# 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.

**2. Notation.**

1. 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.
2. $\langle x \rangle_q := E_q(x)$

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

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

Our goal is to find

$$\begin{aligned} &\hat{g} := \text{argmax}_g\ \ell(g) = \text{argmax}\ log\ p(\boldsymbol{x}|g, \boldsymbol{s}) \\ &p := p(\boldsymbol{\lambda}|\boldsymbol{x}, \hat{g}, \boldsymbol{s}) \end{aligned}$$

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

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

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):

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

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 $\boldsymbol{l}, \boldsymbol{f}$) is

$$\begin{aligned} l_i &= \frac{\sum_j X_{ij}}{\sum_j f_j} \\ f_j &= \frac{\sum_i X_{ij}}{\sum_i l_i} \end{aligned}$$

Remark. *We can see that the optimal* $\boldsymbol{l}$ *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* $\boldsymbol{f}$.

**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:

$$X_{ij} = \sum_k Z_{ijk}$$
$$Z_{ijk} \sim Pois(l_{ik} f_{jk})$$

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.**

$$X_{ij} \sim Pois(l_i f_j)$$
$$l_i \sim g_L(.), g_L \in \mathcal{G}$$
$$f_j \sim g_F(.), g_F \in \mathcal{G}$$

We assume that data is from a Poisson distribution, and the mean is rank-1. We can impose different priors on elements of $\boldsymbol{l}, \boldsymbol{f}$.
Our goal is to find:

$$\hat{g_L}, \hat{g_F} := \text{argmax } l(g_L, g_F) = \text{argmax } log\ p(X|g_L, g_F)$$

and posterior of $\boldsymbol{l}, \boldsymbol{f}$: $p(\boldsymbol{l}|X, \hat{g_L}, \hat{g_F}),\ p(\boldsymbol{f}|X, \hat{g_F}, \hat{g_F})$.

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

$$log\ p(X|g_L, g_F) = ELBO(q_L, q_F, g_L, g_F) + KL(q(\boldsymbol{l}, \boldsymbol{f})||p(\boldsymbol{l}, \boldsymbol{f}|X, g_L, g_F))$$

where the evidence lower bound (ELBO) is defined by

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

Since KL divergence is nonnegative and obtains 0 when the two distributions are the same, $log\ p(X|g_L, g_F) = max_q ELBO(q_L, q_F, g_L, g_F)$. Thus

$$\max_g\ log\ p(X|g_L, g_F) = \max_{g,q}\ ELBO(q_L, q_F, g_L, g_F)$$

Now we focus on the RHS instead, and optimize ELBO over $q_L, q_F, g_L, g_F$, and use the optimal $q_L, q_F$ as approximation to the true posterior.

**Algorithm 5.1** Alternating Optimization for EBPMF (rank-1)

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

Input: $X$, initial values $q_L^{(0)}, q_F^{(0)}, g_L^{(0)}, g_F^{(0)}$
$t \leftarrow 1$
**while** *not converged* **do**
  $t \leftarrow t + 1$
  $q_L^{(t)}, g_L^{(t)} \leftarrow \text{argmax}_{q_L, g_L}\ ELBO(q_L, q_F^{(t)}, g_L, g_F^{(t)})$
  $q_F^{(t)}, g_F^{(t)} \leftarrow \text{argmax}_{q_F, g_F}\ ELBO(q_L^{(t)}, q_F, g_L^{(t)}, g_F)$
**end**

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

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

Therefore we can rewrite the "EBPMF-rank1" this way:

**Algorithm 5.2** Alternating Optimization for EBPMF (rank-1)

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

Input: $X$, initial values $q_L^{(0)}, q_F^{(0)}, g_L^{(0)}, g_F^{(0)}$
$t \leftarrow 1$
**while** *not converged* **do**
  $t \leftarrow t + 1$
  $q_F^{(t)}, g_F^{(t)} \leftarrow EBPM(\sum_i X_{ij}, (\sum_i \langle l_i \rangle_{q^{(t-1)}})\mathbf{1})$
  $q_L^{(t)}, g_L^{(t)} \leftarrow EBPM(\sum_j X_{ij}, (\sum_j \langle f_j \rangle_{q^{(t)}})\mathbf{1})$
**end**

**5.4. Attain optimality in one iteration.** In fact, this algorithm converges after the 1st iteration, as is the case for the MLE for PMF problem:
Suppose in iteration $t + 1$, we have $\sum_i \langle l_i \rangle_{q^{(t)}} = c \sum_i \langle l_i \rangle_{q^{(t-1)}}$. Then we have $q_L^{(t+1)}(l_i) = q_L^{(t)}(cl_i)$ and $g_L^{(t+1)}(l_i) = g_L^{(t)}(cl_i)$. Then $\sum_i \langle f_j \rangle_{q^{(t+1)}} = \frac{1}{c} \sum_i \langle f_j \rangle_{q^{(t)}}$, so we have $q_F^{(t+1)}(f_j) = q_F^{(t)}(f_j/c)$ and $g_F^{(t+1)}(f_j) = g_F^{(t)}(f_j/c)$. We can see $ELBO^{(t+1)} = ELBO^{(t)}$.
The same trick can be applied to see result from the first iteration is indeed optimal easily.

