Supplementary materials for "Estimation of the complexity of a finite mixture distribution: from well- to less known methods"

Appendix A

Further examples of EM algorithm solutions for finite mixtures

Example: EM solutions for the mixture of geometric distributions. Consider an m-component mixture of geometric distributions with probability mass function

$$f_{\boldsymbol{\theta}}(x) = \pi_1(1-\theta_1)\theta_1^x + \ldots + \pi_m(1-\theta_m)\theta_m^x, \quad x \in \{0,1,\ldots,\}$$

with $\theta_j \in [0,1),\ j = 1,\ldots,m_0$. Then the complete likelihood can be written as

$$L_{\boldsymbol{\theta}}(x) = \prod_{i=1}^{n}[\theta_1^{x_i}(1-\theta_1)]^{z_{i1}} \ldots [\theta_m^{x_i}(1-\theta_m)]^{z_{im}}\pi_1^{z_{i1}} \ldots \pi_m^{z_{im}},$$

with $(z_{i1},\ldots,z_{i1}) \overset{i.i.d.}{\sim} \mathrm{Mult}\big(1,(\pi_1,\ldots,\pi_m)\big)$. the E-step at the s-th iteration yields

$$\hat{z}_{ij}^{(s)} = \frac{\hat{\pi}_j^{(s-1)}(1-\hat{\theta}_j^{(s-1)})[\hat{\theta}_j^{(s-1)}]^{x_i}}{\hat{\pi}_1^{(s-1)}(1-\hat{\theta}_1^{(s-1)})[\hat{\theta}_1^{(s-1)}]^{x_i} + \cdots + \hat{\pi}_m^{(s-1)}(1-\hat{\theta}_m^{(s-1)})[\hat{\theta}_m^{(s-1)}]^{x_i}}.$$

The updates at M-step are then:

$$\hat{\pi}_j^{(s)} = \frac{\sum_{i=1}^{n}\hat{z}_{ij}^{(s)}}{n}, \quad \hat{\theta}_j^{(s)} = \frac{\sum_{i=1}^{n}\hat{z}_{ij}^{(s)x_i}x_i}{\sum_{i=1}^{n}\hat{z}_{ij}^{(s)}(1+x_i)}.$$

Example: EM solutions for the mixture of Poisson distributions. Consider an m-component mixture of Poisson distributions with probability mass function

$$f_{\boldsymbol{\theta}}(x) = \pi_1\frac{e^{-\theta_1}\theta_1^x}{x!} + \ldots + \pi_m\frac{e^{-\theta_m}\theta_m^x}{x!} \quad x \in \{0,1,\ldots,\}$$

with $\theta_1,\ldots,\theta_m \in (0,\infty)$. The complete likelihood function is given by

$$L_{\boldsymbol{\theta}}(x) = \prod_{i=1}^{n}\left(\sum_{j=1}^{m}\pi_j\frac{e^{-\theta_j}\theta_j^{x_i}}{x_i!}\right)$$

Thus, for $j = 1,\ldots,m$ the E-step leads to

$$\hat{z}_{ij}^{(s)} = \frac{\hat{\pi}_j^{(s-1)}\frac{e^{-\hat{\theta}_j^{(s-1)}}[\hat{\theta}_j^{(s-1)}]^{x_i}}{x_i!}}{\hat{\pi}_1^{(s-1)}\frac{e^{-\hat{\theta}_1^{(s-1)}}[\hat{\theta}_1^{(s-1)}]^{x_i}}{x_i!} + \cdots + \hat{\pi}_m^{(s-1)}\frac{e^{-\hat{\theta}_m^{(s-1)}}[\hat{\theta}_m^{(s-1)}]^{x_i}}{x_i!}}.$$

and the M-step updates:

$$\hat{\pi}_j^{(s)} = \frac{\sum_{i=1}^{n}\hat{z}_{ij}^{(s)}}{n}, \quad \hat{\theta}_j^{(s)} = \frac{\sum_{i=1}^{n}\hat{z}_{ij}^{(s)}x_i}{\sum_{i=1}^{n}\hat{z}_{ij}^{(s)}}.$$

Further examples of the estimation of Hankel matrix moments of the mixing distribution for finite mixtures

Example: Mixture of geometric distributions. Consider a finite mixture of geometric distributions with probability mass function

$$f_0(x) = \pi_1(1-\theta_1)\theta_1^x + \ldots + \pi_{m_0}(1-\theta_{m_0})\theta_{m_0}^x, \quad x \in \{0, 1, \ldots, \}$$

with $\theta_j \in [0,1)$, $j = 1, \ldots, m_0$. Then, for any integer $j \geq 1$ we have that

$$\begin{aligned} \sum_{k=0}^{j-1} f_0(k) &= \sum_{r=1}^{m_0} \pi_r(1-\theta_r) \sum_{k=0}^{j-1} \theta_r^k = \sum_{r=1}^{m_0} \pi_r(1-\theta_r^j) \\ &= 1 - c_j \end{aligned}$$

and hence $c_j = 1 - \sum_{k=0}^{j-1} f_0(k) = \mathbb{E}[1 - \mathbb{I}_{X \leq j-1}]$ so that for $j = 1, \ldots, 2m$

$$\psi_j(x) = 1 - \mathbb{I}_{x \leq j-1}, \quad \text{and} \quad f_j(\boldsymbol{t}_{2m}) = t_1. \tag{0.1}$$

Example: Mixture of Poisson distributions. Consider a finite mixture of Poisson distributions with probability mass function

$$f_0(x) = \pi_1 \frac{e^{-\theta_1}\theta_1^x}{x!} + \ldots + \pi_{m_0} \frac{e^{-\theta_{m_0}}\theta_{m_0}^x}{x!} \quad x \in \{0, 1, \ldots, \}$$

with $\theta_1, \ldots, \theta_{m_0} \in (0, \infty)$. Using the moment generating function of a Poisson distribution we can show that if $X \sim f_0$ then $j \geq 1$ that

$$c_j = \mathbb{E}[X(X-1)\ldots(X-j+1)].$$

Thus, for $j = 1, \ldots, 2m$

$$\psi_j(x) = x(x-1)\ldots(x-j+1), \quad \text{and} \quad f_j(\boldsymbol{t}_{2m}) = t_1. \tag{0.2}$$

Appendix B

Simulation results

In reporting the results, we use acronyms of Table 27.

Table 27: Acronyms used in the simulations.

BIC	Bayesian information criterion
ICL	Integrated complete-data likelihood
HM$_{bt}$	Hankel matrix approach with bootstrap
HM$_{btsc}$	Hankel matrix approach with bootstrap and scaling
L_2E_{AIC}	L_2 estimation with AIC-based threshold
L_2E_{SBC}	L_2 estimation with SBC-based threshold
MHD$_{AIC}$	Hellinger estimation with AIC-based threshold
MHD$_{SBC}$	Hellinger estimation with SBC-based threshold
MHD$_{bt}$	Estimtion with Hellinger distance and bootstrap
LRT	Likelihood ratio test estimation
NN	Neural network

To make the reading easy, the frequency of recovering the true complexity for each mixture is marked in bold. The cases where a method estimates the number of components correctly in more than 50% of the instances are additionally marked with an asterisk. The results for the Minimum Hellinger distance with bootstrap and the sequential likelihood ratio tests are not given for all sample sizes due to computational complexity which increases rapidly with the number of observations. For NN's trained on 10000 samples (and the features were computed based on sample size 10000) the predicted class probabilities for the selected mixtures are reported.

Table 28: Relative frequencies of $\hat{m}_n$ for 2-comp. Gaussian mixtures based on500 replications.

n \ Method	$\hat{m}_n$	BIC	ICL	HM$_{bt}$	HM$_{btsc}$	MHD$_{AIC}$	MHD$_{SBC}$	MHD$_{bt}$	LRT	NN
	$\hat{m}_n$	$\boldsymbol{\theta}_2 = (0.5, 0.5; 1, 7)$								
50	1			1.000	0.024		0.002	0.002	0.030	
	2	**0.994***	**0.998***		**0.970***	**1.000***	**0.998***	**0.998***	**0.918***	
	3	0.006	0.002		0.004				0.052	
	4				0.002					
	≥5									
100	1			1.000	0.002				0.012	
	2	**0.996***	**1.000***		**0.992***	**1.000***	**1.000***	**1.000***	**0.948***	
	3	0.004			0.006				0.040	
	4									
	≥5									
500	1			1.000	0.002				0.004	
	2	**1.000***	**1.000***		**0.996***	**1.000***	**1.000***	**1.000***	**0.960***	
	3				0.002				0.036	
	4									
	≥5									
1000	1			1.000						
	2	**1.000***	**1.000***		**0.998***	**1.000***	**1.000***	**1.000***	**0.966***	
	3				0.002				0.034	
	4									
	≥5									
5000	1			1.000						
	2	**1.000***	**1.000***		**1.000***	**1.000***	**1.000***	**1.000***		
	3									
	4									
	≥5									
10000	1			1.000						0.00
	2	**1.000***	**1.000***		**1.000***	**1.000***	**1.000***	**1.000***		**0.96***
	3									0.04
	4									0.00
	≥5									0.00
	$\hat{m}_n$	$\boldsymbol{\theta}_2 = (0.05, 0.95; 1, 7)$								
n=50	1	0.100	0.106	1.000	0.888	0.756	0.998	0.980	0.250	
	2	**0.896***	**0.892***		**0.112**	**0.244**	**0.002**	**0.020**	**0.702***	
	3	0.004	0.002						0.048	
	4									
	≥5									
n=100	1	0.006	0.012	1.000	0.336	0.384	0.848	0.848	0.010	
	2	**0.986***	**0.988***		**0.664***	**0.616***	**0.152**	**0.152**	**0.938***	
	3	0.008							0.052	
	4									
	≥5									
n=500	1			1.000		0.016	0.018	0.016	0.004	
	2	**1.000***	**1.000***		**0.998***	**0.974***	**0.982***	**0.984***	**0.948***	
	3				0.002	0.010			0.048	
	4									
	≥5									
n=1000	1			1.000		0.006	0.006	0.006		
	2	**1.000***	**1.000***		**0.994***	**0.984***	**0.992***	**0.992***	**0.956***	
	3				0.006	0.010	0.002	0.002	0.044	
	4									
	≥5									
n=5000	1			1.000		0.002	0.002	0.002		
	2	**1.000***	**1.000***		**1.000***	**0.988***	**0.988***	**0.988***		
	3					0.010	0.010	0.010		
	4									
	≥5									
n=10000	1			1.000						0.00
	2	**1.000***	**1.000***		**0.998***	**0.990***	**0.990***	**0.990***		**0.74***
	3				0.002	0.008	0.010	0.010		0.22
	4					0.002				0.04
	≥5									0.00
	$\hat{m}_n$	$\boldsymbol{\theta}_2 = (0.4, 0.6; 1, 2.5)$								
50	1	0.632	0.954	1.000	1.000	1.000	1.000	1.000	0.952	
	2	**0.366**	**0.046**						**0.048**	
	3	0.002								
	4									
	≥5									
100	1	0.348	0.988	1.000	1.000	1.000	1.000	1.000	0.958	
	2	**0.650***	**0.012**						**0.040**	
	3	0.002							0.002	
	4									
	≥5									
500	1	0.002	1.000	1.000	0.990	1.000	1.000	1.000	0.938	
	2	**0.998***			**0.010**				**0.058**	
	3								0.004	
	4									
	≥5									
1000	1		1.000	1.000	0.920	1.000	1.000	1.000	0.892	
	2	**1.000***			**0.078**				**0.106**	
	3				0.002				0.002	
	4									
	≥5									
5000	1		1.000	1.000	0.582	0.996	1.000	1.000		
	2	**1.000***			**0.418**	**0.004**				
	3									
	4									
	≥5									
10000	1		1.000	1.000	0.458	0.974	1.000	1.000		0.00
	2	**1.000***			**0.542***	**0.026**				**0.91***
	3									0.09
	4									0.00
	≥5									0.00

