MLE using GOP

December 12, 2017

1 Model

We will try to maximize likelihood for simple model with only one hidden variable and n observed variables.

$$p(x) = \theta_h(x_h) \prod_i \theta_{ih}(x_i, x_h)$$

Likelihood of M samples is,

$$L(\theta) = \prod_{m=1}^{M} \sum_{x_h} p(x_1^m, ..., x_n^m, x_h)$$

Taking log of likelihood and normalizing it, we get,

$$\log L(\theta) = \frac{1}{M} \sum_{m=1}^{M} \log \sum_{x_h} p(x_1^m, ..., x_n^m, x_h)$$

Writing it in variational form,

$$\begin{aligned}
\log L(\theta) &= \max_b \frac{1}{M} \sum_m \langle b_h^m, \log \theta_h' \rangle + H(b_h^m) \\
&\quad \text{where, } \theta_h'(x_h) = \theta_h(x_h) * \prod_i \theta_{ih}(x_i^m, x_h) \\
\log L(\theta) &= \max_b \frac{1}{M} \sum_m \sum_{x_h} b_h^m(x_h) \log(\theta_h(x_h)) + \frac{1}{M} \sum_m \sum_{x_h} \sum_i b_h^m(x_h) \log(\theta_{ih}(x_i^m, x_h)) + \frac{1}{M} \sum_m H(b_h^m)
\end{aligned}$$

Define $\bar{b}_h(x_h)$ and $\bar{b}_{ih}(x_i, x_h)$ and rewrite the formulation,

$$\begin{aligned}
\log L(\theta) &= \max_b \sum_{x_h} \bar{b}_h(x_h) \log(\theta_h(x_h)) + \sum_i \sum_{x_i, x_h} \bar{b}_{ih}(x_i, x_h) \log(\theta_{ih}(x_i, x_h)) + \frac{1}{M} \sum_m H(b_h^m) \\
&\quad \text{where, } \bar{b}_h(x_h) = \frac{1}{M} \sum_m b_h^m(x_h) \\
&\quad \bar{b}_{ih}(x_i, x_h) = \frac{1}{M} \sum_{\substack{m \\ x_i^m = x_i}} b_h^m(x_h)
\end{aligned}$$

2 Problem

We want to solve following problem,

$$-\log L = \min_{\theta, b} - \sum_{x_h} \bar{b}_h(x_h) \log(\theta_h(x_h)) - \sum_i \sum_{x_i, x_h} \bar{b}_{ih}(x_i, x_h) \log(\theta_{ih}(x_i, x_h)) - \frac{1}{M} \sum_m H(b_h^m)$$

with respect to following constraints,

$$
\begin{aligned}
&\bar{b}_h(x_h) = \frac{1}{M}\sum_m b_h^m(x_h), \forall x_h \\
&\bar{b}_{ih}(x_i, x_h) = \frac{1}{M}\sum_{\substack{m \\ x_i^m = x_i}} b_h^m(x_h), \forall x_i, x_h \\
&\sum_{x_h} \theta_h(x_h) = 1 \\
&\sum_{x_i} \theta_{ih}(x_i, x_h) = 1 \\
&\sum_{x_h} b_h^m(x_h) = 1, \forall m \in M
\end{aligned}
$$

3 Primal formulation

3.1 Primal problem

We need to solve following primal problem for fixed θ ,

$$
\begin{aligned}
P(\theta) = \min_{\bar{b}, b} \Bigg\{ & -\sum_{x_h} \bar{b}_h(x_h)\log(\theta_h(x_h)) - \sum_i \sum_{x_i, x_h} \bar{b}_{ih}(x_i, x_h)\log(\theta_{ih}(x_i, x_h)) - \frac{1}{M}\sum_m H(b_h^m) \\
& + \sum_{x_h} \lambda_h(x_h)\left(\bar{b}_h(x_h) - \frac{1}{M}\sum_m b_h^m(x_h)\right) + \sum_i \sum_{x_i, x_h} \lambda_{ih}(x_i, x_h)\left(\bar{b}_{ih}(x_i, x_h) - \frac{1}{M}\sum_{\substack{m \\ x_i^m = x_i}} b_h^m(x_h)\right) \\
& + \sum_{m=1}^{M} \lambda_m\left(\sum_{x_h} b_h^m(x_h) - 1\right) + \lambda_{\theta_h}\left(\sum_{x_h} \theta_h(x_h) - 1\right) + \sum_{i, x_h} \lambda_{\theta_{ih}}(x_h)\left(\sum_{x_i} \theta_{ih}(x_i, x_h) - 1\right) \Bigg\}
\end{aligned}
$$

