Modelling the Distribution of Stamp Paper Thickness via Finite Normal Mixtures: The 1872 Hidalgo Stamp Issue of Mexico Revisited

K.E. BASFORD[1], G.J. McLACHLAN[2] and M.G. YORK[2]

Summary

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 normal mixture model with three components with unequal variances. This conflict is examined further by investigating the most appropriate number of components when fitting a normal mixture of components with equal variances.

Key Words: Philatelic mixtures; Normal mixture models; Tests for number of components; Likelihood ratio test.

---

[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) considered the modelling of the distribution of stamp thickness for the printing of a given stamp issue from different types of paper. 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 with a parametric approach, Izenman and Sommer (1988) fitted by maximum likelihood a mixture of normal components with unequal variances to the data. They used the likelihood ratio statistic as modified by the rule of thumb of Wolfe (1971) to test for the smallest number g of components compatible with the data. It yielded $g = 3$ components, which is in apparent conflict with their nonparametric solution of seven modes.

In an attempt to resolve this conflict, we consider further here the fitting of a normal mixture model to the data. It will be seen that by restricting the normal components in the mixture model to have a common variance, the likelihood ratio test for the smallest number of components in the mixture ultimately leads to results consistent with the nonparametric approach.

2 Normal Mixture Model

It is assumed under a normal mixture model that the data may be viewed as having come from a mixture of, say g, component normal distributions in unknown proportions $\pi_1, \ldots, \pi_g$, where the mixing proportions π_i are nonnegative and sum to one. We let

$$f(x; \boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_i \phi(x; \mu_i, \sigma_i^2) \tag{1}$$

denote the g-component normal mixture density, where $\phi(x; \mu_i, \sigma_i^2)$ denotes the univariate normal density with mean μ_i and variance σ_i^2 $(i = 1, ..., g)$. Here

$$\mathbf{\Psi} = (\pi_1, ..., \pi_{g-1}, \mu_1, ..., \mu_g, \sigma_1^2, ..., \sigma_g^2)'$$

is the vector of all unknown parameters, which can be estimated by maximum likelihood. In a later section, notation for (1) is abused by writing $f(x; \mathbf{\Psi})$ as

$$\sum_{i=1}^{g} \pi_i N(\mu_i, \sigma_i^2).$$

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 entities 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 ith component of the mixture is given by

$$\hat{\tau}_i(x_j) = \hat{\pi}_i \phi(x_j; \hat{\mu}_i, \hat{\sigma}_i^2) / f(x_j; \hat{\mathbf{\Psi}}). \tag{2}$$

It can be seen from above that the number of components g must be specified before the normal mixture model can be fitted. An obvious way to assess g is to test for its smallest value compatible with the data, using the likelihood ratio test statistic λ. Unfortunately with mixture models, regularity conditions breakdown for $-2 \log \lambda$ to have its usual null distribution of chi-squared with degrees of freedom equal to the difference between the number of parameters under the null and alternative hypotheses; see, for example, Ghosh and Sen (1985), Hartigan (1985), McLachlan and Basford (1988, Chapter 1), McLachlan (1987), Quinn, McLachlan, and Hjort (1987), Mendell, Thode, and

Finch (1991), and Mendell, Finch and Thode (1993). More recent attempts to resolve this problem by using criteria other than the likelihood ratio statistic include those by Bozdogan (1992), Windham and Cutler (1992), and Celex and Soremenho (1994).

Izenman and Sommer (1988) determined g by using the rule of thumb of Wolfe (1971) to approximate the null distribution of $-2\log\lambda$ as a chi-squared with degrees of freedom equal to the difference between the number of parameters under the null and alternative hypotheses, excluding the mixing proportions.

3 The 1872 Hidalgo Stamp Issue of Mexico

Izeman and Sommer (1988) gave a detailed account of the content and historical data available on the 1872 Hidalgo Stamp issue of Mexico. These data, which are displayed in histogram form in Figure 1, are the thicknesses of 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. 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 (1988) 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 (1981) 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 (1988) 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; that is, it is plausible that paper of seven different thicknesses was used in the production of this stamp issue.

