Modelling a Particular Philatelic Mixture using the Mixture Method of Clustering

by

Kaye E. Basford and Geoffrey J. McLachlan

BU-1181-M November 1992

ABSTRACT

Izenman and Sommer (1988) used a nonparametric kernel density estimation technique to fit a seven component model to the paper thickness of the 1872 Hidalgo Stamp Issue of Mexico. They observed an apparent conflict when fitting a three component normal mixture model. This conflict is resolved by further investigation into the determination of the most appropriate group number to represent the data when fitting parametric mixture models. The likelihood ratio test is not always reliable and allocation of the entities to the components in the underlying mixture must be considered. This leads to a seven component, equal variance, model consistent with that found by the nonparametric assessment of multimodality. It illustrates that the finite mixture model can be a useful tool for describing real data such as in the philatelic mixture in question.

MODELLING A PARTICULAR PHILATELIC MIXTURE USING THE MIXTURE LIKELIHOOD METHOD OF CLUSTERING

K.E. BASFORD[1] AND G.J. MCLACHLAN[2]

The University of Queensland

Summary

The mixture likelihood method of clustering is applied to the Philatelic mixture discussed by Izenman and Sommer (1988). A seven component mixture of normal homoscedastic distributions is determined to be an appropriate description of the paper thicknesses of the 1872 Hidalgo Stamp Issue of Mexico. This resolves the apparent conflict of seven modes (from the nonparametric approach to identify components in the density estimate) compared with a mixture of three heteroscedastic normals (from a parametric finite mixture model approach), as observed in the last section of Izenman and Sommer's paper.

Key Words: Mixture likelihood method; Clustering; Postage stamps.

---

[1]Department of Agriculture, The University of Queensland, Queensland 4072, Australia.

[2]Department of Mathematics, The University of Queensland, Queensland 4072, Australia

Acknowledgement. The authors thank Professor J.W. Tukey of Princeton University for his helpful comments and suggestions.

1. Introduction

Izenman and Sommer (1988) recently wrote an article on "Philatelic Mixtures and Multimodal Densities" where the fitting of a mixture of distributions is most appropriate to the mixture of paper types with different thicknesses used for printing a given stamp issue. Their main concern was the application of the nonparametric approach to identify components by the resulting placement of modes in the density estimate. The specific example of a philatelic mixture, the 1872 Hidalgo issue of Mexico, was used as a particularly graphic demonstration of the combination of a statistical investigation and extensive historical data to reach conclusions regarding the mixture components.

For comparison, the parametric finite mixture model was also fitted to the data with Izenman and Sommer using the likelihood ratio test (Wolfe, 1970) to determine the appropriate number of underlying distributions. However, recent work (McLachlan, Basford and Green 1992a,b) has indicated that Wolfe's approximation is not always reliable. Other considerations, such as the component variances and the allocation of the entities to these components in the underlying mixture, should also be taken into account.

If a homoscedastic mixture model is fitted to the Hidalgo data, then a seven component model consistent with that found by the nonparametric assessment of multimodality is preferred. This resolves the conflict of seven modes (from the nonparametric approach) versus a mixture of three normals (from the parametric approach), as observed in the last section of Izenman and Sommer's paper.

2. Mixture Method of Clustering

The mixture method of clustering uses the measurements on a set of elements to partition these elements into g clusters for a specified value of g. Under the mixture model imposed, it is assumed that the data have been sampled from a mixture of g distributions in some unknown proportions. This mixture model can be fitted parametrically, whereby the component distributions are specified up to a manageable number of unknown parameters. The underlying distributions are, in this instance, assumed to be univariate normal. Thus

$$f(x_j) = \sum_{i=1}^{g} \pi_i f_i(x_j) \tag{1}$$

where $f_i(x_j)$ is the normal density function with mean μ_i and variance σ_i^2, π_i $(i=1,...,g)$ represent the mixing proportions, and x_j $(j=1,...,n)$ are the individual data values. In the subsequent section, the notation is abused by writing $f(x)$ as $\sum \pi_i N(\mu_i, \sigma_i^2)$. The unknown parameters, i.e. means, variances and mixing proportions, are estimated by maximum likelihood. The likelihood equation is solved here using the EM algorithm of Dempster, Laird and Rubin (1977). Note that the chosen solution of the likelihood equation is referred to as the maximum likelihood estimate, although in the case of unrestricted component variances the likelihood is unbounded. For a detailed explanation of this approach, the reader is referred to McLachlan and Basford (1988).

The fitting of the finite mixture model (1) provides a probabilistic clustering of the n elements in terms of their posterior probabilities of membership of the individual g components of the mixture of distributions. Thus the estimated probability that x_j (really the element with observation x_j) comes from the i^{th} component of the mixture is given by

$$\tau_i(x_j) = \frac{\pi_i f_i(x_j)}{f(x_j)} \tag{2}$$

for $i=1,...,g$ and $j=1,...,n$. An outright assignment of the data into g nonoverlapping clusters can be effected by assigning each x_j to the component to which it has the highest posterior probability of belonging.

