Large Cluster Approximation to the Information Matrix Using Complete Data

With an Application to Meta-Analysis

Andrew M. Raim

(Now with U.S. Census Bureau)

Department of Mathematics and Statistics
University of Maryland, Baltimore County
Baltimore, MD, USA

2014 Joint Statistical Meetings
Boston, MA, USA

Joint work with Nagaraj K. Neerchal (UMBC) and Jorge G. Morel (UMBC)

Introduction

- Finite mixture densities are weighted sums of simpler densities

$$f(\mathbf{x} \mid \boldsymbol{\theta}) = \sum_{\ell=1}^{s} \pi_\ell f(x \mid \phi_\ell).$$

  Useful for analyzing data with multiple modes or extra variation.

- The Fisher information matrix (FIM) of $\mathbf{X} \sim f(\mathbf{x} \mid \boldsymbol{\theta})$

$$\mathcal{I}(\boldsymbol{\theta}) = \mathrm{E}\left[\left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{X} \mid \boldsymbol{\theta}) \right\} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{X} \mid \boldsymbol{\theta}) \right\}^T \right]$$

  is routinely used in statistical analysis: scoring, standard errors, etc.

- The FIM under a finite mixture does not have a simple analytical form.

Overview of the Talk

- For the finite mixture of binomials, Blischke (1964) used a simple block-diagonal matrix to approximate the inverse FIM. Morel and Nagaraj (1993) extended it to multinomial finite mixtures.

- In both cases, the block-diagonal matrix was shown to become close to the actual FIM as the number of trials increase.

- Raim, Liu, Neerchal, and Morel (2014) noted it is the FIM of the complete data: the observed $\mathbf{X}$ and missing subpopulation indicator Z.

- In this talk, we present:
  1. A convergence result for exponential family finite mixtures. It requires m observations, "grouped" like binomial.
  2. An example using MVN.
  3. An application in meta-analysis.

Assumption

- **(Grouped Sampling)**: Suppose $\mathbf{X}_1, \ldots, \mathbf{X}_m$ are independent and identically distributed from one of s exponential family densities

$$f(\mathbf{x} \mid \boldsymbol{\eta}_1), \quad \ldots, \quad f(\mathbf{x} \mid \boldsymbol{\eta}_s)$$

- Let $Z = \ell$ (not observed) indicate that the ℓth density was used, and suppose

$$Z = \begin{cases} 1 & \text{w.p. } \pi_1, \\ & \vdots \\ s & \text{w.p. } \pi_s. \end{cases}$$

- The density of the sufficient statistic $\mathbf{T}$ can be written as

$$f(\mathbf{t} \mid \boldsymbol{\theta}) \propto \sum_{\ell=1}^{s} \pi_\ell \exp\left\{ \boldsymbol{\eta}_\ell^T \mathbf{t} + m \cdot a(\boldsymbol{\eta}_\ell) \right\}, \quad \boldsymbol{\theta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_s, \boldsymbol{\pi}).$$

Result

- Complete data FIM of $(\mathbf{T}, Z)$ is $\widetilde{\mathcal{I}}_m(\boldsymbol{\theta}) = \text{Blockdiag}\left(\pi_1 \mathbf{F}_1, \ldots, \pi_s \mathbf{F}_s, \mathbf{F}_\pi\right)$

$$\begin{aligned} \mathbf{F}_\ell &= \text{Var}(\mathbf{T} \mid Z = \ell), & \longleftarrow \text{ FIM under the } \ell\text{th subpopulation}, \\ \mathbf{F}_\pi &= \mathbf{D}_\pi^{-1} + \pi_s^{-1} \mathbf{1}\mathbf{1}^T, & \longleftarrow \text{ FIM of } \text{Mult}_s(1, \boldsymbol{\pi}). \end{aligned}$$

- Raim, Neerchal, and Morel (Submitted 2014) prove the following.

Theorem

(a) $\widetilde{\mathcal{I}}_m(\boldsymbol{\theta}) - \mathcal{I}_m(\boldsymbol{\theta}) \to \mathbf{0}$ *as* $m \to \infty$. *Rate is* $O(m^2 e^{-m \cdot (const)})$.

