EMPIRICAL BAYES POISSON MEAN (EBPM)

ZIHAO WANG

1. Overview. Here we want to solve the Empirical Bayes Poisson Mean (EBPM) problem, a problem analogous to the Empirical Bayes Normal Mean problem. We derive the algorithms for solving EBPM problem with two prior families: mixture of gamma, and spike-and-slab prior. The algorithms are implemented in the R package: ebpm

2. Model. 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 $\hat{g}, p$ where

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

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

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

Next, I will explore different types of prior family $\mathcal{G}$. The naming for the algorithms would be "EBPM-{prior name}".

3. mixture of gamma. The prior is of the form:

$$
\begin{aligned}
g(\lambda) &= \sum_k \pi_k Gamma(\lambda; a_k, b_k) \\
&= \sum_k \pi_k \frac{b_k^{a_k}}{\Gamma(a_k)} \lambda^{a_k - 1} e^{-b_k \lambda}
\end{aligned}
$$

where a_k, b_k are known (in a grid) and mixture weights, $\boldsymbol{\pi}$, are to be estimated. ($\sum_k \pi_k = 1, \pi_k \geq 0$).

3.1. Useful lemmas.

LEMMA 3.1. *If $x|\lambda \sim Pois(\lambda)$, and $\lambda \sim Gamma(a, b)$, then $x \sim NB(.; size = a, prob = \frac{b}{1+b})$, where NB is parameterized the same way as R function "rnbinom".*

Proof.

$$
\begin{aligned}
p(x) &= \int p(x|\lambda) p(\lambda) d\lambda \\
&= \int \frac{e^{-\lambda} \lambda^x}{\Gamma(x+1)} \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} d\lambda \\
&= \frac{b^a}{\Gamma(x+1)\Gamma(a)} \int e^{-(b+1)\lambda} \lambda^{x+a-1} d\lambda \\
&= \frac{\Gamma(x+a)}{\Gamma(x+1)\Gamma(a)} \frac{b^a}{(b+1)^{x+a}} = NB(x; a, \frac{b}{b+1})
\end{aligned}
$$

∎

Lemma 3.2. *If* $\lambda \sim Gamma(a, b)$*, then* $s\lambda \sim Gamma(a, b/s)$

Lemma 3.3. *If* $x|\lambda \sim Pois(s\lambda), \lambda \sim Gamma(a, b)$*, then* $x \sim NB(.; r = a, p = \frac{b}{b+s})$

Lemma 3.4. *If* $x|\lambda \sim Pois(\lambda), \lambda \sim Gamma(a, b)$*, then* $\lambda|x \sim Gamma(a+x, b+1)$

Lemma 3.5. *If* $x|\lambda \sim Pois(s\lambda), \lambda \sim Gamma(a, b)$*, then*

$$p(x, \lambda) = p(x|\lambda)p(\lambda) = NB(x; a, \frac{b}{b+s})Gamma(\lambda; a+x, b+s)$$

3.2. MLE.

$$\ell(\boldsymbol{\pi}) = \sum_i log\ p(x_i|\boldsymbol{\pi}) = \sum_i log\ \sum_k p(z_i = k|\boldsymbol{\pi})p(x_i|z_i = k) = \sum_i log\ \sum_k \pi_k p(x_i|z_i = k)$$

where $z_i = k$ indicates $\lambda_i \sim Gamma(a_k, b_k)$.
Now let's look at $p(x_i|z_i = k)$. Since $x_i|(z_i = k) \stackrel{d}{=} x_i|\lambda \sim Pois(s_i\lambda)$ with $\lambda \sim Gamma(a_k, b_k)$. By Lemma 3.3, we have $x_i \sim NB(r = a_k, p = \frac{b_k}{s_i+b_k})$. Therefore, we have

$$\ell(\boldsymbol{\pi}) = \sum_i log\ \sum_k \pi_k L_{ik}$$

where

$$L_{ik} = NB(x_i; r = a_k, p = \frac{b_k}{s_i + b_k})$$

This problem can be solved efficiently by algorithms like mixsqp.
Note that we can set $a_k = 1$ for all k, which guarantees the prior mode to be 0. In this case, the prior is a mixture of exponential, and $L_{ik} = \frac{s_i^{x_i} b_k}{(b_k+s_i)^{x_i+1}}$

3.3. Posterior computation. By lemma 3.5, we get:

$$\begin{aligned} p(\lambda|x_i, \boldsymbol{\pi}) &\propto p(x_i|\lambda)g(\lambda; \boldsymbol{\pi}) \\ &\propto \sum_k \pi_k NB(x_i, a_k, \frac{b_k}{b_k + s_i})Gamma(\lambda; a_k + x_i, b_k + s_i) \\ &\propto \sum_k \pi_k L_{ik} Gamma(\lambda; a_k + x_i, b_k + s_i) \end{aligned}$$

Thus we have

$$p(\lambda|x_i, \hat{\boldsymbol{\pi}}) = \sum_k \tilde{\Pi}_{ik} Gamma(\lambda; a_k + x_i, b_k + s_i)$$

where $\tilde{\Pi}_{ik} \propto L_{ik}\hat{\pi}_k$, $\sum_k \tilde{\Pi}_{ik} = 1$.
Posterior mean: $E(\lambda) = \sum_k \tilde{\Pi}_{ik} \frac{x_i+a_k}{s_i+b_k}$.
Posterior log mean: $E(log\ \lambda) = \sum_k \tilde{\Pi}_{ik}(\psi(a_k + x_i) - log(b_k + s_i))$.

4. Spike-and-slab. The prior family is $\mathcal{G} = \{\pi_0\delta(.) + (1 - \pi_0)Gamma(a, b) : \pi \in [0, 1], a, b > 0\}$.

4.1. MLE.

$$
\begin{aligned}
p(x) &= \pi_0 p(x|\lambda = 0) + (1-\pi_0)p(x|\lambda \neq 0) \\
&= \pi_0 1_{\{x=0\}} + (1-\pi_0)NB(x; a, \frac{b}{b+s}) \\
&= \pi_0 c(a,b) + d(a,b)
\end{aligned}
$$

where

$$
\begin{aligned}
d(a,b) &= NB(x; a, \frac{b}{b+s}) \\
c(a,b) &= 1_{\{x=0\}} - d(a,b)
\end{aligned}
$$

Then

$$
l(\pi_0, a, b) = \sum_i log\{c_i(a,b)\pi_0 + d_i(a,b)\}
$$

This can be solved by softwares like nlm in R.

4.2. Posterior computation.

$$
\begin{aligned}
p(\lambda|x, \pi_0, a, b) &\propto p(x|\lambda)g(\lambda, \pi_0, a, b) \\
&= \pi_0.\delta(\lambda)p(x|\lambda) + (1-\pi_0)Gamma(\lambda; a, b)p(x|\lambda) \\
&= \pi_0 1_{x=0}\delta(\lambda) + (1-\pi_0)NB(x; a, \frac{b}{b+s})Gamma(\lambda; a+x, b+s)
\end{aligned}
$$

Thus

$$
p(\lambda|x, \pi_0, a, b) = \hat{\pi}_0\delta(\lambda) + (1-\hat{\pi}_0)Gamma(\lambda; a+x, b+s)]
$$

where

$$
\hat{\pi}_0 = \frac{\pi_0 1_{x=0}}{\pi_0 1_{x=0} + (1-\pi_0)NB(x; a, \frac{b}{b+s})}
$$

Therefore, posterior mean is $(1-\hat{\pi}_0)\frac{a+x}{b+s}$. The posterior log mean $E(log(\lambda))$ is $-\infty$ if $x = 0$, and $\psi(a+x) - log(b+s)$ if $x \neq 0$.