

EMPIRICAL BAYES POISSON MATRIX FACTORIZATION

ZIHAO WANG

1. Overview. We introduce an Empirical Bayes approach to Poisson Matrix Factorization (EBPMF), whose key feature is that it uses the observed data to estimate prior distributions on matrix elements. We solve this problem using Variational Inference, which reduces the fitting of EBPMF to solving “Poisson Means” problem.

The writeup is organized as follows.

- * Section 2 introduces notation.
- * Section 3 introduces the Empirical Bayes Poisson Mean problem (EBPM)
- * Section 4 derives the Maximum Likelihood Estimation for Poisson Matrix Factorization using Expectation-Maximization (EM), which is the same as [lee’s multiplicative update](#)
- * Section 5 introduces the rank-1 version of EBPMF, and shows its connection to the EBPM problem
- * Section 6 introduces the rank-k version of EBPMF, and shows its connection to the EBPM problem

2. Notation.

- * For a 3d tensor Z_{ijk} with shape (n, p, K) , I use capital character to indicate a slice along that axis. For example, Z_{Ijk} means $Z_{1:n,jk}$. Similar notation applies to 2d arrays.
- * $\langle x \rangle_q := E_q(x)$

3. Empirical Bayes Poisson Mean Problem (EBPM). Suppose we have observations \mathbf{x} and scale \mathbf{s} , and we assume the following generating process.

$$\begin{aligned}x_i | \lambda_i &\sim \text{Pois}(s_i \lambda_i) \\ \lambda_i &\sim g(\cdot) \\ g &\in \mathcal{G}\end{aligned}$$

Our goal is to find

$$\begin{aligned}\hat{g} &:= \arg\max_g \ell(g) = \arg\max \log p(\mathbf{x}|g, \mathbf{s}) \\ p &:= p(\boldsymbol{\lambda}|\mathbf{x}, \hat{g}, \mathbf{s})\end{aligned}$$

Suppose we can solve this type of problem, and use EBPM to denote the mapping:

$$EBPM(\mathbf{x}, \mathbf{s}) = (p, \hat{g})$$

REMARK. The MLE for the Poisson mean problem is $\hat{\lambda}_i = \frac{x_i}{s_i}$.

4. MLE for Poisson Matrix Factorization (PMF). Let’s first look at an algorithm of a related problem - MLE for Poisson Matrix Factorization (PMF):

$$X_{ij} \sim \text{Pois}(\sum_k l_{ik} f_{jk})$$

and we want to find the MLE for L, F . The most popular algorithm is **lee's multiplicative update**, which is essentially an EM algorithm.

4.1. rank 1 case. We set $K = 1$. It is easy to see that the optimal l_i, f_j (update to a scaling factor between \mathbf{l}, \mathbf{f}) is

$$l_i = \frac{\sum_j X_{ij}}{\sum_j f_j}$$

$$f_j = \frac{\sum_i X_{ij}}{\sum_i l_i}$$

REMARK. We can see that the optimal \mathbf{l} is the MLE solution to the Poisson mean problem where $x_i := \sum_j X_{ij}$ and $s_i := \sum_j f_j$. Similar remark for \mathbf{f} .

4.2. rank k case. In EM we use the data augmentation trick which can reduce solving the rank-k problem to solving rank-1 problems:

$$X_{ij} = \sum_k Z_{ijk}$$

$$Z_{ijk} \sim \text{Pois}(l_{ik} f_{jk})$$

In E-step, we compute $p(Z_{ijk}|X_{ij}, l_{iK}, f_{jK})$, which gives us $\langle Z_{ijk} \rangle_p = X_{ij} \frac{l_{ik} f_{jk}}{\sum_k l_{ik} f_{jk}}$. In M-step, we update L, F by optimizing $E[\log p(X, Z|L, F)]$, which gives us

$$l_{ik} = \frac{\sum_j \langle Z_{ijk} \rangle_p}{\sum_j f_{jk}}$$

$$f_{jk} = \frac{\sum_i \langle Z_{ijk} \rangle_p}{\sum_i l_{ik}}$$

REMARK. We can expand $\langle Z_{ijk} \rangle_p$ and get the famous lee's multiplicative update, but this form is more similar to our Poisson Mean problem: each column of the updated L is the MLE for Poisson Mean problem with $x_i := \sum_j \langle Z_{ijk} \rangle_p$ and $s_i := \sum_j f_{jk}$. Similar remark for F .

