BIOSTAT M280, Homework 6

Due Fri Mar 11 @ 11:59PM, 2016

Consider again the MLE of the Dirichlet-multinomial model. In HW5, we worked out a Newton's method. In this homework, we explore the EM and MM approach.

1. Show that, given iid observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, the log-likelihood can be written as
$$L(\boldsymbol{\alpha}) \quad = \quad \sum_{i=1}^{n} \ln \binom{|\boldsymbol{x}_i|}{\boldsymbol{x}_i} + \sum_{i=1}^{n} \sum_{j=1}^{d} \sum_{k=0}^{x_{ij}-1} \ln(\alpha_j + k) - \sum_{i=1}^{n} \sum_{k=0}^{|\boldsymbol{x}_i|-1} \ln(|\boldsymbol{\alpha}| + k).$$
Hint: $\Gamma(a+k)/\Gamma(a) = a(a+1)\cdots(a+k-1)$.

2. Suppose $(P_1, \ldots, P_d) \in \Delta_d = \{\boldsymbol{p} : p_i \geq 0, \sum_i p_i = 1\}$ follows a Dirichlet distribution with parameter $(\alpha_1, \ldots, \alpha_d)$. Show that
$$\mathbf{E}(\ln P_j) = \Psi(\alpha_j) - \Psi(|\boldsymbol{\alpha}|),$$
where $\Psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function and $|\boldsymbol{\alpha}| = \sum_{j=1}^{d} \alpha_j$. (Hint: Differentiate the identity $1 = \int_{\Delta_d} \frac{\Gamma(|\boldsymbol{\alpha}|)}{\prod_{j=1}^{d} \Gamma(\alpha_j)} \prod_{j=1}^{d} p_j^{\alpha_j - 1} \, d\boldsymbol{p}$.)

3. The admixture representation of the Dirichlet-multinomial distribution suggests that we can treat the unobserved multinomial parameters $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ as missing data and derive an EM algorithm. Show that the Q function is
$$Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(t)}) = \sum_{j=1}^{d} \sum_{i=1}^{n} \alpha_j \left[\Psi(x_{ij} + \alpha_j^{(t)}) - \Psi(|\boldsymbol{x}_i| + |\boldsymbol{\alpha}^{(t)}|)\right] - n \sum_{j=1}^{d} \ln \Gamma(\alpha_j) + n \ln \Gamma(|\boldsymbol{\alpha}|) + c^{(t)},$$
where $c^{(t)}$ is a constant irrelevant to optimization. Comment on why it is not easy to maximize the Q function.

4. We derive an MM algorithm for maximing L. Consider the first formulation of the log-likelihood that contains terms $\ln(\alpha_j + k)$ and $-\ln(|\boldsymbol{\alpha}| + k)$. Applying Jensen's inequality to the concave term $\ln(\alpha_j + k)$ and supporting hyperplane inequality to the convex term $-\ln(|\boldsymbol{\alpha}| + k)$, show that a minorizing function to $L(\boldsymbol{\alpha})$ is
$$g(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(t)}) \quad = \quad - \sum_{k=0}^{\max_i |\boldsymbol{x}_i| - 1} \frac{r_k}{|\boldsymbol{\alpha}^{(t)}| + k} |\boldsymbol{\alpha}| + \sum_{j=1}^{d} \sum_{k=0}^{\max_i x_{ij} - 1} \frac{s_{jk} \alpha_j^{(t)}}{\alpha_j^{(t)} + k} \ln \alpha_j + c^{(t)},$$
where $s_{jk} = \sum_{i=1}^{n} 1_{\{x_{ij} > k\}}$, $r_k = \sum_{i=1}^{n} 1_{\{|\boldsymbol{x}_i| > k\}}$, and $c^{(t)}$ is a constant irrelevant to optimization. Maximizing the surrogate function $g(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(t)})$ is trivial since α_j are separated. Show that the MM updates are
$$\alpha_j^{(t+1)} = \frac{\sum_{k=0}^{\max_i x_{ij} - 1} \frac{s_{jk}}{\alpha_j^{(t)} + k}}{\sum_{k=0}^{\max_i |\boldsymbol{x}_i| - 1} \frac{r_k}{|\boldsymbol{\alpha}^{(t)}| + k}} \alpha_j^{(t)}, \quad j = 1, \ldots, d.$$
The quantities s_{jk}, r_k, $\max_i x_{ij}$ and $\max_i |\boldsymbol{x}_i|$ only depend on data and can be pre-computed. Comment on whether the MM updates respect the parameter constraint $\alpha_j > 0$.

5. Write a function for finding MLE of Dirichlet-multinomial distribution given iid observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, using MM algorithm. The interface should be `dirmultfit.MM(x, alpha0 = NULL, maxiters = 1000, tolfun = 1e-6)`. The arguments are: `x` a n-by-d matrix of counts, `alpha0` a d vector of starting point (optional), `maxiters` the maximum allowable MM iterations (default 1000), `tolfun` the tolerance for relative change in objective values (default 1e-6). The return value should be a list containing: `maximum` the log-likelihood at MLE, `estimate` the MLE, `gradient` the gradient at MLE, `hessian` the Hessian at MLE, `se` a d vector of standard errors, `iterations` the number of iterations performed.

6. Re-do HW5 Q10 using your new `dirmultfit.MM` function. Compare the number of iterations and run time by MM algorithm to those by Newton's method. Comment on the efficiency of Newton's algorithm vs MM algorithm for this problem.

7. Finally let us re-consider the EM algorithm. The difficulty with the M step in EM algorithm can be remedied. Discuss how we can further minorize the $\ln \Gamma(|\boldsymbol{\alpha}|)$ term in the Q function to produce a minorizing function with all α_j separated. For this homework, you do *not* need to implement this EM-MM hybrid algorithm. (Hint: $z \mapsto \ln \Gamma(z)$ is a convex function.)