Empirical Bayes Multiscale Poisson Matrix Factorization

Dongyue Xie

Oct 2019

1 Introduction

A smoothed Poisson Matrix Factorization method is introduced. Factors are smooth and loadings might be sparse.

2 Empirical Bayes Poisson Mean

3 Empirical Bayes Binomial Probability

4 Empirical Bayes Multiscale Poisson Smoothing

Suppose we observe a Poisson sequence $X \in R^n$, $n = 2^S$ and $X_i \sim Poisson(\lambda_i)$. A Haar wavelet like decomposition of X is defined as follows. Let $s = 0, 1, ..., S-1$ denote scale and $l = 0, 1, ..., 2^s - 1$ denote location. X follows a recursive dyadic partition(RDP):

$$X_{S,l} = X_{kl},$$

$$X_{s,l} = X_{s+1,2l} + X_{s+1,2l+1}.$$

Since the sum of two independent Poisson distributed random variables is also Poisson distributed,

$$X_{s,l} \sim Poisson(\lambda_{s,l}),$$

$$\lambda_{s,l} = \lambda_{s+1,2l} + \lambda_{s+1,2l+1}$$

Define

$$R_{s,l} = \frac{\lambda_{s+1,2l}}{\lambda_{s,l}},$$

then

$$X_{s+1,2l} | X_{s,l}, R_{s,l} \sim Binomial(X_{s,l}, R_{s,l}).$$

This multiscale decomposition reparameterizes λ to $\mu := \lambda_{0,0}$ and $R := \{R\}_{s,l}$

Following this decomposition, the likelihood of sequence X is

$$p(X|\lambda) = \prod_{i=0}^{n-1} p(X_i|\lambda_i) = p(X_{0,0}|\mu) \times \prod_{s=0}^{S-1} \prod_{l=0}^{2^s-1} p(X_{s+1,2l}|X_{s,l}, R_{s,l}). \tag{1}$$

An obvious estimate of μ is $\hat{\mu} = X_{0,0} = \sum_i X_i$. In wavelet shrinkage estimation, this preserves the energy and the estimate is unaffected by perturbations in the data at original scale.

We put a multiscale prior on $R_{s,l}$, $R_{s,l} \sim g_{R_s}(\cdot)$. One choice is a mixture of Beta distributions

$$R_{s,l} \sim p_s \frac{1}{2} + (1 - p_s) Beta(a_s, a_s). \tag{2}$$

An empirical Bayes procedure first estimates $g_{R_s}(\cdot)$ by maximizing the marginal likelihood, then obtains posterior $p(R|X, \hat{g})$.

The log marginal likelihood is

$$\begin{aligned}
\log p(X|g_R) &= \mathbb{E}_{q_R} \log p(X|g_R) \\
&= \mathbb{E}_{q_R} \log \frac{p(X, R|g_R)}{p(R|X, g_R)} \\
&= \mathbb{E}_{q_R} (\log \frac{p(X, R|g_R)}{q_R} + \log \frac{q_R}{p(R|X, g_R)}) \\
&= \mathbb{E}_{q_R} \log p(X|R) + \mathbb{E}_{q_R} \log \frac{g_R}{q_R} + KL(q_R||p(R|X, g_R)) \\
&= \mathcal{F}(q_R, g_R; X) + KL(q_R||p(R|X, g_R)) \\
&\geq \mathcal{F}(q_R, g_R; X).
\end{aligned} \tag{3}$$

$$\mathcal{F}(q_R, g_R; X) = \mathbb{E}_{q_R} (\sum_{s=0}^{S-1} \sum_{l=0}^{2^s-1} \log p(X_{s+1,2l}|X_{s,l}, R_{s,l})) + \mathbb{E}_{q_R} \log \frac{g_R}{q_R}. \tag{4}$$

The bound is tight if choose $q_R = p(R|X, g_R)$ so $\arg\max_{g_R} \log p(X|g_R) = \arg\max_{g_R} \mathcal{F}(g_R; X)$.

5 Rank-1 Empirical Bayes Multiscale Poisson Matrix Factorization

Let $Z_{ij} \sim Poisson(l_i f_j)$, $i = 1, 2, ..., N$, $j = 1, 2, ..., p$, where $l \sim g_l(\cdot)$. Assume $f = (f_1, ..., f_p)$ is smooth or spatially structured. Then each row of Z, denoted as $z_{i:}$, is a sequence with smooth mean $l_i f$.