5. EBPMF: rank-1. I will solve this problem using Mean-Field Variational Inference, which can be shown to be equivalent to the EBPM problem. This naturally provides us with an algorithm to solve EBPMF (rank-1) problem. Then it is also easy to see the connection between our algorithm and the EM for PMF (rank-1).

5.1. Model.

$$X_{ij} \sim \text{Pois}(l_i f_j)$$

$$l_i \sim g_L(\cdot), g_L \in \mathcal{G}$$

$$f_j \sim g_F(\cdot), g_F \in \mathcal{G}$$

We assume that data is from a Poisson distribution, and the mean is rank-1. We can impose different priors on elements of \mathbf{l}, \mathbf{f} .

Our goal is to find:

$$\hat{g}_L, \hat{g}_F := \operatorname{argmax} l(g_L, g_F) = \operatorname{argmax} \log p(X|g_L, g_F)$$

and posterior of \mathbf{l}, \mathbf{f} : $p(\mathbf{l}|X, \hat{g}_L, \hat{g}_F), p(\mathbf{f}|X, \hat{g}_F, \hat{g}_L)$.

5.2. Variational Inference. The marginal $p(X|g_L, g_F)$ is intractable. Therefore, we use mean-field variational inference, by using approximating distribution $q(\mathbf{l}, \mathbf{f}) = q_L(\mathbf{l})q_F(\mathbf{f})$. Then we have:

$$\begin{aligned} \log p(X|g_L, g_F) &= ELBO(q_L, q_F, g_L, g_F) + KL(q(\mathbf{l}, \mathbf{f})||p(\mathbf{l}, \mathbf{f}|X, g_L, g_F)) \\ &\geq ELBO(q_L, q_F, g_L, g_F) \end{aligned}$$

where the evidence lower bound (ELBO) is defined by

$$\begin{aligned} ELBO(q_L, q_F, g_L, g_F) &:= E_q(\log p(X, \mathbf{l}, \mathbf{f})) - E_q(\log q(\mathbf{l}, \mathbf{f})) \\ &= E_q[\log p(X|\mathbf{l}, \mathbf{f}) + \log \frac{g_L(\mathbf{l})}{q_L(\mathbf{l})} + \log \frac{g_F(\mathbf{f})}{q_F(\mathbf{f})}] \end{aligned}$$

Now the goal is to maximize ELBO over q_L, q_F, g_L, g_F .

5.3. Optimize ELBO by solving EBPM problem. By proposition A.1, we can see

$$\operatorname{argmax}_{q_F, g_F} ELBO(q_L, q_F, g_L, g_F) = EBPM(\sum_i X_{ij}, (\sum_i \langle l_i \rangle) \mathbf{1})$$

$$\operatorname{argmax}_{q_L, g_L} ELBO(q_L, q_F, g_L, g_F) = EBPM(\sum_j X_{ij}, (\sum_j \langle f_j \rangle) \mathbf{1})$$

Next, by observing that in $ebpm(\mathbf{x}, s\mathbf{1})$, the estimated prior and approximate posterior (\hat{g}, \hat{q}) for $s\mathbf{1}$ is invariant to the value of s , we can see the equations above are sufficient and necessary conditions for optimality. Therefore we have the simple algorithm for "EBPMF-rank1" this way:

Algorithm 5.1 Alternating Optimization for EBPMF (rank-1)

