A STUDY ON RELATIVE EFFICIENCY OF KERNEL TYPE ESTIMATORS OF SMOOTH DISTRIBUTION FUNCTIONS

JEE EUN SOOK

1. Introduction

Let P be a probability measure on the real line with Lebesque-density f . The usual estimator of the distribution function ($\equiv$ df) of P for the sample $x_1, \cdots, x_n$ is the empirical *df*:

$$F_n(t) = n^{-1} \sum_{i=1}^{n} 1_{[-\infty,t]}(x_i).$$

But this estimator does not take into account the smoothness of F, that is, the existence of a density f. Therefore, one should expect that an estimator which is better adapted to this situation beats the empirical df with respect to a reasonable measure of performance.

Since we assume a nonparametric situation an estimator of the form:

$$\hat{F}_n(t) = \int_{-\infty}^{t} f_n(y)dy,$$

where f_n is an estimator of the density f, suggests itself. We shall study such $\hat{F}_n$ for kernel-type density estimators f_n.

Kernel-type estimators of a density were introduced by ROSENBLATT (1956) and PARZEN(1962). They are of the form:

$$f_n(t) := (n\alpha_n)^{-1} \sum_{i=1}^{n} k((t - x_i)/\alpha_n),$$

Typeset by $\mathcal{AMS}$-TEX

where the kernel k fulfills $\int k(x)dx = 1$ and the so-called bandwiths or window-withs $n \in \boldsymbol{N}$, are positive numbers which tend to zero as n increases. The use of kernels k which may take negative values was suggested by BARTLETT (1963).

This estimator is motivated by the law of large numbers, which implies:

$$f_n(t) \sim \alpha_n^{-1} \int (k(t-x)/\alpha_n) f(x)dx$$

if n is large, and the fact that under appropriate conditions on f and k:

$$\alpha_n^{-1} \int k((t-x)/\alpha_n) f(x)dx \rightarrow_{\alpha_n \rightarrow 0} f(t)$$

which is BOCHNER's Theorem.

Because of their simple forms as sums of independent identically distributed random variables there has been a great interest in kernel density estimators. See for example the survey articles by SCOTT et al. (1977) and WERTZ (1978). One should further mention that besides their appealing concept rigorous arguments recommending kernel density estimates over computing procedures were given by FARRELL(1972) and MEYER(1977).

Define therefore the estimator $\hat{F}_n$ of F for the sample $x_1, \cdots, x_n$ by:

$$\hat{F}_n(t) := \int_{-\infty}^{t} f_n(t)dt = n^{-1} \sum_{i=1}^{n} K((t-x_i)/\alpha_n)$$

where

$$\alpha_n > 0 \text{ and } K(x) := \int_{-\infty}^{x} k(y)dy, \quad \int k(x)dx = 1$$

Several properties of $\hat{F}_n$ are well-known for a number of years. Its uniform convergence to F almost sure, for example, was stated by NADARAYA (1964), WINTER(1973) and YAMATO (1973). Moreover, WATSON and LEADBETTER(1964, Theorem 6) proved the asymptotic normality of $\hat{F}_n$, and WINTER (1979) showed that it has the CHUNG-SMIRNOV property, that is,

$$\limsup_{n \in \boldsymbol{N}} (2n/\log \log n)^{\frac{1}{2}} \sup_{t \in \boldsymbol{R}} |\hat{F}_n(t) - F(t)| \leq 1$$

almost sure.

But the question whether $\hat{F}_n(t)$ is a better estimator of $F(t)$ than the empirical df remained open since it was partially answered by REISS (1981) and AZZALINI (1981). A complete solution is given by the present article.

2. Relative efficiency and deficiency.

Let $m \geq 1$ and t be a fixed real number. Assume that $F(t) \in (0,\ 1)$ and that the dfF fulfills the following smoothness condition at the point t:

$$|F(x) - \sum_{i=0}^{m} \alpha_i (x-t)^i| \leq A|x-t|^{m+1} \tag{1}$$

for every real x where α_i, $i = 0,\ 1,\ \cdots,\ m$, are real numbers and $A > 0$. Obviously, $a_{ij} = F(t)$ and $a_1 = f(t)$. Further, if the kernel k has a bounded support, then it suffices, throughout the article, to assume that (1) holds for all x in a neighborhood of t.

We shall consider the following class of kernels:

$$K_m := \{k \in l_1(\boldsymbol{R}) : \int k(x)dx = 1,$$
$$\int x^t k(x)dx = 0,\ i = 1,\ \cdots,\ m,\ \int |x^{m+1}k(x)|dx < \infty\}$$

where $L_p(I) := \{g : I \to \boldsymbol{R},\ \int |g(x)|^p dx < \infty\}$ for $p > 0$ and I being an interval in $\mathbf{R}$. Put $g(x) := 2k(x)K(x)$ and $\mu_n := \int \hat{F}_n(t)dp^n$. If $m \geq 2$ it is clear that k cannot fulfill the condition $k \geq 0$. However, the bias $\mu_n - F(t)$ is reduced in absolute value with m increasing. Next we define the concept within which we shall compare $\hat{F}_n$ and F_n.

It was alredy proved by NADARAYA ((1964). Theorem 1) that $\hat{F}_n(t)$ has asymptotically the same mean and variance as $F_n(t)$. Thus, a sharper result is necessary in order to be able to distinguish the performance of these estimators.

To this end we denote by $i(n)$ the sample size which is needed so that the empirical df has the same (or a smaller) mean square error as $\hat{F}_n(t)$, that is

$$i(n) := \min\{k \in \boldsymbol{N} : \int (F_k(t) - F(t))^2 dp^k \leq \int (\hat{F}_n(t) - F(t))^2 dp^n\}$$

where P^n denotes the n-fold independent product of P.