(b) *If* $\mathcal{I}_m(\boldsymbol{\theta})$ *and* $\widetilde{\mathcal{I}}_m(\boldsymbol{\theta})$ *are nonsingular, then* $\mathcal{I}_m^{-1}(\boldsymbol{\theta}) - \widetilde{\mathcal{I}}_m^{-1}(\boldsymbol{\theta}) \to \mathbf{0}$ *as* $m \to \infty$.

Example: Multivariate Normal (Σ known)

- Suppose $\mathbf{X}_1, \ldots, \mathbf{X}_m$ are iid from one the following MVN densities:

$$\mathrm{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}), \ldots, \mathrm{N}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}).$$

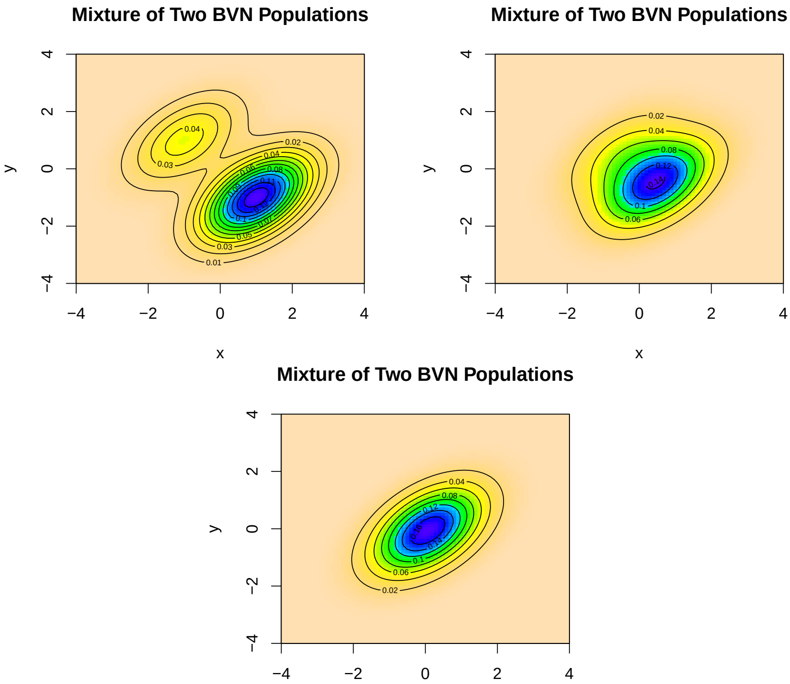
- To compare the approximate and true FIM w.r.t. $\boldsymbol{\psi} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_s, \boldsymbol{\pi})$.

- We obtain $\widetilde{\mathcal{I}}(\boldsymbol{\psi}) = \mathrm{Blockdiag}(\pi_1 \mathbf{F}_1, \ldots, \pi_s \mathbf{F}_s, \mathbf{F}_\pi)$ with

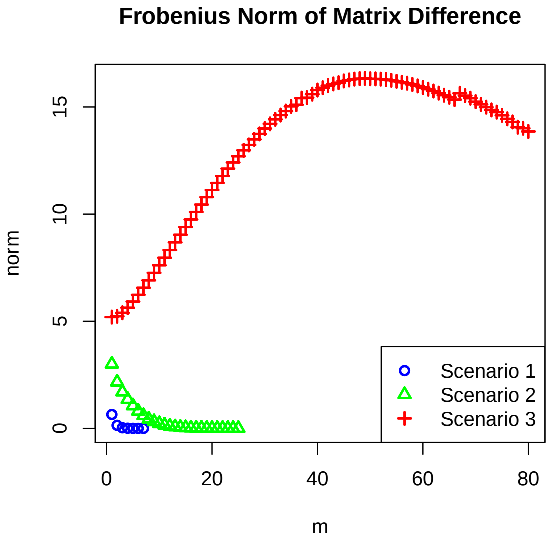
$$\mathbf{F}_\ell = m\boldsymbol{\Sigma}^{-1} \quad \text{and} \quad \mathbf{F}_\pi = \mathbf{D}_\pi^{-1} + \pi_s^{-1}\mathbf{1}\mathbf{1}^T$$

- We will consider three scenarios with mixture of two bivariate normals:
  1. $\boldsymbol{\mu}_1 = (-1, 1)$, $\boldsymbol{\mu}_2 = (1, -1)$.
  2. $\boldsymbol{\mu}_1 = (-0.5, 0.5)$, $\boldsymbol{\mu}_2 = (0.5, -0.5)$.
  3. $\boldsymbol{\mu}_1 = (-0.125, 0.125)$, $\boldsymbol{\mu}_2 = (0.125, -0.125)$.

