

Empirical Bayes Multiscale Poisson Matrix Factorization

Dongyue Xie

Oct 2019

1 Empirical Bayes Multiscale Poisson Smoothing

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

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

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

Similarly, RDP also holds for λ .

Define

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

then

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

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

Following this decomposition, the likelihood of X is

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

Further we put priors on $\lambda_{0,0}$ and $R_{s,l}$: $\mu \sim g_\mu(\cdot)$ and $R_{s,l} \sim g_{R_s}(\cdot)$.

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

$$\begin{aligned}
\log p(X|g_\mu, g_R) &= \mathbb{E}_{q_R q_\mu} \log p(X|g_\mu, g_R) \\
&= \mathbb{E}_{q_R q_\mu} \log \frac{p(X, \mu, R|g_\mu, g_R)}{p(R, \mu|X, g_\mu, g_R)} \\
&= \mathbb{E}_{q_R q_\mu} \left(\log \frac{p(X, \mu, R|g_\mu, g_R)}{q_R q_\mu} + \log \frac{q_R q_\mu}{p(R, \mu|X, g_\mu, g_R)} \right) \\
&= \mathbb{E}_{q_R q_\mu} \log p(X|R, \mu) + \mathbb{E}_{q_R q_\mu} \log \frac{g_\mu g_R}{q_R q_\mu} + KL(q_R q_\mu || p(R, \mu|g_\mu, g_R)) \\
&= \mathcal{F}(q_R, q_\mu, g_\mu, g_R; X) + KL(q_R q_\mu || p(R, \mu|g_\mu, g_R)) \\
&\geq \mathcal{F}(q_R, q_\mu, g_\mu, g_R; X).
\end{aligned} \tag{2}$$

$$\begin{aligned}
\mathcal{F}(q_R, q_\mu, g_\mu, g_R; X) &= \mathbb{E}_{q_R q_\mu} \left(\sum_{s=0}^{S-1} \sum_{l=0}^{2^s-1} \log p(X_{s+1,2l}|X_{s,l}, R_{s,l}) + \log p(X_{0,0}|\mu, c) \right) \\
&\quad + \mathbb{E}_{q_R} \log \frac{g_R}{q_R} + \mathbb{E}_{q_\mu} \log \frac{g_\mu}{q_\mu}.
\end{aligned} \tag{3}$$

2 Rank-1 EB Multiscale Poisson Matrix Factorization

Let $Z_{ij} \sim \text{Poisson}(l_i f_j)$, $i = 1, 2, \dots, N$, $j = 1, 2, \dots, p$. Priors: $l \sim g_l(\cdot)$. Assume $f \in R^p$ has RDP defined in section 1 and $R_{s,l} \sim g_{R_s}(\cdot)$, $\mu \sim g_\mu(\cdot)$.

Using variational method, the evidence lower bound(ELBO) is

$$\begin{aligned}
ELBO &= \mathbb{E}_q \log p(Z, l, R, \mu|g_l, g_\mu, g_R) - \mathbb{E}_q \log q(l, R, \mu) \\
&= \mathbb{E}_q \log p(Z|l, R, \mu) + \mathbb{E}_q \log \frac{g_l}{q_l} + \mathbb{E}_q \log \frac{g_R}{q_R} + \mathbb{E}_q \log \frac{g_\mu}{q_\mu}
\end{aligned} \tag{4}$$

1. Update q_R, g_R, q_μ, g_μ : given q_l, g_l ,

$$\begin{aligned}
ELBO(q_R, g_R) &= \mathbb{E}_q \left(\sum_{i=1}^N \sum_{s=0}^{S-1} \sum_{l=1}^{2^s-1} \log p(Z_{i,s+1,2l}|Z_{i,s,l}, R_{s,l}) + \sum_{i=1}^N \log p(Z_{i,0,0}|\mu, l_i) \right) \\
&\quad + \mathbb{E}_q \log \frac{g_R}{q_R} + \mathbb{E}_q \log \frac{g_\mu}{q_\mu} + \text{Constant}
\end{aligned} \tag{5}$$

This first line of ELBO in (5) is

$$\begin{aligned}
&\sum_s \sum_l \mathbb{E}_{q_R q_\mu, g_R, g_\mu} ((\sum_i Z_i)_{s+1,2l} \log R_{s,l} + (\sum_i Z_i)_{s+1,2l+1} \log(1 - R_{s,l}) \\
&\quad + \sum_i Z_{i,0,0} \log \mu - \mu \sum_i \mathbb{E}_{q_l} l_i) + \text{Constant}
\end{aligned} \tag{6}$$

Combine (4), (5) and (6), and compare with (3), we can conclude that updating q_R, g_R, q_μ, g_μ in rank-1 case is equivalent to solve Empirical Bayes Multiscale Poisson Smoothing with sequence $\sum_i Z_i$ and an extra scale factor $c = \sum_i \mathbb{E}_{q_l} l_i$.