Define the row sums of Z as

$$z_{\cdot j} = \sum_i Z_{ij},$$

then

$$z_{\cdot j} \sim Poisson(f_j \sum_i l_i).$$

A multiscale decomposition of $z = \{z_{\cdot 1}, ..., z_{\cdot p}\}$ is

$$z_{S,l} = z_l, l = 1, 2, ..., p$$

$$z_{s,l} = z_{s+1,2l} + z_{s+1,2l+1}, s = 0, 1, ..., S-1, l = 0, 1, ..., 2^s - 1.$$

Then

$$z_{S,l} \sim Poisson(\sum_i l_i \times f_{S,l})$$

$$z_{s,l} \sim Poisson(\sum_i l_i \times f_{s,l}),$$

where $f_{S,l} = f_l$ for $l = 1, 2, ..., p$ and $f_{s,l} = f_{s+1,2l} + f_{s+1,2l+1}$ for $s = 0, 1, ..., S-1$ and $l = 0, 1, ..., 2^s - 1$.

Define

$$R_{s,l} = \frac{f_{s+1,2l}}{f_{s,l}},$$

then

$$z_{s+1,2l} | z_{s,l}, R_{s,l} \sim Binomial(z_{s,l}, R_{s,l}).$$

We put a multiscale prior on $R_{s,l}$, $R_{s,l} \sim g_{R_s}(\cdot)$.

Empirical Bayes procedure first estimates priors via $\hat{g}_l, \hat{g}_R = \arg\max_g \log p(Z|g_l, g_R)$, then obtains posterior of l, R. The marginal likelihood is

$$\begin{aligned}
\log p(Z|g_l, g_R) &= \log \int p(Z|l, R) g_l g_R dl dR \\
&= \log \int q(l, R) \frac{p(Z|l, R) g_l g_R}{q(l, R)} dl dR \\
&\geq \int q(l, R) \log \frac{p(Z|l, R) g_l g_R}{q(l, R)} dl dR \\
&= E_{q(l,R)} \log \frac{p(Z|l, R) g_l g_R}{q(l, R)}
\end{aligned} \tag{5}$$

Using mean field variational method, $q(l, R) = q_l q_R$, the evidence lower bound(ELBO) is

$$ELBO = \mathbb{E}_q \log p(Z|l, R) + \mathbb{E}_q \log \frac{g_l}{q_l} + \mathbb{E}_q \log \frac{g_R}{q_R}. \tag{6}$$

1. Update q_R, g_R, given q_l, g_l.

The evidence lower bound is

$$ELBO(q_R, g_R) = \mathbb{E}_q(\sum_{i=1}^{N} \sum_{s=0}^{S-1} \sum_{l=1}^{2^s-1} \log p(Z_{i,s+1,2l}|Z_{i,s,l}, R_{s,l})) + \mathbb{E}_q \log \frac{g_R}{q_R} + Constant \tag{7}$$

The expected sum of log likelihood in (7) is

$$
\begin{aligned}
&\sum_s \sum_l \mathbb{E}_{q_R}((\Sigma_i Z_i)_{s+1,2l} \log R_{s,l} + (\Sigma_i Z_i)_{s+1,2l+1} \log(1-R_{s,l}) + Constant \\
&= \mathbb{E}_{q_R}(\sum_{s=0}^{S-1} \sum_{l=0}^{2^s-1} \log p((\Sigma_i Z_i)_{s+1,2l} | (\Sigma_i Z_i)_{s,l}, R_{s,l})) + Constant
\end{aligned} \tag{8}
$$

Combine (7) and (8),

$$
\begin{aligned}
\hat{q}_R, \hat{g}_R &= \arg\max_{q_R, g_R} ELBO(q_R, g_R) \\
&= \arg\max_{q_R, g_R} \{\mathbb{E}_{q_R}(\sum_{s=0}^{S-1} \sum_{l=0}^{2^s-1} \log p((\Sigma_i Z_i)_{s+1,2l} | (\Sigma_i Z_i)_{s,l}, R_{s,l})) + \mathbb{E}_{q_R} \log \frac{g_R}{q_R}\}
\end{aligned} \tag{9}
$$

Compare (9) with (4), we can conclude that updating q_R, g_R is equivalent to solve Empirical Bayes Multiscale Poisson Smoothing with sequence $\sum_i Z_i$. Notice that there is an additional parameter $\mu := f_{0,0}$ appears in $Z_{0,0} \sim Poisson(\sum_i l_i f_{0,0})$. As mentioned in section 3, we simply estimate μ as $\hat{\mu} = \frac{Z_{0,0}}{\mathbb{E}(\sum_i l_i)} = \frac{\sum_j z_{\cdot j}}{\mathbb{E}(\sum_i l_i)}$.

