King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 090

November 1986

On Asymptotic Properties of an Estimate of a Functional of a Probability Density

Khaled I. Abdul-Al

DHAHRAN 31261 ● SAUDI ARABIA ● www.kfupm.edu.sa/math/ ● E-mail: mathdept@kfupm.edu.sa

ON ASYMPTOTIC PROPERTIES OF AN ESTIMATE OF A FUNCTIONAL OF A PROBABILITY DENSITY

Khaled I. Abdul-Al
Department of Mathematical Sciences
University of Petroleum & Minerals
Dhahran, Saudi Arabia

Abstract

Bhattacharyya & Roussas (1969) proposed an estimate of the functional $\Delta = \int f^2(x)dx$ by $\tilde{\Delta} = \int f_n^2(x)dx$ where $f_n(x)$ is a kernel estimate of the probability density $f(x)$. Schuster (1974) proposed an alternative estimate $\hat{\Delta} = \int f_n(x)dF_n(x)$ of Δ, where $F_n(x)$ is the sample distribution function, and showed that the two estimates attain the same rate of strong convergence to Δ. Ahmad (1976) presented two large sample properties of $\hat{\Delta}$; first being the strong convergence of $\hat{\Delta}$ to Δ, and second is the asymptotic normality of $\hat{\Delta}$. In this note, it is proposed to estimate $\theta = E[\gamma(X)] = \int \gamma(x)f(x)dx$ by $\theta_n = \int \gamma(x)f_n(x)dx$, and show the weak and strong convergence of θ_n to θ and establish the asymptotic normality of θ_n.

AMS Subject Classification: 62G05

Keywords: Density Estimation, Characteristic Function, Density Functional.

1. Introduction

Let X be a random variable with distribution function (d.f.) $F(x)$ and probability density function (p.d.f.) $f(x)$, and let the functional be defined as

$$\theta = \int \gamma(x) f(x)\, dx \qquad (1,1)$$

where $\gamma(X)$ is real measurable function of random variable X.

The functional θ is important in many estimation problems as the estimate of the characteristic function $\phi(t)$, moments of any order and any mathematical expectations of the form $E[g(X)]$ when $f(x)$ is unknown.

Let $X_1, \ldots, X_n$ be identically independent distributed (iid) random variables with d.f. $F(x)$ and p.d.f. $f(x)$. Let $k(u)$ be a known symmetric p.d.f. satisfying the following condition:

$$\sup_{-\infty<u<\infty} k(u) < \infty \quad \text{and} \quad \lim_{|u|\to\infty} |u| k(u) = 0 \qquad (1.2)$$

Also let $\{a_n\}$ be a sequence of real positive numbers such that

$$a_n \to 0 \quad \text{as} \quad n \to \infty \qquad (1.3)$$

The kernel estimate of $f(x)$ using $k(u)$ is given by

$$\begin{aligned} f_n(x) &= \frac{1}{a_n} \int k\left(\frac{x-u}{a_n}\right) dF_n(u) \\ &= \frac{1}{na_n} \sum_{i=1}^{n} k\left(\frac{x - X_i}{a_n}\right) \end{aligned} \qquad (1.4)$$

where $F_n(x)$ is the sample distribution function.

In this paper, we examine the conditions under which

$$\theta_n = \int \gamma(x)\ f_n(x)\ dx \tag{1.5}$$

is consistent (weak as well as strong) and asymptotically normal.

All integrals in this paper will be understood to be Lebesgue integrals. Where the limits of integrations over the entire line is considered, they will be omitted.

2. Consistency

We first examine the conditions under which θ_n is asymptotically unbiased in the sense if $a_n \to 0$ as $n \to \infty$, then

$$\lim_{n\to\infty} E(\theta_n) = \theta \tag{2.1}$$

Now

$$E(\theta_n) = E\ \frac{1}{a_n} \iint \gamma(x) k(\frac{x-u}{a_n})\ d\,F_n(u)\ dx$$

$$= \frac{1}{a_n} \iint \gamma(x)\ k(\frac{x-u}{a_n})\ dF(u)\ dx \tag{2.2}$$

In order for (2.1) to hold, the last expression for (2.2) must tend to $\int \gamma(x)\ f(x)\ dx$. Conditions under which this happens are given by the following theorem.

