It is of interest to compare the results of this Bayesian approach applied to the three real data sets with those obtained working in a frequentist framework. Consequently, we fitted k-component univariate normal mixture models with unrestricted component variances by maximum likelihood via the EM algorithm of Dempster, Laird, and Rubin (1977); see also McLachlan and Krishnan (1997). The choice of the number of components k was made in terms of the increase in the log likelihood, $\log L$, as k was increased sequentially from $k = 1$.

In Table 1, we report for the three real data sets, the value of $\log L$ and its modified value under the criteria of AIC (Akaike's information criterion), and BIC (Bayesian information criterion). As is well known, regularity conditions do not hold for the usual asymptotic justification of these criteria. With the AIC and BIC criteria, the choice of the number of components is that value of k that maximizes $2 \log L - am$, where m is the the number of parameters in the model, and $a = 2$, and $\log n$ for AIC and BIC, respectively. In the case of the test based on the likelihood ratio λ, the P-value was assessed by using the resampling approach as in McLachlan (1987).

		Number of components k					
	Data set	1	2	3	4	5	6
Likelihood	Acidity	–225.79	–184.64	–178.75	–175.76		–170.65
	Enzyme	–230.76	–54.64	–47.83	–40.95	–37.78	–36.02
	Galaxy	–240.42	–220.10	–203.48	–197.71	–190.45	–182.57
AIC	Acidity	-455.57	-379.29	-373.51	-373.53		-375.30
	Enzyme	–465.52	–119.28	–111.65	–103.90	–103.56	–106.03
	Galaxy	–484.84	–450.38	–422.96	–417.42	–408.90	–399.15
BIC	Acidity	–461.66	–394.51	–397.86	–407.01		–427.03
	Enzyme	–472.52	–136.79	–139.66	–142.41	–152.58	–165.55
	Galaxy	–489.65	–462.41	–442.21	–443.89	–442.59	–440.06
$-2 \log \lambda$ (P-value) for k vs k+1	Acidity	82.28	11.78	5.98			
		(0.01)	(0.08)	(0.44)			
	Enzyme	352.24	13.63	13.75	6.33		
		(0.01)	(0.02)	(0.06)	(0.39)		
	Galaxy	40.45	33.42	11.54	14.52	15.75	5.21
		(0.01)	(0.01)	(0.09)	(0.04)	(0.02)	(0.22)

Table 1: Various criteria for the number of components

It can been seen that in a number of cases in Table 1 the assessed P-value lies between 5% and 10%. Performing further bootstrap replications in these cases did not clarify the situation. Also, it should be noted that these assessed P-values should not be interpreted too finely, as it is the bootstrap rather than the actual P-value that is being assessed in these cases since the null value of k is greater than one.

On the basis of these P-values interpreted rigidly at the 5% level of significance, k would be chosen to be equal to 2, 3,and 3, for the acidity, enzyme and galaxy data sets, respectively. If the significance level were increased to 10%, it would result in k being chosen equal to 3, 4, and 6 for the acidity, enzyme and galaxy data sets, respectively. This is in general agreement with the results obtained here in a Bayesian framework for k, except in part for the acidity data set. For this set, the Bayesian approach provides essentially equal support for k = 2-6, whereas according to the P-values assessed by resampling, one would not go beyond $k = 3$ components.

References

McLachlan, G.J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics* **36**, 318–324.

McLachlan, G.J. and Krishnan, T. (1997). *The EM Algorithm and Extensions*. New York: Wiley.