Derivatives for a Two-component Binomial Mixture Model

Jason D. M. Rennie
jrennie@csail.mit.edu

July 1, 2004*

We observe many samples of a binomial. h_i and n_i are the numbers of heads and total flips for the i^{th} sample. For each sample, there are two binomials that could have produced it. Our goal is to learn the mixing parameter and parameters for the two binomials that are most likely to have produced the samples. Our model is:

$$p(D|\theta) = \prod_i \binom{n_i}{h_i} \left[\lambda \phi_1^{h_i}(1-\phi_1)^{n_i-h_i} + (1-\lambda)\phi_2^{h_i}(1-\phi_2)^{n_i-h_i}\right] \quad (1)$$

For numerical stability reasons, we don't simply maximize likelihood in order to find the best parameters. We re-parameterize using (unconstrained) natural parameters, and minimize the negative log-odds ratio of the mixture and (simple) binomial likelihoods. Let $g(x) = (1+\exp(-x))^{-1}$, the logistic function. We reparameterize as follows:

$$\lambda = g(u) \qquad \phi_1 = g(v_1) \qquad \phi_2 = g(v_2) \quad (2)$$

Let $\phi^* = \frac{\sum_i h_i}{\sum_i n_i}$ be the maximum-likelihood (simple) binomial parameter. We define the following ratios which appear in the log-likelihood ratio:

$$r_{11} = \frac{\phi_1}{\phi^*} \qquad r_{12} = \frac{1-\phi_1}{1-\phi^*} \qquad r_{21} = \frac{\phi_2}{\phi^*} \qquad r_{22} = \frac{1-\phi_2}{1-\phi^*} \quad (3)$$

We define the following quantities which are useful for writing the log-odds and its derivatives:

$$p_{i1} = r_{11}^{h_i} r_{12}^{n_i-h_i}, \quad (4)$$

$$p_{i2} = r_{21}^{h_i} r_{22}^{n_i-h_i}, \quad (5)$$

$$z_i = \lambda p_{i1} + (1-\lambda) p_{i2}, \quad (6)$$

*Updated May 17, 2005

Then the log-odds is simply

$$l(D|\theta) = -\sum_i \log(z_i) \tag{7}$$

Note that $\frac{\partial g(x)}{\partial x} = g(x)(1 - g(x))$. The partial derivatives are:

$$\frac{\partial l}{\partial u} = -\sum_i \frac{\frac{\partial \lambda}{\partial u} p_{i1} + \frac{\partial (1-\lambda)}{\partial u} p_{i2}}{z_i} \qquad = -\lambda(1-\lambda) \sum_i \frac{p_{i1} - p_{i2}}{z_i} \tag{8}$$

$$\frac{\partial l}{\partial v_1} = -\sum_i \frac{\lambda}{z_i} \frac{\partial p_{i1}}{\partial v_1} \qquad = -\sum_i \frac{\lambda p_{i1}}{z_i} (h_i - \phi_1 n_i) \tag{9}$$

$$\frac{\partial l}{\partial v_2} = -\sum_i \frac{(1-\lambda)}{z_i} \frac{\partial p_{i2}}{\partial v_2} \qquad = -\sum_i \frac{(1-\lambda) p_{i2}}{z_i} (h_i - \phi_2 n_i) \tag{10}$$

One can use general-purpose optimization software to solve for (locally) maximum-likelihood parameters.