One difficulty with the practical application of this clustering method is that the number, g, of underlying distributions is assumed to be known. As stated by Izenman and Sommer (1988), attempts at formulating criteria to test for the number of components in a parametric mixture of distributions have not been overly successful. The test procedure to receive the most attention in the literature is that of Wolfe (1970, 1971), where minus twice the difference in the log likelihoods for different values of g $(-2log\lambda)$ is taken to be distributed approximately as chi-square with degrees of freedom equal to twice the difference between the number of parameters under the null and alternative hypotheses, excluding the mixing proportions.

Recent empirical evidence (Milligan and Cooper, 1985 and McLachlan, Basford and Green, 1992b) suggests that at least for medium sized samples, Wolfe's approximation tends to overestimate the number of components present in a mixture, although this is not a problem with the Hidalgo stamp data. Many authors (Aitkin and Rubin, 1985; Ghosh and Sen, 1985; Hartigan, 1985; Quinn, McLachlan and Hjort, 1987) have noted that the regularity conditions break down, so asymptotically the null distribution of chi-square does not hold. Even if regularity conditions did hold, it is unlikely that the consequent asymptotic results would be applicable in most practical situations. This is because with mixture models, the sample size has to be extremely large for large sample theory to be of practical relevance. For the two component normal mixture with equal variances, Mendell, Thode and Finch (1991) determined that in order to detect a mean difference of three standard deviations for mixing proportions between 0.2 and 0.8, sample sizes of 40 and 90 were needed to achieve 50% and 90% power, respectively.

Practical considerations are also important if a sensible description or summary of the data is to be obtained. If, for instance, a mixture model is used to describe a univariate multimodal frequency distribution, it may be impractical to have the tail of one distribution (say the one centred on the highest mean) accounting for data at the lower end (below that of the lowest mean) by allowing unequal variances. Then, elements which have the lowest value would be said to belong to the group with the highest mean, rather than to the group to which it is closest. Therefore, it may be more realistic to impose the condition of equal variance for the underlying component distributions.

3. The 1872 Hidalgo Stamp Issue of Mexico

Izenman and Sommer (1988) gave a detailed account of the content and historical data available on the particular philatelic mixture to be discussed; i.e. the 1872 Hidalgo Stamp Issue of Mexico. In accordance with these authors, we consider the stamp-thickness data (which they listed) on the 485 unwatermarked used white wove stamps, of which 289 had an 1872 overprint and 196 had either an 1873 or 1874 overprint. As noted by Izenman and Sommer, there is some clustering around the values 0.07mm, 0.08mm, 0.09mm, 0.10mm, 0.11mm, 0.12mm, and 0.13mm with about half the data between 0.06mm and 0.08mm and the other half spread over the larger interval from 0.08mm to 0.13mm (Figure 1). These two portions essentially correspond to the assertion of bimodality in Wilson (1983), although he had analysed a slightly different data set (see Izenman and Sommer, 1988, for details).

Izenman and Sommer tested for multimodality, as suggested by Silverman (1981), using nonparametric kernel density estimation techniques to determine the most probable number of modes in the underlying density. Silverman's test has been viewed primarily as an exploratory data-analytic technique and window-widths near to the critical one must be investigated for the differing number of modes. Using 100 bootstrap replications and a flexible stopping rule with a nominal P-value of 0.40, Izenman and Sommer found that the stamp thickness data were consistent with an underlying density having seven modes. Using the extensive historical information, plus analysis of some related data (the 1868 issue), they showed that seven modes was a sensible description of these data; i.e. it is plausible that paper of seven different thicknesses was used in the production of this stamp issue.

To compare with the nonparametric mode fitting technique, Izenman and Sommer fitted a mixture of seven normal distributions with unrestricted variances. This produced a completely consistent result in that the parametrically fitted mixture density had modes at almost the same locations as the seven modes previously determined. The estimated mixture distribution was given by

$$0.310\,N(0.0723, 0.00000863) + 0.323\,N(0.0797, 0.00000359)$$
$$+\,0.090\,N(0.0905, 0.00000625) + 0.132\,N(0.1002, 0.00000632)$$
$$+\,0.101\,N(0.1095, 0.00000738) + 0.032\,N(0.1208, 0.00000659)$$
$$+\,0.012\,N(0.1293, 0.00000088).$$

But using Wolfe's test, they determined that the number of normal components in the mixture was three, with the estimated mixture given by

$$0.196\,N(0.0712, 0.00000176) + 0.367\,N(0.0786, 0.00000564)$$
$$+\,0.437\,N(0.0989, 0.00019666).$$

This assessment followed Wolfe (1971) where the smallest value of g compatible with the data is determined by sequentially testing g against $g + 1$ groups, starting from $g = 1$.

The graphs of these density estimates are given (Figures 2 and 3, respectively). For completeness, the values of the log likelihood for $g = 1$ *to* 7 for both unrestricted and equal variances are also presented (Table 1).