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\pi} = \begin{pmatrix} 0.25 \\ 0.75 \end{pmatrix}.$$


Mixture of Two BVN Populations
Mixture of Two BVN Populations
Mixture of Two BVN Populations
x
y
−4
−2
0
2
4
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.1
0.12
0.14
0.16

Example: Bivariate Normal (Σ known)



Scenario	m	$\|\widetilde{\mathcal{I}} - \mathcal{I}\|_F$
1	6	0.0002
2	25	0.0005
3	80	13.8565

Finite Mixture Model for Meta-Analysis

Amlodipine Data

- $n = 8$ trials of an angina drug (Amlodipine) vs. placebo.
- Studied by (Hartung and Knapp, 2001) and used as an example in the book by Hartung, Knapp, and Sinha (2008).
- Subject's outcome: $\log\left(\frac{\text{Exercise time } \textbf{after} \text{ treatment}}{\text{Exercise time } \textbf{before} \text{ treatment}}\right)$.
  Objective: inference on $\mu_{\text{AMLO}} - \mu_{\text{PLA}}$.

Study	$m_{\texttt{AMLO}}$	$\bar{y}_{\texttt{AMLO}}$	$s^2_{\texttt{AMLO}}$	$m_{\texttt{PLA}}$	$\bar{y}_{\texttt{PLA}}$	$s^2_{\texttt{PLA}}$
1	46	0.2316	0.2254	48	-0.0027	0.0007
2	30	0.2811	0.1441	26	0.0270	0.1139
3	75	0.1894	0.1981	72	0.0443	0.4972
4	12	0.0930	0.1389	12	0.2277	0.0488
5	32	0.1622	0.0961	34	0.0056	0.0955
6	31	0.1837	0.1246	31	0.0943	0.1734
7	27	0.6612	0.7060	27	-0.0057	0.9891
8	46	0.1366	0.1211	47	-0.0057	0.1291

Finite Mixture Model for Meta-Analysis

- Let $\mathcal{D}_i = (\bar{y}_{Ti}, \bar{y}_{Ci}, s^2_{Ti}, s^2_{Ci})$ represent data from ith study for $i = 1, \ldots, n$.
- Among the treatment/control pairs, assume there are J common subpopulations
- Given that jth subject of the ith study belongs to subpop'n $z_i = \ell$, assume the fixed effect model:

$$y_{Tij} \overset{\text{iid}}{\sim} \mathrm{N}(\mu_{T\ell}, \sigma^2_{T\ell}), \quad j = 1, \ldots, m_{Ti}$$
$$y_{Cij} \overset{\text{iid}}{\sim} \mathrm{N}(\mu_{C\ell}, \sigma^2_{C\ell}), \quad j = 1, \ldots, m_{Ci}$$

Finite Mixture Model for Meta-Analysis

- Given $z_i = \ell$, the density of $\mathcal{D}_i$ is

$$f(\mathcal{D}_i \mid z_i = \ell) = f(\bar{y}_{Ti}) \cdot f(\bar{y}_{Ci}) \cdot f(s^2_{Ti}) \cdot f(s^2_{Ci})$$

- Likelihood wrt $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J, \boldsymbol{\pi})$, for $\boldsymbol{\theta}_\ell = (\mu_{T\ell}, \sigma_{T\ell}, \mu_{C\ell}, \sigma_{C\ell})$, is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left[\sum_{\ell=1}^{J} \pi_j f(\mathcal{D}_i \mid z_i = \ell) \right].$$

