*IOSR Journal of Mathematics (IOSR-JM)*
*e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 14, Issue 1 Ver. III (Jan. - Feb. 2018), PP 37-44*
*www.iosrjournals.org*

# Nonparametric Estimator for the Standardized sum using Edgeworth Expansions

Kololi Moses[1], Orwa George[2]
[1]*(Mathematics Department, Kibabii University,Kenya)*
[2]*(Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Kenya)*
*Corresponding Author: Kololi Moses*

***Abstract:*** *Constructing an estimator for functional estimation is one of the problems in statistical inference. In this paper, the problem of constructing an estimator for a studentized sum is considered in the nonparametric set-up. In this set-up, data are used to infer to an unknown quantity while making as few assumptions as possible. This non-smooth functional lack some degree of properties traditionally relied upon in estimation. Smooth functionals converge at the rate of* $n^{-1/2}$*while non-smooth functionals converge at the rate slower than* $n^{-1/2}$. *This highlights the reason why standard techniques fail to give sharp results. A clear and accurate approximation is obtained by using an approximation that admits cumulant generating function; saddle point approximation. An optimal estimator is obtained using the MiniMax criterion where the lower and upper bounds are constructed. While working in the context of MiniMax estimation, the lower bounds are most important. The MiniMax lower bound is obtained by applying the general lower bound technique based on testing two composite hypotheses. The quality of an estimator is evaluated with the MiniMax risk. Best polynomial approximation of an absolute value function and Hermite polynomials are used to construct an optimal estimator when the means are bounded by a given value $M>0$.*
***Keywords:*** *Edgeworth Expansion, Nonparametric, Non-smooth, Saddle point, Standardized sum*

## I. Introduction

Statistical inference is one of the most important challenges in performing nonparametric statistics. Various nonparametric statistics have been proposed and discussed over the course for many years. However, finding the exact critical test statistic value for small sample sizes is the most challenging testing problem. It is also difficult to find the exact critical value when sample sizes are moderately large. In such circumstances, the exact critical value is estimated with an approximation method. Hence, approximation for evaluating the density of the test statistic remains one of the most important topics in statistics.

In modern statistical analysis, many quantities of interest are either non-smooth functionals of the observed data or original generative distribution, or both. In this paper, non-smooth functionals are considered; however, there are few results on the problem of estimating non-smooth functionals. The problem of estimating the $L_1$ norm of $f$ is studied in [1]. In [11], some functional estimation problems in the image model were considered. [2] focused on estimating $L_r$ norm $||f||_\infty$ where $r \geq 1$. The case of $r = \infty$ corresponds to estimating the maximum of $f$.

Estimating smooth functionals involves estimating functionals that are differentiable and have continuous derivatives. More specifically, this is first-order smoothness. Second-order smoothness means that the second-derivatives exist and are continuous, while infinite smoothness refers to continuous derivatives of all orders (Rockafellar et al, 1994). Smoothness is understood as differentiability of $F$ in $L^2$ norm. The $L^2$ norm is an example of an analytic distance induced by an inner product. When smooth functionals are estimated, their convergence rate is $n^{-1/2}$ [2]. Nonparametric estimates converge at a rate slower than $n^{-1/2}$.

The disparity between parameter, $\theta$ and its estimator $\hat{\theta}$ can be specified by a real valued loss function $L(\theta, \hat{\theta})$ which quantifies the amount by which the prediction deviates from the actual values. The "optimal" estimate, $p(\hat{\theta})$ is chosen so as to minimize the expected loss $E[L(\theta, p)]$. The error is measured on the function space by using a norm while on the vector space,$V$ a norm is a mapping associated to any $f \in V$. A norm is denoted $||f||$ and satisfies the homogeneity condition, $||cf|| = |c|||f||$ where$c \in \mathbb{R}$, and $f \in V$, the triangle inequality $||f + g|| \leq ||f|| + ||g||$, and which is strictly positive for all non-zero $f$

The value of a norm of a function may vary based on the function and the choice of the norm. The choice of the norm also has a significant impact on the solution of the approximation problem. A loss function that is globally continuous and differentiable is desired in optimization algorithms. The two commonly used loss functions are the squared loss $L(a) = a^2$ and the absolute loss $L(a) = |a|$. The absolute loss function is not differentiable at $a = 0$ whereas the square loss function has the tendency to be dominated by outliers.