In an attempt to reconcile the nonparametric (seven modes) and parametric (three component mixture) approaches, Izenman and Sommer postulated that size effects (the presence of the two major modes at 0.07mm and 0.08mm) might be dominating the result for the parametric mixture model. They suggested that further study of this result was warranted.

Using the historical evidence presented by Izenman and Sommer, there could be some explanation for the three component mixture in terms of paper types. They separated the data from the 1872 consignment from that of the 1873 and 1874 consignment, both of which make up the 1872 Hildago Stamp Issue. This showed that the stamps were printed on two different paper types, one of which conforms to the characteristics of the unwatermarked white wove paper used for the 1868 issue and a second much thicker paper (possibly the unwatermarked portions of the Papel Sellado paper) that disappeared completely by the end of 1872. From the data available on the watermarked portions of the Papel Sellado paper (Izenman and Sommer, 1988), its thickness was quite variable (ranging from 0.81mm to 0.130mm) with a mean of 0.1033mm and variance of 0.00011674mm^2. Unwatermarked portions of this thicker Papel Sellado paper could be largely responsible for the third component with mean 0.0989mm and variance 0.00019666mm^2. However, both types of paper had multimodal thickness features (Izenman and Sommer, 1988).

Given that the aim here is to adequately describe the distribution of the thickness measurements as displayed in the histogram (Figure 1), independent of the paper types, it may be appropriate to look further than the three component mixture.

Consider the influence of the large variance of the third component distribution centred on the highest mean. As a consequence of the long left hand tail of this distribution, the five thinnest stamps (ranging from 0.60mm to 0.66mm) are assigned to this component which has a mean of 0.0989mm and not the nearest one which has a mean of 0.0712mm (Table 2). This would not be accepted as a satisfactory grouping of the stamps based on their thickness. Thus, the impracticability of the three group solution as an adequate description or summary of the data is apparent. This implies that consideration of the log likelihood values may not be sufficient in determining the appropriate number of groups used to summarize univariate data. It is informative to also consider the allocation of the elements to the groups, or more explicitly, the estimated posterior probabilities of group membership (McLachlan and Basford, 1988).

Given the unreliability of Wolfe's approximation for the likelihood ratio test, other methods for determining the appropriate number of underlying groups in the data should be considered. Professor John Tukey suggested that a plot of the log likelihood minus the number of estimated parameters (including the proportions) against the group number could be useful. If the mixing proportions are not included, this is equivalent to using Akaike's Information Criterion (Akaike, 1973) which is minus twice the value presented in Figure 4 (see also Bozdogan, 1986). Unfortunately, this procedure is similar to Wolfe's in that it only considers the likelihood. Using the further information on the allocation of the stamps to the underlying groups, the otherwise satisfactory solutions for unrestricted variances at $g = 3, 4,$ and 5 would be rejected. The severe overlap of distributions causes an impractical allocation of stamps to groups and consequent summary description as demonstrated for $g = 3$ above.

From consideration of this plot and the estimates of the posterior probabilities of group membership, both the $g = 6$ and $g = 7$ solutions for unrestricted and equal variances would appear acceptable. The common value for the log likelihood minus the estimated parameters for the unrestricted and equal variances mixture models for $g = 7$ would appear to further support either of these seven group solutions.

Finally, it seems reasonable to assume that the variability of stamp thickness would be consistent about any particular mean in the production process across the range of measurements observed. Even when the heteroscedastic model was fitted, the estimates of the variances of the normal components did not vary greatly, as evidenced by the equation of the seven component mixture distribution displayed above. Only the group corresponding to the largest mean had a variance different from (substantially smaller than) the others. Given that this component also contributed the smallest proportion (about 0.01) to the overall mixture, this is far from being unexpected. The equal variance model would also appear to be more consistent with the non-parametric approach used by Izenman and Sommer (1988).

Given the above arguments, a seven component normal mixture with equal variances would appear to be the most appropriate representation of the data (Figure 5). The resulting maximum likelihood estimate of the mixture is

$$0.272\,N(0.0716, 0.00000588) + 0.363\,N(0.0792, 0.00000588)$$
$$+ 0.089\,N(0.0907, 0.00000588) + 0.136\,N(0.1003, 0.00000588)$$
$$+ 0.097\,N(0.1096, 0.00000588) + 0.027\,N(0.1202, 0.00000588)$$
$$+ 0.016\,N(0.1285, 0.00000588).$$

The estimates of the posterior probabilities of group membership for this solution (Table 3) indicate a satisfactory allocation of the stamps to groups without any of the overlap apparent in the three component solution.

We believe that this resolves the apparent conflict expressed by Izenman and Sommer and illustrates that the finite mixture model can be a useful tool for describing the philatelic mixture under investigation.

References

AITKIN, M. & RUBIN, D.B. (1985). Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society Series B* **47**, 67-75.

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, B.N. Petrov and F. Csake (Eds.). Budapest: Akademiai Kaido, pp 267-281.