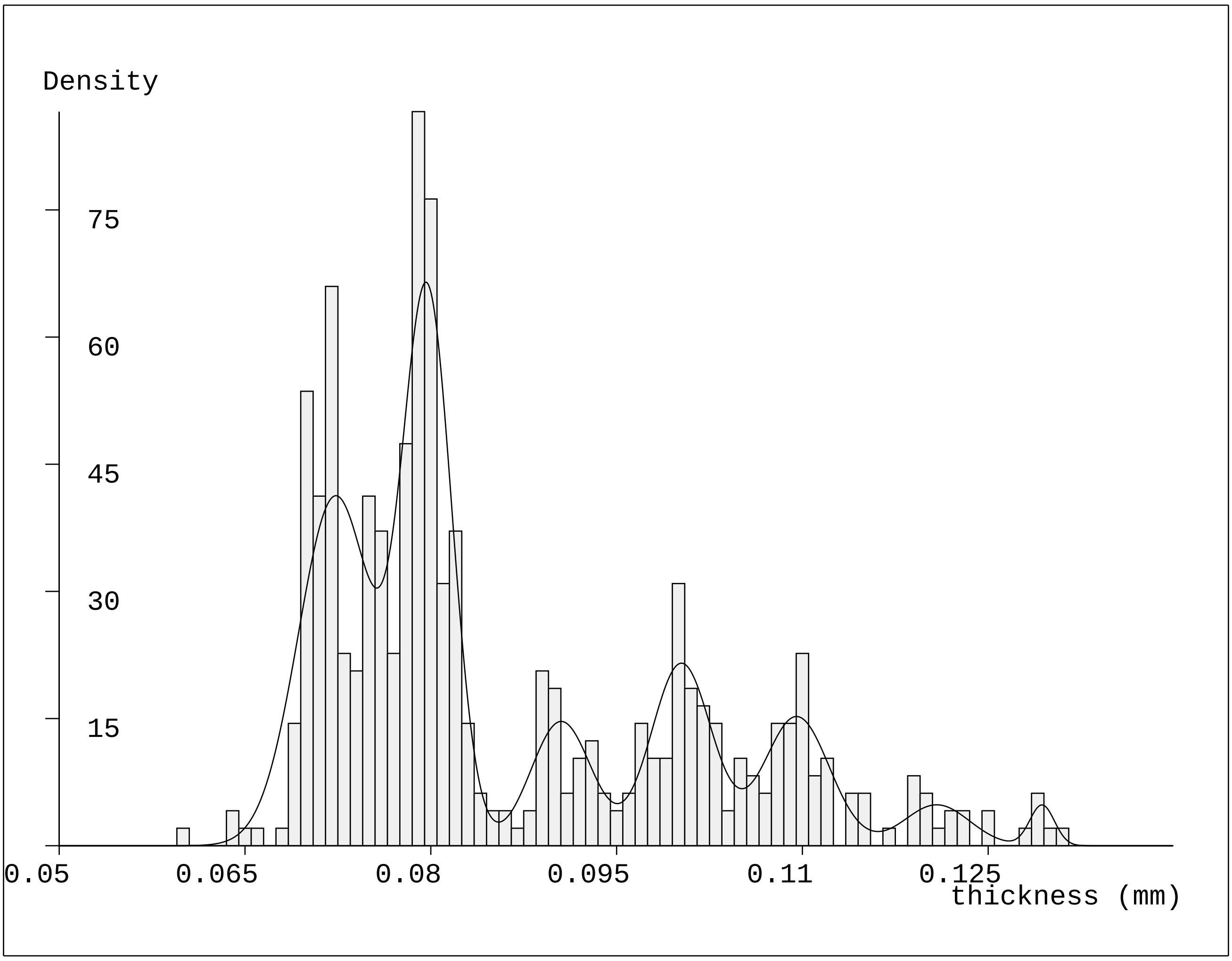

Figure 1. Plot of estimate from Izenman and Sommer (1988) of mixture density of seven unrestricted normals fitted to the 485 measurements of the 1872 Hidalgo Issue of Mexico

For comparison with the nonparametric mode fitting technique, Izenman and Sommer (1988) 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 fitted normal mixture density, which is displayed in Figure 1, is given by

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

Actually, we located another $g = 7$ solution corresponding to a larger local maximum of the likelihood function than the one corresponding to the solution found by Izenman and Sommer (1988). We have plotted the density of this fit in Figure 2. It can be seen that it differs from the latter by devoting a component to model the outlying observations in the lower tail of the data. It thus uses only one rather than two components to model the modes in the histogran at 0.10 and 0.11mm.

By applying Wolfe's (1971) test to sequentially test g against $g + 1$, starting at g=1, Izenman and Sommer (1988) determined that only $g = 3$ components in the heteroscedastic normal mixture model were needed. This fit (Figure 3) is given by

$$\begin{aligned}&0.196N(0.0712,\ 0.00000176) + 0.367N(0.0786,\ 0.00000564)\\&\quad + 0.437N(0.0989,\ 0.00019666).\end{aligned}$$

The values of the log likelihood for $g = 1\ to\ 9$ are listed in Table 1 for both unrestricted and equal variances for the normal components.