Taking derivatives to get λ values,

$$
\frac{\partial P(\theta)}{\partial \bar{b}_h(x_h)} = -\log(\theta_h(x_h)) + \lambda_h(x_h) = 0 \Rightarrow \lambda_h(x_h) = \log(\theta_h(x_h))
$$

$$
\frac{\partial P(\theta)}{\partial \bar{b}_{ih}(x_i, x_h)} = -\log(\theta_{ih}(x_i, x_h)) + \lambda_{ih}(x_i, x_h) = 0 \Rightarrow \lambda_{ih}(x_i, x_h) = \log(\theta_{ih}(x_i, x_h))
$$

$$
\frac{\partial P(\theta)}{\partial b_h^m(x_h)} = 0 \Rightarrow 1 + \log b_h^m(x_h) - \frac{1}{M}\lambda_h(x_h) - \frac{1}{M}\sum_i \lambda_{ih}(x_i^m, x_h) + \lambda_m = 0 \;, \forall x_h
$$

so $b_h^m(x_h)$ has closed form solution,

$$
b_h^m(x_h) = \frac{\exp(\frac{1}{M}\left(\lambda_h(x_h) + \sum_i \lambda_{ih}(x_i^m, x_h)\right))}{\exp(\lambda_m + 1)}
$$

Here $exp(\lambda_m + 1)$ is normalization factor for $b_h^m(x_h)$, such that,

$$
\exp(\lambda_m + 1) = \sum_{x_h} \exp\left(\frac{1}{M}\left(\lambda_h(x_h) + \sum_i \lambda_{ih}(x_i^m, x_h)\right)\right)
$$

$$
\frac{\partial P(\theta)}{\partial \theta_h(x_h)} = 0 \Rightarrow \theta_h(x_h) = \frac{\bar{b}_h(x_h)}{\lambda_{\theta_h}} \;, \forall x_h
$$

Summing over x_h on both sides,

$$\sum_{x_h} \theta_h(x_h) = \frac{\sum_{x_h} \bar{b}_h(x_h)}{\lambda_{\theta_h}}$$

Both sums equal 1, so

$\lambda_{\theta_h} = 1$

Similarly,

$$\frac{\partial P(\theta)}{\partial \theta_{ih}(x_i, x_h)} = 0 \Rightarrow \lambda_{\theta_{ih}} = \frac{\bar{b}_{ih}(x_i, x_h)}{\theta_{ih}(x_i, x_h)} \ , \forall x_i, x_h$$

Substituting $\bar{b}$ for θ Primal problem can be written as

$$\begin{aligned}
P = \min_b \Bigg\{ & -\sum_{x_h} \bar{b}_h(x_h)\log(\bar{b}_h(x_h)) - \sum_i \sum_{x_i, x_h} \bar{b}_{ih}(x_i, x_h)\log(\bar{b}_{ih}(x_i, x_h)) - \frac{1}{M}\sum_m H(b_h^m) \\
& + \sum_{x_h} \lambda_h(x_h)\left(\bar{b}_h(x_h) - \frac{1}{M}\sum_m b_h^m(x_h)\right) + \sum_i \sum_{x_i, x_h} \lambda_{ih}(x_i, x_h)\left(\bar{b}_{ih}(x_i, x_h) - \frac{1}{M}\sum_{\substack{m \\ x_i^m = x_i}} b_h^m(x_h)\right) \\
& + \sum_{m=1}^{M} \lambda_m \left(\sum_{x_h} b_h^m(x_h) - 1\right) + \lambda_{\bar{b}_h}\left(\sum_{x_h} \bar{b}_h(x_h)(x_h) - 1\right) + \sum_{i, x_h} \lambda_{\bar{b}_{ih}}(x_h)\left(\sum_{x_i} \bar{b}_{ih}(x_i, x_h) - 1\right)\Bigg\} \\
= & \min_{\bar{b}, b} f(\bar{b}, b)
\end{aligned}$$

$$\begin{aligned}
P = \min_b \Bigg\{ & -\sum_{x_h} \frac{1}{M}\sum_m b_h^m(x_h)\log(\frac{1}{M}\sum_m b_h^m(x_h)) - \sum_i \sum_{x_i, x_h} \frac{1}{M}\sum_{\substack{m \\ x_i^m = x_i}} b_h^m(x_h)\log(\frac{1}{M}\sum_{\substack{m \\ x_i^m = x_i}} b_h^m(x_h)) - \frac{1}{M}\sum_m H(b_h^m) \\
& + \sum_{m=1}^{M} \lambda_m \left(\sum_{x_h} b_h^m(x_h) - 1\right) + \lambda_{b_h}\left(\sum_{x_h} \frac{1}{M}\sum_m b_h^m(x_h) - 1\right) + \sum_{i, x_h} \lambda_{b_{ih}}(x_h)\left(\sum_{x_i} \frac{1}{M}\sum_{\substack{m \\ x_i^m = x_i}} b_h^m(x_h) - 1\right)\Bigg\} \\
= & \min_b f(b)
\end{aligned}$$

