



Analysis of (cDNA) Microarray Data: Part VI. Bayesmix

A software program for Bayesian analysis of mixture models
with an application to model-based clustering
of microarray gene expression data



Bayesmix

The screenshot shows the Adobe Reader interface displaying the title page of a PDF document. The title is "BAYESMIX: A SOFTWARE PROGRAM FOR BAYESIAN ANALYSIS OF MIXTURE MODELS WITH AN APPLICATION TO MODEL-BASED CLUSTERING OF MICROARRAY GENE EXPRESSION DATA". The authors listed are A. Reverter, K.A. Byrne and B.P. Dalrymple, from CSIRO Livestock Industries, Queensland Bioscience Precinct, 306 Carmody Road, St Lucia, QLD 4067. The document includes a SUMMARY section and an INTRODUCTION section. The Adobe Reader window title is "Adobe Reader - [AAABG03_Bayesmix.pdf]".



Bayesmix

Contents

✘ Introduction:

- ✘ Concept & Analysis possibilities
- ✘ Challenges for microarray

✘ Technical Concerns:

- ✘ Software development
- ✘ Software comparison
- ✘ Final remarks

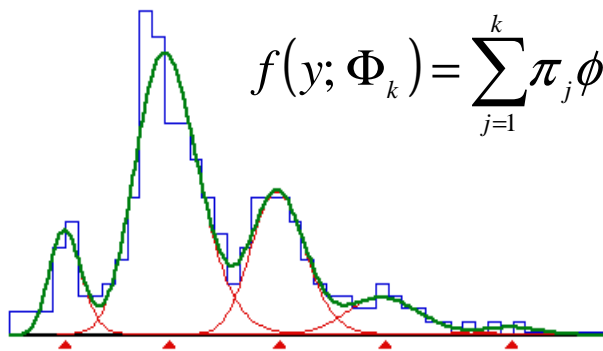
A. Reverter - Sept. 2006, UAB, Barcelona, Spain



Bayesmix

Concept & Analysis Possibilities

$$f(y; \Phi_k) = \sum_{j=1}^k \pi_j \phi(y; \mu_j, V_j)$$



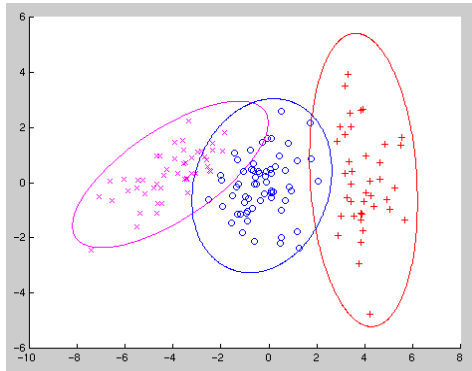
Plot #007 Data: Cassie's Example Components: Lognormal

A. Reverter - Sept. 2006, UAB, Barcelona, Spain



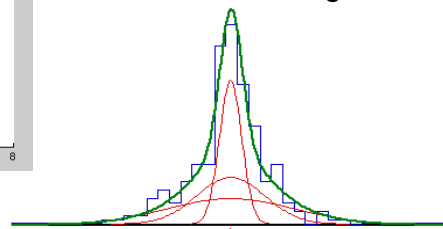
Bayesmix

Challenges for Microarrays



1. Class Comparison
2. Class Discovery

Detecting Outliers



Plot #009 Data: Means Equal Components: Normal

A. Reverter - Sept. 2006, UAB, Barcelona, Spain



Bayesmix

BAYESMIX: Software development

$$f(y; \Phi_k) = \sum_{j=1}^k \pi_j \phi(y; \mu_j, V_j)$$

$$p(a_i = j | \pi, k, \theta, d) \propto \pi_j N(\mu_j, V_j)$$

$$p(\mu_j | \pi, k, u, \mu_{-j}, V_j, d) \propto N((n_j V_j^{-1} + k)^{-1} (n_j V_j^{-1} \bar{d}_j + k \xi_j), (n_j V_j^{-1} + k)^{-1})$$

