Model Selection Criterion for Multivariate Bounded Asymmetric Gaussian Mixture Model

Zixiang Xian [1] Muhammad Azam [1] Manar Amayri [1,2]
Nizar Bouguila [1]

[1] *Concordia Institute for Information Systems Engineering Concordia University, Montreal, Canada*

[2] *Grenoble Institute of Technology, Grenoble, France*

EUSIPCO June 15, 2021

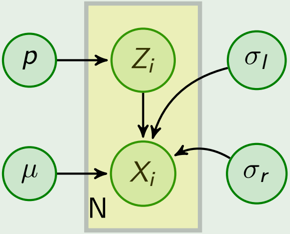
Outline

1. **Proposed Model**
2. **Model Selection**
3. **Experimental Results**
   - Synthetic Datasets
   - Real Datasets
   - Occupancy Detection
4. **Conclusion**

Proposed Model

Mixture of Asymmetric Gaussian Distributions

Asymmetric Gaussian Mixture Model (AGMM)



Graphical representation of an asymmetric Gaussian mixture model

Mathmatical Definition

$$p(\mathcal{X}, \mathcal{Z}|\Theta) = \prod_{i=1}^{N}\prod_{j=1}^{K}\left(p(\vec{X}_i|\xi_j)p_j\right)^{Z_{ij}} \quad (1)$$

- $p\left(\vec{X}_i|\xi_j\right)$ is the PDF of AGMM
- Asymmetric property
- **Unbounded support**

Mixture of Bounded Asymmetric Gaussian Distributions

Add bounded support

$$p(\vec{X}|\xi_j) = \frac{f(\vec{X}|\xi_j)\mathrm{H}(\vec{X}|\Omega_j)}{\int_{\partial_j} f(\vec{u}|\xi_j)du}, \text{where } \mathrm{H}(\vec{X}|\Omega_j) = \begin{cases} 1 & \text{if } \vec{X} \in \partial_j \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Definition

$$f(\vec{X}|\xi_j) = \prod_{d=1}^{D} \frac{2}{\sqrt{2\pi}(\sigma_{l_{jd}} + \sigma_{r_{jd}})} \begin{cases} \exp\left[-\frac{(X_d - \mu_{jd})^2}{2\sigma_{l_{jd}}^2}\right] & \text{if } X_d < \mu_{jd} \\ \exp\left[-\frac{(X_d - \mu_{jd})^2}{2\sigma_{r_{jd}}^2}\right] & \text{if } X_d \geq \mu_{jd} \end{cases} \quad (3)$$

where $f(\vec{X}|\xi_j)$ is the PDF of AGMM

Minimum Message Length (MML)

$$\text{Mess Len}(K) \simeq -\log\left(p\left(\Theta_K\right)\right) - \mathcal{L}\left(\Theta_K, Z, \mathcal{X}\right) + \frac{1}{2}\log\left|F\left(\Theta_K\right)\right| + \frac{N_p}{2}\left(1 + \log\left(\frac{1}{12}\right)\right) \tag{4}$$

Θ_K	set of parameters when mixture contains K components
$p(\Theta_K)$	prior probability
$\mathcal{L}\left(\Theta_K, Z, \mathcal{X}\right)$	log-likelihood of mixture model
$\lvert F(\Theta_K)\rvert$	determinant of Fisher information matrix
N_p	number of parameters (equal to $K(3D+1)$)

Derivation of the prior $p(\Theta)$

$$p(\Theta) = p(\pi)p(\mu)p(\sigma_l)p(\sigma_r) \tag{5}$$

$$p(\pi) = \frac{\Gamma(\sum_{j=1}^{K} \eta_j)}{\sum_{j=1}^{K} \Gamma(\eta_j)} \sum_{j=1}^{K} {p_j}^{\eta_j - 1} \tag{6}$$

$$p(\mu_{jd}) = \frac{1}{\sigma_{ld} + \sigma_{rd}} \implies p(\vec{\mu}_j) = \prod_{d=1}^{D} \frac{1}{\sigma_{ld} + \sigma_{rd}} \tag{7}$$

$$p(\sigma_l) = \prod_{j=1}^{K} p(\vec{\sigma}_{l_j}), \quad p(\sigma_r) = \prod_{j=1}^{K} p(\vec{\sigma}_{r_j}) \tag{8}$$