Table 29: Relative frequencies of $\hat{m}_n$ for 3-comp. Gaussian mixtures based on 500 replications.

Method / n	$\hat{m}_n$	BIC	ICL	HM$_{bt}$	HM$_{btsc}$	MHD$_{AIC}$	MHD$_{SBC}$	MHD$_{bt}$	LRT	NN
	$\hat{m}_n$	$\boldsymbol{\theta}_3 = (0.3, 0.3, 0.4; 10, 17, 13)$								
50	1			1.000	0.560	0.580	0.972	0.970	0.746	
	2	0.020	0.092		0.430	0.420	0.028	0.030	0.236	
	3	**0.926***	**0.874***		**0.010**				**0.018**	
	4	0.048	0.028							
	≥5	0.006	0.006							
100	1			1.000	0.246	0.124	0.084	0.868	0.008	
	2				0.700	0.852	0.644	0.132	0.754	
	3	**1.000***	**1.000***		**0.044**	**0.024**	**0.272**		**0.160**	
	4				0.010					
	≥5									
500	1			1.000	0.010	0.002		0.002		
	2				0.394	0.092	0.968	0.984	0.012	
	3	**1.000***	**1.000***		**0.502***	**0.904***	**0.032**	**0.014**	**0.962***	
	4				0.076	0.002			0.026	
	≥5				0.018					
1000	1			1.000	0.006					
	2				0.138		0.054	0.580		
	3	**1.000***	**1.000***		**0.772***	**1.000***	**0.946***	**0.420***		
	4				0.072					
	≥5				0.012					
5000	1			1.000						
	2				0.008	0.002	0.042	0.002		
	3	**1.000***	**1.000***		**0.946***	**0.996***	**0.958***	**0.998***		
	4				0.044	0.002				
	≥5				0.002					
10000	1			1.000						0.00
	2				0.014		0.002			0.00
	3	**1.000***	**1.000***		**0.968***	**0.996***	**0.998***	**1.000***		**0.51***
	4				0.018	0.004				0.36
	≥5									0.13
	$\hat{m}_n$	$\boldsymbol{\theta}_3 = (0.05, 0.45, 0.5; 17, 13, 10)$								
50	1			1.000	0.644	0.920	1.000	0.970	0.736	
	2	0.554	0.604		0.354	0.080		0.030	0.248	
	3	**0.324**	**0.296**		**0.002**				**0.016**	
	4	0.090	0.070							
	≥5	0.032	0.030							
100	1		0.480	1.000	0.444	0.752	0.998	0.868	0.014	
	2	0.430	0.368		0.552	0.228	0.002	0.132	0.880	
	3	**0.392**	**0.100**		**0.004**	**0.020**			**0.102**	
	4	0.116	0.100						0.004	
	≥5	0.062	0.052							
500	1			1.000	0.056		0.690	0.002		
	2	0.182	0.206		0.940	0.266	0.280	0.984	0.286	
	3	**0.478**	**0.472**		**0.004**	**0.732***	**0.030**	**0.014**	**0.690***	
	4	0.228	0.196			0.002			0.024	
	≥5	0.112	0.126							
1000	1			1.000	0.034		0.034	0.034		
	2	0.082	0.112		0.962	0.202	0.408	0.408		
	3	**0.588***	**0.576***		**0.004**	**0.794***	**0.558***	**0.558***		
	4	0.210	0.170			0.004				
	≥5	0.120	0.142							
5000	1			1.000	0.010					
	2		0.002		0.752	0.162	0.162	0.162		
	3	**0.804***	**0.794***		**0.228**	**0.806***	**0.832***	**0.832***		
	4	0.166	0.144		0.010	0.032	0.006	0.006		
	≥5	0.030	0.058							
10000	1			1.000						0.00
	2				0.590	0.164	0.164	0.162		0.00
	3	**0.880***	**0.884**		**0.396**	**0.810***	**0.814***	**0.816***		**0.47**
	4	0.106	0.082		0.010	0.026	0.022	0.022		0.35
	≥5	0.014	0.034		0.004					0.18
	$\hat{m}_n$	$\boldsymbol{\theta}_3 = (0.2, 0.3, 0.5; 1, 2.5, 5)$								
50	1			1.000	0.846	0.946	1.000	1.000	0.966	
	2	0.962	0.996		0.150	0.054			0.028	
	3	**0.038**	**0.004**		**0.004**				**0.006**	
	4									
	≥5									
100	1			1.000	0.638	0.768	1.000	1.000	0.696	
	2	0.920	0.990		0.344	0.232			0.280	
	3	**0.080**	**0.010**		**0.016**				**0.024**	
	4				0.002					
	≥5									
500	1			1.000	0.140	0.002	0.506	0.506	0.004	
	2	0.698	1.000		0.860	0.998	0.494	0.494	0.916	
	3	**0.296**							**0.080**	
	4	0.006								
	≥5									
1000	1			1.000	0.062		0.008	0.008		
	2	0.448	1.000		0.938	1.000	0.992	0.992		
	3	**0.454**								
	4	0.096								
	≥5	0.002								
5000	1			1.000	0.022					
	2	0.006	1.000		0.976	1.000	1.000	1.000		
	3	**0.390**			**0.002**					
	4	0.410								
	≥5	0.194								
10000	1			1.000	0.006					0.00
	2		1.000		0.966	1.000	1.000	1.000		0.00
	3	**0.334**			**0.028**					**0.71***
	4	0.426								0.16
	≥5	0.240								0.03

Table 30: Relative frequencies of $\hat{m}_n$ for 2-comp. geometric mixtures based on 500 replications

n \ Method		BIC	ICL	HM$_{bt}$	HM$_{btsc}$	L_2E_{AIC}	L_2E_{SBC}	MHD$_{AIC}$	MHD$_{SBC}$	MHD$_{bt}$	LRT	NN
	$\hat{m}_n$	$\boldsymbol{\theta}_2 = \left(0.5, 0.5; 0.4, 0.5\right)$										
50	1	0.994	1.000	0.938	0.916	0.952	1.000	0.998	1.000	0.918	0.938	
	2	**0.006**		**0.056**	**0.082**	**0.048**		**0.002**		**0.082**	**0.062**	
	3			0.006	0.002							
	4											
	≥ 5											
100	1	0.998	1.000	0.932	0.918	0.930	1.000	1.000	1.000	0.914	0.912	
	2	**0.002**		**0.050**	**0.082**	**0.070**				**0.084**	**0.088**	
	3			0.014						0.002		
	4			0.004								
	≥ 5											
500	1	0.998	1.000	0.948	0.926	0.922	1.000	1.000	1.000	0.910	0.896	
	2	**0.002**		**0.042**	**0.074**	**0.078**				**0.088**	**0.104**	
	3			0.008								
	4			0.002						0.002		
	≥ 5											
1000	1	1.000	1.000	0.936	0.916	0.900	1.000	0.994	1.000	0.874	0.876	...
	2			**0.052**	**0.082**	**0.100**		**0.006**		**0.114**	**0.124**	
	3			0.010	0.002					0.008		
	4			0.002						0.004		
	≥ 5											
5000	1	0.996	1.000	0.926	0.918	0.900	1.000	0.980	1.000	0.768		
	2	**0.004**		**0.068**	**0.080**	**0.092**		**0.020**		**0.216**		
	3			0.002	0.002	0.008				0.008		
	4			0.004						0.008		
	≥ 5											
10000	1	0.992	1.000	0.888	0.892	0.864	1.000	0.964	1.000	0.602		0.39
	2	**0.008**		**0.104**	**0.104**	**0.130**		**0.036**		**0.382**		**0.44**
	3			0.006	0.004	0.006				0.008		0.13
	4									0.008		0.03
	≥ 5			0.002								0.01
	$\hat{m}_n$	$\boldsymbol{\theta}_2 = \left(0.1, 0.9; 0.8, 0.3\right)$										
50	1	0.546	0.848	0.942	0.862	0.950	1.000	0.988	0.998	0.708	0.380	
	2	**0.454**	**0.152**	**0.056**	**0.134**	**0.050**		**0.012**	**0.002**	**0.280**	**0.598***	
	3			0.002	0.004					0.010	0.022	
	4											
	≥ 5									0.002		
100	1	0.356	0.800	0.932	0.840	0.906	1.000	0.942	0.996	0.536	0.152	
	2	**0.644***	**0.200**	**0.066**	**0.156**	**0.094**		**0.058**	**0.004**	**0.450**	**0.816***	
	3			0.002	0.004					0.012	0.032	
	4									0.002		
	≥ 5											
500	1	0.002	0.826	0.744	0.572	0.606	1.000	0.056	0.588	0.004		
	2	**0.998***	**0.174**	**0.232**	**0.386**	**0.394**		**0.944***	**0.412**	**0.964***	**0.944***	
	3			0.014	0.042					0.032	0.056	
	4			0.008								
	≥ 5			0.002								
1000	1		0.876	0.506	0.366	0.242	0.996	0.004	0.030	0.004		
	2	**1.000***	**0.124**	**0.462**	**0.586***	**0.756***	**0.004**	**0.996***	**0.970***	**0.934***	**0.944***	
	3			0.026	0.048	0.002				0.058	0.054	
	4			0.004						0.002	0.002	
	≥ 5			0.002						0.002		
5000	1		0.980	0.012	0.008	0.002	0.676					
	2	**1.000***	**0.020**	**0.930***	**0.924***	**0.988***	**0.324**	**1.000***	**1.000***	**0.960***		
	3			0.036	0.066	0.010				0.038		
	4			0.014	0.002					0.002		
	≥ 5			0.008								
10000	1		0.998			0.002	0.014					0.00
	2	**1.000***	**0.002**	**0.932***	**0.940***	**0.992***	**0.986***	**1.000***	**1.000***	**0.942***		**0.56***
	3			0.054	0.060	0.006				0.056		0.30
	4			0.008						0.002		0.11
	≥ 5			0.006								0.03