Theorem 1. Suppose $k(u)$ is a Borel function satisfying the condition (1.2) and

(i) $\int |k(y)| dy < \infty$ and (ii) $\int k(y) dy = 1$

Let $\gamma(y)$ and $f(y)$ satisfy

$$\int |\gamma(y)\ f(y)| dy < \infty. \tag{2.2a}$$

Let $\{a_n\}$ be a sequence of positive constants satisfying (1.3). Define

$$g_n(x) = \frac{1}{a_n} \iint \gamma(x)\ k(\frac{y}{a_n})\ f(x - y)\, dy\, dx .$$

Then, at every point x of continuity of $f(\cdot)$,

$$\lim_{n\to\infty} g_n(x) = \int \gamma(x)\ f(x)\ dx. \tag{2.3}$$

Proof. In view of Theorem 1A, Pursen (1962) and (2.2a)

$$\lim_{n\to\infty} g_n(x) = \int \gamma(x)\ f(x)\ dx$$

The equation (2.3) implies that

$$\lim_{n\to\infty} E(\theta_n) = \theta$$

i.e. θ_n is asymptotically unbiased.

Theorem 2. Assume that a_n satisfy (1.3), and $Var[\gamma(x)] < \infty$, then

$$E|\theta_n - \theta| \to 0 \quad \text{as} \quad n \to \infty \tag{2.4}$$

Proof.

$$E|\theta_n - \theta| \leq E|\theta_n - E\theta_n| + |E\theta_n - \theta|$$

$$= I_{n_1} + I_{n_2}$$

By Theorem 1

$$I_{n_2} \to 0 \quad \text{as} \quad n \to \infty \tag{2.5}$$

and

$$I_{n_1} = E|\theta_n - E\theta_n| \leq [E|\theta_n - E\theta_n|^2]^{\frac{1}{2}} = [Var\ \theta_n]^{\frac{1}{2}} \to 0 \tag{2.6}$$

by Theorem 4 (proved later).

From (2.5) and (2.6) we have that

$$E|\theta_n - \theta| \to 0 \quad \text{as} \quad n \to \infty$$

i.e.,

$$\theta_n \xrightarrow{P} \theta \quad \text{as} \quad n \to \infty$$

Theorem 3. Assume that a_n satisfies (1.3) and suppose that $\gamma(x)$ is absolutely continuous, and $Var[\gamma(x)] < \infty$, then

$$\theta_n \xrightarrow{W.P.1} \theta \quad \text{as} \quad n \to \infty \tag{2.7}$$

Proof. $|\theta_n - \theta| \leq |\theta_n - E\theta_n| + |E\theta_n - \theta|$

By Theorem 1

$$|E\theta_n - \theta| \to 0 \quad \text{as} \quad n \to \infty \tag{2.8}$$

Now

$$\theta_n - E\theta_n = \frac{1}{a_n}\left[\iint \gamma(x)\, k\left(\frac{x-u}{a_n}\right) dF_n(u)\,dx - \iint \gamma(x) k\left(\frac{x-u}{a_n}\right) dF(u)\,dx\right]$$

$$= \iint \gamma(a_n z + u) k(z)\, dF_n(u)\, dz - \iint \gamma(a_n z + u) k(z) f(u)\, du\, dz$$

$$= \frac{1}{n}\int \sum_{i=1}^{n} \gamma(a_n z + X_i) k(z)\, dz - \iint \gamma(a_n z + u) k(z) f(u)\, du\, dz$$

$$= E[\frac{1}{n} \sum_{i=1}^{n} \gamma(a_n Z + X_i) - E_X E_Z \gamma(a_n Z + X)$$

$$= \frac{1}{n} \sum_{i=1}^{n} E_Z \gamma(a_n Z + X_i) - E_X E_Z \gamma(a_n Z + X)$$

Let

$$g_n(x) = E_Z \gamma(a_n Z + x)$$

So

$$\theta_n - E\theta_n = \frac{1}{n} \sum_{i=1}^{n} [g_n(x_i) - Eg_n(X_i)]$$

$$= \frac{1}{n} \sum_{i=1}^{n} V_{ni}, \quad \text{say.}$$

Note $V_{n1}, \ldots, V_{nn}$ are iid random variables and that $EV_{ni} = 0$.

