CONSISTENT ESTIMATOR OF SOME FUNCTIONALS OF DENSITY

Brigida A. Roscom*
& Daniel C. Bonzo**

ABSTRACT

Nonlinear functionals of unknown density f, such as the integral of the square of f, $T(f) = \int f^2 dx$ and the Shannon entropy

$$I(f) = -\int f \log f \, dx$$

are estimated using kernel estimates f_n of the probability density f. It is shown that the strong consistent property of the estimates of these functionals follows from the conistency property of f_n. Illustrations of the convergence are given using simulated density functions.

Keywords and phrases. Nonlinear functionals, asymptotic relative efficiency(ARE), kernel, kernel density estimators, consistent estimator, smoothing parameter, simulation, rate of convergence.

1. Introduction

Some nonlinear functionals of density are of practical interest in nonparametric statistical inference. In comparing two statistical procedures, the asymptotic relative efficiency (ARE) is used. This ARE depends on the efficacy parameters defined in terms of the underlying probability density of the data. One of these efficacy related functionals is $T(f) = \int f^2 dx$, which is included in Pitman asymptotic efficacy of tests based on Wilcoxon scores. The functional T(f) also appears in the standardizing constant of test statistics in linear models. Thus, it is important to have a consistent estimator to standardize certain test statistics. The Shannon entropy $I(f) = \int -f \log f \, dx$ for both discrete and absolutely continuous distributions has applications to information theory and engineering sciences.

Demetriev and Tarasenko (1972), Schuster (1974) Ahmad (1976), Ahmad and Lin (1976) and Ching and Serfling (1976) have shown convergence of various estimates of the functionals to T(f) and I(f). In this paper, we will show a consistent estimate for T(f) and I(f) using a basic lemma based on the asymptotic unbiasedness of the kernel estimate f_n as estimator of f. This consistency result is graphically illustrated using simulated density function of a mixture of two normally distributed populations. Two cases of kernel estimators of

* Associate Professor of Mathematics, Iligan Institute of Technology, Mindanao State University.

** Instructor in Statistics, Statistical Center, University of the Philippines, Diliman.

density are presented using Epanechnikov and normal density functions.

2. Kernel Density Estimates

Motivated by the notion of histograms, Rosenblatt (1956) and Parzen (1962) worked on kernel density estimates. In the first detailed explication of kernel estimate, Parzen considered as an estimator for an unknown density f(x),

$$(1.1) \qquad \hat{f}(x) = f_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{a_n} K\left[\frac{x-X_i}{a_n}\right] =$$

$$\int_{-\infty}^{\infty} \frac{1}{a_n} K\left[\frac{x-y}{a}\right] dF_n(y)$$

with K, the kernel function, satisfying

$$(1.2) \quad K(t) \geq 0, \ \sup |K(t)| < \infty, \ \int_{-\infty}^{\infty} K(t)dt = 1$$

and $\lim_{y \to \infty} |t\, K(t)| = 0$

Some choice for the function K, may come from the following:

(i) the standard normal

$$K(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2},$$

(ii) the triangular function;
$K(t) = 1 - |t|$ for $|t| < 0$,
0, otherwise and

(iii) the Epanechnikov function;

$$K(t) = (3/4)(1-(1/5)t^2)/\sqrt{5} \quad \text{for } |t| < 5,$$
$$0 \text{ otherwise.}$$

Below, we state without proof the asymptotic unbiasedness of f_n due to Parzen.

Lemma 1. (Parzen). The estimates defined by (1.1) are asymptotically unbiased at all points, x at which the probability density function is continuous if the function K(t) satisfies all the conditions in (1.2) and (1.3) $\lim_{n \to \infty} a_n = 0$.

3. Strong Consistency of $T(f_n)$ and $I(f_m)$

In this paper, we show the strong consistency property of the estimates $T(f_n) = \int f_n^2\, dx$ and $I(f_n) = \int - f_n \log f_n\, dx$ for the functionals $T(f) = \int f^2 dx$ and

$$I(f) = \int -f \log f\, dx,$$

respectively, using the following lemma.

