

IMPLEMENTING SEMIPARAMETRIC DENSITY ESTIMATION

Julian FARAWAY

Department of Statistics, University of Michigan, Ann Arbor, MI 48109-1027, USA

Received July 1989

Revised August 1989

Abstract: A semiparametric estimate of a density may be formed via the convex combination of a parametric and a nonparametric density estimate. It is shown that the some trimming is often necessary to obtain an appropriate proportion of these estimates.

Keywords: Nonparametric density estimation, bandwidth selection, trimming.

1. Introduction

Olkin and Spiegelman (1987) describe a semiparametric approach to density estimation. Their estimate of the density $f(x)$ is given by

$$g(x, \pi) = \pi f(x, \hat{\theta}) + (1 - \pi) \hat{f}(x)$$

where $f(x, \hat{\theta})$ is a parametric estimate of the density where $\hat{\theta}$ may be obtained for example by maximum likelihood and $\hat{f}(x)$ is a nonparametric estimate of the density given for example by a kernel estimator. π is estimated by maximum likelihood subject to $0 \leq \pi \leq 1$. It is hoped that $\hat{\pi}$ will express the suitability of the parametric model. That is to say that $\hat{\pi}$ should be close to 1 if the parametric model is good and close to 0 if it is not. So $\hat{\pi}$ would provide some valuable intuition if not a test statistic as well. Also, if the parametric model were to hold a superior rate of convergence for the semiparametric estimator would be had, whilst not sacrificing consistency if that model did not hold. Schuster and Yakowitz (1985) describe essentially the same method with some applications.

However, there are some difficulties in implementing this idea. Essentially, if the nonparametric density estimate is constructed in a reasonable manner and the parametric density is not sup-

ported on a compact and bounded away from zero then, when the parametric model is true, $\hat{\pi}$ tends to be closer to 0 than to 1 for small to moderate sample sizes. We shall see that this difficulty has little to do with bandwidth selection and is in fact caused by the tails of the density.

2. An example

Olkin and Spiegelman (1987) give data on 20 yearly maximum wind speeds. The proposed parametric model for the data is a Gumbel distribution. The parameters of this distribution may be estimated by maximum likelihood. The nonparametric density estimate may be calculated using a kernel estimator for data X_1, \dots, X_n of the form

$$\hat{f}(x) = \frac{1}{nb} \sum_{i=1}^n K\left(\frac{x - X_i}{b}\right)$$

where K is the standard normal distribution and b is the bandwidth. As Olkin and Spiegelman (1987) remark, $\hat{\pi}$ is sensitive to the choice of b and so b should be chosen in some reasonable way. The most popular method of selecting the bandwidth is least-square crossvalidation. However, this method gives a very small bandwidth in this case which may be a result of the small

sample size and the ties in the data. An alternative is likelihood crossvalidation which gives $b = 0.49s$ where s is the standard deviation of the data. This provides a plausible density estimate when plotted. This results in a $\hat{\pi} = 0$. Olkin and Spiegelman (1987) chose $b = 0.7s$ and obtained $\hat{\pi} \approx 0.8$. Clearly the interpretation will be different.

3. Simulations

In this section, I describe some simulations which illustrate the small to moderate sample size properties of the semiparametric density estimator. Suppose we generate data from the standard normal distribution so that we know the correct parametric model. We use a nonparametric estimator based on the Epanechnikov kernel (for speed of computation) and choose the bandwidth by least-square crossvalidation. For purposes of comparison, we also generate data from a bimodal distribution $\frac{1}{2}N(0.8, 0.36) + \frac{1}{2}N(-0.8, 0.36)$. With 1000 replications the results in Table 1 were obtained giving the average $\hat{\pi}$ for each sample size. The estimated SE's are less than 5%. $\hat{\pi}$ decreases with sample size for the normal case which is the opposite of what one would hope and what would be suggested by the results of Olkin and Spiegel (1987). One should note that the normal distribution does not fulfill the conditions that their results require, among these being that the parametric density is not supported on a compact and bounded away from zero. In fact, good results were obtained when a truncated normal was used as the parametric model. Nevertheless, the normal is an important case so these results are hardly encouraging. However, when the parametric model does not hold $\hat{\pi}$ does at least go to zero fairly swiftly.

Table 1

Sample size	25	50	100	200	400	1000	2000
Normal	0.351	0.372	0.293	0.286	0.242	0.211	0.186
Bimodal	0.310	0.245	0.084	0.026	0.002	0	0

Table 2

Sample size	25	50	100	200	400	1000	2000	5000	10000
Normal	0.595	0.517	0.379	0.301	0.227	0.199	0.176	0.139	0.120
Bimodal	0.421	0.252	0.111	0.024	0	0	0	0	0

One possible explanation for this behaviour is that the method of bandwidth selection is untrustworthy. To investigate this, I tried to eliminate the inherent variability in these methods by choosing the bandwidth so as to minimize the L_1 distance between the true model and the nonparametric estimate for each simulated dataset. Table 2 gives the results of that simulation (based on 1000 replications for samples sizes 25–2000 and 400 for the rest).

The results are not qualitatively different from those above so the difficulties can not be attributed to variability in the bandwidth selection methods since this method is the best one can do.

However, a closer investigation of the maximum likelihood estimation of π reveals that the estimate is highly sensitive to the contribution from data in the tails. To demonstrate this I calculated $\hat{\pi}$ this time excluding 5% of the data from each tail from the likelihood function. Under the same conditions as above Table 3 was obtained.

This gives results like one would hope. In fact, there would appear to be some scope for letting the amount of trimming decrease as the sample size increases. However, for any given dataset, I would prefer simulation from the proposed parametric model to determine the appropriate amount of trimming. Returning to the wind speed data set discussed earlier, $\hat{\pi}$ was recalculated, excluding the smallest and the largest observations. $\hat{\pi} = 0.85$ was obtained which is more believable result than the $\hat{\pi} = 0$ calculated earlier.

Table 3

Sample size	25	50	100	200	400	1000	2000	5000	10000
Normal	0.802	0.935	0.965	0.981	0.982	0.977	0.982	0.997	0.990
Bimodal	0.530	0.445	0.337	0.176	0.045	0	0	0	0

4. Conclusion

The semiparametric density estimator as considered here must be used with care. If the proposed parametric density has tails, trimming must be performed, otherwise inappropriate values of $\hat{\pi}$ will be obtained. Failure to do this may lead to poor conclusions regarding the suitability of the parametric model and a poorly performing semiparametric estimate of the density.

Acknowledgements

I acknowledge Dr. Clifford Spiegelman for a helpful conversation.

Reference

- Olkin, I. and C. Spiegelman (1987), A semiparametric approach to density estimation, *J. Amer. Statist. Assoc.* **82**, 858–865.
- Schuster, E. and S. Yakowitz (1985), Parametric/nonparametric mixture density estimation with application to flood frequency analysis, *Water Resources Bull.* **21**, 797–804.