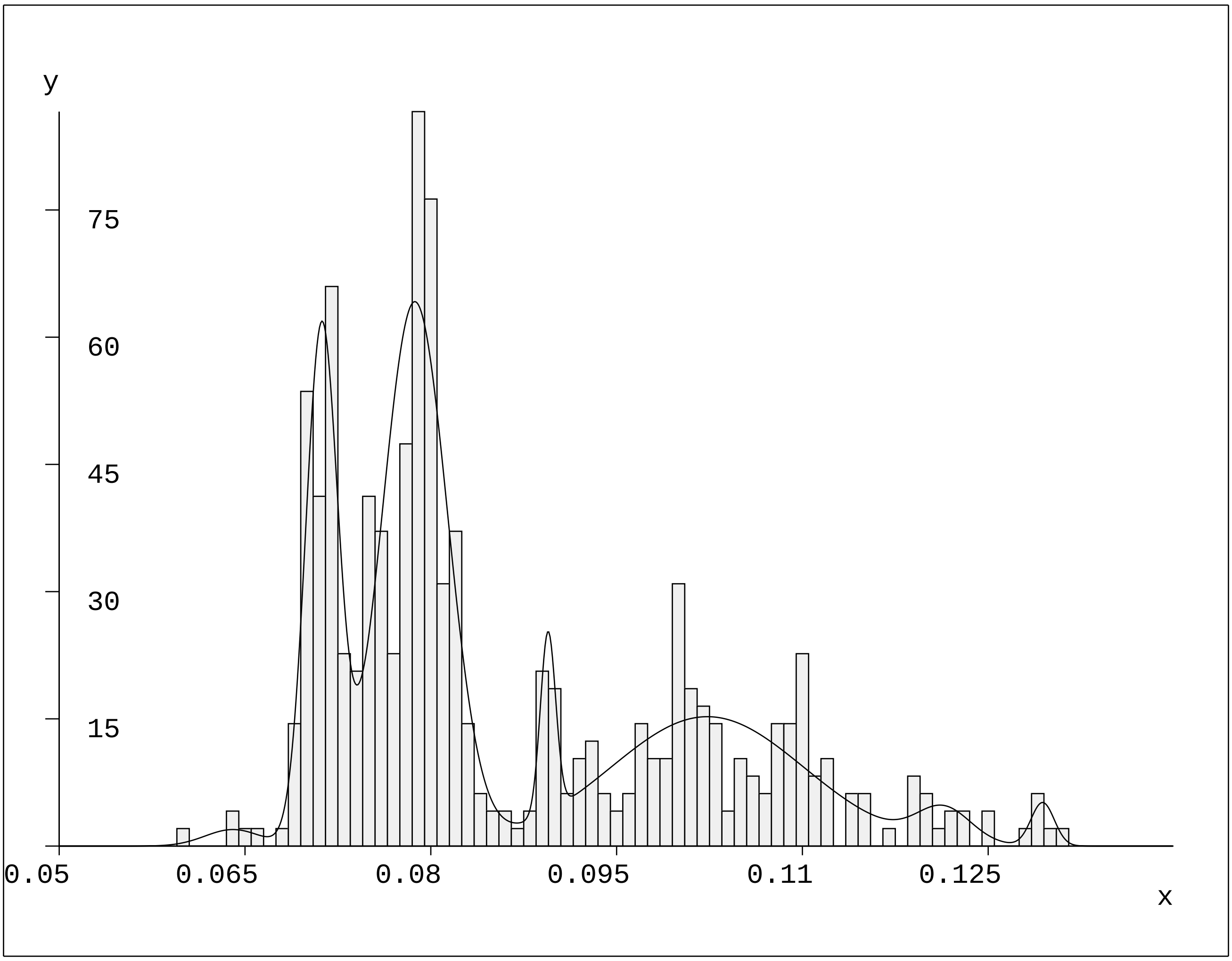

Figure 2. Plot of maximum likelihood estimate of mixture density of seven unrestricted normals fitted to the 485 measurements of the 1872 Hidalgo Issue of Mexico

