Estimation of the density ratio of two distributions

Toshinari Kamakura
Chuo University
1-13-27 Kasuga, Bunkyo-ku
Tokyo 112-8551, Japan
kamakura@indsys.chuo-u.ac.jp

1. Introduction

This work in this paper is motivated by the spacial distributions of two different types of points. For example, we are conserned with differences of distributions two types fast-food franchise shops scattered on some area compared with distances from the nearest railroadstation. In this study it is a very interesting to estimate the ratio of two distributions from the nearest stations.

Kelsall and Diggle (1995) considers this type of problem by estimating a relative risk function using a ratio of two kernel density estimates, concentrating on the problem of choosing the smoothing parameters and also a croos-validation method is proposed.

Suppose tha f and g are probability density functions defined within an interval $I \subset R$ and from each distribution we have samples x_i,$i = 1, \ldots, n_1$, and y_j,$j = 1, \ldots, n_2$. Let

$$\gamma(x) = \log \frac{f(x)}{g(x)}.$$

Kelsall and Diggle (1995) estimates nonparametrically the densities using the kernel function $K(x)$;

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - z_i}{h}\right),$$

where n is n_1 or n_2 and z_i is x_i or y_j for the fisrt or second sample.

2. Choice of smoothing parameters

In kernel density estimation it is not easy to choose the smoothing parameter automatically even for only one distribution. Silverman (1985) describes in detail the methods of choosing the smoothing parameter: subjective choice by plotting out several curves and choosing the estimate that is most in accordance with prior ideas about the density, asymptotically optimal window techniques,least-squares cross-validations, and likelihood cross-validation.

Kelsall and Diggle (1995) proposes the method which minimize the integrated squre error (ISE):

$$ISE\left\{\hat{\gamma}_{h1,h2}(x)\right\} = \int_I \left\{\hat{\gamma}_{h1,h2}(x) - \gamma(x)\right\}^2 dx.$$

They devise cross-valication technique using the usual method of 'leave-one-out' averagin in place of expectation and give recommendations of constraining the bandwidths to be equal considering the underlying two densities to be nealy equal in epidemiological applications. However in our spatial distributions of stores they are sometimes much different. Then we are required to minimize the approximation function of the integrated squre error for both smoothing parameters jointly. When we calculate the ratio of distributions, we encounter the difficulties for very small values of tail distributions. In this case we may succeed in estimation of density ratio using the appropriate weight function which has very light tail.

$$WISE\left\{\hat{\gamma}_{h1,h2}(x)\right\} = \int_I \left\{\hat{\gamma}_{h1,h2}(x) - \gamma(x)\right\}^2 w(x)dx.$$

3. Other restrictions

In the previous section we pointed out the requirements of estimation of two smoothing parameters jointly and less weighting the tail distributions. When we would like to detect location shifting of distribution. It is very interesting to estimate the ratio of distribution under the following restrictions:

$$\frac{\hat{f}_k}{\hat{g}_k} \leq \frac{\hat{f}_{k+1}}{\hat{g}_{k+1}} \qquad (k = 1, \ldots, n_1 + n_2 - 1)$$

where $\hat{f}_k$ and $\hat{g}_k$ are nonparametric estimates of f and g, respectively, which are corresponding to $k - th$ order statistic of jointly sorted sample.

4. Discussion

We considered the problems of estimating the ratio of two distributions and pointed out that we shall handle two smoothing parameter jointly and cautiously calculate tail distributions. It is mometimes very useful that to use the weighting integrated square error as the objective function and order restrictions.

REFERENCES

Kelsall, J. E. and Diggle, P. J. (1995). Kernel estimation of relative risk. *Bernoulli*, **1**, 3-16.

Kelsall, J. E. and Diggle, P. J. (1998). Spatial variation in risk of disease: a nonparametric binary regression approach. *Appl. Statist.*, **47**, 559-573.

Silverman, B. W. (1985). *Density estimation for statistics and data analysis*, Chapma & Hall, London.