A Note on Edgeworth Expansion

REZA HABIBI[1]

Abstract. This paper is concerned with application of Euler-Lagrange equation in Edgeworth expansion. The method is proposed and error analysis shows that the method is accurate.

1. Introduction and Main Results

Let $\{X_i, i \geq 1\}$ be a sequence of independent and identically distributed random variables with zero mean and unit variance and $E(X_1^3) = \gamma$ and $E(X_1^4) = \tau$. Let $S_n = n^{-1/2} \sum_{i=1}^{n} X_i$. The Edgeworth technique is a true asymptotic expansion for approximating the distribution function (df) of S_n (denoted by G). The first and second order of this expansion has a general form given by

$$\widehat{G}(x) = \Phi(x) - \Phi'(x)u(x),$$

where $u(x) = \frac{\gamma H_2(x)}{6\sqrt{n}}$, and it is

$$\frac{\gamma}{6\sqrt{n}} H_2(x) - \frac{\tau - 3}{24n} H_3(x) + \frac{\gamma^2}{72n} H_5(x),$$

for the first and second orders cases, respectively (see [G02]). Here, Φ is the df of standard normal distribution and Φ' is the related density function. Also, H_2 (for example) is the second order Hermite polynomials. The sum of squared errors is given by

$$\begin{aligned} \int_a^b E^2(x)dx &= \int_a^b |G(x) - \widehat{G}(x)|^2 dx \\ &= \int_a^b H(x, \Phi(x), \Phi'(x))dx. \end{aligned}$$

which we hope that it is negligible over every interval $[a, b]$ of support G. The approach in the usual applications of Edgeworth method is to include more terms into the asymptotic expansion or to increase the sample size n to make sure that the errors are vanished. However, it this paper we use an alternative method, which is described as follows.

Here, we replace Φ by an arbitrary df say F. Let $y = F(x)$. Then, we search for a F which minimizes the sum of squared errors. The Euler-Lagrange equation implies that

$$\frac{\partial H}{\partial y} = \frac{d}{dx}[\frac{\partial H}{\partial y'}].$$

[1]2000 Mathematics Subject Classiffication. 62G20, 62E20.

Key words and phrases. Edgeworth expansion, Error analysis, Euler-Lagrange equation.
Author is partially supported by Central Bank of Iran

This equation (see [P02]) states that $E(x) = -\frac{d}{dx}[u(x)E(x)]$. Therefore,

$$E(x) = \frac{c}{u(x)} \exp\{-\int_a^x \frac{dt}{u(t)}\},$$

where c is a constant. As follows, we show that this error is negligible under some mild conditions.

Proposition 1. Suppose that for every $x \in [a, b]$, we have $M_1 \leq |u(x)| \leq M_2$,then

$$\sup_{a \leq x \leq b} |E(x)| = O(\frac{|c|}{M_1}).$$

Proof. Let

$$q(x) = \exp\{-\int_a^x \frac{dt}{u(t)}\}.$$

For simplicity reasons, suppose that $u(x)$ is positive on $[a, b]$. then using the mean value theorem for integrals, we have

$$q(x) = \exp\{-\frac{(x-a)}{u(t_0)}\},$$

for some $t_0 \in [a, x]$. Therefore, $q(x) \leq \exp\{-\frac{(x-a)}{M_2}\} \leq 1$. Thus, we have $|E(x)| \leq \frac{|c|}{M_1}$. This completes the proof.

As follows, we propose a method for finding F. Suppose that a function $u(x)$ with two suitable upper and lower bounds is chosen and the $E(x)$ is computed.

Proposition 2. For given $u(x)$ and $E(x)$, then

$$F(x) = \frac{-1}{q(x)} \int_a^x \frac{G(t)q(t)}{u(t)} dt + \frac{F(a)}{q(x)}.$$

Proof. One can note that

$$F'(x) - \frac{1}{u(x)} F(x) + \frac{G(x)}{u(x)} = 0.$$

This differential equation has solution of $F(x)$ presented in the above proposition.

However, the expression for F contains unknown G. here, we suggest to replace G with its the bootstrap estimate.

REFERENCES

[G02] J. E. Gentle *Elements of computational statistics,* (2002), Springer.

[P02] P. N. Paraskevopoulos *Modern control engineering*, (2002), Marcel Dekker.

Department of Statistics, Central Bank of Iran, Tehran, Iran.
rezahabibi2681@yahoo.com