- The complete data information matrix is given by

$$\widetilde{\mathcal{I}}(\boldsymbol{\theta}) = \text{Blockdiag}(\pi_1 \mathbf{F}_1, \dots, \pi_J \mathbf{F}_J, \mathbf{F}_\pi),$$

$$\mathbf{F}_\ell = \text{Diag}\left(\sigma^{-2}_{T\ell} \sum_{i=1}^{n} m_{Ti},\ 2\sigma^{-2}_{T\ell} \sum_{i=1}^{n} m_{Ti},\ \sigma^{-2}_{C\ell} \sum_{i=1}^{n} m_{Ci},\ 2\sigma^{-2}_{C\ell} \sum_{i=1}^{n} m_{Ci} \right)$$

$$\mathbf{F}_\pi = n \left[\mathbf{D}_\pi^{-1} + \pi_J^{-1} \mathbf{1}\mathbf{1}^T \right].$$

Finite Mixture Model for Meta-Analysis

- The score function is composed of the entries

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \mu_{T\ell}} = \sum_{i=1}^{n} \frac{\pi_\ell f(\mathcal{D}_i \mid z_i = \ell)}{f(\mathcal{D}_i)} \left[m_{Ti} \frac{\bar{y}_{Ti} - \mu_{T\ell}}{\sigma^2_{T\ell}} \right]$$

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \mu_{C\ell}} = \sum_{i=1}^{n} \frac{\pi_\ell f(\mathcal{D}_i \mid z_i = \ell)}{f(\mathcal{D}_i)} \left[m_{Ci} \frac{\bar{y}_{Ci} - \mu_{C\ell}}{\sigma^2_{C\ell}} \right]$$

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \sigma_{T\ell}} = \sum_{i=1}^{n} \frac{\pi_\ell f(\mathcal{D}_i \mid z_i = \ell)}{f(\mathcal{D}_i)} \left[-\frac{m_{Ti}}{\sigma_{T\ell}} + m_{Ti} \frac{(\bar{y}_{Ti} - \mu_{T\ell})^2}{\sigma^3_{T\ell}} + \frac{(m_{Ti} - 1) s^2_{Ti}}{\sigma^3_{T\ell}} \right]$$

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \sigma_{C\ell}} = \sum_{i=1}^{n} \frac{\pi_\ell f(\mathcal{D}_i \mid z_i = \ell)}{f(\mathcal{D}_i)} \left[-\frac{m_{Ci}}{\sigma_{C\ell}} + m_{Ci} \frac{(\bar{y}_{Ci} - \mu_{C\ell})^2}{\sigma^3_{C\ell}} + \frac{(m_{Ci} - 1) s^2_{Ci}}{\sigma^3_{C\ell}} \right]$$

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \pi_\ell} = \sum_{i=1}^{n} \frac{f(\mathcal{D}_i \mid z_i = \ell) - f(\mathcal{D}_i \mid z_i = J)}{f(\mathcal{D}_i)}$$

- Mixture fit by approximate scoring (Raim, Liu, Neerchal, and Morel, 2014)

$$\boldsymbol{\theta}^{(g+1)} = \boldsymbol{\theta}^{(g)} + \widetilde{\mathcal{I}}^{-1}(\boldsymbol{\theta}^{(g)}) S(\boldsymbol{\theta}^{(g)}), \quad \text{until } |\log L(\boldsymbol{\theta}^{(g+1)}) - \log L(\boldsymbol{\theta}^{(g)})| < \varepsilon_0.$$

  Then Newton-Raphson was used until final convergence. Standard errors are computed from $\widetilde{\mathcal{I}}(\hat{\boldsymbol{\theta}})$.