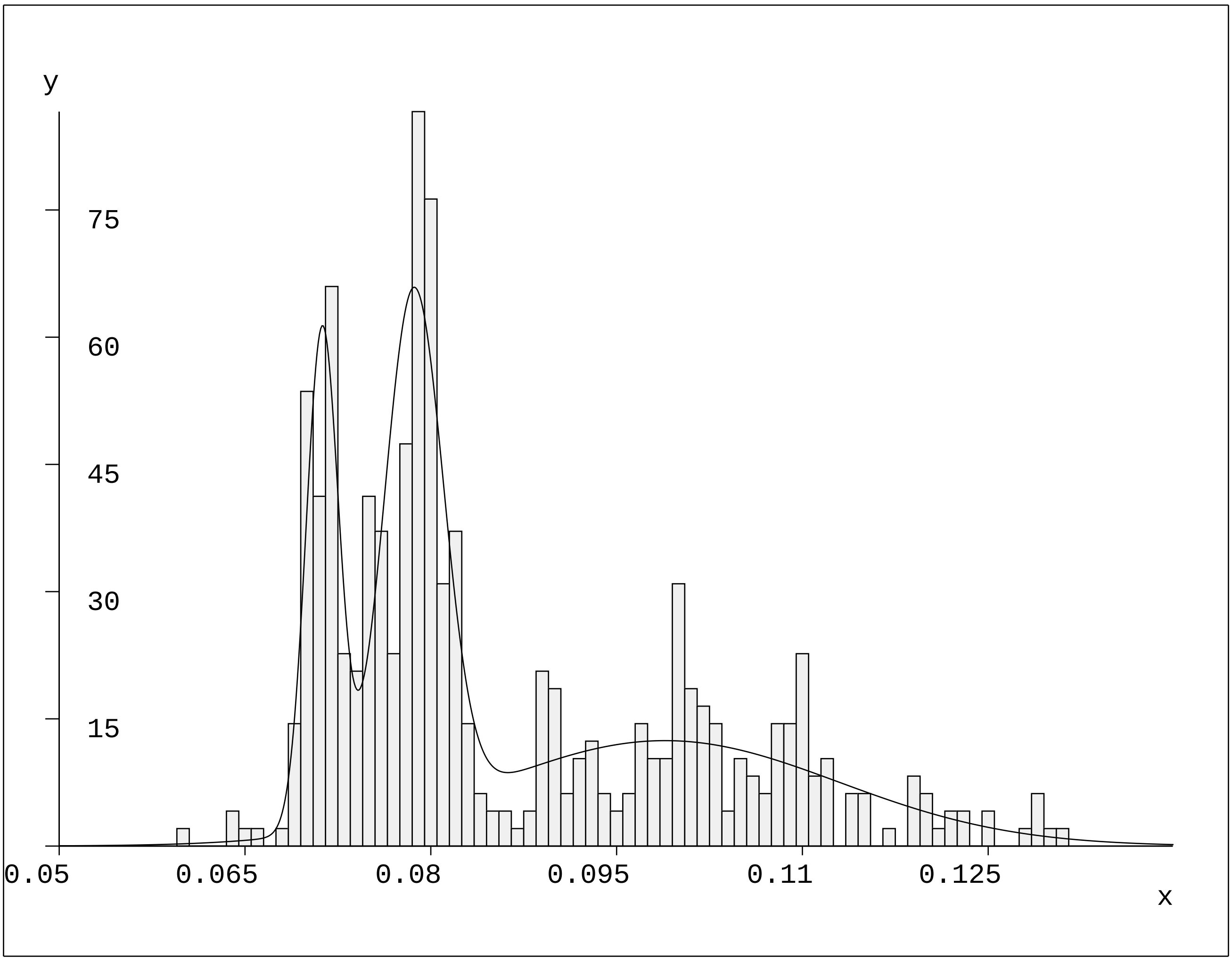

Figure 3. Plot of maximum likelihood estimate of mixture density of three unrestricted normals fitted to the 485 measurements of the 1872 Hidalgo Issue of Mexico

TABLE 1

Value of the log likelihood for $g = 1$ to 9 for the fitted mixture of normal distributions with unrestricted and equal variances: 1872 Hidalgo Issue of Mexico

Number of Components	Unrestricted Variances	Equal Variances
1	1350.3	1350.3
2	1484.8	1442.6
3	1518.8	1475.7
4	1521.9	1487.4
5	1527.1	1489.5
6	1535.2	1512.9
7	1538.7	1525.3
8	1544.1	1535.1
9	1552.2	1536.5

In an attempt to reconcile the nonparametric (seven modes) and parametric (three-component mixture) approaches, Izenman and Sommer (1988) 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 (1988), 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).

Besides the assessment of g by testing as above for its smallest value as compatible with the data, other practical considerations are important if a sensible description or summary of the data is to be reached. For instance, if a mixture model is used to describe a univariate multimodal density, it may be unacceptable from a practical point of view to have the tail of one component (say, the one centred on the highest mean) accounting for data at the lower end (below that of the smallest mean) by allowing its variance to be much greater than the other component variances. This is the case here where the variance of the third component in the $g = 3$ heteroscedastic normal mixture solution of Izenman and Sommer (1988) is appreciately greater than the variances for the other two components. As a consequence of the long left-hand tail of the third component, 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; see Table 2. 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. Hence the $g = 3$ heteroscedastic normal mixture model is impractical for physical considerations. We now consider the fitting of a normal mixture model with the components restricted to having a common variance.

TABLE 2

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

Thickness x_j	Estimated posterior probability $\hat{\tau}_1(x_j)$	$\hat{\tau}_2(x_j)$	$\hat{\tau}_3(x_j)$
.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

4 Fitting of Homoscedastic Normal Mixture Model

It can be seen from Table 1 that, if on the basis of $-2\log\lambda$ we were to test g versus $g+1$ sequentially, starting with $g=1$, then testing would terminate at $g=4$, as there is only a slight increase in the log likelihood in preceding to $g=5$. This is because the common variance of the components in the mixture model is not sufficiently small in the case of $g=5$ to provide any further modes in the fit; see Figures 4 and 5.