Replacing parameters in Eq. (5) by Eqs. (6 & 7 & 8)

$$p(\Theta) = (M-1)! \prod_{d=1}^{D} \frac{1}{{\sigma_{l_d}}^K {\sigma_{r_d}}^K (\sigma_{ld} + \sigma_{rd})^K} \tag{9}$$

Derivation of the Fisher information matrix $|F(\Theta)|$

$$|F(\Theta)| = |F(\pi)||F(\mu)||F(\sigma_l)||F(\sigma_r)| \quad (10)$$

Approximate the Hessian matrix by complete-data Fisher information matrix

$$|F(\pi)| = \frac{N^{K-1}}{\sum_{j=1}^{K} p_j}, \quad F(\vec{\mu}_j)_{k_1,k_2} = \frac{\partial^2 \mathcal{L}(\Theta, Z, \mathcal{X}_j)}{\partial \mu_{jk_1} \partial \mu_{jk_2}} \quad (11)$$

$$F(\vec{\sigma}_{l_j})_{k_1,k_2} = \frac{\partial^2 \mathcal{L}(\Theta, Z, \mathcal{X}_j)}{\partial \sigma_{l_{jk_1}} \partial \sigma_{l_{jk_2}}}, \quad F(\vec{\sigma}_{r_j})_{k_1,k_2} = \frac{\partial^2 \mathcal{L}(\Theta, Z, \mathcal{X}_j)}{\partial \sigma_{r_{jk_1}} \partial \sigma_{r_{jk_2}}} \quad (12)$$

Model Learning Algorithm

Algorithm 1 Model Learning for BAGMM

```
Input: Dataset 𝒳 = {X⃗_1, ..., X⃗_N}, t_min, K_min, K_max.
Output: Θ, 𝒵, K*.
for K_min ≤ K ≤ K_max do
    {Initialization}:
    K-Means (Compute μ⃗_1, ..., μ⃗_K & cluster assignment)
    for all 1 ≤ j ≤ K do
        Computation of p_j and {(σ⃗_{l_j} & σ⃗_{r_j})=σ⃗_j}
    {Expectation Maximization}:
    while relative change in log-likelihood ≥ t_min or iterations ≤ epoch_max or relative
  changes of parameters ≥ t_min do
        {[E Step]}:
        for all 1 ≤ j ≤ K do
            Compute p(j|X⃗_i) for i = 1, ..., N.
        {[M step]}:
        update bounded support range
        for all 1 ≤ j ≤ K do
            Estimate p_j, μ⃗_j, σ⃗_{l_j} & σ⃗_{r_j}
    end while
    Compute K* = arg min MML(K)
end for
```

Model Selection Results

Table: The model selection and clustering results for synthetic dataset

Synthetic Dataset(2,000 instances in each cluster)									
clusters	μ, σ_l, σ_r	AIC	BIC	CAIC	MDL	MMDL	MML like	LEC	MML
2	(2, -4) , (2, 3) , (1, 5) (5, 4), (3, 6), (2.1, 3.8)	2	2	2	2	2	2	2	2
3	(2, -4), (2, 3), (1, 5) (5, 4), (3, 6), (2.1, 3.8) (-10, 12), (3, 3.7), (3.4, 5.9)	3	3	3	3	3	3	3	3
4	(2, -4), (2, 3), (1, 5) (5, 4), (3, 6), (2.1, 3.8) (-10, 12), (3, 3.7), (3.4, 5.9) (-13, 14), (1, 2.1), (3, 3)	4	4	4	4	4	4	4	4
5	(2, -4), (2, 3), (1, 5) (5, 4), (3, 6), (2.1, 3.8) (-10, 12), (3, 3.7), (3.4, 5.9) (-13, 14), (1, 2.1), (3, 3) (-15, 16.6),(3.3, 4.4), (2.8, 2.7)	5	5	5	5	5	5	5	5