However, finding the best estimate which minimizes the risk function is not easy since the risk depends on unknown parameter. Therefore, it is important to find an additional criteria for finding an optimal estimator; the MiniMax criterion. MiniMax estimators are estimators whose maximal risk is minimal among all estimators [3] and [6] and perform best in the "worst" possible case allowed in the problem. The MiniMax Risk is used as a benchmark for evaluating the performance of any estimation method [4].

An effective approach to consider is to use the near MiniMax approximation. MiniMax polynomial approximation is very close to Chebyshev polynomial approximation. Chebyshev polynomial approximation exists and is unique when the function is continuous on the interval $[a, b]$. In order to improve the accuracy of the testing procedure, the saddle point approximation is considered since it has relative error. Saddle point approximation is an important tool in asymptotic analysis and statistical analysis.

The saddle point approximation can be obtained for any statistic that admits a cumulant generating function, (CGF). Cumulants are obtained by taking derivatives of the log moment generating function and evaluating at $x = 0$. It can be applied in finance and insurance, Physics and Biology. The saddle point literature focus on the approximation of densities and tail probabilities [5] and [6] and provide accurate approximations of the distribution of some statistic even in small samples.

## II. Chebyshev And Hermite Polynomials

Chebyshev and Hermite polynomials are orthogonal. A sequence of orthogonal polynomials can be defined recursively and used to improve control of the interpolation error on the interpolation interval [15]. The orthogonal polynomials also have the property of bounded variation and their local maxima and minima on $[-1,1]$ are exactly equal to 1 and -1 respectively regardless of the order of the polynomial.

Orthogonality properties are good for polynomial basis. The Chebyshev polynomials and Hermite polynomials are used to make the coefficients uncorrelated and minimize the sensitivity of calculations of round-off error. The Hermite polynomials are uncorrelated when $X$ is standard normal on$(-\infty, \infty)$. They also agree with function values and their derivatives at the node points.

However, the Chebyshev polynomials are not easy to compute. A more effective approach to consider is to use a near MiniMax approximation. MiniMax polynomial approximation is very close to Chebyshev polynomial approximation. This polynomial approximation exist and is unique when the function $f(x)$ is continuous on an interval $[a, b]$. This is stated and proofed in the Weierstrass Approximation Theorem.

The Hermite polynomials form an orthogonal basis of the Hilbert space of functions satisfying:

$$\int_{-\infty}^{\infty} |f(x)|^2 \omega(x) dx < \infty \quad (1)$$

in which the inner product is given by the integral including the Gaussian weight function $\omega(x)$. Suppose $\phi$ is the density function of the standard normal variable, then the Hermite polynomials $H_k(x)$ with respect to $\phi$ for positive integers $k$ are defined by the equation:

$$\frac{d^k}{dx}\phi(x) = -1^k H_k(y)\phi(x) \quad (2)$$

Thus we can obtain $H_1(x) = x, H_2(x) = x^2 - 1, H_3(x) = x^3 - 3x$ and so on. By differentiating (2) we obtain:

$$\frac{d}{dx}[H_k(x)] = H_{k+1}(x)\phi(x) \quad (3)$$

## III. Saddle point Approximation

As earlier noted, the commonly used loss functions are the squared error loss and the absolute error loss function. When the error is absolute, the error holds in the tails as the modes of the density. Therefore, an approximation whose error gets small with the density in the tails of the distribution is required. A clear and accurate approximation can be obtained by using saddle point approximation. These techniques provide accurate approximations of the distribution of some statistic even in small samples.