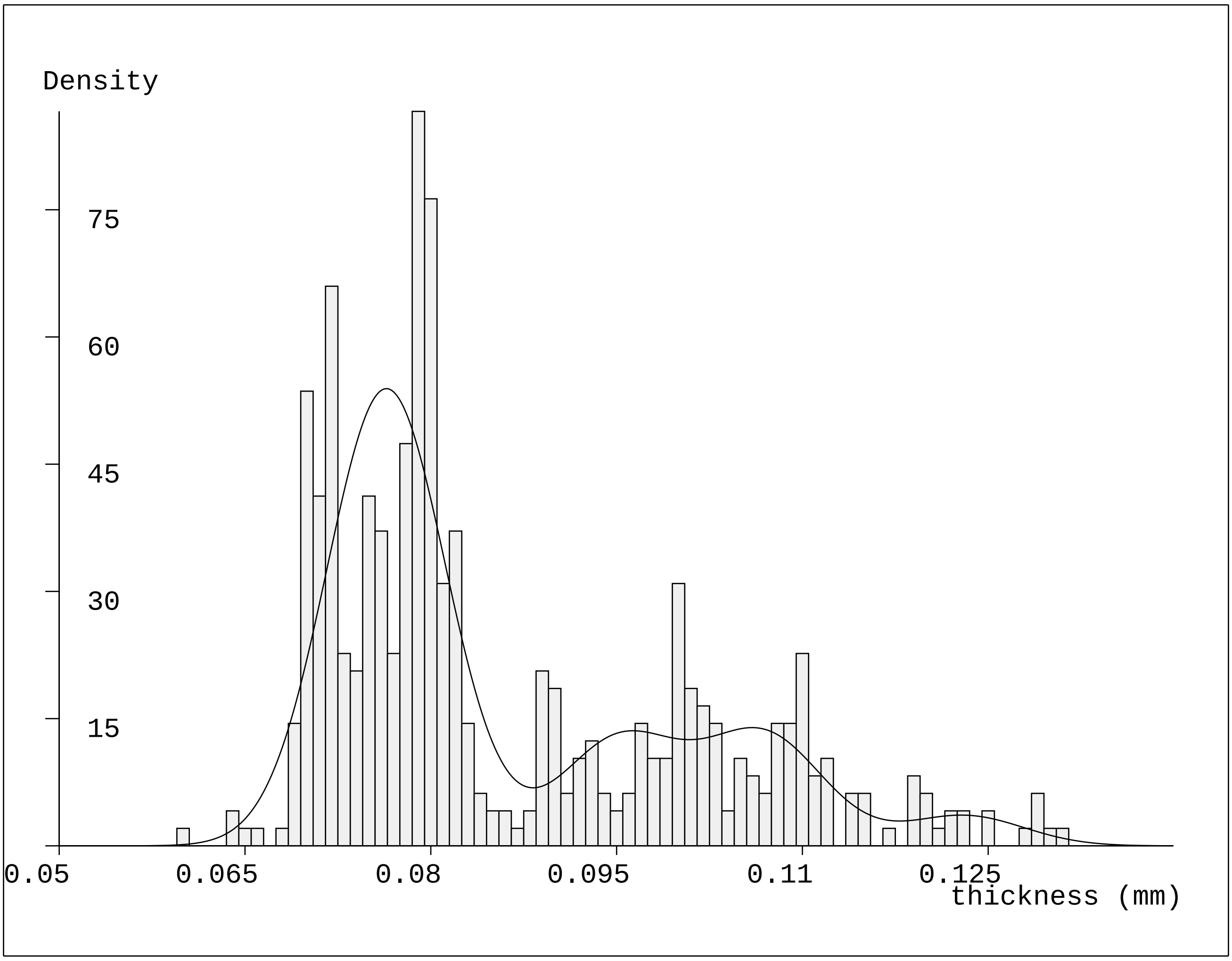

Figure 4. Plot of maximum likelihood estimate of mixture density of four normals with equal variances fitted to the 485 measurements of the 1872 Hidalgo Issue of Mexico