Table 31: Relative frequencies of $\hat{m}_n$ for 3-comp. geometric mixtures based on 500 replications.

n \ Method		BIC	ICL	HM$_{bt}$	HM$_{btsc}$	L_2E_{AIC}	L_2E_{SBC}	MHD$_{AIC}$	MHD$_{SBC}$	MHD$_{bt}$	LRT	NN
	$\hat{m}_n$	$\boldsymbol{\theta}_3 = (0.2, 0.3, 0.5; 0.4, 0.2, 0.5)$										
	1	0.984	1.000	0.924	0.890	0.942	1.000	0.994	1.000	0.838	0.866	
	2	0.016		0.070	0.104	0.058		0.006		0.156	0.134	
50	3			**0.004**	**0.006**					**0.002**		
	4									0.002		
	≥5			0.002						0.002		
	1	0.986	1.000	0.920	0.868	0.904	1.000	0.994	1.000	0.856	0.796	
	2	0.014		0.068	0.130	0.096		0.006		0.138	0.202	
100	3			**0.008**	**0.002**					**0.004**	**0.002**	
	4			0.004						0.002		
	≥5											
	1	0.944	1.000	0.820	0.762	0.734	1.000	0.930	1.000	0.532	0.434	
	2	0.056		0.170	0.228	0.266		0.070		0.460	0.562	
500	3			**0.008**	**0.010**						**0.004**	
	4			0.002						0.008		
	≥5											
	1	0.804	1.000	0.692	0.642	0.556	1.000	0.764	0.998	0.340	0.216	
	2	0.196		0.290	0.338	0.444		0.236	0.002	0.628	0.774	
1000	3			**0.014**	**0.020**					**0.008**	**0.008**	
	4			0.004						0.020	0.002	
	≥5									0.004		
	1	0.052	1.000	0.104	0.082	0.028	0.950	0.072	0.680	0.076		
	2	0.946		0.842	0.868	0.972	0.050	0.926	0.320	0.898		
5000	3	**0.002**		**0.040**	**0.050**			**0.002**		**0.008**		
	4			0.014						0.016		
	≥5									0.002		
	1		1.000	0.004	0.002		0.624	0.082	0.110	0.084		0.00
	2	0.964		0.914	0.930	0.992	0.376	0.918	0.890	0.888		0.20
10000	3	**0.036**		**0.068**	**0.068**	**0.008**				**0.016**		**0.27**
	4			0.010						0.010		0.25
	≥5			0.004						0.002		0.28
	$\hat{m}_n$	$\boldsymbol{\theta}_3 = (0.1, 0.3, 0.6; 0.8, 0.2, 0.4)$										
	1	0.546	0.848	0.918	0.806	0.932	1.000	0.990	0.998	0.650	0.298	
	2	0.454	0.152	0.070	0.182	0.068		0.010	0.002	0.336	0.670	
50	3			**0.008**	**0.012**					**0.012**	**0.032**	
	4			0.002						0.002		
	≥5			0.002								
	1	0.282	0.778	0.862	0.720	0.850	1.000	0.902	0.994	0.378	0.080	
	2	0.718	0.222	0.132	0.266	0.150		0.098	0.006	0.588	0.848	
100	3			**0.006**	**0.014**					**0.024**	**0.072**	
	4									0.006		
	≥5									0.004		
	1		0.796	0.382	0.222	0.240	1.000	0.010	0.282			
	2	0.996	0.204	0.578	0.730	0.760		0.990	0.718	0.938	0.882	
500	3	**0.004**		**0.036**	**0.048**					**0.062**	**0.118**	
	4			0.004								
	≥5											
	1		0.854	0.128	0.052	0.020	0.988		0.002			
	2	0.998	0.146	0.828	0.900	0.980	0.012	1.000	0.998	0.894	0.864	
1000	3	**0.002**		**0.034**	**0.048**					**0.102**	**0.134**	
	4			0.008						0.002	0.002	
	≥5			0.002						0.002		
	1		0.960				0.016					
	2	0.996	0.040	0.926	0.938	0.986	0.984	0.984	1.000	0.676		
5000	3	**0.004**		**0.058**	**0.058**	**0.014**		**0.016**		**0.292**		
	4			0.010	0.004					0.018		
	≥5			0.006						0.014		
	1		0.994									0.00
	2	0.990	0.006	0.916	0.894	0.958	1.000	0.940	1.000	0.482		0.16
10000	3	**0.010**		**0.072**	**0.100**	**0.042**		**0.060**		**0.490**		**0.25**
	4			0.010	0.006					0.016		0.26
	≥5			0.002						0.012		0.32

Table 32: Relative frequencies of $\hat{m}_n$ for 2-comp. Poisson mixtures based on 500 replications.

n \ Method		BIC	ICL	HM$_{bt}$	HM$_{btsc}$	L_2E_{AIC}	L_2E_{SBC}	MHD$_{AIC}$	MHD$_{SBC}$	MHD$_{bt}$	LRT	NN
	$\hat{m}_n$	$\boldsymbol{\theta}_2 = (0.5, 0.5; 1, 9)$										
50	1					0.016	0.054	0.004	0.004	0.002		
	2	**0.996***	**1.000***	**0.972***	**0.726***	**0.920***	**0.940***	**0.992***	**0.992***	**0.936***	**0.962***	
	3	0.004		0.026	0.222	0.062	0.006	0.004	0.004	0.050	0.038	
	4			0.002	0.036	0.002				0.012		
	≥5				0.016							
100	1					0.008	0.014		0.002			
	2	**1.000***	**1.000***	**0.952***	**0.726***	**0.924***	**0.970***	**0.996***	**0.998***	**0.938***	**0.938***	
	3			0.040	0.236	0.068	0.016	0.004		0.054	0.060	
	4			0.004	0.028					0.004	0.002	
	≥5			0.004	0.010					0.004		
500	1					0.004	0.004					
	2	**0.998***	**1.000***	**0.934***	**0.796***	**0.924***	**0.964***	**1.000***	**1.000***	**0.940***	**0.942***	
	3	0.002		0.058	0.188	0.064	0.032			0.048	0.056	
	4			0.006	0.012	0.008				0.008	0.002	
	≥5			0.002	0.004					0.004		
1000	1					0.014	0.018			0.002		
	2	**1.000***	**1.000***	**0.934***	**0.850***	**0.936***	**0.964***	**0.998***	**0.998***	**0.936***	**0.958***	
	3			0.062	0.136	0.048	0.018	0.002	0.002	0.054	0.040	
	4			0.002	0.014	0.002				0.004	0.002	
	≥5			0.002						0.004		
5000	1					0.008	0.008					
	2	**1.000***	**1.000***	**0.952***	**0.914***	**0.946***	**0.974***	**0.998***	**1.000***	**0.926***		
	3			0.042	0.076	0.042	0.018	0.002		0.068		
	4			0.002	0.010	0.004				0.006		
	≥5			0.004								
10000	1					0.008	0.010					0.00
	2	**1.000***	**1.000***	**0.926***	**0.914***	**0.942***	**0.968***	**1.000***	**1.000***	**0.942***		**0.57***
	3			0.064	0.086	0.050	0.022			0.042		0.27
	4			0.006						0.010		0.11
	≥5			0.004						0.006		0.05
	$\hat{m}_n$	$\boldsymbol{\theta}_2 = (0.2, 0.8; 9, 1)$										
50	1				0.050	0.242	0.728	0.090	0.302	0.006		
	2	**0.992***	**1.000***	**0.988***	**0.906***	**0.710***	**0.270**	**0.910***	**0.698***	**0.956***	**0.966***	
	3	0.008		0.006	0.040	0.048	0.002			0.028	0.032	
	4			0.004	0.002					0.010	0.002	
	≥5			0.002	0.002							
100	1					0.054	0.630	0.004	0.008	0.004		
	2	**0.998***	**1.000***	**0.948***	**0.870***	**0.836***	**0.364**	**0.996***	**0.992***	**0.966***	**0.944***	
	3	0.002		0.032	0.110	0.108	0.006			0.028	0.054	
	4			0.014	0.018	0.002				0.002	0.002	
	≥5			0.006	0.002							
500	1					0.010	0.012					
	2	**0.998***	**1.000***	**0.944***	**0.738***	**0.944***	**0.982***	**1.000***	**1.000***	**0.942***	**0.962***	
	3	0.002		0.034	0.220	0.046	0.006			0.054	0.038	
	4			0.016	0.036					0.002		
	≥5			0.006	0.006					0.002		
1000	1					0.010	0.010					
	2	**1.000***	**1.000***	**0.942***	**0.802***	**0.940***	**0.976***	**1.000***	**1.000***	**0.924***	**0.946***	
	3			0.040	0.180	0.048	0.014			0.068	0.052	
	4			0.012	0.014	0.002				0.004	0.002	
	≥5			0.006	0.004					0.004		
5000	1					0.006	0.006					
	2	**1.000***	**1.000***	**0.938***	**0.892***	**0.940***	**0.974***	**0.994***	**1.000***	**0.922***		
	3			0.046	0.094	0.048	0.018	0.006		0.074		
	4			0.012	0.012	0.006	0.002			0.002		
	≥5			0.004	0.002					0.002		
10000	1					0.004	0.004					0.00
	2	**1.000***	**1.000***	**0.948***	**0.922***	**0.956***	**0.978***	**0.998***	**1.000***	**0.930***		**0.49**
	3			0.032	0.068	0.034	0.018	0.002		0.066		0.30
	4			0.012	0.010	0.004						0.13
	≥5			0.008		0.002				0.004		0.08
	$\hat{m}_n$	$\boldsymbol{\theta}_2 = (0.5, 0.5; 1.1, 1)$										
50	1	0.998	1.000	0.956	0.882	0.826	0.868	1.000	1.000	0.934	0.972	
	2	**0.002**		**0.044**	**0.100**	**0.170**	**0.132**			**0.060**	**0.028**	
	3				0.016	0.004				0.006		
	4				0.002							
	≥5											
100	1	0.998	1.000	0.918	0.902	0.824	0.862	1.000	1.000	0.942	0.936	
	2	**0.002**		**0.080**	**0.086**	**0.168**	**0.136**			**0.048**	**0.064**	
	3			0.002	0.010	0.006	0.002			0.008		
	4				0.002	0.002						
	≥5									0.002		
500	1	1.000	1.000	0.954	0.946	0.828	0.846	1.000	1.000	0.932	0.954	
	2			**0.036**	**0.048**	**0.162**	**0.152**			**0.064**	**0.046**	
	3			0.010	0.004	0.010	0.002			0.004		
	4				0.002							
	≥5											
1000	1	1.000	1.000	0.962	0.950	0.846	0.856	0.998	1.000	0.950	0.958	
	2			**0.036**	**0.046**	**0.150**	**0.142**	**0.002**		**0.044**	**0.040**	
	3			0.002	0.002	0.004	0.002			0.006	0.002	
	4				0.002							
	≥5											
5000	1	1.000	1.000	0.942	0.954	0.864	0.874	1.000	1.000	0.912		
	2			**0.056**	**0.042**	**0.126**	**0.124**			**0.070**		
	3				0.004	0.010	0.002			0.016		
	4			0.002						0.002		
	≥5											
10000	1	1.000	1.000	0.932	0.930	0.858	0.868	1.000	1.000	0.922		0.86
	2			**0.056**	**0.062**	**0.136**	**0.132**			**0.070**		**0.12**
	3			0.010	0.008	0.006				0.008		0.02
	4			0.002								0.00
	≥5											0.00