Since $\gamma(x)$ is absolutely continuous and $\mathrm{Var}[\gamma(x)] < \infty$,

$$|\theta_n - E\theta_n| \le \left|\frac{1}{n}\sum_{i=1}^{n} E_Z\gamma(a_nZ + X_i) - \right.$$

$$\left. - E_XE_Z(a_nZ + X)\right| \to 0 \qquad (2.9)$$

From (2.8) and (2.9), we have that

$$\theta_n \xrightarrow{W.P.1} \theta \quad \text{as} \quad n \to \infty.$$

Next, we discuss the asymptotic behavior of the variance of the estimate θ_n. It is given by

$$\mathrm{Var}(\theta_n) = E\theta_n^2 - E^2\theta_n$$

Now,

$$E\theta_n^2 = \frac{1}{n^2a_n^2} E[\sum_{i=1}^{n} (\int\gamma(x)k(\frac{x - X_i}{a_n})dx)^2]$$

$$+ \frac{1}{n^2a_n^2} E[\sum_{i\neq j} \int\gamma(x)k(\frac{x - X_i}{a_n})dx\int\gamma(x)k(\frac{x - X_j}{a_n})dx]$$

$$= A_{n_1} + A_{n_2}$$

$$A_{n1} = \frac{1}{na_n^2} E[\int\gamma(x)k(\frac{x - X}{a_n})dx]^2$$

$$= \frac{1}{na_n^2} E[\iint\gamma(x_1)k(\frac{x_1 - X}{a_n})\gamma(x_2)k(\frac{x_2 - X}{a_n})dx_1dx_2]$$

$$= \frac{1}{na_n^2} \iiint \gamma(x_1)\gamma(x_2)k(\frac{x_1 - u}{a_n})k(\frac{x_2 - u}{a_n})f(u)dx_1dx_2du$$

$$= \frac{1}{n} \iiint \gamma(a_nz_1 + u)\gamma(a_nz_2 + u)f(u)k(z_1)k(z_2)dz_1dz_2$$

and

$$A_{n2} = \frac{n(n - 1)}{n^2a_n^2} E[\int \gamma(x)k(\frac{x - X}{a_n})dx]^2$$

$$= \frac{(n - 1)}{n} [\iint \gamma(a_nz + u)k(z)f(u)dzdu]^2$$

which imply that

$$\text{Var } \theta_n = \frac{1}{n} [\iiint \gamma(a_nz_1 + u)\gamma(a_nz_2 + u)k(z_1)k(z_2)f(u)dz_1dz_2du$$

$$- [\iint \gamma(a_nz + u)k(z)f(u)dzdu]^2$$

then

$$n \text{ Var } \theta_n \to \int \gamma^2(u)f(u)du - [\int \gamma(u)f(u)du]^2$$

$$= E[\gamma^2(X)] - [E\gamma(X)]^2$$

In view of the above we have proved the following theorem.

Theorem 4. The estimates θ_n have variance satisfying

$$\lim_{n \to \infty} n \text{ Var } \theta_n = E[\gamma^2(X)] - [E\gamma(X)]^2 = \text{Var}[\gamma(X)]$$

and if $\text{Var}[\gamma(X)] < \infty$, then

$$\lim_{n \to \infty} \text{Var } \theta_n = 0$$

at all points x of continuity of $f(\cdot)$ if $a_n \to 0$.

From Theorem 4 one can state conditions under which the estimates θ_n are consistent in quadratic mean in the sense that $E|\theta_n - \theta|^2 \to 0$ as $n \to \infty$.

The mean square error may be written as

$$E|\theta_n - \theta|^2 = \text{Var } \theta_n + |E\theta_n - \theta|^2$$

Consequently, if $a_n \to 0$ as $n \to \infty$, it then follows that θ_n is a consistent estimate of θ.

3. Asymptotic Normality

Since the estimate θ_n may be written as $\theta_n = \frac{1}{n} \sum_{j=1}^{n} V_{nj}$, where $V_{nj} = \frac{1}{a_n} \int \gamma(x) k(\frac{x - X_j}{a_n}) dx$ and are independent and identically distributed random variables for all j, i.e. $V_n = \frac{1}{a_n} \int \gamma(x) k(\frac{x - X}{a_n}) dx$, it is easy to state conditions under which sequance θ_n is asymptotically normal, in the sense that

$$\sqrt{n}(\theta_n - \theta) \to N(0,\sigma^2) \quad \text{as} \quad n \to \infty$$

where $\sigma^2 = \text{Var}[\gamma(X)]$.