2. Update q_l, g_l, given $q_R, g_R, \hat{\mu}$.

We can write

$$
f_j = \mu \prod_{s=0}^{S-1} (R_{s,s(j)})^{\epsilon_j(s)} (1 - R_{s,s(j)})^{1-\epsilon_j(s)},
$$

where $\epsilon_j(s) = 1$ if the jth element of f goes to left children node at scale s, and $s(j)$ is the location of jth element of f at scale s.

Intuitively, updating q_l, g_l should be equivalent to solve Empirical Bayes Poisson mean with sequence $\sum_j Z_j$ and a scale factor $\hat{\mu} \sum_j \prod_s \mathbb{E}_{q_{R_s}}((R_{s,s(j)})^{\epsilon_j(s)}(1-R_{s,s(j)})^{1-\epsilon_j(s)})$. The derivations below show this equivalence.

The ELBO with respect to q_l, g_l is

$$
\begin{aligned}
ELBO(q_l, g_l) &= \mathbb{E}_{q_R q_l} \sum_i \sum_j (Z_{ij} \log(l_i f_j) - l_i f_j) + \mathbb{E}_q \log \frac{g_l}{q_l} + Constant \\
&= \mathbb{E}_{q_R q_l} \sum_i \sum_j (Z_{ij} \log(l_i \hat{\mu} \prod_s (R_{s,s(j)})^{\epsilon_j(s)} (1 - R_{s,s(j)})^{1-\epsilon_j(s)}) \\
&\quad - l_i \hat{\mu} \prod_s (R_{s,s(j)})^{\epsilon_j(s)} (1 - R_{s,s(j)})^{1-\epsilon_j(s)}) + \mathbb{E}_q \log \frac{g_l}{q_l} + Constant \\
&= \mathbb{E}_{q_l}(\sum_i (\sum_j Z_{ij} \log l_i - l_i \hat{\mu} \, \mathbb{E}_{q_R} \sum_j \prod_{s=0}^{S-1} (R_{s,s(j)})^{\epsilon_j(s)} (1 - R_{s,s(j)})^{1-\epsilon_j(s)})) \\
&\quad + \mathbb{E}_q \log \frac{g_l}{q_l} + Constant.
\end{aligned} \tag{10}
$$

The objective function for solving Empirical Bayes Poisson mean given sequence $\sum_j Z_j$ and a scale factor $f = \hat{\mu} \sum_j \prod_s \mathbb{E}_{q_{R_s}}((R_{s,s(j)})^{\epsilon_j(s)}(1 - R_{s,s(j)})^{1-\epsilon_j(s)})$ is

$$
\begin{aligned}
\mathbb{E}_{q_l} \log p(\sum_j Z_j | l, f) &= \mathbb{E}_{q_l} \sum_i (\sum_j Z_j \log l_i - l_i f) + Constant \\
&= \mathbb{E}_{q_l} \sum_i (\sum_j Z_j \log l_i - l_i \hat{\mu} \sum_j \prod_s \mathbb{E}_{q_{R_s}}((R_{s,s(j)})^{\epsilon_j(s)}(1 - R_{s,s(j)})^{1-\epsilon_j(s)}) \\
&\quad + Constant \\
\mathcal{F}(q_l, g_l) &= \mathbb{E}_{q_l} \log p(\sum_j Z_j | l, f) + \mathbb{E}_q \log \frac{g_l}{q_l}
\end{aligned}
\tag{11}
$$

The equivalence holds since $\arg\max_g ELBO(q_l, g_l)$ and $\arg\max_g \mathcal{F}(q_l, g_l)$ are equivalent.