$$\begin{aligned}
P = \min_b \Bigg\{ & -\sum_{x_h} \frac{1}{M}\sum_m b_h^m(x_h)\log(\frac{1}{M}\sum_m b_h^m(x_h)) - \sum_i \sum_{x_i, x_h} \frac{1}{M}\sum_{\substack{m \\ x_i^m = x_i}} b_h^m(x_h)\log(\frac{1}{M}\sum_{\substack{m \\ x_i^m = x_i}} b_h^m(x_h)) - \frac{1}{M}\sum_m H(b_h^m) \\
& + \sum_{m=1}^{M} \log(\frac{1}{M}\sum_m b_h^m(x_h))\left(\sum_{x_h} b_h^m(x_h) - 1\right) - \log(b_h^m(x_h))\left(\sum_{x_h} \frac{1}{M}\sum_m b_h^m(x_h) - 1\right) \\
& + \sum_{i, x_h}(\frac{1}{M} + \log(\frac{1}{M}\sum_{\substack{m \\ x_i^m = x_i}} b_h^m(x_h)))\left(\sum_{x_i} \frac{1}{M}\sum_{\substack{m \\ x_i^m = x_i}} b_h^m(x_h) - 1\right)\Bigg\} \\
= & \min_b f(b)
\end{aligned}$$

3.2 Calculate first Conjugate

We can calculate the first Conjugate for f as follows

$$f^*(\bar{b}^*, b^*) = sup_{\bar{b}, b}(< \bar{b}^*, \bar{b} > + < b^*, b > - f(\bar{b}, b))$$

Let g($\bar{b}$, b) $= < \bar{b}^*, \bar{b} > + < b^*, b > - f(\bar{b}, b)$, Taking the derivatives to calculate $\bar{b}$, b values

$$\frac{\partial g(\bar{b},b)}{\partial \bar{b}_h(x_h)} = \bar{b}_h^*(x_h) + \log(\bar{b}_h(x_h)) - \lambda_h(x_h) - \lambda_{\bar{b}_h} + 1 = 0 \Rightarrow -\bar{b}_h^*(x_h) + \lambda_h(x_h) + \lambda_{\bar{b}_h} - 1 = \log(\bar{b}_h(x_h)) \Rightarrow$$
$$\exp(-\bar{b}_h^*(x_h) + \lambda_h(x_h) + \lambda_{\bar{b}_h} - 1) = \bar{b}_h(x_h)$$

$$\frac{\partial g(\bar{b},b)}{\partial \bar{b}_{ih}(x_i,x_h)} = \bar{b}_{ih}^*(x_i,x_h) + \log(\bar{b}_{ih}(x_i,x_h)) - \lambda_{ih}(x_i,x_h) - \lambda_{\bar{b}_{ih}} + 1 = 0 \Rightarrow$$
$$-\bar{b}_{ih}^*(x_i,x_h) + \lambda_{ih}(x_i,x_h) + \lambda_{\bar{b}_{ih}} - 1 = \log(\bar{b}_{ih}(x_i,x_h)) \Rightarrow$$
$$\exp(-\bar{b}_{ih}^*(x_i,x_h) + \lambda_{ih}(x_i,x_h) + \lambda_{\bar{b}_{ih}} - 1) = \bar{b}_{ih}(x_i,x_h)$$

$$\frac{\partial g(\bar{b},b)}{\partial b_h^m(x_h)} = {b_h^*}^m(x_h) - \frac{1}{M} - \frac{\log(b_h^m(x_h))}{M} + \frac{\lambda_h(x_h)}{M} - \lambda_m - \frac{1}{M} * \sum_i \sum_{x_i,x_h} \sum_{\substack{m \\ x_i^m = x_i}} \lambda_{ih}(x_i,x_h) \Rightarrow$$
$$M * {b_h^*}^m(x_h) - 1 + \lambda_h(x_h) - M * \lambda_m - \sum_i \sum_{x_i,x_h} \sum_{\substack{m \\ x_i^m = x_i}} \lambda_{ih}(x_i,x_h) = \log(b_h^m(x_h)) \Rightarrow$$
$$\exp(M * {b_h^*}^m(x_h) - 1 + \lambda_h(x_h) - M * \lambda_m - \sum_i \sum_{x_i,x_h} \sum_{\substack{m \\ x_i^m = x_i}} \lambda_{ih}(x_i,x_h)) = b_h^m(x_h)$$