$$p(V_j | \pi, k, u, \mu, d) \propto \chi^{-2} \left(2\alpha + n_j, \left(2v_j + \sum_{i:d_i=j} (d_i - \mu_j)^2 \right)^{-1} \right)$$

$$p(\pi | k, u, \theta, d) \propto Dir(t + n_1, \dots, t + n_k)$$

- FORTRAN 90
- Up to 5 Components
- Gibbs Sampling
- Chain Length 12,000

A. Reverter - Sept. 2006, UAB, Barcelona, Spain



Bayesmix

BAYESMIX: Software development

$$f(y; \Phi_k) = \sum_{j=1}^k \pi_j \phi(y; \mu_j, V_j)$$

$$AIC = -2 \log L(\hat{\Phi}_k) + 2v_k$$

$$BIC = -2 \log L(\hat{\Phi}_k) + v_k \log(n) \quad v_k = 3k - 1$$

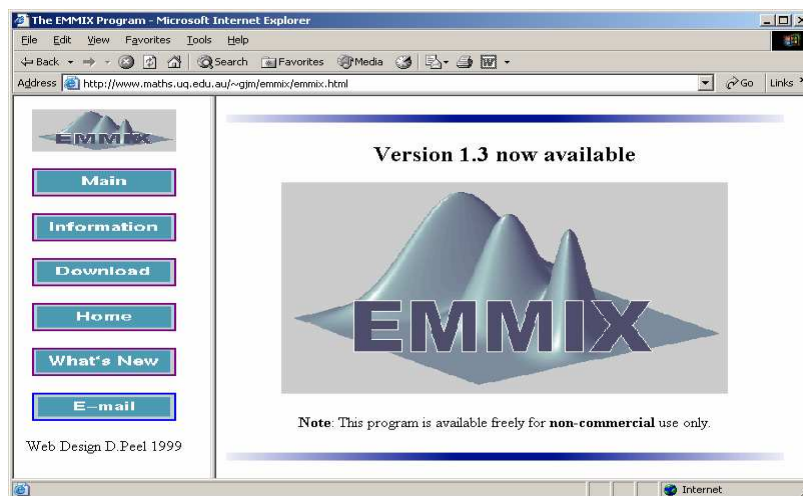
$$\tau_{ij}^{(m)} = \frac{\pi_i^{(m)} \phi(y_j; \mu_i^{(m)}, V_i^{(m)})}{f(y_j; \Phi^{(m)})}$$

A. Reverter - Sept. 2006, UAB, Barcelona, Spain



Bayesmix

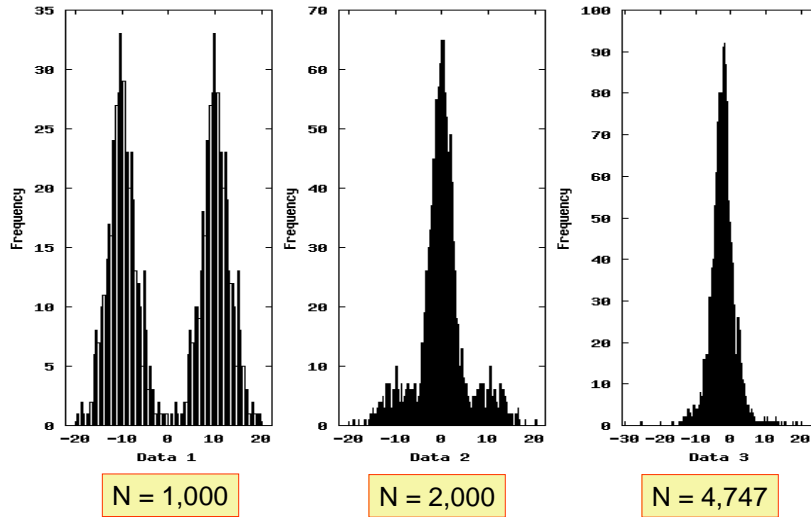
BAYESMIX: Software comparison (EMMIX)



A. Reverter - Sept. 2006, UAB, Barcelona, Spain



BAYESMIX: Software test



A. Reverter - Sept. 2006, UAB, Barcelona, Spain



BAYESMIX: Software test (results)