6 Rank-K Empirical Bayes Multiscale Poisson Matrix Factorization

We extend the rank-1 Empirical Bayes Multiscale Poisson Matrix Factorization to rank-K case. The model is

$$
\begin{aligned}
X &= \sum_k Z_k, \\
Z_k &\sim Poisson(l_k f_k^T), l_k \in R^N, f_k \in R^p, \\
f_{kj} &= \mu_k \prod_{s=0}^{S-1} (R_{k,s,s(j)})^{\epsilon_j(s)}(1 - R_{k,s,s(j)})^{1-\epsilon_j(s)}, \\
l_k &\sim g_{l_k}(\cdot), \\
R_{k,s} &\sim g_{R_{k,s}}(\cdot),
\end{aligned}
\tag{12}
$$

Factorization of joint distribution is

$$
p(X, Z, L, R | g, \mu) = p(X|Z) p(Z|L, R, \mu) p(R|g_R) p(L|g_L). \tag{13}
$$

Again, we follow Empirical Bayes procedure and use variational inference. The Variational lower bound is

$$
\begin{aligned}
\mathcal{F}(q,g) &= \mathbb{E}_q \log p(X,Z,L,R|g,\mu) - \mathbb{E}_q \log q(Z,L,R) \\
&= \mathbb{E}_q \log p(X,Z,R,L|g_R,g_L,\mu) - \mathbb{E}_{q_L} \log q_L(L) - \mathbb{E}_{q_R} \log q_R(R) - \mathbb{E}_{q_Z} \log q_Z(Z) \\
&= \mathbb{E}_q \log p(Z|R,L,\mu) + \mathbb{E}_q \log \delta(X - \sum_k Z_k) + \mathbb{E}_q \log \frac{g_R}{q_R} + \mathbb{E}_q \log \frac{g_L}{q_L} - \mathbb{E}_q \log q_Z \\
&= \sum_k (\mathbb{E}_q \log p(Z_k|R_k,l_k,\mu_k) + \mathbb{E}_q \log \frac{g_{R_k}}{q_{R_k}} + \mathbb{E}_q \log \frac{g_{L_k}}{q_{L_k}}) \\
&+ \mathbb{E}_q \log \delta(X - \sum_k Z_k) - \mathbb{E}_q \log q_Z.
\end{aligned}
\tag{14}
$$

We update $q_Z, (q_l, g_l), (q_R, g_R)$ iteratively.

1. Update q_Z: Take functional derivatives of the lower bound with respect to q_Z, then

$$
\begin{aligned}
q_Z(Z) &\propto \exp(\int\int \log p(X,Z|R,L,\mu,g_R,g_L) q_R q_L dR dL) \\
&\propto \exp(\int\int \log p(Z|X,R,\mu,L,g_R,g_L) q_R q_L dR dL) \\
&\propto \exp(\mathbb{E}_{q_R,q_L} \log p(Z|X,R,L,\mu,g_R,g_L)).
\end{aligned}
\tag{15}
$$

This leads to

$$q_Z(Z_{1:K,i,j}) \sim Multinomial(X_{ij}, \pi_{1:K,i,j}),$$

where

$$\pi_{k,i,j} = \frac{\exp(\mathbb{E}\log l_{ik} + \mathbb{E}\log f_{kj})}{\sum_k \exp(\mathbb{E}\log l_{ik} + \mathbb{E}\log f_{kj})},$$

$$f_{kj} = \mu_k \prod_{s=0}^{S-1} R_{k,s,s(j)}^{\epsilon_j(s)} (1 - R_{k,s,s(j)})^{1-\epsilon_j(s)}.$$

The expectation $\mathbb{E}\log f_{kj}$ is

$$\mathbb{E}\log f_{kj} = \log \mu_k + \sum_{s=0}^{S-1} \mathbb{E}_{R_{k,s}} (\epsilon_j(s) \log R_{k,s,s(j)} + (1-\epsilon_j(s)) \log(1 - R_{k,s,s(j)})). \tag{16}$$

From (14), updating distributions related to R, L in rank-K problem reduces to solving K rank-1 problems.

2. Update $q_{l_k}(\cdot)$ and $g_{l_k}(\cdot)$:

$$
\begin{aligned}
q_L &\propto \exp(\mathbb{E}_{q_R,q_Z} \log p(Z|L,R,\mu) p(L|g_L)) \\
&\propto \exp(\sum_k (\mathbb{E}_q \log p(Z_k|L_k,R_k,\mu_k) p(L_k|g_{L_k}))).
\end{aligned}
\tag{17}
$$

We update q_{l_k} for each topic k. The q_{l_k} and g_{l_k} can be updated by solving the following Empirical Bayes Poisson mean problem:

$$\begin{aligned}\sum_j \mathbb{E}_{q_Z}(Z_{k,i,j}) \sim Poisson(l_{ik}\sum_j \mathbb{E}_{q_{R_k}}(f_{kj})), i = 1, 2, ..., N,\\ \mathbb{E}_{q_{R_k}}(f_{kj}) = \mu_k \times \prod_{s=0}^{S-1} \mathbb{E}_{R_{k,s}}(R_{k,s,s(j)}^{\epsilon_j(s)}(1 - R_{k,s,s(j)})^{1-\epsilon_j(s)}),\\ l_k \sim g_{l_k}(\cdot).\end{aligned} \tag{18}$$