Saddle point approximation can be obtained for any statistic that admits a cumulant generating function. Saddle point approximation uses relative error which gets small with the density in the tails of the distribution. When the underlying distribution is unknown and the empirical distribution is used, the empirical saddle point approximation results in a relative error of order $N^{-1/2}$. The relative error results in an improvement over the absolute error in the tails of the distribution. This is important in the calculation of p-values. For instance, if $f(x)$denotes the density being approximated, the asymptotic approximation can be written:

$$\hat{f}_A(x) = f(x) + \mathcal{O}(N^{-1/2}) \quad (4)$$

This means $\hat{f}_A(x) - f(x)/(N^{-1/2})$ is bounded as $x$ approaches some limiting value. The empirical saddle point approximation can be written:

$$\hat{f}_S(x) = f(x)\left\{1 + O(N^{-1/2})\right\} \tag{5}$$

The Saddle point density replaces the CLT in the traditional general method of moment's distribution theory. The CLT uses information about the location and convexity of the general method of moment's objective function at the global minimum. The saddle point approximation uses information about the shape of the objective function at each point in the parameter space.

The CLT is usually presented in terms of estimating the mean of a distribution. It shows that the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ has a distribution that is approximately$N(\mu, \frac{\delta^2}{n})$. The function $M_X(t) = e^{tX}$ is called the moment generating function of $X$. The MGF can be used to prove how fast the distribution converges to normality. It is also possible to make adjustments to increase the accuracy of the approximations. Edgeworth expansions are one method of using information about higher order moments to increase accuracy.

MGF uniquely identify a particular distribution. One problem with the MGFs is that they do not always exist. This is solved by using the characteristic function which is the complex extension of the MGF and is defined $\phi_X(t) = E(e^{iXt})$, where $i^2 = -1$. The characteristic function uniquely identify distributions and can be used to show limiting results.

As seen earlier, the CLT is built on a two-term Taylor series expansion, i.e., a linear approximation of the characteristic function about the mean. A higher order expansion of Taylor series about the mean is used to obtain additional precision. This results in Edgeworth expansion. The Edgeworth expansion gives a significantly better approximation at the mean of a distribution. However, the quality of the approximation can deteriorate significantly for the values away from the mean. The saddle point exploits these characteristics of the Edgeworth expansion in order to get an improved approximation.

The improved approximation of the Edgeworth expansion only occurs at the mean of the distribution. In order to obtain such an improvement at an arbitrary value in the parameter space, the original distribution is transformed to a conjugate distribution. A specific conjugate distribution is selected so that its mean is transformed back to the original distribution at the value of interest. To develop the estimator, the following important mathematical concepts are considered; The normal Means, Hilbert sample space, Polynomial approximation and MiniMax lower bounds and saddle point approximation.

### 3.1 Normal Means Model

The problem of density estimation can be connected to the Normal means problem. The normal means problem unifies some NP problems. Let $Z^n = (Z_1, \cdots, Z_n)$ where:

$$Z_i = \theta_i + \delta_n \epsilon_i, i = 1,2,\dots,n \tag{6}$$

$\epsilon_i, \dots, \epsilon_n$ are independent, $N(0,1)$ random variables, $Z_i = n^{-1}\sum_{j=1}^{n} X_{ij}$

$$\theta^n = (\theta_1, \cdots, \theta_n \in \mathbb{R}^n$$

is a vector of unknown parameters and $\delta_n$ is assumed known. The model appears to be parametric but the number of parameters increases at the same rate as the number of data points. The model has all the complexities of NP problem. An infinite-dimensional version of (3) is given as:

$$Z_i = \theta_i + \delta_n \epsilon_i, i = 1,2,\cdots \tag{7}$$

where now the unknown parameter is $\theta = \theta_1, \theta_2, \dots$

| $\theta_1$ | $\theta_2$ | $\cdots$ | $\theta_i$ | $\cdots$ | $\theta_n$ |
|---|---|---|---|---|---|
| $X_{11}$<br>$\vdots$<br>$X_{1j}$<br>$\vdots$<br>$X_{1n}$ | $X_{21}$<br><br>$X_{2j}$<br><br>$X_{2n}$ | $\cdots$<br>$\vdots$<br>$\cdots$<br>$\vdots$<br>$\cdots$ | $X_{i1}$<br><br>$X_{ij}$<br><br>$X_{in}$ | $\cdots$<br><br>$\cdots$<br>$\cdots$<br>$\cdots$ | $X_{n1}$<br>$\vdots$<br>$X_{nj}$<br>$\vdots$<br>$X_{nn}$ |
| $Z_1$ | $Z_2$ | $\cdots$ | $Z_i$ | $\cdots$ | $Z_n$ |