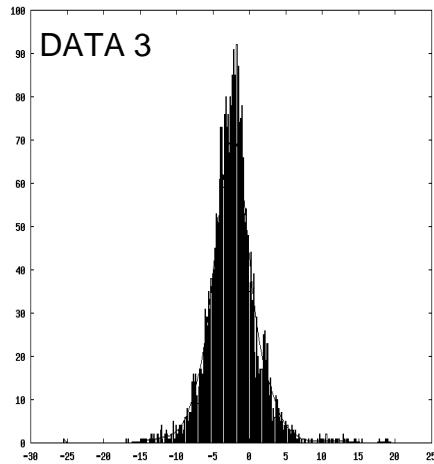
Data	N	Procedure	LogL	Parameter			
				Cluster	%	Mean	Var.
1	1,000	True	-	1	50.0	-10.0	10.0
				2	50.0	10.0	10.0
		ML		1	50.0	-9.97	9.46
				2	50.0	10.02	9.47
		BAYESMIX		1	50.0	-9.61	9.59
				2	50.0	9.66	9.61
2	2,000	True	-	1	10.0	-10.0	10.0
				2	80.0	0.0	5.0
				3	10.0	10.0	10.0
		ML		1	10.5	-9.52	9.93
				2	79.0	0.01	4.55
				3	10.5	9.98	10.34
		BAYESMIX		1	10.3	-8.80	8.93
				2	79.6	0.04	4.62
				3	10.1	9.20	9.50

A. Reverter - Sept. 2006, UAB, Barcelona, Spain



Bayesmix

BAYESMIX: Software test (results)



EMMIX: (logL = -11,864)

$$0.044 \times N(-0.87, 67.46) \\ + 0.590 \times N(-2.30, 10.42) \\ + 0.366 \times N(-2.41, 2.32)$$

BAYESMIX: (logL = -11,944)

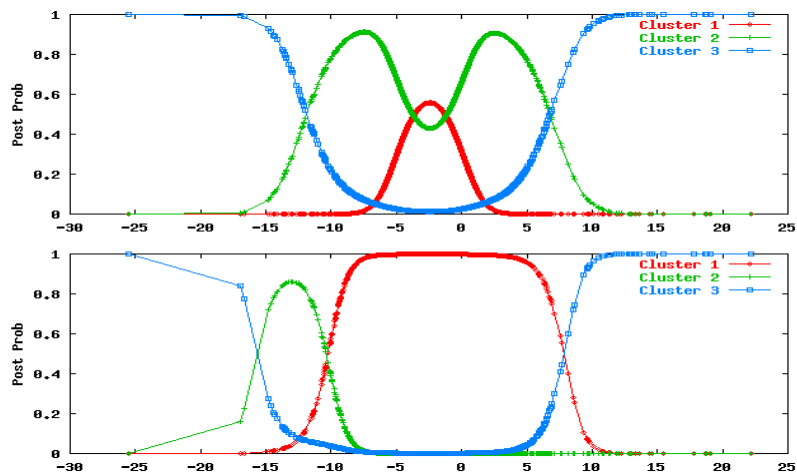
$$0.008 \times N(-1.02, 208.79) \\ + 0.981 \times N(-2.26, 7.61) \\ + 0.011 \times N(-11.18, 3.63)$$

A. Reverter - Sept. 2006, UAB, Barcelona, Spain



Bayesmix

BAYESMIX: Software test (results)



A. Reverter - Sept. 2006, UAB, Barcelona, Spain



BAYESMIX: Software comparison (EMMIX)

		CPU Time (sec.)	
		BAYESMIX	EMMIX
DATA 1	1,000	74	314
DATA 2	2,000	154	2,265
DATA 3	4,747	360	55,359 (15.4hr!)

NB: EMMIX can be modified to make it faster



Conclusions

- ✗ BAYESMIX works
- ✗ Some features require further development:
 - ✗ Flexibility in Chain Length (CODA)
 - ✗ Unknown Number of Components
 - ✗ Multivariate
- ✗ EMMIX is a far more complete software
- ✗ Mixtures have other applications (eg. Selective Genotyping)
- ✗ Both softwares are availableand will be used in this course