Amlodipine Data

Posterior Probabilities

- Estimate of $\mathrm{P}(Z_i = \ell \mid \mathcal{D}_i)$, for $i = 1, \ldots, n$ and $\ell = 1, \ldots, J$

(a) $J = 2$

Study	Group 1	Group 2
1	1.00E+00	2.73E−24
2	1.00E+00	4.32E−14
3	1.00E+00	2.79E−09
4	1.00E+00	4.04E−08
5	1.00E+00	7.20E−20
6	1.00E+00	1.07E−15
7	8.23E−32	1.00E+00
8	1.00E+00	5.11E−25

(b) $J = 3$

	Group 1	Group 2	Group 3
1	1.00E+00	1.24E−31	1.04E−17
2	1.00E+00	8.49E−15	9.85E−06
3	1.61E−41	1.35E−17	1.00E+00
4	1.00E+00	8.52E−09	4.71E−04
5	1.00E+00	1.01E−21	1.10E−08
6	9.96E−01	3.04E−14	3.71E−03
7	4.95E−62	1.00E+00	2.63E−16
8	1.00E+00	1.89E−25	5.97E−08

Amlodipine Data

Estimates under Finite Mixture

(a) $J = 2$

	Est.	SE
μ_{T1}	0.1896	0.0247
σ_{T1}	0.3989	0.0174
μ_{C1}	0.0346	0.0277
σ_{C1}	0.4462	0.0196
μ_{T2}	0.6612	0.1349
σ_{T2}	0.8245	0.0954
μ_{C2}	-0.0057	0.1602
σ_{C2}	0.9759	0.1133
π	0.8750	0.1169
ϕ_1	0.1550	0.0371
ϕ_2	0.6669	0.2094
ϕ_{avg}	0.2190	0.0729

	$J = 2$	$J = 3$
LogLik	-374.9294	-333.7822
AIC	767.8589	695.5644
AICC	677.8589	635.5644
BIC	768.5739	696.6766

(b) $J = 3$

	Est.	SE
μ_{T1}	0.1897	0.0255
σ_{T1}	0.3811	0.0180
μ_{C1}	0.0310	0.0204
σ_{C1}	0.3051	0.0145
μ_{T2}	0.6612	0.1349
σ_{T2}	0.8245	0.0954
μ_{C2}	-0.0057	0.1602
σ_{C2}	0.9759	0.1133
μ_{T3}	0.1894	0.0721
σ_{T3}	0.4420	0.0510
μ_{C3}	0.0444	0.1146
σ_{C3}	0.6998	0.0810
π_1	0.7495	0.1532
π_2	0.1250	0.1169
ϕ_1	0.1587	0.0327
ϕ_2	0.6669	0.2094
ϕ_3	0.1450	0.1354
ϕ_{avg}	0.2205	0.0717

Amlodipine Data

Comparison of Diagonal FIM Entries

(a) $J = 2$.

	$-H(\hat{\boldsymbol{\theta}})$	$\widetilde{\mathcal{I}}(\hat{\boldsymbol{\theta}})$	$\mathcal{I}(\hat{\boldsymbol{\theta}})$	$\hat{V}_{\texttt{boot}}$
μ_{T1}	1709.7	1644.4	1648.2	1447.1
σ_{T1}	3419.3	3288.9	3278.7	3030.2
μ_{C1}	1356.5	1305.5	1304.4	1180.3
σ_{C1}	2712.6	2610.9	2607.9	2401.2
μ_{T2}	39.7	55.0	55.1	54.2
σ_{T2}	79.4	110.0	109.4	107.5
μ_{C2}	36.0	39.0	39.0	38.5
σ_{C2}	56.7	78.0	77.4	76.4
π	73.1	73.1	73.0	119.3

(b) $J = 3$.

	$-H(\hat{\boldsymbol{\theta}})$	$\widetilde{\mathcal{I}}(\hat{\boldsymbol{\theta}})$	$\mathcal{I}(\hat{\boldsymbol{\theta}})$	$\hat{V}_{\texttt{boot}}$
μ_{T1}	1355.2	1542.6	1547.7	1195.2
σ_{T1}	2709.9	3085.2	3089.9	2445.8
μ_{C1}	2123.7	2391.5	2393.8	1856.6
σ_{C1}	4224.7	4783.1	4744.7	2968.9
μ_{T2}	39.7	55.0	54.6	52.8
σ_{T2}	79.4	110.0	109.4	104.5
μ_{C2}	32.9	39.0	38.9	37.2
σ_{C2}	56.7	78.0	77.6	71.6
μ_{T3}	384.6	192.4	190.5	184.9
σ_{T3}	766.4	384.3	379.0	341.2
μ_{C3}	147.3	76.1	75.8	76.0
σ_{C3}	291.3	152.2	149.8	116.8
π_1	74.1	74.4	73.3	128.5
π_2	127.4	127.7	126.1	228.7