Continued on next page

Table 32 – *Continued from previous page*

n \ Method		BIC	ICL	HM$_{bt}$	HM$_{btsc}$	L_2E_{AIC}	L_2E_{SBC}	MHD$_{AIC}$	MHD$_{SBC}$	MHD$_{bt}$	LRT	NN
	$\hat{m}_n$	$\boldsymbol{\theta}_2 = (0.05, 0.95; 10, 1)$										
50	1	0.088	0.112	0.098	0.792	0.774	0.822	0.938	0.996	0.310	0.094	
	2	**0.910***	**0.888***	**0.882***	**0.208**	**0.218**	**0.178**	**0.062**	**0.004**	**0.642***	**0.866***	
	3	0.002		0.018		0.008				0.038	0.040	
	4			0.002						0.008		
	≥5									0.002		
100	1	0.010	0.018	0.010	0.478	0.764	0.836	0.634	0.946	0.132	0.010	
	2	**0.986***	**0.982***	**0.962***	**0.514***	**0.226**	**0.164**	**0.366**	**0.054**	**0.830***	**0.946***	
	3	0.004		0.028	0.008	0.010				0.024	0.044	
	4									0.012		
	≥5									0.002		
500	1					0.480	0.858	0.010	0.014	0.006		
	2	**1.000***	**1.000***	**0.966***	**0.902***	**0.472**	**0.142**	**0.990***	**0.986***	**0.976***	**0.976***	
	3			0.016	0.078	0.046				0.018	0.024	
	4			0.008	0.016	0.002						
	≥5			0.010	0.004							
1000	1					0.238	0.836	0.002	0.002			
	2	**1.000***	**1.000***	**0.956***	**0.808***	**0.656***	**0.164**	**0.996***	**0.998***	**0.936***	**0.958***	
	3			0.032	0.148	0.094		0.002		0.058	0.040	
	4			0.004	0.040	0.012				0.002	0.002	
	≥5			0.008	0.004					0.004		
5000	1					0.056	0.398	0.004	0.004			
	2	**1.000***	**1.000***	**0.942***	**0.876***	**0.582***	**0.572***	**0.996***	**0.996***	**0.930***		
	3			0.050	0.098	0.272	0.030			0.068		
	4			0.004	0.026	0.070						
	≥5			0.004		0.020				0.002		
10000	1					0.056	0.158	0.002	0.002	0.004		0.00
	2	**1.000***	**1.000***	**0.922***	**0.858***	**0.510***	**0.706***	**0.998***	**0.998***	**0.928***		**0.63***
	3			0.062	0.124	0.282	0.134			0.062		0.27
	4			0.004	0.016	0.112	0.002			0.002		0.08
	≥5			0.012	0.002	0.040				0.004		0.02

Table 33: Relative frequencies of $\hat{m}_n$ for 3-comp. Poisson mixtures based on 500 replications

n \ Method		BIC	ICL	HM$_{bt}$	HM$_{btsc}$	L_2E_{AIC}	L_2E_{SBC}	MHD$_{AIC}$	MHD$_{SBC}$	MHD$_{bt}$	LRT	NN
	$\hat{m}_n$	$\boldsymbol{\theta}_3 = (0.1, 0.45, 0.45; 10, 5, 1)$										
50	1		0.014		0.022	0.010	0.238	0.028	0.114	0.010		
	2	0.890	0.952	0.772	0.910	0.926	0.762	0.958	0.886	0.878	0.626	
	3	**0.010**	**0.034**	**0.214**	**0.058**	**0.064**		**0.014**		**0.108**	**0.370**	
	4			0.012	0.008					0.002	0.004	
	≥5			0.002	0.002					0.002		
100	1					0.004	0.012					
	2	0.800	0.958	0.626	0.892	0.870	0.986	0.950	0.998	0.708	0.414	
	3	**0.200**	**0.042**	**0.354**	**0.104**	**0.126**	**0.002**	**0.050**	**0.002**	**0.290**	**0.568***	
	4			0.016							0.018	
	≥5			0.004	0.004					0.002		
500	1											
	2	0.118	0.988	0.088	0.464	0.550	0.990	0.182	0.830	0.050	0.014	
	3	**0.882***	**0.012**	**0.876***	**0.512***	**0.440**	**0.010**	**0.810***	**0.170**	**0.916***	**0.936***	
	4			0.030	0.022	0.010		0.008		0.022	0.048	
	≥5			0.006	0.002					0.012	0.002	
1000	1											
	2	0.002	1.000		0.138	0.298	0.996	0.030	0.284	0.012		
	3	**0.998***		**0.972***	**0.728***	**0.676***	**0.004**	**0.970***	**0.716***	**0.946***		
	4			0.016	0.112	0.024				0.026		
	≥5			0.012	0.022	0.002				0.016		
5000	1							0.002	0.002			
	2		1.000			0.066	0.400	0.012	0.012	0.008		
	3	**1.000***		**0.966***	**0.630***	**0.624***	**0.600***	**0.974***	**0.978***	**0.928***		
	4			0.020	0.318	0.268		0.012	0.008	0.042		
	≥5			0.014	0.052	0.042				0.022		
10000	1											0.00
	2		1.000			0.046	0.154	0.002	0.002	0.016		0.06
	3	**1.000***		**0.948***	**0.622***	**0.518***	**0.792**	**0.988***	**0.990***	**0.934***		**0.32**
	4			0.030	0.308	0.320	0.054	0.010	0.008	0.012		0.29
	≥5			0.022	0.070	0.116				0.038		0.33
	$\hat{m}_n$	$\boldsymbol{\theta}_3 = (0.34, 0.33, 0.33; 10, 5, 1)$										
50	1							0.134	0.002	0.006		
	2	0.888	0.988	0.876	0.946	0.778	0.862	0.960	0.994	0.744	0.572	
	3	**0.112**	**0.012**	**0.116**	**0.054**	**0.222**	**0.004**	**0.038**		**0.250**	**0.416**	
	4			0.006							0.012	
	≥5			0.002						0.006		
100	1						0.002			0.002		
	2	0.734	0.984	0.790	0.914	0.568	0.986	0.812	0.994	0.434	0.316	
	3	**0.266**	**0.016**	**0.194**	**0.086**	**0.430**	**0.012**	**0.188**	**0.006**	**0.552***	**0.666***	
	4			0.016		0.002				0.006	0.018	
	≥5									0.006		
500	1						0.002			0.002		
	2	0.026	1.000	0.408	0.656	0.050	0.658	0.044	0.476	0.026		
	3	0.974		**0.560***	**0.292**	**0.898***	**0.340**	**0.950***	**0.520***	**0.922***		
	4			0.018	0.040	0.050		0.006	0.004	0.036		
	≥5			0.014	0.012	0.002				0.014		
1000	1											
	2		1.000	0.130	0.258	0.038	0.150	0.016	0.036	0.008		
	3	**1.000***		**0.842***	**0.528***	**0.930***	**0.850***	**0.968***	**0.952***	**0.938***		
	4			0.024	0.180	0.032		0.016	0.012	0.034		
	≥5			0.004	0.034					0.020		
5000	1					0.002	0.002	0.002	0.002	0.002		
	2		1.000		0.002	0.032	0.032	0.022	0.022	0.008		
	3	**1.000***		**0.928***	**0.706***	**0.906***	**0.964***	**0.966***	**0.968***	**0.954***		
	4			0.046	0.264	0.060	0.002	0.010	0.008	0.022		
	≥5			0.026	0.028					0.014		
10000	1											0.00
	2		1.000			0.048	0.048	0.014	0.014	0.016		0.05
	3	**1.000***		**0.938***	**0.776***	**0.890***	**0.944***	**0.970***	**0.976***	**0.920***		**0.30**
	4			0.050	0.188	0.058	0.008	0.016	0.010	0.034		0.28
	≥5			0.012	0.036	0.004				0.030		0.37
	$\hat{m}_n$	$\boldsymbol{\theta}_3 = (0.2, 0.4, 0.4; 3.1, 3, 1)$										
50	1	0.730	0.952	0.474	0.660	0.534	0.960	0.920	0.994	0.616	0.388	
	2	0.270	0.018	0.518	0.334	0.466	0.040	0.080	0.006	0.354	0.586	
	3			**0.008**	**0.006**					**0.022**	**0.026**	
	4									0.004		
	≥5									0.004		
100	1	0.556	0.998	0.204	0.346	0.288	0.870	0.742	0.962	0.476	0.148	
	2	0.444	0.002	0.758	0.604	0.706	0.130	0.258	0.038	0.488	0.802	
	3			**0.038**	**0.032**	**0.006**				**0.014**	**0.050**	
	4				0.018					0.022		
	≥5											
500	1		1.000			0.100	0.190	0.232	0.324	0.236		
	2	1.000		0.938	0.814	0.856	0.808	0.764	0.676	0.726	0.936	
	3			**0.062**	**0.164**	**0.040**	**0.002**	**0.004**		**0.022**	**0.062**	
	4				0.018	0.004				0.016		
	≥5				0.004						0.002	
1000	1		1.000			0.096	0.096	0.170	0.172	0.142		
	2	1.000		0.962	0.786	0.876	0.900	0.826	0.826	0.822	0.954	
	3			**0.036**	**0.186**	**0.026**	**0.004**	**0.004**	**0.002**	**0.022**	**0.044**	
	4			0.002	0.018	0.002				0.014	0.002	
	≥5				0.010							
5000	1		1.000			0.080	0.080	0.024	0.024	0.016		
	2	1.000		0.944	0.878	0.914	0.920	0.974	0.974	0.928		
	3			**0.048**	**0.104**	**0.006**		**0.002**	**0.002**	**0.036**		
	4			0.008	0.010					0.020		
	≥5				0.008							
10000	1		1.000			0.074	0.074	0.010	0.010	0.002		0.42
	2	1.000		0.936	0.904	0.918	0.924	0.984	0.984	0.940		0.45
	3			**0.062**	**0.086**	**0.008**	**0.002**	**0.006**	**0.006**	**0.048**		**0.11**
	4				0.010					0.008		0.02
	≥5			0.002						0.002		0.00