BOZDOGAN, H. (1986). Multi-sample cluster analysis as an alternative to multiple comparison procedures. *Bulletin of Informatics and Cybernetics*, **22**, 95-124.

DEMPSTER, A.P., LAIRD, N.M. & RUBIN, D.N. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society Series B* **39**, 1-38.

GHOSH, J.H. & SEN, P.K. (1985). On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer (Vol. ii)*, L.M. Le Cam and R.A. Olshen (Eds.). Monterey: Wadsworth, pp 789-806.

HARTIGAN J.A. (1985). A failure of likelihood asymptotics for normal mixtures. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer Vol. ii)*, L.M. Le Cam and R.A. Olshen (Eds.). Monterey: Wadsworth, pp 807-810.

IZENMAN, A.J. & SOMMER, C.J. (1988). Philatelic mixtures and multimodal densities. *Journal of the American Statistical Association* **83**, 941-953.

MENDELL, N.R., THODE, H.C. JR. & FINCH, S.J. (1991). The likelihood ratio test for the two-component normal mixture problem: Power and sample size. *Biometrics* **47**, 1143-1148.

MILLIGAN, G.W. & COOPER, M.C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika* **50**, 159-179.

MCLACHLAN, G.J. & BASFORD, K.E. (1988). *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker.

MCLACHLAN, G.J., BASFORD, K.E. & GREEN, M. (1992a). On the null distribution of the likelihood ratio test statistic for a mixture of two multivariate normals. Unpublished manuscript.

MCLACHLAN, G.J., BASFORD, K.E. & GREEN, M. (1992b). On the unreliability of Wolfe's approximation for the number of components in a finite mixture model. Unpublished manuscript.

QUINN, B.G., MCLACHLAN, G.J. & HJORT, N.L. (1987). Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association* **73**, 730-738.

SILVERMAN, B.W. (1981). Density estimation for univariate and bivariate data. In *Interpreting Multivariate Data*, V. Barnett (Ed.). New York: John Wiley, pp 37-53.

WILSON, I.G. (1983). Add a dimension to your philately. *The American Philatelist* **97**, 342-349.

WOLFE, J.H. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research* 5, 329-350.

WOLFE, J.H. (1971). A Monte Carlo study of the sampling distribution of the likelihood ratio for mixtures of multinormal distributions. Technical Bulletin STB 72-2. San Diego: U.S. Naval Personnel and Training Research Laboratory.

TABLE 1

Value of the log likelihood for g = 1 to 7 for the fitted mixture of normal distributions with unrestricted and equal variances to the 1872 Hidalgo Issue of Mexico.

Number of Groups	Unrestricted Variances	Equal Variances
1	1350.3	1350.3
2	1484.8	1442.6
3	1518.8	1475.7
4	1521.9	1487.5
5	1527.4	1489.6
6	1529.3	1512.9
7	1531.3	1525.4

TABLE 2

Estimated posterior probabilities of group membership for the fitted mixture of three normal distributions with unrestricted variances to the 1872 Hidalgo Issue of Mexico.

Thickness	Estimated posterior probability		
(x_j)	$\hat{t}_1(x_j)$	$\hat{t}_2(x_j)$	$\hat{t}_3(x_j)$
.060			1.00
.064			1.00
.065			1.00
.066	.03		.97
.068	.73		.27
.069	.92		.08
.070	.96		.04
.071	.96	.01	.03
.072	.94	.02	.04
.073	.80	.12	.08
.074	.36	.50	.14
.075	.05	.83	.12
.076		.91	.09
.077		.93	.07
.078		.94	.06
.079		.93	.07
.080		.91	.09
.081		.87	.13
.082		.79	.21
.083		.63	.37
.084		.40	.60
.085		.18	.82
.086		.06	.94
.087		.01	.99
.088			1.00
.089 - .112			1.00
.114 - .115			1.00
.117			1.00
.119 - .123			1.00
.125			1.00
.128 - .131			1.00

TABLE 3

Estimated posterior probabilities of group membership for the fitted mixture of seven normal distributions with equal variances to the 1872 Hidalgo Issue of Mexico.

Thickness	Estimated posterior probability						
(x_j)	$\hat{\tau}_1(x_j)$	$\hat{\tau}_2(x_j)$	$\hat{\tau}_3(x_j)$	$\hat{\tau}_4(x_j)$	$\hat{\tau}_5(x_j)$	$\hat{\tau}_6(x_j)$	$\hat{\tau}_7(x_j)$
.060	1.00						
.064 - .066	1.00						
.068 - .071	1.00						
.072	.98	.02					
.073	.94	.06					
.074	.81	.19					
.075	.54	.46					
.076	.24	.76					
.077	.08	.92					
.078	.02	.98					
.079	.01	.99					
.080 - .083		1.00					
.084		.97	.03				
.085		.80	.20				
.086		.36	.64				
.087		.07	.93				
.088		.01	.99				
.089 - .092			1.00				
.093			.98	.02			
.094			.90	.11			
.095			.61	.39			
.096			.23	.77			
.097			.05	.94			
.098			.01	.99			
.099 - .101				1.00			
.102				.99	.01		
.103				.97	.03		
.104				.86	.14		
.105				.56	.44		
.106				.21	.79		
.107				.05	.95		
.108				.01	.99		
.109 - .112					1.00		
.114					.95	.05	
.115					.73	.27	
.117					.07	.93	
.119 - .121						1.00	
.122						.98	.02
.123						.93	.07
.125						.46	.54
.128						.01	.99
.129 - .131							1.00