Lemma 2. Let (Ω, A, P) be a probability space and $\{Z_n\}$ be a sequence of random variables that are in $L^1(\Omega, A, P)$. Let $\{Z_n\}$ be uniformly bounded, i.e., $|Z_n(x)| \leq M$ for all n and for all x. Suppose Z is in $L^1(\Omega, A, P)$ and has compact support, then $Z_n \to Z$ a.s. [P] as $n \to \infty$ if and only if Z_n converges to Z in L^1.

Proof. Let Z_n be a sequence of random variables. Z_n

converging to Z in L^1 implies, by definition, that

$$\lim_{n\to\infty} E|Z_n - Z| = 0 \qquad (1.4)$$

Now, by the dominated convergence theorem (DCT)

$$\int_\Omega \lim |Z_n - Z| \, dP = 0 \qquad (1.5)$$

if and only if (1.4) is true. Also,

$$\lim_{n\to\infty} |Z_n - Z| = 0 \text{ a.s. } [P] \qquad (1.6)$$

if and only if (1.5) holds. Hence

$$Z_n \to Z \text{ a.s. } [P] \qquad (1.7)$$

if and only if (1.6) is satisfied.

Theorem 1. For a suitable choice of the kernel function K and smoothing parameter a_n satisfying the Parzen conditions in (1.2) and assuming that f is bounded, then f_n converges to f with probability 1 if the kernel density estimates $\{f_n(x)\}$ in (1.1) are uniformly bounded, that is, $|f_n(x)| \le M$ for all n, for all x.

Proof. By lemma 1. $\lim E|f_n - f| = 0$. In view of Lemma 2 the kernel density estimates f_n converges almost surely to f.

Let $T(f_n) = \int f_n^2 dx$ and

$I(f_n) = \int -f_n \log f_n dx$ be a

kernel based estimates for T(f) and I(f), respectively, where f_n is the kernel estimate satisfying the Parzen condition in (1.2) and (1.3).

We formalize the strong consistency results for $T(f_n)$ and $I(f_n)$ with the following theorems:

Theorem 2. Under the conditions in Theorem 1,

$$\lim_{n\to\infty} E\,|f_n^2 - f^2| = 0 .$$

Proof. By the dominated convergence theorem (DCT),

$$\lim_{n\to\infty} E\,|f_n^2 - f^2|$$

$$= E \lim_{n\to\infty} |f_n^2 - f^2|$$

$$= E\,|\lim_{n\to\infty} f_n^2 - f^2|$$

$$= E\,|(\lim_{n\to\infty} f_n)^2 - f^2|$$

$= 0$ by corollary 1.

Theorem 3. Under the conditions in Theorem 1. $T(f_n)$ converges to T(f) in L^1.

Proof.

$$\lim_{n\to\infty} E\,\left| \int_a^b f_n^2 dx - \int f^2 dx \right| \le$$

$$\lim_{n\to\infty} E \int_a^b |f_n^2 - f^2| dx$$

$$= \lim_{n\to\infty} \int_a^b (E\,|f_n^2 - f^2|) dx,$$

by Fubini's theorem

$$= \int_a^b \lim_{n\to\infty} E\,|f_n^2 - f^2| dx,$$

by the dominated convergence theorem

$= 0$,by Theorem 2.

Theorem 4. Under the conditions in Theorem 1, $T(f_n)$

converges to $T(f)$ with probability 1.

Proof. Applying Lemma 2 for $Z_n = T(f_n)$ and with the use of Theorem 2, Theorem 3 is proved.

Theorem 5. Under the conditions in Corollary 1,

$$\lim_{n\to\infty} E\,|f_n \log f_n - f \log f| = 0$$

Proof. By the dominated convergence theorem (DCT),

$$\lim_{n\to\infty} E\,|f_n \log f_n - f \log f| =$$

$$E \lim_{n\to\infty} |f_n \log f_n - f \log f|$$

$$= E|\lim_{n\to\infty} f_n \log f_n - f \log f|$$

$= E|f \log f - f \log f|$, by corollary 1

$$= 0$$

Theorem 6. Under the conditions in Theorem 1, $I(f_n)$ converges to $I(f)$ in L^1.