Result: $\hat{q}_L, \hat{q}_F, \hat{g}_L, \hat{g}_F$

- 1 Input: X , initial values $q_L^{(0)}, q_F^{(0)}, g_L^{(0)}, g_F^{(0)}$
 $q_F^*, g_F^* \leftarrow EBPM(\sum_i X_{ij}, (\sum_i \langle l_i \rangle_{q_L^{(0)}}) \mathbf{1})$
 $q_L^*, g_L^* \leftarrow EBPM(\sum_j X_{ij}, (\sum_j \langle f_j \rangle_{q_F^*}) \mathbf{1})$
-

REMARK. *It is clear that both algorithms solve the same Poisson Means problems. In PMF-MLE we use MLE for the Poisson Means problem; in EBPMF we solve it with an Empirical Bayes approach instead.*

6. EBPMF: rank-k. Like in the rank-k case for PMF-MLE, we introduce hidden variable Z_{ijk} :

$$\begin{aligned} X_{ij} &= \sum_k Z_{ijk} \\ Z_{ijk} &\sim \text{Pois}(l_{ik}f_{jk}) \\ l_{ik} &\sim g_{L,k}(\cdot), g_{L,k} \in \mathcal{G} \\ f_{jk} &\sim g_{F,k}(\cdot), g_{F,k} \in \mathcal{G} \end{aligned}$$

We introduce hidden variable Z so that it is easy to see the connection to "rank-1" case. Independent priors are placed for each column of F, L , both for flexible modeling, and computational convenience.

6.1. Variational Inference. Similarly, we use factorized variational inference:

$$\begin{aligned} q(L, F, Z) &= q_L(L)q_F(F)q_Z(Z) \\ q_Z(Z_{ij,1:K}) &= MN(X_{ij}, \zeta_{ij,1:K}) \end{aligned}$$

The optimal parametric family of $q_L(L), q_F(F)$ is dependent on the choice of \mathcal{G} .

6.2. Reduce rank-k to rank-1. The ELBO for rank-k case is:

$$\begin{aligned} (6.1) \quad ELBO(q, g) &= E_q[\log p(X, L, F, Z | g_{L,1:K}, g_{F,1:K})] - E_q \log q_L(L) - E_q \log q_F(F) - E_q \log q_Z(Z) \\ &= E_q \{ \log p(Z | L, F) + \log \frac{g_L(L)}{q_L(L)} + \log \frac{g_F(F)}{q_F(F)} - \log q_Z(Z) \} \\ &= E_q \{ \sum_{ijk} [-l_{ik} f_{jk} + Z_{ijk} \log(l_{ik} f_{jk}) - \log(Z_{ijk}!)] \\ &\quad + \log \frac{g_L(L)}{q_L(L)} + \log \frac{g_F(F)}{q_F(F)} \\ &\quad - \sum_{ij} [\sum_k Z_{ijk} \log(\zeta_{ijk}) + \log(X_{ij}!) - \sum_k \log(Z_{ijk}!)] \} \\ &= \sum_k \{ [\sum_{ij} (-\langle l_{ik} \rangle_q \langle f_{jk} \rangle_q + \langle Z_{ijk} \rangle_q \langle \log(l_{ik}) \rangle_q + \langle \log(f_{jk}) \rangle_q)] + E_q \left(\log \frac{g_L(L_{Ik})}{q_L(L_{Ik})} + \log \frac{g_F(F_{Jk})}{q_F(F_{Jk})} \right) \} \\ &\quad - \sum_{ijk} \langle Z_{ijk} \rangle_q \log(\zeta_{ijk}) - \sum_{ij} \log(X_{ij}!) \end{aligned}$$

REMARK. In the second equation, X is ignored as $\sum_k Z_{ijk} = X_{ij}$ with probability 1 under $q_Z(\cdot)$.