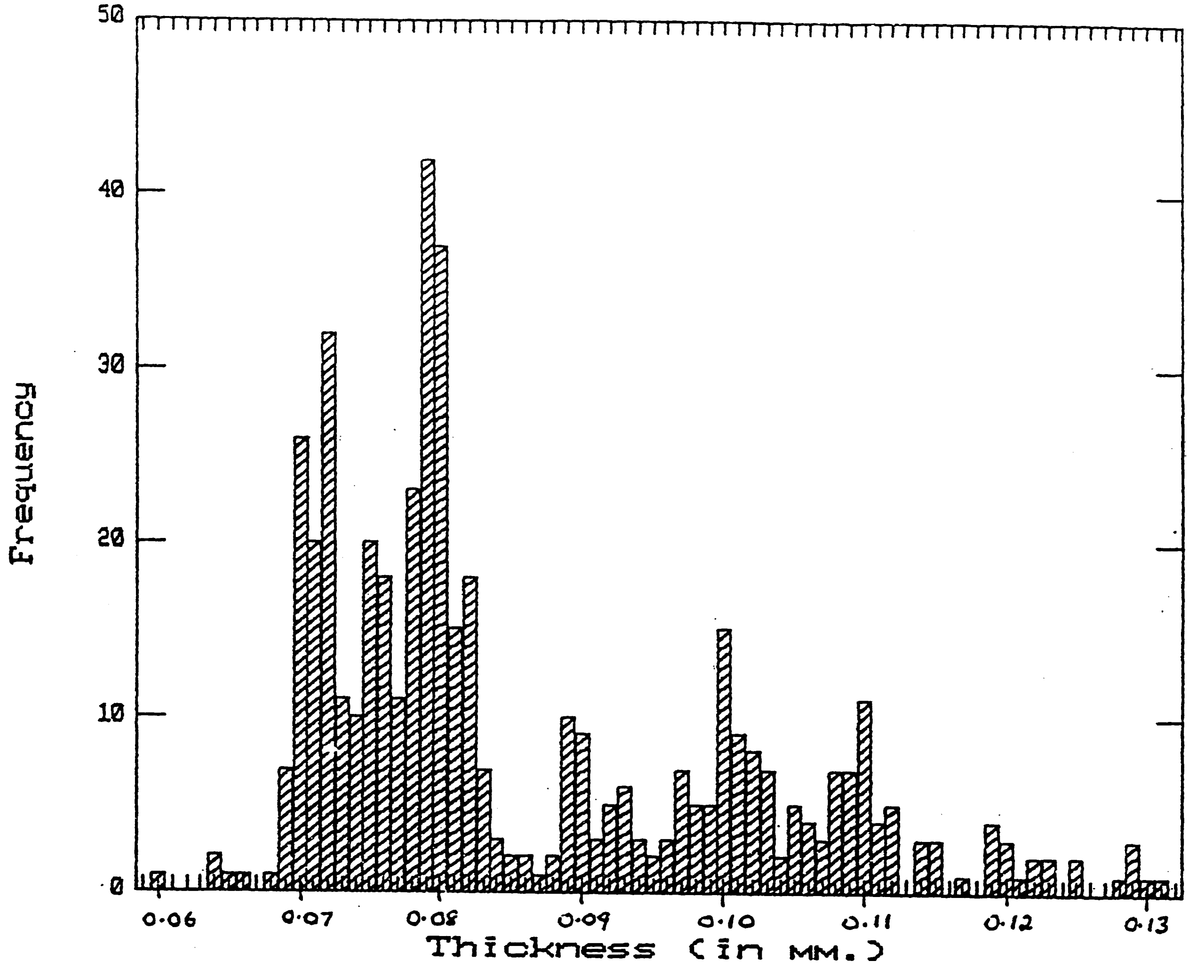

Figure 1. Histogram of all 485 measurements from the 1872 Hidalgo Issue of Mexico (from Izenman and Sommer, 1988).