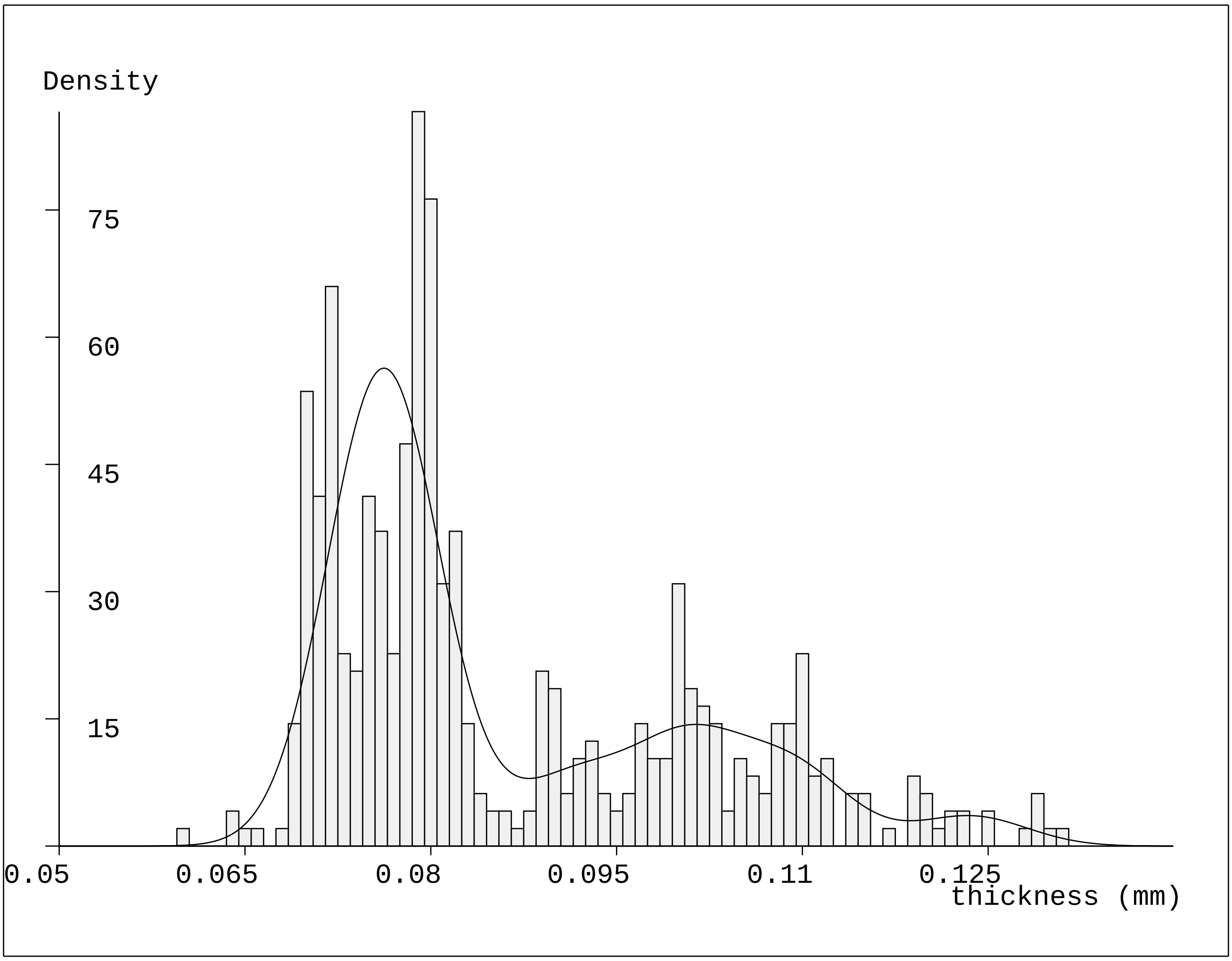

Figure 5. Plot of maximum likelihood estimate of mixture of five normals with equal variances fitted to the 485 measurements of the 1872 Hidalgo Issue of Mexico

If we continue on past $g = 5$ and monitor the increase in the log likelihood, then it is clear from Table 1 that there is no need to proceed beyond $g = 8$. On whether there is a need to proceed from $g = 7$ to $g = 8$, the value of $-2\log\lambda$ is 19.6, which would appear to be significant. For example, if Wolfe's approximation were adopted which would take $-2\log\lambda$ to be χ^2_2, then the P-value is practically zero. The $g = 7$ and $g = 8$ homoscedastic normal mixture fits for the density are displayed in Figures 6 and 7, respectively. It can be seen that the additional component in the $g = 8$ component normal mixture model exists to accommodate the outlying observations in the lower tail of the data. Elsewhere, it practically gives the same fit as the $g = 7$ homoscedastic solution.

The chi-squared goodness-of-fit test statistic for the $g = 8$ homoscedastic normal mixture fit is 57.4. While this is highly significant, it has fallen from a value of 158.14 for the $g = 4$ fit. This high value would appear to be due to the rounding of the stamp measurements to the third decimal place. If allowance is made for this rounding by the judicious grouping of intervals in forming the chi-squared goodness-of-fit statistic, the latter is reduced to 20.1 with a P-value of 27%.

If we relax the assumption of equal variances now in fitting the $g = 7$ or $g = 8$ component mixture models, then we obtain essentially the same fit as under homoscedasticity. The only difference, really, is that a smaller variance is obtained for the component with the largest fitted mean.

In Figure 8, we have plotted a quantity equal to the log likelihood minus the number of parameters (including the mixing proportions) in the normal mixture model versus g for the homoscedastic and heteroscedastic cases. The common value for this quantity in both cases for $g = 8$ provides further support in favour of $g = 8$ components.

5 Conclusions

We have considered a parametric approach via normal mixture models to the modelling of the distribution of the thicknesses of stamps in the 1872 Hidalgo Stamp issue of Mexico. From physical considerations of the problem, we argue that it is reasonable to proceed under the restriction that the normal components have equal variances. It is shown here this approach leads to a final solution that is not in conflict with the nonparametric solution of Izenman and Sommer (1988).

Recently, Efron and Tibshirani (1994) considered the problem of how many modes in the fitting of a Gaussian kernel density estimate to this data set. In their application of the smooth bootstrap to this problem, they first rescaled the kernel density estimate so that it had a reduced variance approximately equal to the sample variance. Proceeding sequentially from $g = 1$ in testing the null hypothesis of g modes versus the alternative of more than g modes, they concluded that the null hypothesis of two modes would not be rejected. But they also noted from their Table 16.1 of P-values that if one "were willing to entertain more exotic alternative hypotheses, then there is also a suggestion that the population might have seven modes".