Now the last equation tells us how to optimize the ELBO. We separate parameters into two parts (g_L, g_F, q_L, q_F) and q_Z , then optimize alternately. If we fix $q(Z)$, The first part is just k ELBOs for rank-1 problem (X_{ij} replaced with $\langle Z_{ijk} \rangle_q$), which we can solve independently. If we fix the rest and optimize $q(Z)$: we can decompose the problem into np independent subproblems, each can be solved analytically (using Lagrange multiplier). The optimizer w.r.t ζ is:

$$\begin{aligned} \zeta_{ijk} &\propto \exp(\langle \ln(l_{ik}) \rangle_q + \langle \ln(f_{jk}) \rangle_q) \\ \sum_k \zeta_{ijk} &= 1 \end{aligned}$$

The algorithm is summarized below.

Algorithm 6.1 Alternating Optimization for EBPMF (rank-k)

Result: $\hat{q}_L, \hat{q}_F, \hat{g}_L, \hat{g}_F$

2 Input: X , initial values $q_{L,1:K}^{(0)}, q_{F,1:K}^{(0)}, g_{L,1:K}^{(0)}, g_{F,1:K}^{(0)}$
 $t \leftarrow 1$
 while *not converged* **do**
3 $t \leftarrow t + 1$
 for $k \leftarrow 1$ **to do**
4 $q_{L,k}^{(t)}, g_{L,k}^{(t)}, q_{F,k}^{(t)}, g_{F,k}^{(t)} = \text{EBPMF-rank1}(\langle Z_{ijk} \rangle_q, q_{L,k}^{(t-1)}, g_{L,k}^{(t-1)}, q_{F,k}^{(t-1)}, g_{F,k}^{(t-1)})$
 update q_Z
5 **end**
6 **end**

Appendices

A. Connecting EBPM problem to EBPM problem.

PROPOSITION A.1.

$$\operatorname{argmax}_{q_F, g_F} ELBO(q_L, q_F, g_L, g_F) = EBPM(\sum_i X_{ij}, (\sum_i \langle l_i \rangle) \mathbf{1})$$

where $\langle l_i \rangle_q := E_{q_L}(l_i)$.

Proof. The idea is that LHS is also the ELBO for the RHS. In order to see this, we define the ELBO for RHS as F^{PM} , then show LHS can be expressed as F^{PM} .

Let's define

$$\begin{aligned} F^{PM}(q_\lambda, g; \mathbf{x}, \mathbf{s}) &:= E_q[\log p(\mathbf{x}|\boldsymbol{\lambda}, \mathbf{s}) + \log(\frac{g(\boldsymbol{\lambda})}{q_\lambda(\boldsymbol{\lambda})})] \\ &= \sum_t E_q(-s_t \lambda_t + x_t \log(s_t \lambda_t)) + E_q(\log(\frac{g(\boldsymbol{\lambda})}{q_\lambda(\boldsymbol{\lambda})})) - \sum_t \log(x_t!) \end{aligned}$$

By the following lemma, we can see F^{PM} is the ELBO of the RHS:

LEMMA A.2.

$$\operatorname{argmax}_{q_\lambda, g} F^{PM}(q_\lambda, g; \mathbf{x}, \mathbf{s}) = EBPM(\mathbf{x}, \mathbf{s})$$

Proof.

$$\begin{aligned} \ell(g) &= \log p(\mathbf{x}|g) \\ &= E_q[\log p(\mathbf{x}|g)] \\ &= E_q[\log \frac{p(\mathbf{x}, \boldsymbol{\lambda}|g)}{p(\boldsymbol{\lambda}|\mathbf{x}, g)}] \\ &= E_q[\log \frac{p(\mathbf{x}, \boldsymbol{\lambda}|g)}{q_\lambda(\boldsymbol{\lambda})} + \log \frac{q_\lambda(\boldsymbol{\lambda})}{p(\boldsymbol{\lambda}|\mathbf{x}, g)}] \\ &= E_q[\log p(\mathbf{x}|\boldsymbol{\lambda}) + \log \frac{g(\boldsymbol{\lambda})}{q_\lambda(\boldsymbol{\lambda})}] + KL(q_\lambda || p_{\boldsymbol{\lambda}|\mathbf{x}, g}) \\ &= F^{PM}(q_\lambda, g; \mathbf{x}, \mathbf{s}) + KL(q_\lambda || p_{\boldsymbol{\lambda}|\mathbf{x}, g}) \\ &\geq F^{PM}(q_\lambda, g; \mathbf{x}, \mathbf{s}) \end{aligned}$$

