한국통계학회논문집 제2권 2호
1995년 12월 pp. 227-230

Utilizing Order Statistics in Density Estimation[1]

W.C. Kim and B.U. Park[2]

Abstract

In this paper, we discuss simple ways of implementing non-basic kernel density estimators which typically need extra pilot estimation. The methods utilize order statistics at the pilot estimation stages. We focus mainly on variable location and scale kernel density estimator (Jones, Hu and McKay, 1994), but the same idea can be applied to other methods too.

Key words and phrases. Kernel smoothing, reduced bias.

1. Introduction.

The kernel density estimator

$$\hat{f}(x)=n^{-1}\sum_{i=1}^{n}h^{-1}K\{h^{-1}(x-X_i)\} \tag{1}$$

of a density f based on an i.i.d. sample $X_1,X_2,\cdots,X_n$ has bias of order h^2 as $h=h(n)\rightarrow 0$, and variance of order $(nh)^{-1}$ as $n\rightarrow\infty$ and $nh\rightarrow\infty$. Here, h is the smoothing parameter that controls the degree of smoothing applied to the data, and K is the kernel function which will be taken to be a symmetric probability density hereafter. Good discussion of many important practical aspects of kernel density estimation may be found in Silverman (1986).

There have been many proposals for improving the bias of kernel density estimators. Among them, variable bandwidth kernel density estimation (Breiman, Meisel and Purcell, 1977, and Abramson, 1982) is the most long lived one. The idea is to use varying degrees of smoothing across the $X-$ space, i.e. replace the constant h in (1) by $b(X_i)$, say. It turns out that it affords bias of order h^4 if $b(x)$ were taken to be $h/f^{1/2}(x)$. Samiuddin and El-Sayyad (1990) have proposed varying the location of each kernel, in contrast to varying its

---

1 This research was supported by the Basic Research Institute Program, Ministry of Education, 1994 Project No. BSRI-94-1418.

2 Department of Computer Science and Statistics, Seoul National University, Seoul, 151-742, KOREA

scale. They showed that adding (constant) $\times h^2 f'(X_i)/f(X_i)$ to each X_i ensures $O(h^4)$ bias. An intuitive explanation of how it improves the bias is that it moves each data point a little bit in the direction of increasing density and this helps out with capturing features like modes in the true density function.

Jones, Hu and McKay (1994) put the aforementioned ideas in more general framework. They describe a variety of ways in which scale and location variation can be combined to good theoretical effect. However, all these estimators need extra pilot estimation of f since the varying location and scale depend on f and its derivatives. In this paper, we describe two simple ways of utilizing order statistics to estimate these f dependent quantities. The advantage of these methods is that it does not need extra smoothing for the pilot estimation and is very easy to implement.

2. The methods.

The general formula which Jones, Hu and McKay (1994) considered is

$$n^{-1}\sum_{i=1}^{n}\alpha^{-1}(X_i)K\{\alpha^{-1}(X_i)(x-X_i-\beta(X_i))\}. \tag{2}$$

If f were known at the pilot stage, one of the many choices in Jones, Hu and McKay (1994) that affords bias is

$$\alpha(X_i)=h/f(X_i),\quad \beta(X_i)=-h^2 f'(X_i)/(2f^3(X_i)) \tag{3}$$

provided $\int x^2K(x)\,dx=1$. In what follows, we describe two simple ways of estimating the above α and β's.

Method 1. Let F denote the cumulative distribution function of f. Then the first two derivatives of the quantile function $Q(u)=F^{-1}(u)$ are

$$Q'(u) = \{F^{-1}(u)\}' = 1/f\{F^{-1}(u)\},$$

$$Q''(u) = \{F^{-1}(u)\}'' = -f'\{F^{-1}(u)\}/f^3\{F^{-1}(u)\}.$$

Note that these derivatives have the same functional forms as those of α and β's in (3). This observation essentially reduces the estimation of the unknown quantities to that of the derivatives of the quantile functions. In particular, one can use estimates of $Q'(i/n)$ and $Q''(i/n)$ to estimate $1/f(X_{(i)})$ and $-f'(X_{(i)})/f^3(X_{(i)})$ where $X_{(i)}$ denotes the ith order statistic. Now, $Q'(i/n)$ can be estimated by