**Figure 1**: The normal means $X_{ij} = \theta_i + N(0, \delta^2)$ and $Z_i = n^{-1}\sum_{j=1}^{n} X_{ij} = \theta_i + \delta_n \epsilon_i$ where $\delta_n = \frac{\delta}{\sqrt{n}}$.

Estimating the parameters $\theta_1 \cdots \theta_n$ from the $n$ column means $Z_1 \cdots Z_n$ leads to the model: $Z_i = \theta_i + \delta_n \epsilon_i$ with $\delta_n = \frac{\delta}{\sqrt{n}}$.
Given an estimator the squared error loss:

$$L(\hat{\theta}^n, \theta^n) = \sum_{i=1}^{n} (\hat{\theta}_i - \theta_i)^2 = ||\hat{\theta}^n - \theta^n||^2 \tag{8}$$

is used with risk function:

$$R\left(\hat{\theta}^n, \theta^n\right) = \mathbb{E}_\theta (L(\hat{\theta}^n, \theta^n)) = \sum_{i=1}^{n} \mathbb{E}_\theta (\hat{\theta}_i - \theta_i)^2 \tag{9}$$

One of the choices for an estimator of $\theta^n$ is $\hat{\theta}^n = Z^n$. This estimator is the maximum likelihood estimator, the minimum variance unbiased estimator and the Bayes estimator under a flat prior. However, $\hat{\theta}^n$ is a poor estimator as its risk is: $R(Z^n, \theta_n) = \sum_{i=1}^{n} \mathbb{E}_\theta (Z_i - \theta_i)^2 = \sum_{i=1}^{n} \delta_n^2 = n\delta_n^2$
There are other estimators with significantly smaller risks. To improve this estimator, the normal means problem relation to NP regression and density estimation is important. In order to do this, a review on the Hilbert function space is required.

### 3.2 Hilbert Space

Nonparametric estimation is by nature infinite-dimensional. When the form of the functions in the true model are unknown, the most efficient use of data is to allow the estimated functions to depend on the size of the sample tending to infinity with the sample size [7]. In this research, the Hilbert sample space is a natural choice because it is the closest analog to the Euclidean space, $\mathbb{R}^n$ in infinite dimension.

The Hilbert space is a complete inner product space denoted by $(L\langle .,. \rangle)$ and abbreviated by $L_2$ [9]. The $L_2$ denotes the set of square integrable functions, but does not specify the selection of metric, norm or inner product. The set, together with the specific inner product $\langle .,. \rangle$ specify the inner product space. The square integrable functions form a complete metric space under the metric induced by the inner product. A metric allows a discussion on convergence of sequences. The importance of completeness is to allow Cauchy sequences converge, and find a point to which they converge within space.

In vector spaces, elements can be added and multiplied by scalars with rules of $\mathbb{R}^n$ and $\mathbb{C}^n$. Vector spaces with inner product structure allow length and angle to be measured. This inner product in an abstract space therefore opens the door to geometric arguments that are remarkably similar to those used in Euclidean space $\mathbb{R}^n$. The normed vector space idea is away to forget that the function spaces being dealt with actually arise as functions, and treat them as elements of a vector space with a norm that gives the correct topology [9].