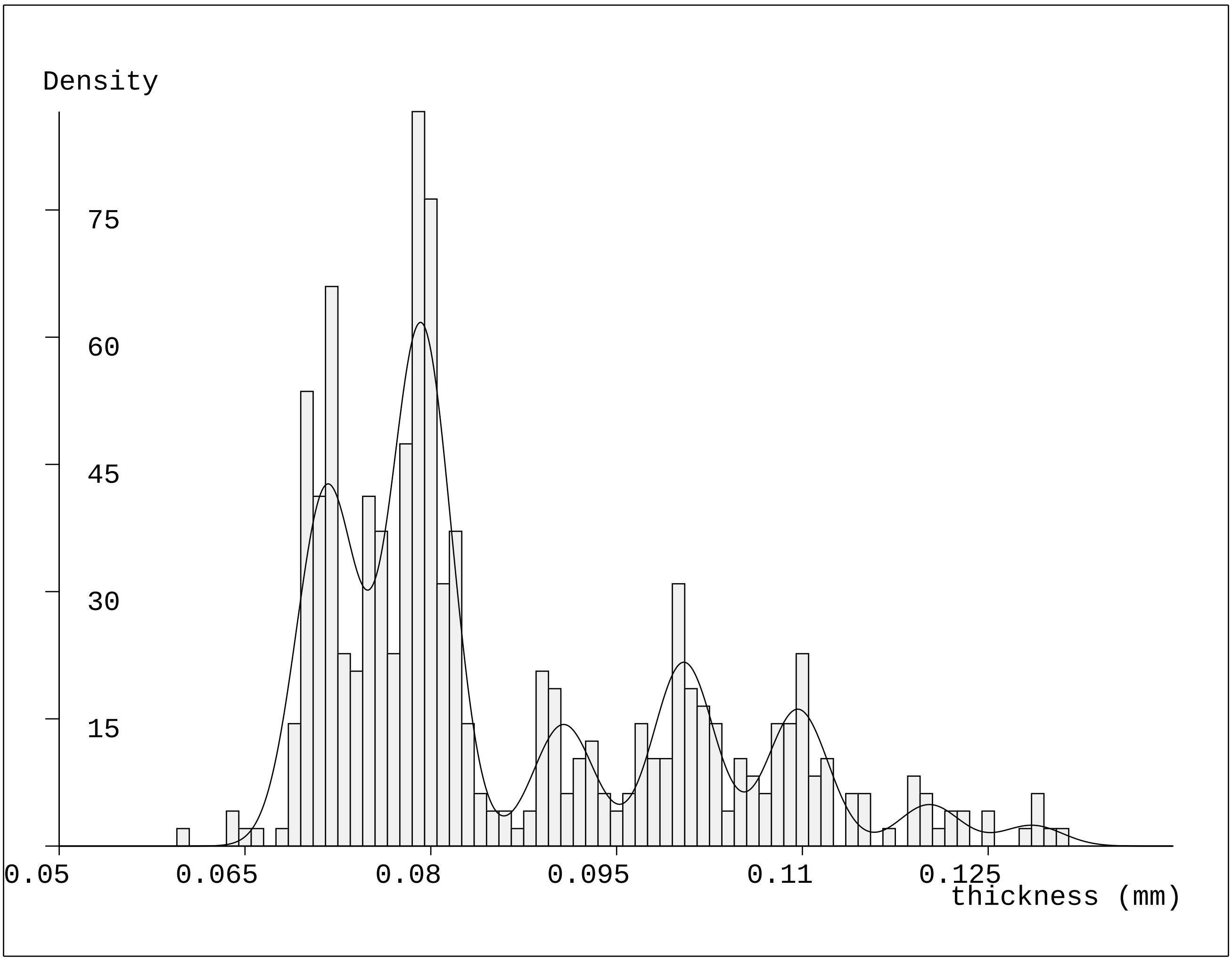

Figure 6. Plot of maximum likelihood estimate of mixture density of seven normals with equal variances fitted to the 485 measurements of the 1872 Hidalgo Issue of Mexico