Table 34: Relative frequencies of $\hat{m}_n$ for 4-comp. Poisson mixtures based on 500 replications

n \ Method		BIC	ICL	HM$_{bt}$	HM$_{btsc}$	L_2E_{AIC}	L_2E_{SBC}	MHD$_{AIC}$	MHD$_{SBC}$	MHD$_{bt}$	LRT	NN
	$\hat{m}_n$	$\boldsymbol{\theta}_4 = (0.05, 0.25, 0.3, 0.4; 15, 9, 1, 5)$										
50	1		0.004		0.008	0.016	0.266	0.006	0.062	0.008		
	2	0.638	0.850	0.536	0.864	0.800	0.734	0.946	0.934	0.716	0.288	
	3	0.362	0.146	0.446	0.126	0.182		0.048	0.004	0.266	0.672	
	4			**0.018**	**0.002**	**0.002**				**0.008**	**0.040**	
	≥5									0.002		
100	1						0.014	0.002	0.002			
	2	0.330	0.780	0.316	0.788	0.602	0.984	0.680	0.978	0.320	0.098	
	3	0.670	0.220	0.630	0.210	0.396	0.002	0.310	0.020	0.656	0.814	
	4			**0.054**	**0.002**	**0.002**		**0.008**		**0.016**	**0.088**	
	≥5									0.008		
500	1											
	2		0.758		0.188	0.056	0.840	0.014	0.110	0.006		
	3	0.976	0.242	0.900	0.754	0.882	0.160	0.960	0.890	0.908	0.620	
	4	**0.024**		**0.088**	**0.052**	**0.062**		**0.026**		**0.076**	**0.372**	
	≥5			0.012	0.006					0.010	0.008	
1000	1						0.002			0.036		
	2		0.804		0.024	0.034	0.258	0.010	0.010	0.908		
	3	0.918	0.196	0.880	0.866	0.886	0.740	0.908	0.990	0.046		
	4	**0.082**		**0.100**	**0.096**	**0.078**		**0.082**		**0.010**		
	≥5			0.020	0.014	0.002						
5000	1					0.002	0.002					
	2		0.912			0.026	0.026	0.014	0.014	0.006		
	3	0.152	0.088	0.694	0.922	0.764	0.966	0.250	0.888	0.354		
	4	**0.848***		**0.272**	**0.070**	**0.190**	**0.006**	**0.706***	**0.098**	**0.550***		
	≥5			0.034	0.008	0.018		0.030		0.090		
10000	1							0.002	0.002			0.00
	2		0.984			0.026	0.028	0.016	0.016			0.00
	3		0.016	0.498	0.848	0.718	0.950	0.090	0.454			0.00
	4	**1.000***		**0.462**	**0.128**	**0.240**	**0.022**	**0.740***	**0.528***			**0.16**
	≥5			0.040	0.024	0.016		0.152				0.78
	$\hat{m}_n$	$\boldsymbol{\theta}_4 = (0.2, 0.2, 0.3, 0, 3; 9, 5, 1.2, 1)$										
50	1				0.008	0.008	0.430	0.004	0.058			
	2	0.978	0.998	0.908	0.930	0.896	0.570	0.986	0.942	0.882	0.802	
	3	0.022	0.002	0.078	0.054	0.092		0.008		0.112	0.196	
	4			**0.014**	**0.006**	**0.004**		**0.002**			**0.002**	
	≥5				0.002					0.006		
100	1					0.004	0.052					
	2	0.954	1.000	0.862	0.904	0.844	0.946	0.986	1.000	0.776	0.644	
	3	0.046		0.122	0.086	0.152	0.002	0.014		0.214	0.350	
	4			**0.010**	**0.010**					**0.006**	**0.006**	
	≥5			0.006						0.004		
500	1											
	2	0.680	1.000	0.718	0.924	0.596	0.986	0.666	0.994	0.318	0.176	
	3	0.320		0.250	0.076	0.382	0.014	0.326	0.006	0.666	0.784	
	4			**0.030**		**0.022**		**0.008**		**0.014**	**0.038**	
	≥5			0.002						0.002	0.002	
1000	1					0.004	0.004	0.002	0.002			
	2	0.362	1.000	0.462	0.680	0.360	0.988	0.226	0.936	0.044		
	3	0.638		0.518	0.288	0.588	0.008	0.766	0.062	0.928		
	4			**0.008**	**0.026**	**0.048**		**0.006**		**0.012**		
	≥5			0.012	0.006					0.016		
5000	1					0.002	0.002					
	2		1.000	0.004	0.012	0.154	0.548	0.026	0.036	0.024		
	3	1.000		0.948	0.688	0.582	0.448	0.948	0.960	0.938		
	4			**0.026**	**0.260**	**0.224**	**0.002**	**0.026**	**0.004**	**0.028**		
	≥5			0.022	0.040	0.038				0.010		
10000	1											0.00
	2		1.000			0.126	0.192	0.016	0.018	0.032		0.23
	3	1.000		0.946	0.692	0.536	0.790	0.946	0.978	0.942		0.37
	4			**0.026**	**0.274**	**0.272**	**0.018**	**0.038**	**0.004**	**0.008**		**0.23**
	≥5			0.028	0.034	0.066				0.018		0.17
	$\hat{m}_n$	$\boldsymbol{\theta}_4 = (0.25, 0.25, 0.25, 0, 25; 15, 10, 5, 1)$										
n=50	1					0.002	0.204	0.004	0.004	0.002		
	2	0.438	0.800	0.712	0.924	0.428	0.786	0.650	0.926	0.464	0.168	
	3	0.562	0.200	0.258	0.074	0.562	0.010	0.342	0.070	0.510	0.800	
	4			**0.030**	**0.002**	**0.008**		**0.004**		**0.016**	**0.032**	
	≥5									0.008		
n=100	1						0.006	0.002	0.002	0.006		
	2	0.106	0.788	0.518	0.818	0.120	0.880	0.218	0.730	0.284	0.024	
	3	0.888	0.212	0.440	0.180	0.850	0.114	0.778	0.266	0.682	0.864	
	4	**0.006**		**0.030**	**0.002**	**0.030**		**0.002**	**0.002**	**0.016**	**0.112**	
	≥5			0.012						0.012		
n=500	1											
	2		0.740	0.012	0.082	0.018	0.022	0.018	0.018	0.002		
	3	0.970	0.260	0.912	0.712	0.796	0.978	0.926	0.976	0.858	0.566	
	4	**0.030**		**0.064**	**0.166**	**0.184**		**0.056**	**0.006**	**0.128**	**0.420**	
	≥5			0.012	0.040	0.002				0.012	0.014	
n=1000	1					0.002	0.002	0.002	0.002	0.002		
	2		0.738		0.006	0.018	0.018	0.016	0.016	0.006		
	3	0.912	0.262	0.902	0.758	0.632	0.980	0.820	0.980	0.734	0.348	
	4	**0.088**		**0.078**	**0.220**	**0.340**		**0.162**	**0.002**	**0.228**	**0.636***	
	≥5			0.020	0.016	0.008				0.030	0.016	
n=5000	1											
	2		0.782			0.012	0.012	0.012	0.012	0.008		
	3	0.084	0.218	0.820	0.926	0.198	0.964	0.130	0.834	0.436		
	4	**0.916***		**0.154**	**0.062**	**0.650***	**0.024**	**0.822***	**0.154**	**0.514***		
	≥5			0.026	0.012	0.140		0.036		0.042		
n=10000	1											0.00
	2		0.836			0.012	0.012	0.008	0.008			0.00
	3	0.002	0.164	0.758	0.928	0.086	0.784	0.054	0.298			0.08
	4	**0.998***		**0.218**	**0.044**	**0.640***	**0.204**	**0.762***	**0.694***			**0.17**
	≥5			0.024	0.028	0.262		0.176				0.75