Proof.

$$\lim_{n\to\infty} E|I(f_n)-I(f)| \leq$$

$$\lim_{n\to\infty} E \int |f_n \log f_n dx - f \log f|\,dx$$

$= \lim_{n\to\infty} \int E|f_n \log f_n - f \log f|\,dx$, by Fubini's theorem

$= \int (\lim_{n\to\infty} E\,|f_n \log f_n - f \log f|\,dx$, by DCT

$= 0$ by Theorem 4.

Finally, Theorem 7, will show the strong consistency of $I(f_n)$.

Theorem 7. Under the conditions in Theorem 1, $I(f_n)$ coverges to $I(f)$ with probability 1.

Proof. We apply Lemma 2 for $Z_n = I(f_n)$ and use Theorem 6.

4. Illustrations on the Strong Consistency Results of $T(f_n)$ and $I(f_n)$

4.1. Simulation Procedure

To demonstrate the results empirically, a simulation from a bimodal distribution was made. Box-Muller's method was used to generate data points from a mixture of two normals: N(-3, 1) with weight 0.3 and N(5, 1) with weight 0.7. For a fixed bandwidth a_n and sample size n, estimates of the functionals of f for 50 data points with equal increments in the range -6 to 8 were then computed using normal and Epanechnikov kernels. a_n and n were made to vary to demonstrate their effect on the rate of convergence; $n^{-1/6}$ and $n^{-1/4}$ were chosen for the bandwidth a_n, and 50, 100, and 200 for the sample size n.

4.2. Analysis of Results

Numerical integration was resorted to approximate $\int f^2 dx$ and $\int -f \log f\,dx$ over the region -6 to 7.2. The approximation yielded 0.81795 and +0.43812 for $\int f^2 dx$ and $\int -f \log f dx$, respectively.

Approximation was to get $\int f_n^2 dx$ and $\int -f_n \log f_n dx$, the results of which are shown below.

Table 1. Values of $\int f_n^2 dx$

	Using Std Normal a_n		Using Epanechnikov a_n	
n	$n^{-1/6}$	$n^{-1/4}$	$n^{-1/8}$	$n^{-1/4}$
50	0.119738	0.129727	0.130017	0.130017
100	0.109859	0.116749	0.115843	0.115843
200	0.104161	0.109929	0.108028	0.108028

Table 2. Values of $\int -f_n \log f_n dx$

	Using Std Normal a_n		Using Epanechnikov a_n	
n	$n^{-1/6}$	$n^{-1/4}$	$n^{-1/6}$	$n^{-1/4}$
50	0.84510	0.80997	0.81108	0.81108
100	0.88113	0.85175	0.84973	0.84973
200	0.91825	0.88521	0.88678	0.88521

From Table 1, it is evident that the integral of the estimates converges to the true value both for the standard normal and Epanechnikov kernels. For the standard normal kernel, the rate of convergence increases as the bandwidth decreases. In the case of the Epanechnikov kernel, the rate of convergence is invariant. Hence it is better to use the Epanechnikov kernel for smaller bandwidth. Table 2 presents an unstable pattern of convergence of the integral of the estimates. Both kernels provide increasing estimates of the integral irrespective of the bandwidth used. However, when the bandwidth is increased, marked decrease in instability is observed. In this case, the Epanechnikov provides a more conservative estimate.

A look at the graph can help explain the reason for the instability. The integral converges more rapidly in the interval (-6, 2.4) than in the interval (2.4, 7.72) which accounts for the greater mass of $l(f_n)$, hence the instability of the estimate. Note also that since the logarithm function has decreasing derivative, this type of functional estimate converges much slower than a polynomial type, which has an increasing derivative.

For purposes of estimating $\int f^2 dx$, a sample size of 200 is recommended with smaller bandwidth for the standard normal kernel and larger bandwidth for the Epanechnikov kernel. More points should be used in estimating $\int -f \log f dx$ to cover up the slow convergence between 2.4 to 7.72.