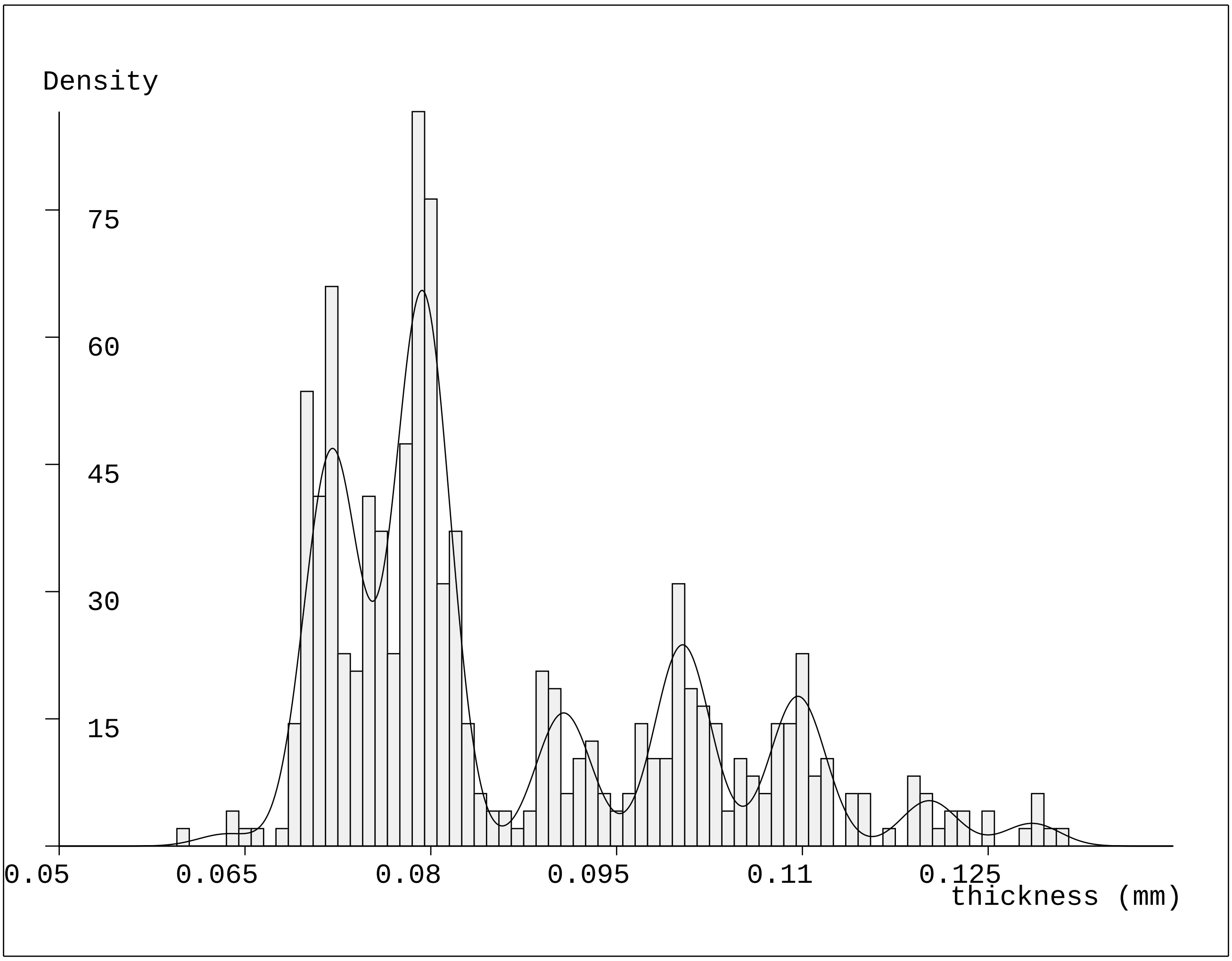

Figure 7. Plot of maximum likelihood estimate of mixture density of eight normals with equal variances fitted to the 485 measurements of the 1872 Hidalgo Issue of Mexico

Figure 8. Plot of the log likelihood minus the number of estimated parameters against the number of components in the normal mixture model (with unrestricted and equal component variances) fitted to the 485 measurements of the 1872 Hidalgo Issue of Mexico

References

Bozdogan, H. (1992). Choosing the number of component clusters in the mixture-model using a new informational complexity criterion of the inverse-Fisher information matrix. In *Studies in Classification, Data Analysis, and Knowledge Organization*, O.Opitz, B. Lausen, and R. Klar (Eds.). Heidelberg: Springer-Verlag.

Celeux, G. and Soromenho, G. (1993). An entropy criterion for assessing the number of clusters in a mixture model. Unpublished manuscript.

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.

Efron, B. and Tibshirani, R.J. (1994). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Ghosh, J.H. and 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. and Sommer, C.J. (1988). Philatelic mixtures and multimodal densities. *Journal of the American Statistical Association* **83**, 941-953.

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 Basford, K.E. (1988). *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker.

Mendell, N.R., Finch, S.J., and Thode, H.C. (1993). Where is the likelihood ratio test powerful for detecting two component normal mixtures? *Biometrics* **49**, 907–915.

Mendell, N.R., Thode, H.C., and Finch, S.J. (1991). The likelihood ratio test for the normal mixture problem: power and sample size analysis. *Biometrics* **47**, 1143–1148. Correction (1992). *Biometrics* **48**, 661.

Quinn, B.G., McLachlan, G.J., and Hjort, N.L. (1987). A note on the Aitkin-Rubin approach to hypothesis testing in mixture models. *Journal of the Royal Statistical Society B* **49**, 311–314.

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.

Windham, M.P. and Cutler, A. (1992). Information ratios for validating mixture analyses. *Journal of the American Statistical Association* **87**, 1188–1192.

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.

Wilson, J.G. (1983). Add a new dimension to your philately. *The American Philatelist* **98**, 442–452.