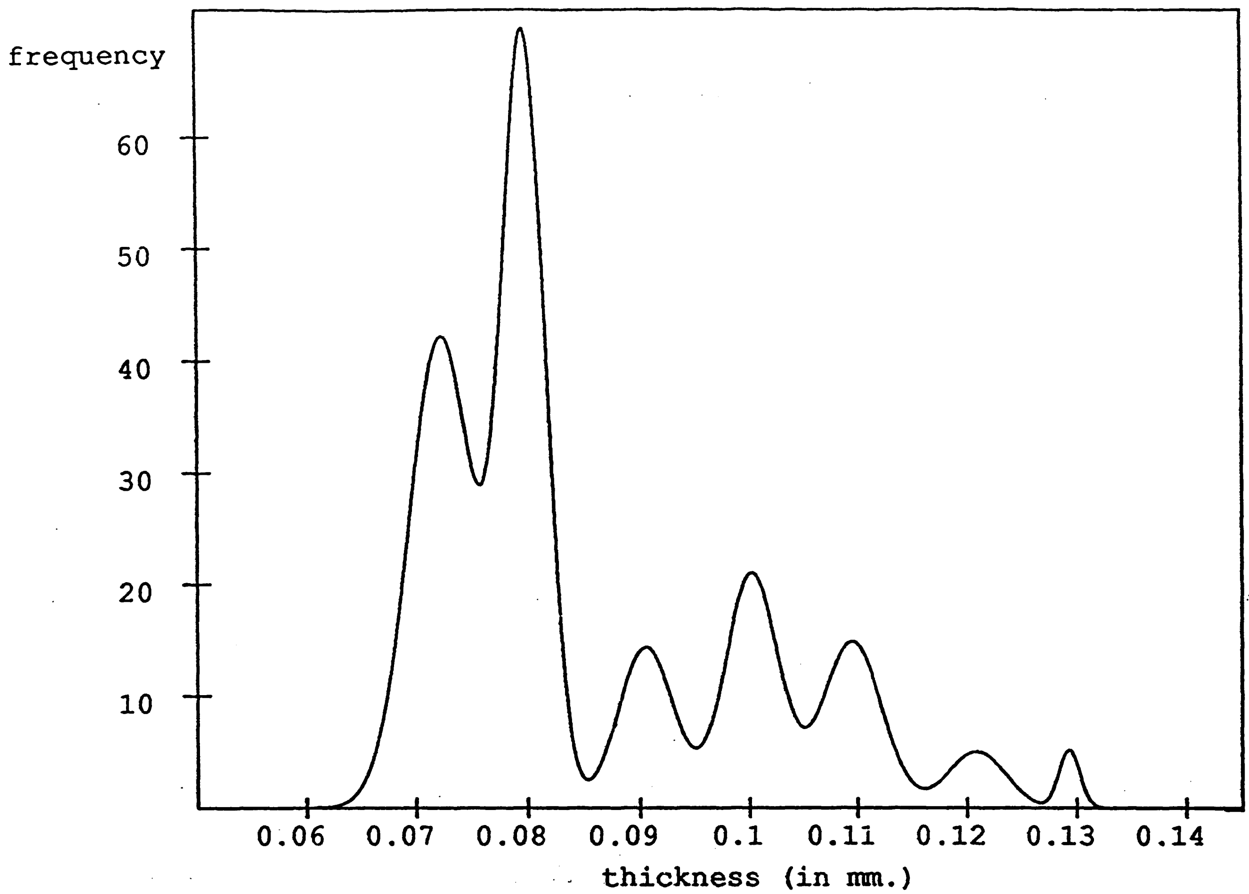

Figure 2. Maximum likelihood density estimate of a mixture of seven unrestricted normals fitted to the 485 measurements of the 1872 Hidalgo Issue of Mexico (from Izenman and Sommer, 1988).