3. Update q_{R_k}, g_{R_k}:

Solve Empirical Bayes Multiscale Poisson Smoothing problem for each topic k with the Poisson sequence $\sum_i \mathbb{E}_{q_Z}(Z_{k,i,j}), j = 1, 2, ..., p$ and an extra scale factor $c = \sum_i E_{q_{l_k}} l_{ik}$.

7 Translation Invariant Wavelet Transform

We have shown that updating q_R, g_R in rank-1 multiscale Poisson matrix factorization is equivalent to solving Poisson smoothing problem with observed sequence $z := \{\sum_i Z_{ij}, j = 1, 2, ..., p\}$.

In order to utilize full location-scale information and make the wavelet transformation translation-invariant, cycle-spinning algorithm is widely used, which averages results over all n possible rotations of the data (effectively treating the observations as coming from a circle, rather than a line). The full Bayesian multiscale likelihood structure results from using model mixing over the set of n rotations.

Following Johnstone and Silverman (2005b), empirical Bayes shrinkage is applied to each scale of translation-variant wavelet transformation by treating the coefficients at this scale "as if independent".

Let t denote each cycle or shift, for $t = 0, 1, ..., n-1$. To estimate the prior at scale s, we find g_{R_s} to maximize $\log p(\{X_{s,:}^{(t)}, t = 0, 1, ..., n-1\}|g_{R_s})$, where $\{X_{s,:}^{(t)}, t = 0, 1, ..., n-1\}$ is a collection of subsequences results from translation-invariant transformation. There are 2^{S-s} such subsequences at scale s and the collection of such subsequences at each scale has length 2^S. In summary,

$$\hat{g}_{R_s} = \arg\max_g \log p(\{X_{s,:}^{(t)}\}|g_{R_s}). \tag{19}$$

Following the derivations in (3), we have $q_{R_s}^{(t)} = p(R_s^{(t)}|X, g_{R_s})$ for $t = 0, 1, ..., n-1$ at each scale s. The log marginal likelihood is now

$$\log \prod_{t=0}^{n-1} p(X^{(t)}|g_R) = \sum_{t=0}^{n-1} \mathcal{F}(q_{R^{(t)}}, g_R; X) + \sum_{t=0}^{n-1} KL(q_{R^{(t)}}||p(R^{(t)}|X^{(t)}, g_R)). \tag{20}$$

In Poisson smoothing problem, to get the point estimate of λ, we average all n estimates of λ over n shifts then use the posterior mean as the estimator. The posterior distribution of

λ is

$$
\begin{aligned}
p(\lambda|X,\mu) &= \frac{1}{n}\sum_{t=0}^{n-1} p(\lambda^{(t)}|X^{(t)},\mu) \\
&= \frac{1}{n}\sum_{t=0}^{n-1}\prod_{j=1}^{p} p(\lambda_j^{(t)}|X^{(t)},\mu) \\
&= \frac{1}{n}\sum_{t=0}^{n-1}\prod_{j=1}^{p} p(\prod_{s=0}^{S-1}(R_{s,s(j)}^{(t)})^{\epsilon_j^{(t)}(s)}(1-R_{s,s(j)}^{(t)})^{1-\epsilon_j^{(t)}(s)}|X^{(t)},\mu) \\
&= \frac{1}{n}\sum_{t=0}^{n-1}\prod_{j=1}^{p}\prod_{s=0}^{S-1} p((R_{s,s(j)}^{(t)})^{\epsilon_j^{(t)}(s)}(1-R_{s,s(j)}^{(t)})^{1-\epsilon_j^{(t)}(s)}|X^{(t)},\mu).
\end{aligned}
\tag{21}
$$

So the posterior mean of λ_j is

$$
\begin{aligned}
\mathbb{E}(\lambda_j|X,\mu) &= \frac{1}{n}\sum_{t=0}^{n-1}\mathbb{E}(\lambda_j^{(s)}|X^{(s)},\mu) \\
&= \frac{1}{n}\sum_{t=0}^{p-1}\mu\prod_{s=0}^{S-1}\mathbb{E}(R_{s,s(j)}^{(t)})^{\epsilon_j^{(t)}(s)}(1-R_{s,s(j)}^{(t)})^{1-\epsilon_j^{(t)}(s)}|X_{s+1,2s(j)}^{(t)},X_{s+1,2s(j)+1}^{(t)}).
\end{aligned}
\tag{22}
$$