Appendix C

Relevant theoretical results for Section 4

Lemma 0.1. *Let $\mathcal{F} = \{f_{\boldsymbol{\phi}_d}, \boldsymbol{\phi}_d \in \Phi_d\}$ be a family of densities with respect to some dominating measure μ such that $\mathcal{F}_m$ is identifiable for any integer $m \geq 1$. For a given $f_0 = f_{\boldsymbol{\theta}_0} \in \mathcal{F}_{m_0}$, let us denote by $\widehat{f}_m$ its L_2-projection on $\mathcal{F}_m$; i.e.,*

$$\int (\widehat{f}_m(x) - f_0(x))^2 d\mu(x) = \min_{f \in \mathcal{F}_m} (f(x) - f_0(x))^2 d\mu(x)$$

(assumed to exist). Then, for all $m < m_0$ it holds true that

$$\int (\widehat{f}_{m+1}(x) - f_0(x))^2 d\mu(x) < \int (\widehat{f}_m(x) - f_0(x))^2 d\mu(x)$$

Proof of Lemma 0.1. Since $f_0 \in \mathcal{F}_{m_0}$, then we can write

$$f_0 = \pi_1 f_{\boldsymbol{\phi}_{d,1}} + \ldots + \pi_{m_0} f_{\boldsymbol{\phi}_{d,m_0}}$$

for $0 < \pi_1, \ldots, 0 < \pi_{m_0} \in \mathcal{S}_{m_0-1}$ and $\boldsymbol{\phi}_{d,j} \in \Phi_d$ for $j = 1, \ldots, m_0$.

As done in [4], we assume that for some $m < m_0$

$$\int (\widehat{f}_{m+1}(x) - f_0(x))^2 s\mu(x) = \int (\widehat{f}_m(x) - f_0(x))^2 s\mu(x).$$

Then, $\widehat{f}_m \in \mathcal{F}_{m+1}$ is also a minimizer of $f \mapsto \int (f(x) - f_0(x))^2 d\mu(x)$ over $\mathcal{F}_{m+1}$. Now, for any $\epsilon \in (0,1)$ consider the density in $\mathcal{F}_{m+1}$.

$$f_{\epsilon,j} = \epsilon f_{\boldsymbol{\phi}_{d,j}} + (1-\epsilon)\widehat{f}_m.$$

Then, the minimizer $\widehat{f}_m$ has to satisfy

$$\begin{aligned} \int (\widehat{f}_m - f_0(x))^2 d\mu(x) &\leq \int (f_{\epsilon,j} - f_0(x))^2 d\mu(x) \\ &= \int \Big(\epsilon f_{\boldsymbol{\phi}_{d,j}}(x) + (1-\epsilon)\widehat{f}_m(x) - f_0(x)\Big)^2 d\mu(x) \end{aligned}$$

which implies after taking the difference, dividing by ϵ and taking the limit as $\epsilon \searrow 0$

$$\int \widehat{f}_m(x)(\widehat{f}_m(x) - f_0(x)) d\mu(x) \leq \int f_{\boldsymbol{\phi}_{d,j}}(x)(\widehat{f}_m(x) - f_0(x)) d\mu(x)$$

which in turn yields

$$\int \widehat{f}_m(x)(\widehat{f}_m(x) - f_0(x)) d\mu(x) \leq \int f_0(x)(\widehat{f}_m(x) - f_0(x)) d\mu(x)$$

after multiplying with ϕ_j and summing up over $j = 1, \ldots, m_0$. But the latter inequality is equivalent to

$$\int (\widehat{f}_m(x) - f_0(x))^2 d\mu(x) \leq 0$$

and hence $f_0 = \widehat{f}_m$ μ-everywhere, which is impossible since $f_0 \in \mathcal{F}_{m_0}$ and $m < m_0$. $\square$

Consistency and choice of the penalty: a missing assumption

Let us focus on the case where $\mathcal{D}$ is the L_2-distance and the sample space is $\mathcal{X} = \{0, 1, \ldots\}$ as in the setting considered by [2]. Let us denote $\widehat{f}_m^{L_2}$ the L_2-projection of the empirical probability mass function, $\hat{f}_n$, on the class $\mathcal{F}_m$ denoted above by $\widehat{f}_m^{\mathcal{D}}$ in (4.4).By definition, we have for all $m \geq m_0$

$$L_2^2(\hat{f}_n, \widehat{f}_m^{L_2}) \leq L_2^2(\hat{f}_n, f_0) = \sum_{x=0}^{\infty} (f_0(x) - \hat{f}_n(x))^2.$$

It can be shown that the right side is $O_p(n^{-1})$. In fact, it is a classical result that the process $\{\sqrt{n}(\hat{f}_n(x)-f_0(x)), x=0,1,\ldots,\}$ converges weakly in the Hilbert space $\{(x_k)_{k\geq 0} : \sum_{k=0}^{\infty} x_k^2 < \infty\}$ (equipped with the L_2-norm) to a well-defined centered Gaussian process; see e.g. Example 1.8.5 in [3] on Central Limit theorems for the mean of i.i.d. centered random variables taking their values in some Hilbert space. As mentioned above, [2] state that the minimum L_2 estimator of the complexity is strongly consistent under the sole condition that $\lim_{n\to\infty} \alpha_{n,m} = 0$. We argue that this is not enough. Recall that

$$\alpha_{n,m} = b_n(v_{m+1} - v_m)$$

and assume for convenience that $m_0 \geq 2$. If $m < m_0 - 1$, then under some regularity assumptions it can be shown that

$$L_2(\widehat{f}_m^{L_2}, f_m^{L_2}) \to 0 \ \text{ a.s}$$

where $f_m^{L_2}$ is defined in (4.5)when $\mathcal{D}$ is the L_2-distance. Note that $f_m^{L_2} \neq f_0$ from which we have that

$$n\left(L_2^2(\hat{f}_n, \widehat{f}_m^{L_2}) - L_2^2(\hat{f}_n, \widehat{f}_{m+1}^{L_2})\right) \approx n\left(L_2^2(f_0, f_m^{L_2}) - L_2^2(f_0, f_{m+1}^{L_2})\right) \to \infty \tag{0.3}$$

with probability 1, for n large enough since $L_2(f_0, f_m^{L_2}) - L_2(f_0, f_{m+1}^{L_2}) > 0$ by (4.7)and the fact that $m < m_0 - 1$. When $m \geq m_0$, it follows from the discussion above that

$$L_2^2(\widehat{f}_m^{L_2}, \hat{f}_n) = O_p(n^{-1}).$$

The latter convergence can be made even more refined for $m = m_0$ by showing that

$$n\left(L_2^2(\hat{f}_n, \widehat{f}_{m_0}^{L_2}) - L_2^2(\hat{f}_n, \widehat{f}_{m_0+1}^{L_2})\right) \to_d Y \tag{0.4}$$

where $Y = \max(0, Z)^2$ for some non-degenerate random variable Z. The proof of this claimed weak convergence is rather involved and will be the subject of another work. Similar results have been proved for the MLE in the context of mixture models where Z is a standard Gaussian, which means that Y is a mixture of a Dirac at 0 and $\chi^2_{(1)}$; see for instance the example treated by [1] in their Case 5. Estimation in mixture models is part of many other non-standard settings where the classical asymptotic theory for M-estimators cannot be directly used. The main issue is that with a non-trivial probability, an M-estimator gets closer to the true parameter while the latter is on the boundary of the parameter space. This phenomenon emanates from the already mentioned nesting $\mathcal{F}_m \subset \mathcal{F}_{m+1}$, which implies that the mixture representation is not unique. We can go from the most economic representation (corresponding to the true complexity m_0) to a less economic one by putting the irrelevant mixture probabilities to 0, which explains why the true parameter is on the boundary. We will come back to this in some more detail in Section 5on the approach based on the Likelihood Ratio Statistic.