$$\min_{\theta,\mu_B} \mu_B$$

Subject to,

$$\mu_B \geq L(b^{B_j},\theta,\lambda^k)|_{b^k}^{lin}$$
$$\Delta_{b_i} L(b,\theta,\lambda^k)|_{b^k} \leq 0 \text{ if } b_i^{B_j} = b_i^U$$
$$\Delta_{b_i} L(b,\theta,\lambda^k)|_{b^k} \geq 0 \text{ if } b_i^{B_j} = b_i^L$$

where, $\forall j \in UL(k,K)$

$k = 1,2,...,K-1$

UL(k,K) is set of lagrange functions from k'th iteration whose qualifying constraints satisfy at the current fixed value θ^K for current primal problem.

$$\mu_B \geq L(b^{B_l},\theta,\lambda^K)|_{b^K}^{lin}$$
$$\Delta_{b_i} L(b,\theta,\lambda^K)|_{b^K} \leq 0 \text{ if } b_i^{B_l} = b_i^U$$
$$\Delta_{b_i} L(b,\theta,\lambda^K)|_{b^K} \geq 0 \text{ if } b_i^{B_j} = b_i^L$$

Linearization of lagrange function around b^k at iteration k is defined as,

$$L^k(b,\theta,\lambda^k)|_{b^k}^{lin} = L^k(b^k,\theta,\lambda^k) + \Delta_b L(b^k,\theta,\lambda^k)(b - b^k)$$

After plugging in values of b^k, λ^k and simplifying,

$$L^k(b,\theta,\lambda^k)|_{b^k}^{lin} = -\sum_{x_h} \bar{b}_h^k(x_h)\log(\theta_h(x_h)) - \sum_i \sum_{x_i,x_h} \bar{b}_{ih}^k(x_i,x_h)\log(\theta_{ih}(x_i,x_h)) + \frac{1}{M}\sum_m \sum_{x_h} b_h^{m^k}(x_h)\log b_h^{m^k}(x_h)$$
$$+ \lambda_{\theta_h}^k \left(\sum_{x_h} \theta_h(x_h) - 1\right) + \sum_{i,x_h} \lambda_{\theta_{ih}}^k(x_h)\left(\sum_{x_i} \theta_{ih}(x_i,x_h) - 1\right)$$
$$- \sum_{x_h}(\bar{b}_h(x_h) - \bar{b}_h^k(x_h))\log\theta_h(x_h) - \sum_i \sum_{x_i,x_h}(\bar{b}_{ih}(x_i,x_h) - \bar{b}_{ih}^k(x_i,x_h))\log(\theta_{ih}(x_i,x_h))$$
$$\frac{1}{M}\sum_m \sum_{x_h}(1 + \log b_h^{m^k}(x_h))(b_h^m(x_h) - b_h^{m^k}(x_h)) + \sum_{x_h} \lambda_h^k(x_h)(\bar{b}_h(x_h) - \bar{b}_h^k(x_h))$$

$$-\sum_{x_h}\frac{1}{M}\lambda_h^k(x_h)(b_h^m(x_h)-b_h^{m^k}(x_h))+\sum_i\sum_{x_i,x_h}\lambda_{ih}^k(x_i,x_h)(\bar{b}_{ih}(x_i,x_h)-\bar{b}_{ih}^k(x_i,x_h))$$
$$-\sum_i\sum_{x_h}\frac{1}{M}\sum_m\lambda_{ih}^k(x_i^m,x_h)(b_h^m(x_h)-b_h^{m^k}(x_h))+\sum_m\lambda_m^k\sum_{x_h}(b_h^k(x_h)-b_h^{m^k}(x_h))$$