Amlodipine Data

Comparison of Standard Errors

(a) $J = 2$

	$-H(\hat{\boldsymbol{\theta}})$	$\widetilde{\mathcal{I}}(\hat{\boldsymbol{\theta}})$	$\mathcal{I}(\hat{\boldsymbol{\theta}})$	$\hat{V}_{\texttt{boot}}$
μ_{T1}	.02419	.02466	.02463	.02631
σ_{T1}	.01710	.01744	.01747	.01817
μ_{C1}	.02715	.02768	.02769	.02912
σ_{C1}	.01920	.01957	.01958	.02042
μ_{T2}	.15868	.13487	.13471	.13587
σ_{T2}	.11220	.09537	.09563	.09676
μ_{C2}	.16657	.16017	.16022	.16137
σ_{C2}	.13281	.11326	.11365	.11477
π	.11693	.11693	.11703	.09208

(b) $J = 3$

	$-H(\hat{\boldsymbol{\theta}})$	$\widetilde{\mathcal{I}}(\hat{\boldsymbol{\theta}})$	$\mathcal{I}(\hat{\boldsymbol{\theta}})$	$\hat{V}_{\texttt{boot}}$
μ_{T1}	.02716	.02546	.02542	.02899
σ_{T1}	.01921	.01800	.01799	.02023
μ_{C1}	.02170	.02045	.02044	.02324
σ_{C1}	.01539	.01446	.01452	.01837
μ_{T2}	.15868	.13487	.13534	.13788
σ_{T2}	.11221	.09537	.09563	.09836
μ_{C2}	.17444	.16017	.16045	.16430
σ_{C2}	.13281	.11326	.11354	.11853
μ_{T3}	.05100	.07214	.07246	.07385
σ_{T3}	.03612	.05101	.05137	.05486
μ_{C3}	.08240	.11462	.11487	.11481
σ_{C3}	.05859	.08105	.08171	.09336
π_1	.15343	.15320	.15363	.11772
π_2	.11695	.11693	.11718	.08834

Conclusions

Under "grouped" sampling with exponential family finite mixtures, the complete data FIM and true FIM become close as $m \to \infty$.

- Rate depends on "distinctness" of subpopulations

Grouped sampling assumption naturally holds in a meta-analysis combining multiple studies.

- Aitkin (1999) fit finite mixtures to meta-analysis data via NPMLE, as a robust alternative to assuming normal random effect.

(Raim, Neerchal, and Morel, Submitted 2014) gives examples where:

- Convergence **does not happen** when sampling is "ungrouped".
- Convergence **does happen** under continuous mixtures of exponential family densities.
- Convergence **does happen** under finite mixtures of non-exponential family densities.

References

Murray Aitkin. Meta-analysis by random effect modelling in generalized linear models. *Statistics in Medicine*, 18(17–18):2343–2351, 1999.

W. R. Blischke. Estimating the parameters of mixtures of binomial distributions. *Journal of the American Statistical Association*, 59(306):510–528, 1964.

Joachim Hartung and Guido Knapp. On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Statistics in Medicine*, 20 (12):1771–1782, 2001.

Joachim Hartung, Guido Knapp, and Bimal K. Sinha. *Statistical Meta-Analysis with Applications*. Wiley, 2008.

Jorge G. Morel and Neerchal K. Nagaraj. A finite mixture distribution for modelling multinomial extra variation. *Biometrika*, 80(2):363–371, 1993.

Andrew M. Raim, Minglei Liu, Nagaraj K. Neerchal, and Jorge G. Morel. On the method of approximate Fisher scoring for finite mixtures of multinomials. *Statistical Methodology*, 18:115–130, 2014.

Andrew M. Raim, Nagaraj K. Neerchal, and Jorge G. Morel. An approximation to the information matrix of exponential family finite mixtures, Submitted 2014.

Contact Information

Andrew M. Raim

araim1@umbc.edu

Thank you!