Now, a comparison of the two estimators $\hat{F}_n(t)$ and $F_n(t)$ is made by comparing i(n) with n. This can be carried out by considering the ratio i(n)/n or the difference i(n)-n. The ratio i(n)/n is usually called relative or first order efficiency while the difference i(n)-n is known as second order efficiency of deficiency. The limit values $\lim_{n\in N} i(n)/n$ and $\lim_{n\in N}(i(n)-n)$ are called, if they exist, asymptotic relative efficiency and asymptotic deficiency, respectively. The concept of deficiency was introduced by HODGES and LEHMANN (1970) while earlier comparisons had been based mainly on the ratio i(n)/n

The following representation of the mean square error of $F_n(t)$ is essential to our considerations:

$$\begin{aligned} n&\int(\hat{F}_n(t)-F(t))^2dp^n \\ &=(n-1)(\mu_n-F(t))^2+\int K((t-x)/\alpha_n)^2P(dx)-F(t) \\ &\quad -2F(t)(\mu_n-F(t))+F(t)(1-F(t)) \end{aligned} \tag{2}$$

This leads to the following lemma, which is immediate from integration by parts and formula (2.2) in REISS (1981).

Lemma 1. *Let F be an absolutely continuous df which fulfills condition (1). Then, if* $\alpha_n \to_{n\in N} 0$,

$$\begin{aligned} &\int(\hat{F}_n(t)-F(t))^2dp^n/\int(F_n(t)-F(t))^2dp^n \\ &=1+<n(\mu_n-F(t))^2-\alpha_n(f(t)\int b(x)xdx> \;/F(t)(1-F(t))+O(\alpha_n^2) \end{aligned}$$

From Lemma 1 one can easily deduce the following result which states that a kernel type estimator of the df F with fixed kernel does not have an asymptotically better performance on the level of efficiency than the empirical df. This was to be expected, for example, by Theorem 1 in NADARAYA (1964).

Proposition 1. *Let P be a probability measure with absolutely continuous df F. Assume that F fulfills condition (1) for some* $m \geq 1$ *at a fixed point t. Let* $\alpha_n \to_{n\in N} 0$, *Then, for every* $k \in K_m$:

(i) $\lim_{n\in N} i(n)/n \leq 1$

and

(ii) $\lim_{n\in N} i(n)/n \leq 1 \quad iff \quad \lim_{n\in N} n(\mu_n - F(t))^2 = 0$

Since $|\mu_n - F(t)| = O(\alpha_n^{m+1})$ *(see formula (2.2) in REISS (1981)), we can always ensure that* $\lim_{n\in N} n(\mu_n - F(t))^2 = 0$ *by letting* α_n *quickly tend to zero. Then Proposition 1 implies* $\lim_{n\in N} i(n)/n = 1$, *no matter, which particular kernel is considered. Thus, at this point of investigations we cannot distinguish between the asymptotic performance of different kernels when compared to the empirical df. To this end we have to do considerations on the level of relative deficiency.*

It was shown by REISS (1981) that under the above conditions on F and if $\int xk(x)K(x)dx > 0$, one can choose α_n, $n \in \boldsymbol{N}$, such that:

$$i(n) - n \geq cn^{1-1/(2m+1)}$$

Here c is a positive constant which depends on F and t. Thus, the relative deficiency of $F_n(t)$ with respect to $\hat{F}_n(t)$ quickly tends to infinity as the sample size increases. A similar result for the case m=1 has been stated independently by AZZALINI (1981).

Thus, our results enable us to decide exactly whether a given kernel estimator should be prefered to the empirical *df* simply by computing the sign of $\int xk(x)K(x)dx$.

REFERENCES

1. AZZALINI, A., *A note on the estimation of a distribution function and quantiles by a kernel method*, Biometrika **68** (1981), 326-328.
2. BARTLETT, M.S., *Statistical estimation of density functions*, Sankhya Ser. A **25** (1963), 245-254.
3. FARRELL, R.H., *On the best obtainable asymptotic rates of convergence in estimation of a density functio at a point*, Ann. Math. Statist. **43** (1972), 170-180.
4. HODGES, J.L. and E.L. LEHMANN, *Deficiency*, Ann. Math. Statist. **41** (1970), 783-801.
5. MEYER, T.G., *Bounds for estimation of density functions and their derivatives*, Ann. Statist. **5** (1977), 136-142.
6. NADARAYA, E.A., *Some new estimates for distribution functions*, Theory Prob. Appl. **9** (1964), 497-500.
7. PARZEN, E., *On the estimation of a probability density and mode*, Ann. Math. Statist. **33** (1962), 1065-1076.
8. REISS, R.D., *Nonparametric estimation of smooth distribution functions*, Scand. J. Statist. **8** (1981), 116-119.

9. ROSENBLATT, M., *Remarks on some non-parametric estimates of a denwity function*, Ann. Math. Statist. **27** (1956), 832-837.
10. SCOTT, E.W., R.A. TAPIA and J.R. THOMPSON, *Kernel density estimation revisited*, Non-linear Anal., Theor. Math. Appl. **1** (1977), 339-372.
11. WATSON, G.S. and M.R. LEADBETTER, *Hazard analysis 11*, Sankhya Ser. A. **26** (1964), 101-116.
12. WERTZ, W., *Statistical density estimation*, A survey, Vandenhoeck and Ruprecht, Gottignen. (1978).
13. WINTER, B.B., *Strong uniform consistency of integrals of density estimators*, Canad. J. Statist. **1** (1973), 247-253.
14. YAMATO, H., *Uniform convergence of an estimator of a distribution function*, Buli. Math. Statist. **15** (1973), 69-78.

Kwangwoon University