Theorem 5. Assume the following conditions:

(i) $na_n^4 \to 0$ as $n \to \infty$,

(ii) $\int zk(z)dz = 0$ and $\int z^2k(z)dz < \infty$

(iii) $f(x)$ is twice differentiable

(iv) $\int \gamma(x) f'(x) dx < \infty$, $\int \gamma(x) f''(x)dx < \infty$ and $E|\gamma(X)|^3 < \infty$, then

$\sqrt{n}(\theta_n - \theta) \to N(0,\sigma^2)$ where $\sigma^2 = \mathrm{Var}\ \gamma(X)$.

Proof. To prove the theorem, we divide the argument into two parts:

(1) $\sqrt{n}(\theta_n - E\theta_n) \to N(0,\sigma^2)$ as $n \to \infty$

(2) $\sqrt{n}(E\theta_n - \theta) \to 0$ as $n \to \infty$

To show (1), it is enough to show that

$$\frac{E|V_n - EV_n|^3}{n^{\frac{1}{2}}\sigma^3} \to 0 \quad \text{as} \quad n \to \infty$$

where

$$V_n = \frac{1}{a_n}\int \gamma(x)k(\frac{x - X}{a_n})dx$$

But

$$E|V_n - EV_n|^3 \leq 2^3(E|V_n|^3 + E^3|V_n|)$$

Now

$$E|V_n|^3 = \int \left|\int \gamma(a_n z + u)k(z)dz\right|^3 f(u)du$$

$$\leq \iiiint |\gamma(a_n z_1 + u)\gamma(a_n z_2 + u)\gamma(a_n z_3 + u)| \times$$

$$\times\ k(z_1)k(z_2)k(z_3)f(u)dz_1dz_2dz_3 \to \int |\gamma^3(u)| du$$

$$= E|\gamma(X)|^3$$

So $\quad E|V_n - EV_n|^3 \leq z^3(E|\gamma(X)|^3 + E^3|\gamma(X)| < \infty$

because $\quad E[V_n] \to E[\gamma(X)] \quad$ as $\quad n \to \infty$

and $\quad E|\gamma(X)|^3 < \infty$

which shows that

$$\frac{E|V_n - EV_n|^3}{n^{\frac{1}{2}}\sigma^3} \to 0 \qquad \text{as} \quad n \to \infty$$

Then Laypanouff condition is satisfied for $\delta = 1$ and $\sqrt{n}(\theta_n - E\theta_n) \to N(0,\sigma^2)$ as $n \to \infty$. Next, we show part (2)

$$\sqrt{n}(E\theta_n - \theta) = \frac{\sqrt{n}}{a_n}\left[\iint \gamma(x)k\left(\frac{x-u}{a_n}\right)f(u)\,du - \int \gamma(x)f(x)\,dx\right)$$

$$= \sqrt{n}\iint \gamma(x)k(z)[f(x - a_n z) - f(x)dxdz]$$

Using Taylor's expansion

$$f(x - a_n z) - f(x) = -a_n z f'(x) + (a_n z)^2 f''(x) + 0(a_n^2)$$

then

$$\sqrt{n}(E\theta_n - \theta) = \sqrt{n}\; a_n^2\left[\int z^2 k(z)dz\right]\left[\int \gamma(x)f''(x)dx\right] + \sqrt{n}\; 0(a_n^2) \to 0$$

$$\text{as} \quad n \to \infty$$

by conditions (i) - (iv).

References

1. Ahmad, I.A. (1976). On the Aymptotic Properties of an Estimate of a Functional of a Probability Density. Scand. Actuarial J., 4, 176-181.

2. Ahmad, I.A. and Lin, P.E. (1977). Nonparametric Estimation of a Vector-valued Bivariate Failure Rate. Ann. Statis., 3, 1027-1038.

3. Ahmad, I.A. and Lin, P.E. (1983). Consistency of a Nonparametric Estimation of a Density Functional. Metrika, 30, 21-29.

4. Bhattacharayya, G.K. and Roussas, G. (1969). Estimation of Certain Functional of Probability Density Function. Skand. Aktuart, 52, 201-206.

5. Bochner, S. (1955). Harmonic Analysis and the Theory of Probability. University of California Press.

6. Parzen, E. (1962). On the Estimation of a Probability Density Function and the Mode. Ann. Math Stat., 33, 1065-1076.

7. Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. Ann. Math. Stat, 27, 832-837.