$$\widehat{Q}'(i/n) = \frac{n}{2s}(X_{(i+s)} - X_{(i-s)})$$

using the difference between two order statistics whose indices are $2s$ apart. This is sometimes called Siddiqui-Bloch-Gastwirth estimate and its asymptotic MSE properties are well known (see Hall and Sheather, 1988). The idea behind this estimate is that $Q'(i/n)$ can be approximated by $\{Q((i+s)/n) - Q((i-s)/n)\}/(2s/n)$. This idea can be extended for the estimation of $Q''(i/n)$. By 'twice differencing' of order statistics, we obtain

$$\widehat{Q}''(i/n) = \frac{n^2}{s^2}(X_{(i+s)} - X_{(i)} + X_{(i-s)}).$$

The estimator of f would then be

$$\hat{f}(x) = n^{-1}\sum_{i=1}^{n}(h\,\widehat{Q}'(i/n))^{-1}K\{(h\,\widehat{Q}'(i/n))^{-1}(x - X_{(i)} - h^2\,\widehat{Q}''(i/n)/2)\}.$$

Method 2. For the second method, let c_i and s_i denote the mean and standard deviation, respectively, of the ith order statistic $X_{(i)}$. Then two standard asymptotic results from the theory of order statistics (see page 80 of David, 1981, for example) are as follows. First,

$$c_i \simeq Q_i + \frac{p_i q_i}{2(n+2)} Q_i'' = Q_i - \frac{p_i q_i}{2(n+2)} \frac{f'(Q_i)}{f^3(Q_i)} \tag{4}$$

where $p_i = i/(n+1), q_i = 1 - p_i, Q_i = F^{-1}(p_i)$ and $Q_i'' = Q''(u)|_{u=p_i}$. Second

$$s_i^2 \simeq \frac{p_i q_i}{(n+2)}(Q_i')^2 = \frac{p_i q_i}{(n+2)}\{f(Q_i)\}^{-2}. \tag{5}$$

Write $\widehat{c_i}$ and $\widehat{s_i}$ for simple estimates of c_i and s_i. For example, the bootstrap estimates of Hall and Martin (1988) might be used. Then (4) and (5) suggest the following estimates of α and β's:

$$\widehat{\alpha}(X_{(i)}) = h_i\,\widehat{s_i} \quad \widehat{\beta}(X_{(i)}) = -h_i^2(\widehat{c_i} - X_{(i)})$$

where $h_i = h\{(n+2)/p_i q_i\}^{1/2}$. Plugging the above estimates into (2) yields another fully practical density estimator

$$\hat{f}(x) = n^{-1}\sum_{i=1}^{n}(h_i\,\widehat{s_i})^{-1}K\{(h_i\,\widehat{s_i})^{-1}(x - X_{(i)} - h_i^2(\widehat{c_i} - X_{(i)}))\}.$$

The major advantage of the above two methods is that they do not need extra smoothing, as opposed to other non-basic kernel estimators. Detailed investigations into their theoretical and practical performance are left for further study, however.

REFERENCES

[1] David, H.A. (1981). *Order Statistics*, 2nd edition, John Wiley & Sons, Inc.

[2] Hall, P. and Martin, M.A. (1988). Exact convergence rate of bootstrap quantile variance estimator, *Probability Theory and Related Fields*, Vol. 80, 261-268.

[3] Hall, P. and Sheather, S.J. (1988). On the distribution of a studentized quantile, *Journal of Royal Statistical Society*, Series B, 50, 381-391.

[4] Jones, M.C., Hu, T.-C. and McKay, I.J. (1994). Variable location and scale kernel density estimation, *Annals of Institute of Statistical Mathematics*, Vol. 46, 521-535.

[5] Samiuddin, M. and El-Sayyad, G.M. (1990). On nonparametric kernel density estimates, *Biometrika*, Vol. 77, 865-874.

[6] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall.