EMPIRICAL BAYES POISSON MEAN (EBPM)

ZIHAO WANG

1. Overview. Here we want to solve the Empirical Bayes Poisson Means (EBPM) problem, a problem analogous to the Empirical Bayes Normal Means problem. We consider 3 prior families: mixture of gamma prior, spike-and-slab (point gamma) prior, and beta-gamma prior.

For mixture of gamma prior, spike-and-slab (point gamma) prior, we find the optimal prior through maximizing marginal log-likelihood, then give the posterior.

For beta-gamma, the marginal log-likelihood is not closed form. So we use Mean-Field Variational Inference and get prior and approximate posterior by maximizing Evidence Lower BOund (ELBO).

The algorithms are implemented in the R package: ebpm.

2. EBPM 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

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

Use EBPM to denote the mapping:

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

3. Useful Lemmas. Since the prior families considered use gamma as a basic component, I list some of the useful lemmas regarding Gamma-Poisson mixture.

Lemma 3.1. *If $\lambda \sim Gamma(a, b)$, then $s\lambda \sim Gamma(a, b/s)$*

Lemma 3.2. *If $x|\lambda \sim Pois(\lambda)$, and $\lambda \sim Gamma(a, b)$, then $x \sim NB(.; size = a, prob = \frac{b}{1+b})$.*

Lemma 3.3. *If $x|\lambda \sim Pois(\lambda), \lambda \sim Gamma(a, b)$, then $\lambda|x \sim Gamma(a + x, b + 1)$.*

4. 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. ($\sum_k \pi_k = 1, \pi_k \geq 0$).

4.1. 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 3.1 and 3.2, 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 is convex, and can be solved efficiently by algorithms like mixsqp.

4.2. 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: $E(log\ \lambda) = \sum_k \tilde{\Pi}_{ik}(\psi(a_k + x_i) - log(b_k + s_i))$.

5. Spike-and-slab (Point Gamma). The prior family is point gamma: $\mathcal{G} = \{\pi_0\delta_0(.) + (1 - \pi_0)Gamma(.; a, b) : \pi \in [0, 1], a, b > 0\}$.

5.1. 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.

5.2. 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 $(1-\hat{\pi}_0)(\psi(a+x) - log(b+s))$ if $x \neq 0$.

6. Beta-Gamma. In spike-and-slab (point-gamma) prior, optimal π_0 is 0, if all x are nonzero. Also, the point mass at 0 doesn't have an effect when $x_i \neq 0$, as likelihood is 0 for $\lambda_i = 0$. To avoid the two issues, we consider a generalization for point gamma: beta-gamma. For simplicity we only consider the case for $s_i = 1, i = 1...n$, but it is easily generalized to different choices of s_i, using lemma 3.1.

6.1. Model.

$$\begin{aligned} x_i &\sim Pois(\lambda_i) \\ \lambda_i &= p_i v_i \\ v_i &\sim Gamma(\alpha, \beta) \\ p_i &\sim Beta(a, b) \end{aligned}$$

The marginal loglikelihood is not closed form, so we consider variational inference below.

6.2. Variational Inference. Introducing latent variable z_i, we reparameterize the model as below:

$$\begin{aligned} x_i &\sim Bin(z_i, p_i) \\ z_i|v_i &\sim Pois(v_i) \\ v_i &\sim Gamma(\alpha, \beta) \\ p_i &\sim Beta(a, b) \end{aligned}$$

Then we can write out log likelihood of all variables:

$$
\begin{aligned}
log\ p(x_i, z_i, p_i, v_i) &= log(p(x_i|z_i, p_i)p(z_i|v_i)p(v_i)p(p_i)) \\
&= log(z_i!) - log((z_i - x_i)!) - log(x_i!) + (x_i + a - 1)log(p_i) + (z_i + b - x_i - 1)log(1 - p_i) \\
&+ (\alpha + z_i - 1)logv_i - (\beta + 1)v_i \\
&+ \alpha log\beta - log(\Gamma(\alpha)) - logB(a, b)
\end{aligned}
$$

Using Mean-Field Variational Inference, we have the following update rule (as well the necessary condition for optimizer of ELBO) (h_j is the j-th latent variable):

$$q^*(h_j) \propto E_{-j} log\ p(x, h)$$

Apply this we can get the Coordinate Ascent for each $q(h_j)$:

$$
\begin{aligned}
q^*(p_i) &= Beta(.; a + x_i, b + \langle z_i \rangle_q - x_i) \\
q^*(v_i) &= Gamma(\alpha + \langle z_i \rangle_q, \beta + 1) \\
q^*(z_i - x_i) &= Poisson(exp(\langle log(1 - p_i) \rangle_q + \langle log\ v_i \rangle_q))
\end{aligned}
$$

I will use $\tilde{a}_i, \tilde{b}_i, \tilde{\alpha}_i, \tilde{\beta}_i, \mu_i$ to denote the variational parameter for $a, b, \alpha, \beta, \lambda$. (Note μ_i is for the shifted Poisson).

Now we have the proper parameterization for q, we can write out ELBO in closed form:

6.3. ELBO in closed form.

$$
\begin{aligned}
ELBO &= \sum_i E_q[log\ p(x_i, z_i, p_i, v_i) - log\ q(z_i, p_i, v_i)] \\
&= \sum_i E_q[log\ p(x_i|z_i, p_i) + log\ \frac{p(z_i|v_i)}{q(z_i)} + log\ \frac{p(v_i)}{q(v_i)} + log\ \frac{p(p_i)}{q(q_i)}] \\
&= \sum_i E_q\{x_i logp_i + (z_i - x_i)log(1 - p_i) - log(x!) \\
&\quad + \mu_i - v_i + z_i(logv_i - log\mu_i) + x_i log\mu_i \\
&\quad - [(\tilde{\alpha}_i - \alpha_i)logv_i - (\tilde{\beta}_i - \beta)v_i + \tilde{\alpha}_i log\tilde{\beta}_i - \alpha log\beta + log\Gamma(\alpha) - log\Gamma(\tilde{\alpha}_i)] \\
&\quad - [(\tilde{a}_i - a)logp_i + (\tilde{b}_i - b)log(1 - p_i) + logB(a, b) - logB(\tilde{a}_i, \tilde{b}_i)]\}
\end{aligned}
$$

Since q is fully factorized and that mean of posterior and posterior-log used here are all closed form, the ELBO is closed form. We can optimize over prior parameters $\Theta := (a, b, \alpha, \beta)$ using some available optimization method, like nlm.

6.4. Coordinate Ascent. To maximize $ELBO(g, q)$, we use coordiante ascent: fixing q, we can find the optimal g using nlm; fixing g, we update q use the update rules above.