To show that the condition $\lim_{n\to\infty} \alpha_{n,m} = 0$ alone is not enough for ensuring consistency, let us take $b_n = n^{-\gamma}$ with some $\gamma > 1$ and $v_m = m$. Then, $n\alpha_{m,n} = n^{1-\gamma} \searrow 0$ and we have for all $m \leq m_0 - 1$,

$$\begin{aligned} n\left(L_2^2(\widehat{f}_m^{L_2}, \hat{f}_n) - L_2^2(\widehat{f}_{m+1}^{L_2}, \hat{f}_n)\right) &\approx n\left(L_2^2(\widehat{f}_m^{L_2}, f_0) - L_2^2(\widehat{f}_{m+1}^{L_2}, f_0)\right) \\ &\to \infty, \ \text{ by (0.3)}. \end{aligned}$$

Now for $m = m_0$

$$n\left(L_2^2(\widehat{f}_{m_0}^{L_2}, \hat{f}_n) - L_2^2(\widehat{f}_{m_0+1}^{L_2}, \hat{f}_n)\right) >> n^{1-\gamma}$$

with some positive probability since by (0.4) the left side converges weakly to a non-degenerate random variable. Thus, with some positive probability, the true m_0 will not be selected as it fails to satisfy

$$L_2^2(\widehat{f}_{m_0}^{L_2}, \hat{f}_n) \leq L_2^2(\widehat{f}_{m_0+1}^{L_2}, \hat{f}_n) + \alpha_{n,m_0} \tag{0.5}$$

while $L_2^2(\widehat{f}_m^{L_2}, \hat{f}_n) > L_2^2(\widehat{f}_{m+1}^{L_2}, \hat{f}_n) + \alpha_{n,m_0}$ for all $m < m_0$. Hence, for such a choice of the threshold consistency does not hold. The picture would be very different had we chosen for example $b_n = \log n$. In that case, m_0 will be indeed the first integer satisfying (0.5) with probability tending 1 as $n \to \infty$. Thus, the missing assumption in [2] is

$$n\alpha_{n,m} \to \infty \tag{0.6}$$

which is required in addition to $\lim_{n\to\infty} \alpha_{n,m} = 0$. Note that this additional assumption is also missing in the consistency theorem given in [5] for count data as well as in Theorem 1 in [4] where the observations are assumed to have an absolutely continuous distribution. In the latter paper, we believe that (0.6) should be replaced by

$$r_n^2 \alpha_{n,m} \to \infty$$

with r_n the rate of convergence of $\mathcal{D}(\widehat{f}^D_{m_0}, \hat{f}_n)$ to 0 while searching for the first m such that

$$\mathcal{D}^2(\widehat{f}^{\mathcal{D}}_m, \hat{f}_n) \le \mathcal{D}^2(\widehat{f}^{\mathcal{D}}_{m+1}, \hat{f}_n) + \alpha_{n,m}$$

where $\widehat{f}^{\mathcal{D}}_m$ is equal to $\widehat{f}^{\mathcal{H}}_m$ in the case $\mathcal{D}$ is the Hellinger distance. The rate r_n depends of course on the bandwidth c_n, the properties of the kernel used in the definition of the kernel estimator $\hat{f}_n$ and smoothness of the true density.

Table 35: Proportion of the time the penalized minimum L_2-distance is equal to $m_0 = 2$ in the example of the finite mixture of geometric probability mass functions given in (3.5).The results are reported for the thresholds $\alpha_{n,m} = \log n/n$ and $1/n^2$. The proportion is computed on the basis of 100 independent replications.

n	100	1000	10000
$\alpha_{n,m} = \log n/n$	0	0.47	1
$\alpha_{n,m} = 1/n^2$	0.82	0.63	0.68

To illustrate the issue mentioned above, let us again consider the geometric mixture with $m_0 = 2$ introduced in (3.5)and focus on the minimum L_2-distance estimator of m_0 with two different thresholds: thresholds $\alpha_{n,m} = \log n/n$ and $\alpha_{n,m} = 1/n^2$. Note that the first threshold satisfies the missing condition $\lim_{n\to\infty} n\alpha_{n,m} = \infty$ while the second one does not. In Table 35, we report the frequency of exact recovery of m_0 with these different thresholds for the samples sizes $n \in \{100, 1000, 10000\}$ on the basis of 100 replications. While consistency is validated with the first threshold, it is clear that this is not the case with the second one although the frequency seems to be quite high for $n = 100$.

As for the first two approaches based on the determinant of the Hankel matrix of moments described in Section 4, the performance of the method described above is very sensitive to the choice of the threshold. This is certainly undesirable unless one resorts to some way of selecting the best threshold among several candidates. Again, cross-validation could be used in this context but, as already mentioned above, this out of the scope of this survey. We will return to the examples with the 3-component Gaussian and 2-component geometric mixtures in the next section, which treats a modification based on bootstrapping from the fitted mixtures.

Appendix D

Pseudoalgorithms for the methods considered in the paper

Algorithm 1 Estimation of the mixture complexity using the Hankel matrix of moments of the mixing distribution

inputs: $\boldsymbol{X}$ = sample, m_{max} = upper bound for the order
$n = \text{count}(\boldsymbol{X})$ ▷ compute the sample size

procedure ESTIMATE$(m, \text{data} = \boldsymbol{X})$
    for $j = 1$ **to** $2m$ **do**
        compute $\hat{c}_{j,n}$
    end
    compute $H(\hat{\boldsymbol{c}}_{2m,n}) = H(\hat{c}_{1,n}, \ldots, \hat{c}_{2m,n})$ ▷ compute the Hankel matrix
    compute $\hat{D}_m = \det(H(\hat{\boldsymbol{c}}_{m,n}))$ ▷ compute the determinant
    compute $J(m, n) = |\hat{D}_m| + \frac{\ln(n)}{\sqrt{n}} m$ ▷ compute the value of the objective function
    return J
end procedure

procedure ORDER$(m_{max}, \boldsymbol{X})$
    for $m = 1$ **to** m_{max} **do**
        $\boldsymbol{J}(m, n) =$ **procedure** ESTIMATE$(m, \text{data} = \boldsymbol{X})$ ▷ a vector of objective function values for each m
    end
    return $\hat{m} = \arg\min_{1 \leq m \leq m_{max}} \{\boldsymbol{J}(m)\}$
end procedure

Algorithm 2 Estimation of mixture complexity using Hankel matrix of moments of the mixing distribution with scaling

inputs: $\boldsymbol{X}$ = sample, m_{max} = upper bound for the order, $B = \#$ generated samples
$n = \text{count}(\boldsymbol{X})$ ▷ compute the sample size

procedure ESTIMATE(m, data)
  for $j = 1$ **to** $2m$ **do**
    compute $\hat{c}_{j,n}$
  end
  compute $H(\hat{\boldsymbol{c}}_{2m,n}) = H(\hat{c}_{1,n}, \dots, \hat{c}_{2m,n})$ ▷ compute the Hankel matrix
  compute $\hat{D}_m = \det(H(\hat{\boldsymbol{c}}_{2m,n}))$ ▷ compute the determinant
  return $\hat{D}_m$
end procedure ▷ returns the value of the Hankel matrix determinant for given data and m

procedure SCALE(m, B)
  generate B n-samples by resampling with replacement from $\boldsymbol{X}$: $\{\boldsymbol{X}_k^*\}_{1 \le k \le B}$
  for $k = 1$ **to** B **do**
    $\hat{\boldsymbol{D}}^* = \hat{D}^*_{m,k} =$ **procedure** ESTIMATE($m, \boldsymbol{X}_k^*$)
  end ▷ obtain asequence of determinants $\{\hat{D}^*_{m,k}\}_{1 \le k \le B}$
  compute $\hat{\Sigma}_m = \text{cov}\left(\hat{\boldsymbol{D}}^*\right)$
  return $\hat{\Sigma}_m$
end procedure

procedure ORDER($m_{max}, \boldsymbol{X}$)
  for $m = 1$ **to** m_{max} **do**
    compute $\hat{D}_m =$ **procedure** ESTIMATE(m, data $= \boldsymbol{X}$)
    compute $\hat{\Sigma}_m =$ **procedure** SCALE(m, B)
    $\boldsymbol{J}(m) = \sqrt{n} \cdot \hat{\Sigma}_m^{-\frac{1}{2}} |\hat{D}_m| + m \ln n$ ▷ a sequence of obj. function values for each m
  end
  return $\hat{m} = \arg\min_{1 \le m \le m_{max}} \{\boldsymbol{J}(m)\}$
end procedure

Algorithm 3 Estimation of mixture complexity using Hankel matrices of moments of mixing distributions with ML parameter estimation

inputs: α = significance level, $\boldsymbol{X}$ = sample, B = number of generated samples
initialize: $m \leftarrow 1$ ▷ start with 1-component mixture
$n = \text{count}(\boldsymbol{X})$ ▷ compute the sample size

procedure ESTIMATE(m, data)
  for $j = 1$ **to** $2m$ **do**
    compute $\hat{c}_{j,n}$
  end
  compute $H(\hat{\boldsymbol{c}}_{2m,n}) = H(\hat{c}_{1,n}, \ldots, \hat{c}_{2m,n})$ ▷ compute the Hankel matrix
  compute $\hat{D}_m = \det(H(\hat{\boldsymbol{c}}_{2m,n}))$ ▷ compute the determinant
  return $\hat{D}_m$
end procedure ▷ returns the value of the Hankel matrix determinant for given data and m

procedure QUANTILES(m, B, α)
  compute $\hat{\boldsymbol{\theta}}_m = \arg\min_{\boldsymbol{\theta}_m \in \Theta} \big\{ -\sum_{i=1}^{n} \ln f_{\boldsymbol{\theta}_m}(X_{i,n}) \big\}$
  estimate $f_{\hat{\theta}_m}$ ▷ estimate the mixture based on the MLE for m
  generate B n-samples from $f_{\hat{\theta}_m}$: $\{\boldsymbol{X}_k^*\}_{1 \le k \le B}$
  for $k = 1$ **to** B **do**
    $\hat{\boldsymbol{D}}^* = \{\hat{D}^*_{m,k}\}$ = **procedure** ESTIMATE($m, \boldsymbol{X}_k^*$)
  end ▷ this loop results in a distribution of determinants $\{\hat{D}^*_{m,k}\}_{1 \le k \le B}$
  estimate $\hat{F}_B(d) = B^{-1} \sum_{k=1}^{B} \mathbb{I}(\hat{D}^{*\,k}_{m,n} \le d)$
  compute $q_l = \inf \big\{ d : \frac{\alpha}{2} \le \hat{F}_B(d) \big\}$ ▷ compute the $(\frac{\alpha}{2})$-quantile of $\hat{F}_B$
  compute $q_u = \inf\{ d : \big(1 - \frac{\alpha}{2}\big) \le \hat{F}_B(d) \}$ ▷ compute the $(1 - \frac{\alpha}{2})$-quantile of $\hat{F}_B$
  return $\big(I_m = [q_l, q_u]\big)$ ▷ returns the $(1-\alpha)$-confidence interval
end procedure

procedure REJECT($B, \alpha, \boldsymbol{X}$)
  $\hat{D}_m$ = **procedure** ESTIMATE($m, \boldsymbol{X}$)
  I_m = **procedure** QUANTILES(m, B, α) ▷ compute the starting values for $\hat{D}_m$ and I_m
  while $\hat{D}_m \notin I_m$ **do**
    $m \leftarrow m + 1$
    $\hat{D}_m \leftarrow$ **procedure** ESTIMATE($m, \boldsymbol{X}$) ▷ update the determinant value based on $\boldsymbol{X}$
    $I_m \leftarrow$ **procedure** QUANTILES(m, B, α) ▷ update the quantiles
  end
  return $(m, \hat{\boldsymbol{\theta}}_m)$
end procedure