Canceling out some terms,

$$\begin{aligned}L^k(b,\theta,\lambda^k)|_{b^k}^{lin}=&\lambda_{\theta_h}^k\left(\sum_{x_h}\theta_h(x_h)-1\right)+\sum_{i,x_h}\lambda_{\theta_{ih}}^k(x_h)\left(\sum_{x_i}\theta_{ih}(x_i,x_h)-1\right)\\&-\sum_{x_h}\bar{b}_h(x_h)\log\theta_h(x_h)-\sum_i\sum_{x_i,x_h}\bar{b}_{ih}(x_i,x_h)\log(\theta_{ih}(x_i,x_h))-\frac{1}{M}\sum_m\sum_{x_h}b_h^{m^k}(x_h)\\&\frac{1}{M}\sum_m\sum_{x_h}(1+\log b_h^{m^k}(x_h))b_h^m(x_h)+\sum_{x_h}\lambda_h^k(x_h)(\bar{b}_h(x_h)-\bar{b}_h^k(x_h))\\&-\sum_{x_h}\frac{1}{M}\lambda_h^k(x_h)(b_h^m(x_h)-b_h^{m^k}(x_h))+\sum_i\sum_{x_i,x_h}\lambda_{ih}^k(x_i,x_h)(\bar{b}_{ih}(x_i,x_h)-\bar{b}_{ih}^k(x_i,x_h))\\&-\sum_i\sum_{x_h}\frac{1}{M}\sum_m\lambda_{ih}^k(x_i^m,x_h)(b_h^m(x_h)-b_h^{m^k}(x_h))+\sum_m\lambda_m^k\sum_{x_h}(b_h^m(x_h)-b_h^{m^k}(x_h))\end{aligned}$$

After Summing up the terms includes both ($b_h^{m^k}(x_h)$, and $b_h^m(x_h)$) and substitute for ($\lambda_h^k(x_h)$, and $\lambda_{ih}^k(x_i,x_h)$)

$$\begin{aligned}L^k(b,\theta,\lambda^k)|_{b^k}^{lin}=&\lambda_{\theta_h}^k\left(\sum_{x_h}\theta_h(x_h)-1\right)+\sum_{i,x_h}\lambda_{\theta_{ih}}^k(x_h)\left(\sum_{x_i}\theta_{ih}(x_i,x_h)-1\right)\\&-\sum_{x_h}\bar{b}_h(x_h)\log\theta_h(x_h)-\sum_i\sum_{x_i,x_h}\bar{b}_{ih}(x_i,x_h)\log(\theta_{ih}(x_i,x_h))+\frac{1}{M}\sum_m\sum_{x_h}b_h^{m^k}(x_h)\log(b_h^{m^k}(x_h))\\&+\sum_{x_h}\log(\theta_h^k(x_h))(\bar{b}_h(x_h)-\bar{b}_h^k(x_h))+\sum_i\sum_{x_i,x_h}\log(\theta_{ih}^k(x_i,x_h))(\bar{b}_{ih}(x_i,x_h)-\bar{b}_{ih}^k(x_i,x_h))\end{aligned}$$

Now we can write the equation in the form

$$\begin{aligned}L^k(b,\theta,\lambda^k)|_{b^k}^{lin}=&\lambda_{\theta_h}^k\left(\sum_{x_h}\theta_h(x_h)-1\right)+\sum_{i,x_h}\lambda_{\theta_{ih}}^k(x_h)\left(\sum_{x_i}\theta_{ih}(x_i,x_h)-1\right)\\&-\sum_{x_h}\log(\theta_h^k(x_h))\bar{b}_h^k(x_h)-\sum_i\sum_{x_i,x_h}\log(\theta_{ih}^k(x_i,x_h))\bar{b}_{ih}^k(x_i,x_h)+\frac{1}{M}\sum_m\sum_{x_h}b_h^{m^k}(x_h)\log(b_h^{m^k}(x_h))\\&+\sum_{x_h}\bar{b}_h(x_h)(\log(\theta_h^k(x_h))-\log\theta_h(x_h))+\sum_i\sum_{x_i,x_h}\bar{b}_{ih}(x_i,x_h)(\log(\theta_{ih}^k(x_i,x_h))-\log(\theta_{ih}(x_i,x_h)))\end{aligned}$$

We can see that first line of the equation depends only on θ, the second line equals to a constant, and the third line depends on both θ and the connected values

Now Let's find the derivatives with respect to θ

$$\frac{\partial L^k(b,\theta,\lambda^k)|_{b^k}^{lin}}{\partial\theta_h(x_h)}=\lambda_{\theta_h}^k-\frac{\bar{b}_h(x_h)}{\theta_h(x_h)}=0\Rightarrow\theta_h(x_h)=\frac{\bar{b}_h(x_h)}{\lambda_{\theta_h}^k}$$
$$\frac{\partial L^k(b,\theta,\lambda^k)|_{b^k}^{lin}}{\partial\theta_{ih}^k(x_i,x_h)}=\lambda_{\theta_{ih}}^k(x_h)-\frac{\bar{b}_h(x_h)}{\theta_{ih}^k(x_i,x_h)}=0\Rightarrow\theta_{ih}^k(x_i,x_h)=\frac{\bar{b}_h(x_h)}{\lambda_{\theta_{ih}}^k(x_h)}$$