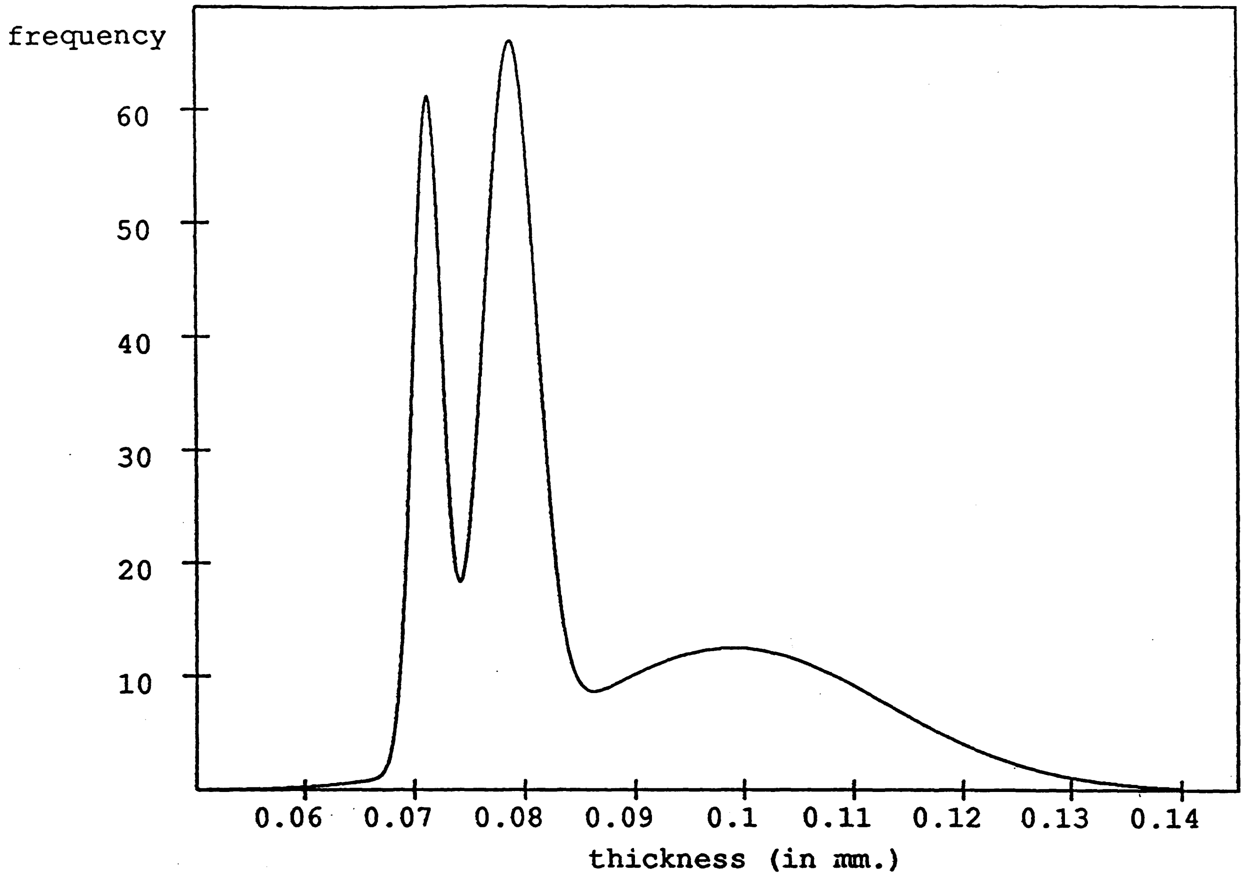

Figure 3. Maximum likelihood density estimate of a mixture of three unrestricted normals fitted to the 485 measurements of the 1872 Hidalgo Issue of Mexico (from Izenman and Sommer, 1988).