Algorithm 4 Estimation of the mixture complexity using minimal L_2 distance between distributions

inputs: $\boldsymbol{X}$ = sample
$n = \text{count}(\boldsymbol{X})$ ▷ compute the sample size

procedure OPTIMIZE(m, data $= \boldsymbol{X}$)
    put $L_2^2(\hat{f}_n, f_{\theta_m}) = \sum_{x=0}^{\infty} \left(f_{\theta_m}(x) - \hat{f}_n(x)\right)^2$
    compute $\hat{\theta}_m = \arg\min_{\theta_m \in \Theta_m} \{L_2^2(\hat{f}_n, f_{\theta_m})\}$ s.t. $\sum_{j=1}^{m} \pi_j = 1, \pi_j \geq 0$ (+ constraints on ϕ_j if any)
    return $\hat{\theta}_m$ ▷ compute the estimate $\hat{\theta}_m$ for the assumed order of the model
end procedure

$m \leftarrow 1$ ▷ initialize the values
compute $L_2^2(\hat{f}_n, f_{\hat{\theta}_{m-1,n}})$, $L_2^2(\hat{f}_n, f_{\hat{\theta}_m})$, $\alpha_{m-1}(n) \leftarrow 0$
procedure REJECT($\boldsymbol{X}$)
    while $L_2^2(\hat{f}_n, f_{\hat{\theta}_m}) - L_2^2(\hat{f}_n, f_{\hat{\theta}_{m-1}}) \geq \alpha_{m-1}(n)$ **do**
        $m \leftarrow m + 1$
        $\hat{\theta}_{m-1} \leftarrow \hat{\theta}_m$ ▷ reassign the values from the previous step
        $\hat{\theta}_m \leftarrow$ **procedure** OPTIMIZE($m, \boldsymbol{X}$)
        compute $L_2^2(\hat{f}_n, f_{\hat{\theta}_m})$
        compute $\alpha_{m-1}(n)$
    end
    return $(m-1, \hat{\theta}_{m-1})$ ▷ returns the estimated order as well as parameters
end procedure

Algorithm 5 Estimation of mixture complexity using the minimum L_2 distance and the bootstrap procedure

inputs: α = significance level, $\boldsymbol{X}$ = sample, B = number of generated samples
initialize: $m \leftarrow 1$ ▷ start with 1-component mixture
$n = \text{count}(\boldsymbol{X})$ ▷ compute the sample size

procedure ESTIMATE(m, data)
  put $L_2^2(\hat{f}_n, f_{\theta_m}) = \sum_{x=0}^{\infty} \left(f_{\theta_m}(x) - \hat{f}_n(x)\right)^2$
  compute $\hat{\theta}_m = \arg\min_{\theta_m \in \Theta_m} \{L_2^2(f_{\theta_m}, \hat{f}_n)\}$
  compute $\hat{\theta}_{m+1} = \arg\min_{\theta_{m+1} \in \Theta_{m+1}} \{L_2^2(f_{\theta_{m+1}}, \hat{f}_n)\}$ ▷ compute the parameter estimates for m, $m+1$
  $\hat{\delta}_m = L_2^2(f_{\hat{\theta}_m}, \hat{f}_n) - L_2^2(f_{\hat{\theta}_{m+1}}, \hat{f}_n)$
  return $(\hat{\delta}_m, \hat{\theta}_m, \hat{\theta}_{m+1})$
end procedure ▷ the difference between squared L_2 distances for m and $m+1$ and the parameter vectors for mixtures with m and $m+1$ components

procedure QUANTILES(m, B, α)
  estimate $f_{\hat{\theta}_m}$ ▷ estimate the respective mixture density
  generate B n-samples from $f_{\hat{\theta}_m}$: $\{\boldsymbol{X}_k^*\}_{1\leq k\leq B}$
  for $k = 1$ **to** B **do**
    $\hat{\boldsymbol{\Delta}}^* = \{\delta_{m,k}^*\}$ = **procedure** ESTIMATE($m, \boldsymbol{X}_k^*$)
  end ▷ distribution of differences $\hat{\boldsymbol{\Delta}}^* = \{\hat{\delta}_{m,k}^*\}_{1\leq k\leq B}$
  estimate $\hat{F}_B(\delta) = B^{-1}\sum_{k=1}^{B} \mathbb{I}(\hat{\delta}_{m,k}^* \leq \delta)$
  compute $q_l = \inf\left\{\delta : \frac{\alpha}{2} \leq \hat{F}_B(\delta)\right\}$ ▷ compute the $(\frac{\alpha}{2})$-quantile of $\hat{F}_B$
  compute $q_u = \inf\{\delta : \left(1 - \frac{\alpha}{2}\right) \leq \hat{F}_B(\delta)\}$ ▷ compute the $(1 - \frac{\alpha}{2})$-quantile of $\hat{F}_B$
  return $Q_m = [q_l, q_u]$ ▷ returns the $(1-\alpha)$-confidence interval
end procedure

procedure REJECT($B, \alpha, \boldsymbol{X}$)
  $\hat{\delta}_m$ = **procedure** ESTIMATE($m, \boldsymbol{X}$)
  Q_m = **procedure** QUANTILES(m, B, α) ▷ compute starting values for $\hat{\delta}_m$ and Q_m
  while $\hat{\delta}_m \in Q_m$ **do**
    $m \leftarrow m + 1$
    $\hat{\delta}_m \leftarrow$ **procedure** ESTIMATE($m, \boldsymbol{X}$) ▷ update the determinant value based on $\boldsymbol{X}$
    $Q_m \leftarrow$ **procedure** QUANTILES(m, B, α) ▷ update the quantiles
  end
  return $(m, \hat{\theta}_m)$
end procedure

Algorithm 6 Estimation of mixture complexity using the LRTS

inputs: α = significance level, $\boldsymbol{X}$ = sample, B = number of generated samples
initialize: $m \leftarrow 1$ ▷ start with 1-component mixture
$n = \text{count}(\boldsymbol{X})$ ▷ compute the sample size

procedure ESTIMATE(m, data)
  compute $\hat{\theta}_m = \arg\min_{\theta\in\Theta_0}\{-l(f_\theta, \hat{f}_n)\}$
  compute $\hat{\theta}_{m+1} = \arg\min_{\theta\in\Theta_1}\{-l(f_\theta, \hat{f}_n)\}$ ▷ compute the parameter estimates for m, $m+1$
  $\hat{\lambda}_m = -2\ln\frac{\mathcal{L}(\hat{\theta}_m)}{\mathcal{L}(\hat{\theta}_{m+1})}$
  return $(\hat{\lambda}_m, \hat{\theta}_m, \hat{\theta}_{m+1})$
end procedure ▷ return the LRTS for m and $m+1$ components

procedure QUANTILES(m, B, α)
  estimate $f_{\hat{\theta}_m}$ ▷ estimate the respective mixture density
  generate B n-samples from $f_{\hat{\theta}_m}$: $\{\boldsymbol{X}_k^*\}_{1\leq k\leq B}$
  for $k = 1$ **to** B **do**
    $\hat{\boldsymbol{\Lambda}}^* = \{\hat{\lambda}_{m,k}^*\} =$ **procedure** ESTIMATE($m, \boldsymbol{X}_k^*$)
  end ▷ produce a sequence of LRTS $\{\hat{\lambda}_{m,k}^*\}_{1\leq k\leq B}$
  estimate $\hat{F}_B(\lambda) = B^{-1}\sum_{k=1}^{B}\mathbb{I}(\hat{\lambda}_{m,k}^* \leq \lambda)$
  compute $q_m = \inf\left\{\lambda : (1-\alpha) \leq \hat{F}_B(\lambda)\right\}$ ▷ compute the $(1-\alpha)$-quantile of $\hat{F}_B$
  return q_m
end procedure

procedure REJECT($B, \alpha, \boldsymbol{X}$)
  $\hat{\lambda}_m =$ **procedure** ESTIMATE($m, \boldsymbol{X}$)
  $q_m =$ **procedure** QUANTILES(m, B, α) ▷ compute the starting values for $\hat{\lambda}_m$ and q_m
  while $\hat{\lambda}_m \leq q_m$ **do**
    $m \leftarrow m + 1$
    $\hat{\lambda}_m \leftarrow$ **procedure** ESTIMATE($m, \boldsymbol{X}$) ▷ update the LRTS based on $\boldsymbol{X}$
    $q_m \leftarrow$ **procedure** QUANTILES(m, B, α) ▷ update the quantile
  end
  return $(m, \hat{\theta}_m)$
end procedure

References

[1] Steven G Self and Kung-Yee Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610, 1987.

[2] T Umashanger and TN Sriram. L2e estimation of mixture complexity for count data. *Computational Statistics & Data Analysis*, 53(12):4243–4254, 2009. doi: 10.1016/j.csda.2009.05.013.

[3] Aad W Vaart and Jon A Wellner. Weak convergence. In *Weak convergence and empirical processes*, pages 16–28. Springer, 1996.

[4] Mi-Ja Woo and TN Sriram. Robust estimation of mixture complexity. *Journal of the American Statistical Association*, 101(476):1475–1486, 2006. doi: 10.1198/016214506000000555.

[5] Mi-Ja Woo and TN Sriram. Robust estimation of mixture complexity for count data. *Computational Statistics & Data Analysis*, 51(9):4379–4392, 2007. doi: 10.1016/j.csda.2006.06.006.