Then by

$$F^{PM}(q_\lambda, g) = \ell(g) - KL(q_\lambda || p_{\boldsymbol{\lambda}|\mathbf{x}, g})$$

we know the maximizer (over q) for $F^{PM}(q_\lambda, g)$ is the posterior $p_{\boldsymbol{\lambda}|\mathbf{x}, g}$, which also makes KL divergence equal 0. Therefore we have

$$\max_{q_\lambda, g} F^{PM}(q_\lambda, g) = \max_g \ell(g)$$

Then our lemma follows. \square

Then we can express ELBO (w.r.t F) in terms of F^{PM} :

$$\begin{aligned}
ELBO(q_F, g_F) &= E_q[\log p(X|\mathbf{l}, \mathbf{f}) + \log \frac{g_L(\mathbf{l})}{q_L(\mathbf{l})} + \log \frac{g_F(\mathbf{f})}{q_F(\mathbf{f})}] \\
&= \sum_j E_q[-(\sum_i \langle l_i \rangle) f_j + (\sum_i X_{ij}) \log f_j + \log \frac{g(f_j)}{q_F(f_j)}] + const \\
&= F^{PM}(q_F, g_F; \sum_i X_{ij}, (\sum_i \langle l_i \rangle) \mathbf{1}) + const
\end{aligned}$$

Therefore, our proposition follows. \square

B. ELBO computation. We do not need to compute ELBO explicitly in our algorithm, but it is useful as an indicator of progress for the algorithm.

B.1. rank 1. ELBO for rank-1 is:

$$\begin{aligned}
ELBO(q_L, q_F, g_L, g_F) &:= E_q[\log p(X|\mathbf{l}, \mathbf{f}) + \log \frac{g_L(\mathbf{l})}{q_L(\mathbf{l})} + \log \frac{g_F(\mathbf{f})}{q_F(\mathbf{f})}] \\
&= \sum_{i,j} [-\langle l_i \rangle_q \langle f_j \rangle_q + X_{ij} (\langle \log(l_i) \rangle_q + \langle \log(f_j) \rangle_q)] \\
&\quad + E_q[\log \frac{g_L(\mathbf{l})}{q_L(\mathbf{l})} + \log \frac{g_F(\mathbf{f})}{q_F(\mathbf{f})}] - \sum_{ij} \log(X_{ij}!)
\end{aligned}$$

However, we can't get a closed form expression for the terms

$$E_q[\log \frac{g_L(\mathbf{l})}{q_L(\mathbf{l})} + \log \frac{g_F(\mathbf{f})}{q_F(\mathbf{f})}] = -KL(q_L(\mathbf{l})||g_L(\mathbf{l})) - KL(q_F(\mathbf{f})||g_F(\mathbf{f}))$$

But fortunately we can use the relationship between ELBO and the corresponding EBPM problem.

After each update of our algorithm, we have $\hat{q}_F(\cdot) = p(\cdot|X, \hat{g}_F)$, thus

$$F^{PM}(\hat{q}_F, \hat{g}_F) = l(\hat{g}_F)$$

and RHS can be easily computed in our EBPM algorithm. Also

$$F^{PM}(\hat{q}_F, \hat{g}_F; \sum_i X_{ij}, \sum_i \langle l_i \rangle_q \mathbf{1}) = E_q[\log p(\sum_i X_{ij}|\mathbf{f}, \sum_i \langle l_i \rangle_q \mathbf{1}) + \log \frac{g_F(\mathbf{f})}{q_F(\mathbf{f})}] - \sum_j \log((\sum_i X_{ij})!)$$

where the first term is easily computable:

$$E_q[\log p(\sum_i X_{ij}|\mathbf{f}, \sum_i \langle l_i \rangle_q \mathbf{1})] = \sum_j [(\sum_i \langle l_i \rangle_q) \langle f_j \rangle_q + (\sum_i X_{ij}) \langle \log f_j \rangle_q]$$