2. Update q_l, g_l : given q_R, g_R, q_μ, g_μ ,

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 j th element of f goes to left children node at scale s , and $s(j)$ is the location of j th element of f at scale s .

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

ELBO for rank-1 case:

$$\begin{aligned}
ELBO(q_l, g_l) &= \mathbb{E}_{q_R q_\mu q_l} \sum_i \sum_j (Z_{ij} \log(l_i \mu \prod_s (R_{s,s(j)})^{\epsilon_j(s)} (1 - R_{s,s(j)})^{1-\epsilon_j(s)})) \\
&\quad - l_i \mu \prod_s (R_{s,s(j)})^{\epsilon_j(s)} (1 - R_{s,s(j)})^{1-\epsilon_j(s)} + \mathbb{E}_q \log \frac{g_l}{q_l} + C_1 \\
&= \mathbb{E}_{q_l} (\sum_i (\sum_j Z_{ij} \log l_i - l_i \mathbb{E}_{q_R q_\mu} \sum_j (\mu \prod_{s=0}^{S-1} (R_{s,s(j)})^{\epsilon_j(s)} (1 - R_{s,s(j)})^{1-\epsilon_j(s)})) + \mathbb{E}_q \log \frac{g_l}{q_l} + C_2.
\end{aligned} \tag{7}$$

Solve Empirical Bayes Poisson mean given sequence $\sum_j Z_j$ and a scale factor $f_j = \mathbb{E}_{q_\mu}(\mu) \sum_j \prod_s \mathbb{E}_{q_{R_s}}((R_{s,s(j)})^{\epsilon_j(s)} (1 - R_{s,s(j)})^{1-\epsilon_j(s)})$:

$$\begin{aligned}
\mathbb{E}_{q_l} \log p(\sum_j Z_j | l, f_j) &= \mathbb{E}_{q_l} \sum_i (\sum_j Z_j \log l_i - l_i f_j) + C_1 \\
&= \mathbb{E}_{q_l} \sum_i (\sum_j Z_j \log l_i - l_i \times \mathbb{E}_{q_\mu}(\mu) \sum_j \prod_s \mathbb{E}_{q_{R_s}}((R_{s,s(j)})^{\epsilon_j(s)} (1 - R_{s,s(j)})^{1-\epsilon_j(s)})) + C_1 \\
\mathcal{F}(q_l, g_l) &= \mathbb{E}_{q_l} \log p(\sum_j Z_j | l, f_j) + \mathbb{E}_q \log \frac{g_l}{q_l} + C_1
\end{aligned} \tag{8}$$

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

3 Rank-K EBMPMF

$$\begin{aligned}
X &= \sum_k Z_k, \\
Z_k &\sim \text{Poisson}(l_k f_k^T), l_k \in R^N, f_k \in R^p, \\
f_{kj} &= \mu_k \prod_{s=0}^{S-1} (R_{k,s,s(j)})^{\epsilon_j(s)} (1 - R_{k,s,s(j)})^{1-\epsilon_j(s)}, \\
l_k &\sim g_{l_k}(\cdot), \\
R_{k,s} &\sim g_{R_{k,s}}(\cdot), \\
\mu_k &\sim g_{\mu_k}(\cdot),
\end{aligned} \tag{9}$$

Factorization of joint distribution is

$$p(X, Z, L, R, \mu|g) = p(X|Z)p(Z|L, R, \mu)p(R|g_R)p(\mu|g_\mu)p(L|g_L). \tag{10}$$

Goal is to estimate g and obtain posterior $p(L, R, \mu|X, \hat{g})$. Two ways to derive variational lower bound: Jensen's Inequality and KL divergence.