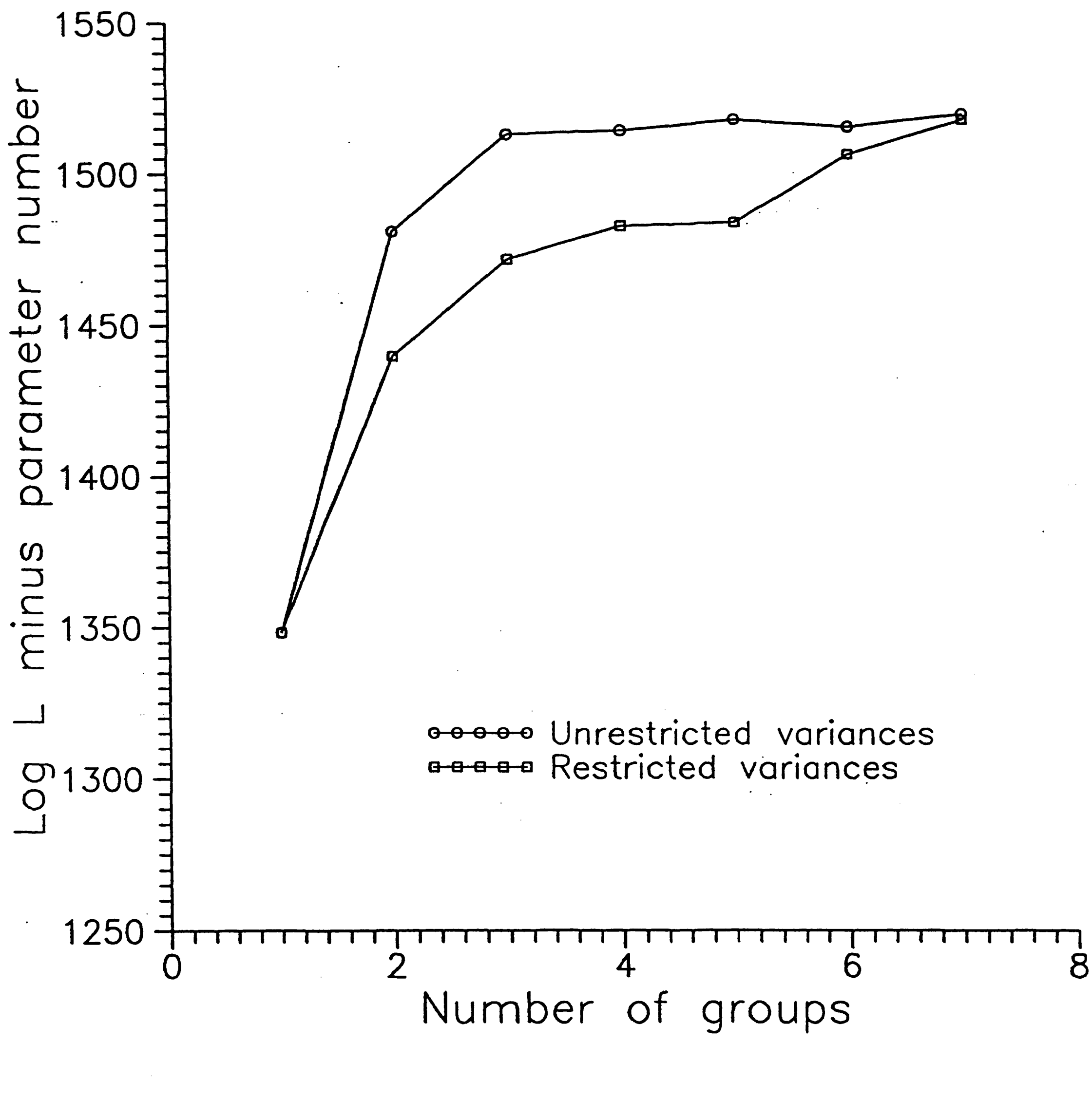

Figure 4. Plot of log likelihood minus the number of estimated parameters against the number of distributions in the mixture of normal distributions (with unrestricted and equal variances) fitted to the 1872 Hidalgo Issue of Mexico.

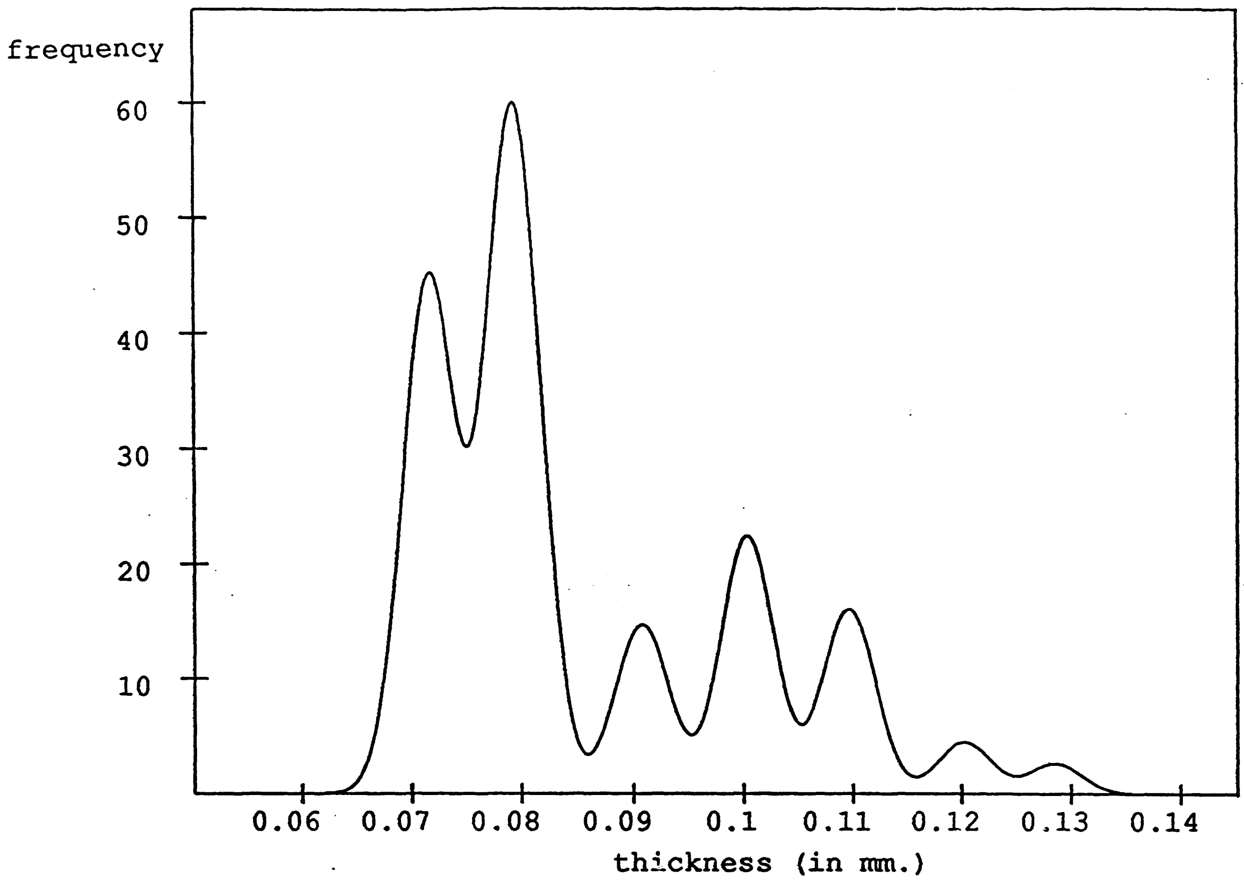

Figure 5. **Maximum likelihood density estimate of a mixture of seven normals with equal variances fitted to the 485 measurements of the 1872 Hidalgo Issue of Mexico (from Izenman and Sommer, 1988).**