j} \gamma_{ij}(\sum_{x_i, x_j} p_{ij}(x_i, x_j) - 1). \qquad (1)
\end{aligned}
$$

What are the fixed point equations for $\{p_{ij}, p_i, \sigma_{ij}, \rho_i\}$? [You'll have to provide the constraints on δ, q_i, r_i, ξ_i in order to establish equivalence.]
Minimizing the objective function w.r.t. σ, we get

$$
-\delta \frac{\sum_{x_j} p_{ij}(x_i, x_j)}{\sigma_{ji}(x_i)} + \delta = 0 \Rightarrow \sigma_{ji}(x_i) = \sum_{x_j} p_{ij}(x_j), \forall ij.
$$

Minimizing the objective function w.r.t. ρ, we get

$$-q_i \frac{p_i(x_i)}{\rho_i(x_i)} + q_i = 0 \Rightarrow \rho_i(x_i) = p_i(x_i), \forall i.$$

If $q_i = (1 - \delta)n_i$ and $r_i = 1 - n_i\xi_i$ we get equivalence to the Bethe free energy when we set $\delta = 0$ and $\xi_i = 1$ since then $q_i = n_i$ and $r_i = 1 - n_i$. The constraints ensure equivalence.