Variational lower bound:

$$\begin{aligned}
\mathcal{F}(q, g) &= \mathbb{E}_q \log p(X, Z, L, R, \mu|g) - \mathbb{E}_q \log q(Z, L, R, \mu) \\
&= \mathbb{E}_q \log p(X, Z, R, L, \mu|g_R, g_\mu, g_L) - \mathbb{E}_{q_L} \log q_L(L) - \mathbb{E}_{q_R} \log q_R(R) - \mathbb{E}_{q_Z} \log q_Z(Z) - \mathbb{E}_q \log q_\mu \\
&= \mathbb{E}_q \log p(Z|R, L, \mu) + \mathbb{E}_q \log \delta(X - \sum_k Z_k) + \mathbb{E}_q \log \frac{g_R}{q_R} + \mathbb{E}_q \log \frac{g_L}{q_L} + \mathbb{E}_q \log \frac{g_\mu}{q_\mu} - \mathbb{E}_q \log q_Z \\
&= \sum_k (\mathbb{E}_q \log p(Z_k|R_k, l_k, \mu_k) + \mathbb{E}_q \log \frac{g_{R_k}}{q_{R_k}} + \mathbb{E}_q \log \frac{g_{L_k}}{q_{L_k}} + \mathbb{E}_q \log \frac{g_{\mu_k}}{q_{\mu_k}}) + \mathbb{E}_q \log \delta(X - \sum_k Z_k) - \mathbb{E}_q \log q_Z.
\end{aligned} \tag{11}$$

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

$$\begin{aligned}
q_Z(Z) &\propto \exp\left(\int \int \log p(X, Z|R, L, \mu, g_R, g_L, g_\mu) q_R q_L q_\mu dR dL d\mu\right) \\
&\propto \exp\left(\int \int \log p(Z|X, R, \mu, L, g_R, g_L, g_\mu) q_R q_L q_\mu dR dL d\mu\right) \\
&\propto \exp(\mathbb{E}_{q_R, q_L, q_\mu} \log p(Z|X, R, L, \mu, g_R, g_L, g_\mu)).
\end{aligned} \tag{12}$$

This leads to

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

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 then

$$\mathbb{E} \log f_{kj} = \mathbb{E}_{q_{\mu_k}} \log \mu_k + \sum_{s=0}^{S-1} \mathbb{E}_{R_{k,s}} (\epsilon_j(s) \log R_{k,s,s(j)} + (1 - \epsilon_j(s)) \log(1 - R_{k,s,s(j)})). \quad (13)$$

From the ELBO formula (11), updating distributions related to R, μ, L in rank-K problem reduces to solving K rank-1 problems.

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

$$\begin{aligned} q_L &\propto \exp(\mathbb{E}_{q_R, q_{\mu}, q_Z} \log p(Z|L, R, \mu) p(L|g_L)) \\ &\propto \exp(\sum_k (\mathbb{E}_q \log p(Z_k|L_k, R_k, \mu_k) p(L_k|g_{L_k}))). \end{aligned} \quad (14)$$

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:

$$\begin{aligned} \sum_j \mathbb{E}_{q_Z}(Z_{k,i,j}) &\sim \text{Poisson}(l_{ik} \sum_j \mathbb{E}_{q_{R_k}, q_{\mu_k}}(f_{kj})), i = 1, 2, \dots, N, \\ \mathbb{E}_{q_{R_k}, q_{\mu_k}}(f_{kj}) &= \mathbb{E}_{q_{\mu_k}} \mu_k \times \prod_{s=0}^{S-1} \mathbb{E}_{R_{k,s}} (R_{k,s,s(j)}^{\epsilon_j(s)} (1 - R_{k,s,s(j)})^{1-\epsilon_j(s)}), \\ l_k &\sim g_{l_k}(\cdot). \end{aligned} \quad (15)$$

3. Update $q_{R_k}, q_{\mu_k}, g_{R_k}, g_{\mu_k}$:

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

4 Practical Concerns

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 μ_k , the simplest way is to set $\hat{\mu}_k = \sum_i \sum_j \mathbb{E}_{q_Z}(Z_{k,i,j}) / \sum_i (l_{ik})$. This retains the property of preservation of energy and results in an estimate relatively unaffected by perturbations in the data at their original scale.

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

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

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.

A common practice in wavelet shrinkage estimation is to use 'cycle spinning' or Non-decimated WT to get translation-invariant property.

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

5 To do

Build code based on EBPMF implementations.

How to deal with many 0's? Is zero-inflated model possible?

Run some simulation examples.

Potential applications.

Other modifications to the model