Evaluation of Approximation of Fisher Information Matrix in Poisson Mixture Model using High Performance Computing

Qing Ji†, Matthias K. Gobbert†, Andrew M. Raim⋆, Nagaraj K. Neerchal†, Jorge G. Morel†
† Department of Mathematics and Statistics, University of Maryland, Baltimore County
⋆ United States Census Bureau

UMBC AN HONORS UNIVERSITY IN MARYLAND

Project Summary

- Fisher information matrix(FIM) is an essential part in the computation of maximum likelihood estimation(MLE) as well as in obtaining their standard errors
- Computation of FIM is resource intensive in some widely used models such as mixture distributions
- Neerchal and Morel(1993), Raim (2014), and Raim et. al.(2014) have provided approximations to the FIM
- In this project, we apply this approximation idea to a mixture of two Poisson distributions
- A program in C with MPI is designed to test the performance of our approximation under various selections of parameter values

Background

$$X_1, \dots, X_n \overset{iid}{\sim} f(x \mid \boldsymbol{\theta}),$$
$$\boldsymbol{\theta} = (\theta_1, \dots, \theta_k) \in \Theta, \quad x \in \mathcal{X}$$

- The likelihood function is

$$\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{x}) = \prod_{i=1}^n f(x_i \mid \boldsymbol{\theta}),$$

where $\boldsymbol{x} = (x_1, \dots, x_n)$ is observed data
- The score function is given by

$$S(\boldsymbol{\theta}) = E\Big\{\frac{\partial}{\partial \boldsymbol{\theta}} log\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{x})\Big\}$$

- The Fisher Information matrix(FIM) is:

$$\boldsymbol{I}(\boldsymbol{\theta}) = -\,E\Big[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} log\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{x})\Big]$$
$$= \mathbf{var}\Big\{\frac{\partial}{\partial \boldsymbol{\theta}} log\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{x})\Big\}$$

- The Fisher scoring algorithm of obtaining MLE is given by:

$$\hat{\boldsymbol{\theta}}^{(i+1)} = \hat{\boldsymbol{\theta}}^{(i)} - [\boldsymbol{I}(\hat{\boldsymbol{\theta}}^{(i)})]^{-1} S(\hat{\boldsymbol{\theta}}^{(i)})$$

Poisson Mixture Model

- The probability function of the mixture of two Poisson distributions is given by:

$$p(x;\boldsymbol{\theta}) = \pi\frac{\lambda_1^x e^{-\lambda_1}}{x!} + (1-\pi)\frac{\lambda_2^x e^{-\lambda_2}}{x!}$$

- The log-likelihood function for $\boldsymbol{x} = [x_1 x_2 \dots x_n]'$ is given by:

$$log\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{x}) = \sum_{i=1}^n log\{\pi\frac{\lambda_1^{x_i} e^{-\lambda_1}}{x_i!} + (1-\pi)\frac{\lambda_2^{x_i} e^{-\lambda_2}}{x_i!}\},$$

- The score function is given by:

$$S(\boldsymbol{\theta}; \boldsymbol{x}) = \frac{\partial}{\partial\boldsymbol{\theta}} log\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{x})$$
$$= \frac{\partial}{\partial\boldsymbol{\theta}} \sum_{i=1}^n log\{\pi\frac{\lambda_1^{x_i} e^{-\lambda_1}}{x_i!} + (1-\pi)\frac{\lambda_2^{x_i} e^{-\lambda_2}}{x_i!}\}$$

- The true FIM for $n = 1$:

$$-\,E\Big\{\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} log\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{x})\Big\}$$
$$= \mathbf{var}\Big\{S(\boldsymbol{\theta}; x)\Big\}$$
$$= -\sum_{x=0}^{\infty}\{S(\boldsymbol{\theta}; x) S'(\boldsymbol{\theta}; x)\} p(x;\boldsymbol{\theta})$$

Why Consider This Model?

- Mixture models are very useful in modeling populations consisting of two or more distinct subgroups but the group labels are not available
- Mixture models are also used, as shown in Raim(2014), to analyze data from multiple studies with a common objective but possibly different target population (Meta Analysis)
- Mixture models are also useful in modeling count data exhibiting more variance than accommodated by the binomial or Poisson models, a phenomenon known as overdispersion

Implementation Details

- Let's consider several combinations of values for the parameters:

$$\lambda_1, \lambda_2 \in \{0, 1, 2, \dots, 500\}$$
$$\pi \in \{0.1, 0.2, \dots, 0.5\}$$

- Compute the true FIM, $\boldsymbol{I}$, by truncation method:

$$\sum_{x=0}^{N}\Big[\frac{\partial}{\partial\theta_i}\ln p(x;\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}\ln p(x;\boldsymbol{\theta})\Big]p(x;\boldsymbol{\theta}),$$

with $p(x;\boldsymbol{\theta}) = \pi\frac{\lambda_1^x e^{-\lambda_1}}{x!} + (1-\pi)\frac{\lambda_2^x e^{-\lambda_2}}{x!}$

where N is a very large number chosen for truncation
- The approximation is given by:

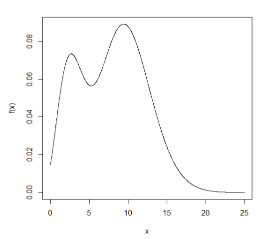
$$\mathbf{I_A} = \begin{bmatrix} \frac{n\pi_1}{\lambda_1} & 0 & 0 \\ 0 & \frac{n\pi_2}{\lambda_2} & 0 \\ 0 & 0 & \frac{1}{\pi(1-\pi)} \end{bmatrix}$$

- Quality of the approximation is measured using Frobenius norm:

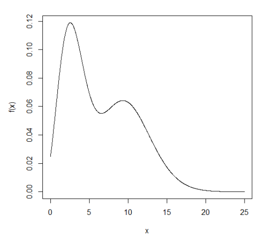
$$\|I_A - I\|_F = \sqrt{\sum_{a=1}^3 \sum_{b=1}^3 (I_A[a,b] - I[a,b])^2}$$

where $I_A[a,b]$ and $I[a,b]$ denote the $(a,b)^{th}$ entry in the approximate and exact FIM

Examples of Poisson Mixture



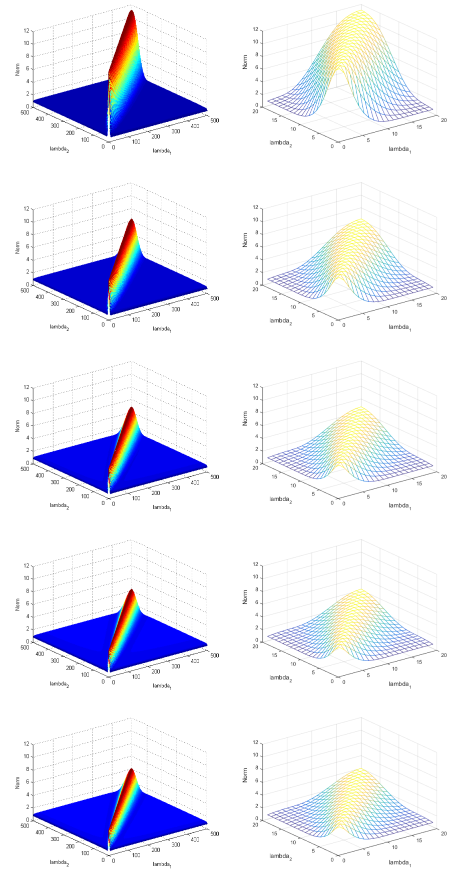
The density of a mixture Poisson for $\lambda_1 = 3, \lambda_2 = 10$ and $\pi = 0.3$



The density of a mixture Poisson for $\lambda_1 = 3, \lambda_2 = 10$ and $\pi = 0.5$

Results

Below are two sets of 3D plots addressing the relationship between $\|I_A - I\|_F$ and (λ_1, λ_2) as π increased from 0.1 to 0.5 (top to bottom). The complete picture of this simulation is displayed, $\lambda \in [0, 500]$ on the left and $\lambda \in [0, 20]$ and the domain of $lambda$ is on the right



- The $\|I_A - I\|_F$ remains very low and almost fixed when λ_1 and λ_2 are far apart
- $\|I_A - I\|_F$ increases dramatically once two λ's are within certain proximity
- $\|I_A - I\|_F$ reaches maximum when $\lambda_1 = \lambda_2$
- This pattern is not affected by the individual λ values but rather the closeness of the two λ's
- The value of π only decreases the maximum value as $\pi \to 0.5$, the pattern remains the same

Acknowledgment

The hardware used in the computational studies is part of the UMBC High Performance Computing Facility (HPCF). The facility is supported by the U.S. National Science Foundation through the MRI program (grant nos. CNS0821258 and CNS1228778) and the SCREMS program (grant no. DMS0821311),with additional substantial support from the University of Maryland, Baltimore County (UMBC). See `www.umbc.edu/hpcf` for more information on HPCF and the projects using its resources.

Reference

- Raim, A.M., Neerchal, N.K., and Morel, G.J., An approximation to the information matrix of exponential family finite mixtures, *Statistical Methodology*, 18:115-130, 2014.
- Raim, A.M., Computational Methods in Finite Mixtures Using Approximate Information and Regression Linked to the Mixture Mean, Unpublished PhD. Thesis, University of Maryland Baltimore County, 2014.
- Morel, J.G., Neerchal, N.K., A finite mixture distribution for modeling multinational extra variation *Biometrika*, 80(2):363-371, 1993.