Oscillation Analysis for the Mixture Complexity

Guenther Walther
Stanford University, Department of Statistics
390 Serra Mall
Stanford, CA 94305, USA
walther@stat.stanford.edu

The problem under consideration is to determine the number of components in a location mixture, in the case where one does not want to make parametric assumptions on the component distribution. Some basic properties of such mixtures are derived, which yield a simple but powerful criterion for setting lower confidence bounds on the mixture complexity. The resulting procedure has the advantage over mode-hunting approaches that it is sensitive to detect mixing in more general unimodal situations, yet it cannot be improved upon in the more restricted situation where mixing manifests itself in multi-modality, even if one is allowed to use that knowledge a priori. This is explained heuristically and made precise in the asymptotic minimax framework.

Inference about the mixture complexity

One important issue in the analysis of mixture data is the inference on the number of components in the mixture. This article focuses on the important case of a location mixture

$$f(x) = \sum_{i=1}^{k} p_i \, g(x - t_i), \quad g \in \mathcal{S}, \tag{1}$$

where the nonnegative weights sum to unity and $\mathcal{S}$ denotes a set of univariate single-component distributions. In this model the weights p_i, the location parameters t_i, and the single-component density g are unknown. The object of the inference is the *mixture complexity* $\inf\{k : \ f(x) = \sum_{i=1}^{k} p_i g(x - t_i), g \in \mathcal{S}\}$. Scientific and statistical interest focuses on lower confidence bounds for the mixture complexity; it is well known that no nontrivial upper confidence bounds exist, see Donoho (1988).

Powerful theoretical results are available in the case where the component class $\mathcal{S}$ is parametric, see e.g. Lindsay and Roeder (1997). There is a large literature on an approach where $\mathcal{S}$ is a nonparametric class of distributions, motivated by the fact that the parametric procedures are quite sensitive to the structure imposed on $\mathcal{S}$. For example, if $\mathcal{S}$ is taken to be the normal family in the analysis but the real g is skewed, then many normal components are required in the mixture to pick up the skewness, which can result in a considerable overestimate of the mixture complexity, see e.g. Roeder (1994). The nonparametric methods usually proceed by mode- or bump-hunting, i.e. by establishing a lower confidence bound on the number of modes of f or of 'bumps' (maxima of the density derivative). One disadvantage of such an approach is that it is not very sensitive to detect mixing. For example, the means of two homoscedastic normal distributions need to be separated by at least two standard deviations before any mixture becomes bimodal. Another problem, which is apparently not widely recognized, is that counting modes and 'bumps' does not always lead to a valid inference: Figure 1 shows a unimodal density g (solid line) and the mixture $1/2(g(x) + g(x - 4))$, which is trimodal.

In fact, it is not hard to modify g in figure 1 such that the two-component mixture possesses any prescribed number m of modes. Thus if the sample size is large enough, then mode-hunting will give a lower bound of 3 (or even m) for the mixture complexity with arbitrary high confidence level, while for unimodal components the mixture complexity is only two. This is a most serious scientific and statistical mistake.

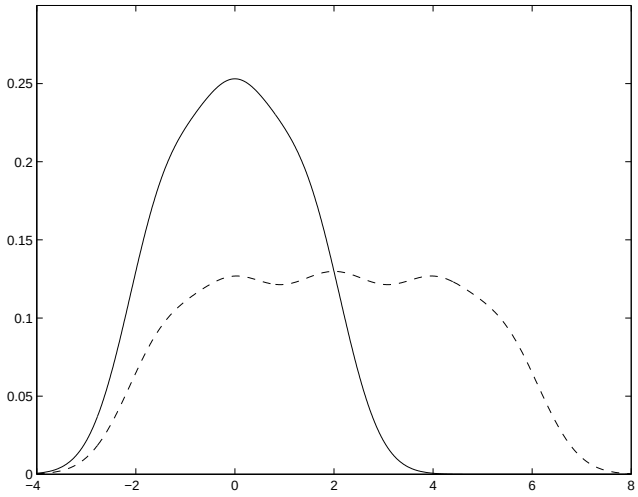

Fig.1. A mixture of two unimodal distributions which is trimodal.

A nonparametric mixture model

To understand what mode- and bump-hunting are doing, recall that a univariate density g is called Polya frequency density ($g \in PF_\infty$) if for all $x_1 < \ldots < x_n$, $y_1 < \ldots < y_n$,

$$\det \parallel g(x_i - y_j) \parallel_{i,j=1}^{n} \geq 0. \tag{2}$$

The nonparametric class PF_∞ contains many standard unimodal parametric densities, such as the normal family, but also skewed distributions such as the gamma family with integer shape parameter. PF_∞ is thus a flexible nonparametric model for single-component distributions. Furthermore, it is essentially the model under which mode- and bump-hunting will work (i.e. the number of modes is not larger than the mixture complexity k):

Proposition 1 *If $\mathcal{S} = PF_\infty$ in the model (1), then mode-hunting works.*

The proposition and the example show that mode-hunting (and likewise bump-hunting) work for a flexible and useful class of component distributions, and that the component model is not the class of unimodal distributions, but the class PF_∞ (indeed, the unimodal component in figure 1 was constructed as a mixture of several normal distributions, and this fact was detected by mode-hunting in this case!) Having identified the model underlying mode- (and bump)-hunting, one may attempt to use these criteria directly, rather than indirectly via the modality approach. There is legitimate hope that such a direct approach will lead to stronger results or simpler inference, or possibly both. This turns out to be indeed the case.

The oscillation analysis

Theorem 1 *Consider the mixture (1) and assume $g \in PF_\infty$, $t_1 < \ldots < t_k$, and all $p_i > 0$. Then there exist $m \geq 1$, real $a_1, \ldots, a_m$, and $c_1 < \ldots < c_k$ such that $\sum_{j=1}^{m} a_j f(\cdot - c_j)$ has more changes of sign than $(a_1, \ldots, a_m)$.*

The conclusion of the theorem is known to imply that $f \notin PF_\infty$. Thus under the same model under which mode-hunting works, it is always possible to detect the presence of a mixture, while mode-hunting usually requires a considerable separation of the locations t_i, see above. This surprisingly powerful result mirrors the situation in the parametric context, see

e.g. Lindsay and Roeder (1997). Moreover, the theorem provides a criterion that is useful for developing appropriate methodology.

It will be shown how this criterion allows to set lower confidence bounds on the mixture complexity that have finite-sample guaranteed coverage. Moreover, while it was shown in the last paragraph that this approach is more sensitive to detect mixing than mode- and bump-hunting, this increased sensitivity is essentially free: There is no loss of power in the case where the mixture does in fact result in a bimodal situation, and where one would hence suspect that a more narrowly focused mode-hunting approach would outperform the more general procedure. A heuristic explanation will be given why this is in fact not so, together with a precise justification in terms of asymptotic minimax results.

REFERENCES

Donoho, D.L. (1988). One-sided inference about functionals of a density. *Ann. Stat.* **16**, 1390–1420

Lindsay, B. G. and Roeder, K. (1997). Moment-based oscillation properties of mixture models. *Ann. Stat.* **25**, 378–386

Roeder, K. (1994). A graphical technique for determining the number of components in a mixture of normals. *J. Amer. Statist. Assoc.* **89**, 487–495