**5.5. Compare with MLE for PMF (rank-1).** It is clear that both algorithms solve the same Poisson Mean problems. In MLE we use MLE for the Poisson Mean problem; in EBPMF we solve the Empirical Bayes Poisson Mean problem instead.

**6. EBPMF: rank-k.** I will solve this problem using Variational Inference, which can be reduced to rank-1 problems using the data augmentation trick. This leads to an iterative algorithm for optimizing ELBO.

**6.1. Model.**

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

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

**6.2. Variational Inference.** Similarly, we use factorized varaitional 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 $\mathcal{G}$.

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

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

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

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 alternatingly.
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 $\zeta$ is:

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

The agorithm is summarized below.

**Algorithm 6.1** Alternating Optimization for EBPMF (rank-k)

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

Input: $X$, initial values $q^{(0)}_{L,1:K}, q^{(0)}_{F,1:K}, g^{(0)}_{L,1:K}, g^{(0)}_{F,1:K}$

$t \leftarrow 1$

**while** *not converged* **do**

  $t \leftarrow t + 1$

  **for** $k \leftarrow 1$ **to do**

    $q^{(t)}_{L,k}, g^{(t)}_{L,k}, q^{(t)}_{f,k}, g^{(t)}_{F,k} =$ EBPMF-rank1($\langle Z_{ijk} \rangle_q, q^{(t-1)}_{L,k}, g^{(t-1)}_{L,k}, q^{(t-1)}_{F,k}, g^{(t-1)}_{F,k}$)

    update $q_Z$

  **end**

**end**

# Appendices

**A. Connecting EBPM problem to EBPM problem.**

PROPOSITION A.1.

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

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

*Proof.* In order to see this, let's introduce the following lemma:

LEMMA A.2.
$argmax_{q_\lambda,g} F^{PM}(q_\lambda, g; \boldsymbol{x}, \boldsymbol{s}) = EBPM(\boldsymbol{x}, \boldsymbol{s})$

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

*Proof.*

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

Then by

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

we know the maximizer (over $q$) for $F^{PM}(q_\lambda, g)$ is the posterior $p_{\boldsymbol{\lambda}|\boldsymbol{x},g}$, 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.

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

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

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:

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

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

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

But fortunately we can use the relationship between ELBO and the corresponding EBPM problem.
After each update of our algorithm, we have $\hat{q}_F(.) = p(.|X, \hat{g}_F)$, thus

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

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

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

where the first term is easily computable:

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

Thus we can compute $E_q[log\ \frac{g_F(\boldsymbol{f})}{q_F(\boldsymbol{f})}]$ (and similarly $E_q[log\ \frac{g_L(\boldsymbol{l})}{q_L(\boldsymbol{l})}]$) 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:

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

The first and the last term together is the "loglikelihood" (compare the first part with $log\ p(X|L,F)$):

$$
\begin{aligned}
log\ p(X|L,F) &= \sum_{ij}(-\sum_k l_{ik} f_{jk} + X_{ij} log(\sum_k l_{ik} f_{jk})) - \sum_{ij} log(X_{ij}!) \\
\text{“loglikelihood”} &= \sum_{ij}(-\sum_k \langle l_{ik}\rangle_q \langle f_{jk}\rangle_q + X_{ij} log\ (\sum_k exp(\langle log(l_{ik})\rangle_q + \langle log(f_{jk})\rangle_q)) - \sum_{ij} log(X_{ij}!) \\
&\leq \sum_{ij}(-\sum_k \langle l_{ik}\rangle_q \langle f_{jk}\rangle_q + X_{ij} log\ (\sum_k (\langle l_{ik}\rangle_q + \langle f_{jk}\rangle)) - \sum_{ij} log(X_{ij}!)
\end{aligned}
$$

(from Jensen inequality $exp(E(log(x))) \leq E(x)$).

The second part is the negative of the KL divergence between $g, q$s, which we compute along with the EBPM algorithm.