Back to the rank-1 matrix factorization, we also apply translation-invariant wavelet transformation when updating g_R, q_R. Now equation (9) becomes

$$
\begin{aligned}
\hat{q}_R,\hat{g}_R &= \arg\max_{q_R,g_R} ELBO(q_R,g_R) \\
&= \arg\max_{q_R,g_R}\{\sum_{t=0}^{p-1}[\mathbb{E}_{q_{R^{(t)}}}(\sum_{s=0}^{S-1}\sum_{l=0}^{2^s-1}\log p((\Sigma_i Z_i)_{s+1,2l}^{(t)}|(\Sigma_i Z_i)_{s,l}^{(t)},R_{s,l}^{(t)}))+\mathbb{E}_{q_{R^{(t)}}}\log\frac{g_R}{q_{R^{(t)}}}]\}
\end{aligned}
\tag{23}
$$

In summary, both (20) and (23) are just sum of shifts of the formulas without translation-invariant transformation so we still have the equivalence of updating g_R, q_R in rank-1 multiscale Poisson matrix factorization and multiscale Poisson smoothing.

When updating g_l, q_l, the scale factor becomes

$$
\hat{\mu}\frac{1}{p}\sum_{t=0}^{p-1}\sum_{j}\prod_{s}\mathbb{E}_{q_{R_s^{(t)}}}((R_{s,s(j)}^{(t)})^{\epsilon_j^{(t)}(s)}(1-R_{s,s(j)}^{(t)})^{1-\epsilon_j^{(t)}(s)}). \tag{24}
$$

8 Algorithms

Algorithm 1: EBMPMF Rank-1 Algorithm

Input: matrix $Z \in R^N$
Initialize: $q_l^{(0)}, q_R^{(0)}, g_l^{(0)}, g_R^{(0)}$
Output: g_l, g_R, q_l, q_R
1. $q_l, g_l = \text{EBPM}(\sum_j Z_{ij}, \sum_j \mathbb{E}_q f_j)$
2. $q_R, g_R = \text{EBPS}(\sum_i Z_{ij}, \sum_i \mathbb{E}_q l_i)$

Algorithm 2: EBMPMF Rank-K Algorithm

Input: matrix $X \in R^N$, rank K
Initialize: $q_{L,1:K}^{(0)}, q_{R,1:K}^{(0)}, g_{L,1:K}^{(0)}, g_{R,1:K}^{(0)}$
Output: g_L, g_R, q_L, q_R
while *not converged* **do**
  $i = i + 1$;
  for $k = 1, 2, ..., K$ **do**
    1. Update $q_Z^{(i)}$.
    2. $q_{L,k}^{(i)}, q_{R,k}^{(i)}, g_{L,k}^{(i)}, g_{R,k}^{(i)} = \text{EBMPMF-Rank1}(\mathbb{E}(Z_{k,i,j}),\ q_{L,k}^{(i-1)}, q_{R,k}^{(i-1)}, g_{L,k}^{(i-1)}, g_{R,k}^{(i-1)})$.
  end
end

Some implementation details:

1. Empirical Bayes Poisson Smoothing.

We have at least two methods, all of which depend on the multiscale decomposition of Poisson sequence. The first one is the standard BMSM models in Kolaczyk(1999). Another one is in Xing et al(2019).

For $R_{k,s}$, one way is to put a mixture of Beta distribution prior separately for each scale

$$R_{k,s,l} \sim \sum_h p_{k,s,h} Beta(\alpha_{k,s,h}, \alpha_{k,s,h}).$$

To force smoothness in each f, we can also put a point mass on $\frac{1}{2}$. Another way is reparameterize $R_{k,s,l}$ to its logit form $\log \frac{R_{k,s,l}}{1-R_{k,s,l}}$, then use normal approximation hence standard ash prior to achieve shrinkage estimate.

To use the mixture of beta distribution, or more generally shrinking the probability directly, we need to develop Empirical Bayes Binomial probability .

2. Empirical Bayes Poisson Mean.

We can put a mixture of Gamma or exponential distributions on l.

3. Initialization of algorithms.

4. Convergence criteria.