Execution Information

Table: Execution information of MML on synthetic dataset

Execution information on synthetic dataset(seconds)				
Mixture Models	Clusters	Time	Accuracy	Iterations
BAGMM	2 clusters	2.35	**71.3%**	5
BAGMM	3 clusters	8.60	**85.7%**	2
BAGMM	4 clusters	12.09	**72.2%**	4
BAGMM	5 clusters	12.58	**65.7%**	5

All experiments are running on a Macbook Pro 2015 with Dual-Core Intel Core i5 CPU. The BAGMM is as relatively fast as the AGMM for 5 clusters or more, but in the case of less than 5 clusters, the AGMM is a little bit faster. The BAGMM always converges faster than the AGMM with less iterations.

Model Selection Results

Real Dataset											
dataset	N	D	K	AIC	BIC	CAIC	MDL	MMDL	MML like	LEC	MML
Indian Liver Patient(AGMM)	583	10	**2**	4	2	2	2	4	4	2	2
Indian Liver Patient(BAGMM)				2	2	2	2	2	2	**2**	**2**
Iris(AGMM)	150	4	**3**	6	3	3	3	3	6	6	6
Iris(BAGMM)				6	6	6	6	6	6	**3**	**3**
Vertebral(AGMM)	310	6	**3**	3	3	3	3	3	3	3	3
Vertebral(BAGMM)				5	3	3	3	5	5	**3**	**3**
Wine Quality(red)(AGMM)	1599	11	**6**	5	5	5	5	5	5	6	6
Wine Quality(red)(BAGMM)				8	8	8	8	8	8	**6**	**6**
Spect Heart(AGMM)	80	44	**2**	6	4	2	4	4	6	2	2
Spect Heart(BAGMM)				5	2	2	2	5	5	**2**	**2**
Cryotherapy(AGMM)	90	6	**2**	2	2	2	2	2	2	2	2
Cryotherapy(BAGMM)				6	2	2	2	6	6	**2**	**2**
Immunotherapy(AGMM)	90	7	**2**	3	2	2	2	3	3	2	2
Immunotherapy(BAGMM)				2	2	2	2	2	2	**2**	**2**
Statlog(Heart)(AGMM)	270	13	**2**	6	6	2	6	6	6	6	6
Statlog(Heart)(BAGMM)				2	2	2	2	2	2	**2**	**2**
Parkinsons(AGMM)	197	22	**2**	6	6	6	6	6	6	6	6
Parkinsons(BAGMM)				2	2	2	2	2	2	**2**	**2**
Haberman Survival(AGMM)	306	3	**2**	2	2	2	2	2	2	2	2
Haberman Survival(BAGMM)				2	2	2	2	2	2	**2**	**2**

Occupancy Detection

Detect room occupancy as a binary classification from CO2, light, Humidity, temperature, and humidity ratio, which were taken every minute.

Table: Occupancy estimation and model selection results

Models	*N*	*D*	*K*	AIC	BIC	CAIC	MDL	MMDL	MML like	LEC	MML	Acc
AGMM	9752	5	**2**	5	5	5	5	5	5	5	5	79%
BAGMM				2	2	2	2	2	2	**2**	**2**	**94.8%**

Conclusion

- Proposed model selection criterion for bounded support asymmetric Gaussian mixture model (BAGMM) using minimum message length (MML)
- Validated using synthetic data, real data and occupancy detection application
- Compared with asymmetric Gaussian mixture model (AGMM)

For Further Reading I

Zixiang Xian, Muhammad Azam, Manar Amayri, Nizar Bouguila
Model Selection Criterion for Multivariate Bounded Asymmetric Gaussian Mixture Model
29th European Signal Processing Conference, *EUSIPCO 2021*