Thus we can compute $E_q[\log \frac{g_F(\mathbf{f})}{q_F(\mathbf{f})}]$ (and similarly $E_q[\log \frac{g_L(\mathbf{l})}{q_L(\mathbf{l})}]$) and plug them in to get EBPM.

B.2. rank k. Similarly the only hard part is the KL divergence between g, q for L, F . We can get them from EBPM results. Thus,

for each rank-1 problem, we can compute $E_q[\frac{g_L(L_{Ik})}{q_L(L_{Ik})} + \log \frac{g_F(F_{Jk})}{q_F(F_{Jk})}]$.

Then we compute $\sum_k \{[\sum_{ij} (-\langle l_{ik} \rangle_q \langle f_{jk} \rangle_q + \langle Z_{ijk} \rangle_q \langle \log(l_{ik}) \rangle_q + \langle \log(f_{jk}) \rangle_q)]$ and

$$E_q(\log \frac{g_L(L_{Ik})}{q_L(L_{Ik})} + \log \frac{g_F(F_{Jk})}{q_F(F_{Jk})}).$$

Add them up we get ELBO up to a constant ($\sum_{ij} \log(X_{ij}!)$). Let's write down the formula explicitly:

$$\begin{aligned} ELBO(q, g) &= \sum_k \{ [\sum_{ij} (-\langle l_{ik} \rangle_q \langle f_{jk} \rangle_q + \langle Z_{ijk} \rangle_q \langle \log(l_{ik}) \rangle_q + \langle \log(f_{jk}) \rangle_q)] + E_q(\log \frac{g_L(L_{Ik})}{q_L(L_{Ik})} + \log \frac{g_F(F_{Jk})}{q_F(F_{Jk})}) \} \\ &\quad - \sum_{ijk} \langle Z_{ijk} \rangle_q \log(\zeta_{ijk}) - \sum_{ij} \log(X_{ij}!) \\ &= \sum_{ijk} [(-\langle l_{ik} \rangle_q \langle f_{jk} \rangle_q + \langle Z_{ijk} \rangle_q (\langle \log(l_{ik}) \rangle_q + \langle \log(f_{jk}) \rangle_q - \log(\zeta_{ijk}))) \\ &\quad + \sum_{ijk} (\log \frac{g_L(L_{Ik})}{q_L(L_{Ik})} + \log \frac{g_F(F_{Jk})}{q_F(F_{Jk})}) - \sum_{ij} \log(X_{ij}!) \\ &= \sum_{ijk} [(-\langle l_{ik} \rangle_q \langle f_{jk} \rangle_q + X_{ij} \zeta_{ijk} (\langle \log(l_{ik}) \rangle_q + \langle \log(f_{jk}) \rangle_q - \log(\zeta_{ijk}))) \\ &\quad + \sum_{ijk} (\log \frac{g_L(L_{Ik})}{q_L(L_{Ik})} + \log \frac{g_F(F_{Jk})}{q_F(F_{Jk})}) - \sum_{ij} \log(X_{ij}!) \\ &= \sum_{ijk} (-\langle l_{ik} \rangle_q \langle f_{jk} \rangle_q + X_{ij} \log (\sum_k \exp(\langle \log(l_{ik}) \rangle_q + \langle \log(f_{jk}) \rangle_q)) \\ &\quad + \sum_{ijk} (\log \frac{g_L(L_{Ik})}{q_L(L_{Ik})} + \log \frac{g_F(F_{Jk})}{q_F(F_{Jk})}) - \sum_{ij} \log(X_{ij}!) \end{aligned}$$