Futschik, A. and Clarke, B.R. (1999) Estimating nonsmooth k-modal densities and their inflection points. In:
52th Session of the International Statistical Institute, 11 - 18 August, Helsinki, Finland

Estimating nonsmooth k–modal densities and their inflection points.

Andreas Futschik
Inst. f. Statistik, Vienna University
Universitätsstr. 5/9
A-1010 Vienna, Austria
andreas.futschik@univie.ac.at

Brenton R. Clarke
Murdoch University
90, South Street
Murdoch, Australia
clarke@prodigal.murdoch.edu.au

1. Estimates for unimodal densities

In literature several estimates have been proposed for unimodal densities. They are typically derived from the Grenander estimate (see Grenander (1956)) for decreasing densities which can be easily extended to the case of unimodal densities with a known mode at θ. The resulting estimator is the nonparametric maximum likelihood estimator (NPMLE). If the mode is unknown, the NPMLE does not exist anymore. One possible solution that has been proposed by Wegman (1969, 1970a,b) is to add the additional constraint of a modal interval of length ϵ, where ϵ has to be chosen by the statistician. More recently Bickel and Fan (1996) and Birgé (1997) proposed methods that are based on an initial mode estimate $\hat{\theta}$ and an application of the NPMLE with mode $\hat{\theta}$. The initial mode estimate requires the calculation of the MLE of f for $O(n)$ candidate points for the mode, where n denotes the sample size. They showed their methods to provide good estimates both for nonsmooth densities and for the mode of nonsmooth densities.

While unimodal density estimates can be generalized to the case of a known number of modes $k > 1$, the initial step of identifying all inflection points requires the evaluation of the NPMLE for $O\left[\binom{n}{2k-1}\right]$ possible combinations of local extrema. For increasing n and/or k the necessary computational effort quickly becomes extremely high.

2. Estimates for k-modal densities

Densities with a known number $k > 1$ of modes occur e.g. in mixture models when the number of mixture components is known, or when estimating the intensity function of inhomogeneous Poisson processes. In such situations it seems natural to require the estimated function to reflect this shape information.

This can be achieved by using kernel density estimates with suitable (random) bandwidths. In particular

$$\hat{f}_{n,H}(x) := \frac{1}{nH} \sum_{i=1}^{n} K\left(\frac{x - X_i}{H}\right),$$

satisfies the requirement with $K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ taken as the Gaussian kernel and H being the smallest bandwidth that yields k modes. (It has been shown by Silverman (1981) that the number of modes of $\hat{f}_{n,H}$ is decreasing and right continuous in H, if the Gaussian kernel is used. Thus H is uniquely defined.)

While the above estimate behaves well for smooth densities, it can be improved for nonsmooth functions. For densities with support $[a, b]$ the following estimate inherits the desirable properties of the NPMLE in the nonsmooth case at greatly reduced computational effort:

Let $(\hat{\theta}_1, \ldots, \hat{\theta}_k)^t$ denote the vector of sorted local maxima of $\hat{f}_{n,H}$ and let $(\hat{\theta}^*_1, \ldots, \hat{\theta}^*_{k-1})^t$ denote the vector of sorted local minima of $\hat{f}_{n,H}$. Define $\hat{\theta}^{(S)}_1, \hat{\theta}^{(S)}_2, \ldots, \hat{\theta}^{(S)}_{2k-1}$ to be the sorted elements of the set $\{\hat{\theta}_1, \ldots, \hat{\theta}_k, , \hat{\theta}^*_1, \ldots, \hat{\theta}^*_{k-1}\}$. Now methods for unimodal (or U-shaped) densities can be applied separately for each of the subintervals

$$I_1 := \left[a,\, (\hat{\theta}^{(S)}_1 + \hat{\theta}^{(S)}_2)/2\right],\, I_2 := \left[(\hat{\theta}^{(S)}_1 + \hat{\theta}^{(S)}_2)/2,\, (\hat{\theta}^{(S)}_2 + \hat{\theta}^{(S)}_3)/2\right], \ldots,$$

$$I_{2k-1} := \left[(\hat{\theta}^{(S)}_{2k-2} + \hat{\theta}^{(S)}_{2k-1})/2,\, b\right].$$

in order to estimate the local extrema. (For nonsmooth asymmetric peaks the estimates by Birge (1997) and Bickel & Fan (1997) are better asymptotically than the initial mode estimates obtained from $\hat{f}_{n,H}$.) The consistency of the NPMLE based on the above two–step estimation procedure of the local extrema can be shown under quite general conditions.

REFERENCES

Birge, L. (1997) Estimation of unimodal densities without smoothness assumptions. Ann. Statist. 25, 970–981.

Bickel, P.J. and Fan, J. (1996) Some problems on the estimation of unimodal densities. Statist. Sinica 6, 23–46.

Grenander, U. (1956) On the theory of mortality measurement II. Skand. Aktuarietidskr. 39, 125–153.

Silverman, B. W. (1981) Using kernel estimates to investigate multimodality. J. R. Stat. Soc. Ser. B 43, 97–99.

Wegman, E.J. (1969) A note on estimating a unimodal density. Ann. Math. Statist. 40, 1661–1667.

Wegman, E.J. (1970a) Maximum likelihood estimation of a unimodal density function. Ann. Math. Statist. 41, 457–471.

Wegman, E.J. (1970b) Maximum likelihood estimation of a unimodal density II. Ann. Math. Statist. 41, 2169–2174.

FRENCH RÉSUMÉ

Nous considérons l'estimation d'une fonction de densité avec k valeurs modales. L'adaption des estimateurs pour les fonctions unimodales paraît impossible considérant le montant des calculs necessaires. Nous proposons deux estimateurs alternatifs pour cette situation.