An inner product always generates a norm. And where there is a norm, there is a metric. A norm defined by the inner product $\langle .,. \rangle$ will define the following metric:

$$d(x, y) = \left||x - y|\right| = \sqrt{\langle x - y, x - y \rangle} \;\; \forall\, x, y \in X \tag{10}$$

Inner product norms satisfy a number of properties that aren't enjoyed by all norms. For instance, a complex vector space $V$ is called an inner product space (or a pre-Hilbert space) if there is a mapping $(.,.): V \times V \to \mathbb{C}$, called an inner product that satisfies:
$\forall x, y, z \in V, \forall \propto \in \mathbb{C}$

1) $(x, x) \geq 0$
2) $(x, x) = 0 \Leftrightarrow x = 0$
3) $(x, y + z) = (x, y) + (x, z)$
4) $x, \propto y = \propto (x, y)$
5) $(x, y) = (y, x)^*$

Hilbert spaces are easy to come by in practice. This research is concerned with only four: First, given a positive Riemann integrable weight function $\omega(x)$ defined on some interval $[a, b]$, the expression:

$$\langle f, g \rangle = \int_a^b f(t) g(t) \omega(t) dt \tag{11}$$

Defines an inner product on $C[a, b]$, the space of all continuous, real-valued functions $f: [a, b] \to \mathbb{R}$, with associated norm:

$$||f||_2 = \left(\int_a^b |f(t)|^2 \omega(t) dt\right)^{\frac{1}{2}} \tag{12}$$

Secondly, the sequence space $\ell^2$, where $x = (x_1, x_2, \ldots) \in \ell^2$ implies that $\sum_{i=1}^{n} x_i^2 < \infty$ .For $x, y \in \ell^2$, the inner product is defined by as:

$$\langle x, y \rangle = \sum_{i=1}^{\infty} x_i y_i \tag{13}$$

The $\ell^2$ is the only $\ell^p$ space for which an inner product exists.

Thirdly, the space of (real- or complex-valued) square-integrable $L^2[a, b]$$ where:

$$||f||^2 = \langle f, f \rangle = \int_a^b |f(x)|^2 dx < \infty \tag{14}$$

The inner product in this space is given by:

$$\langle f, g\rangle = \int_a^b f(x)\overline{g(x)}dx < \infty \tag{15}$$

In this case of sequence spaces, $L^2$ is the only $L^p$ space for which an inner product exists

Lastly, a Banach space is also a Hilbert space since a Hilbert space is complete with respect to the induced metric [12]. A Banach space is a normed space with associated metric $d(x, y) = ||x - y||$ such that every Cauchy sequence in $B$ has a limit in $B$ [13]. The difference between Banach space and Hilbert space is the source of the norm. In the Hilbert space case the norm is defined via the inner product, whereas in the Banach space case the norm is defined directly. Thus a Hilbert space is a Banach space, but the inverse is not true, because in some cases the norm cannot be related with an inner product [13] and [14].

A Banach space, $\int_{-\infty}^{\infty} |f(x)|^2 dx < \infty$, which is a space of square integrable functions gives the Hilbert space further property of inner product under the metric induced by the norm which in turn is induced by the inner product. A part from the classical Euclidean spaces and spaces of square-integrable functions, other examples of Hilbert spaces include spaces of sequences, Sobolev spaces consisting of generalized functions and Hardy spaces of holomorphic functions.

The coordinates of an element of the Hilbert space can be specified uniquely with respect to a set of coordinate axes (an orthonormal basis), which is in analogy with the Cartesian coordinates in the plane. A sequence of functions $\phi_1, \phi_2, \ldots$ is called orthonormal if $||\phi_j|| = 1$ for all $j$ (normalized) and $\int_a^b \phi_i(x)\phi_j(x)dx = 0$ for $i \neq j$ (orthogonal). This sequence is complete if the only function that is orthogonal to each $\phi_j$ is the zero function. A complete, orthogonal set of functions forms a basis, meaning that if $f \in L_2(a, b)$ then $f$ can be expanded in the basis. An example of an orthonormal basis is the Chebyshev and Hermite polynomials defined on $(-1,1)$. These polynomials are discussed in the next section.

We observe that [8] pointed out that there is a limitation for the use of Hilbert space on the real line. The standard Hilbert space such as $L_2(\mathbb{R})$ requires that the unknown function approaches zero at infinity. This makes it unreasonable to be used on some models like the economic model, since it excludes widely used functions such as constant increasing and cyclical functions on a line. To overcome this, weighted Hilbert spaces are used, since they impose weaker limiting requirements at infinity.

The Hilbert space is restricted to bounded sample spaces in nonparametric estimation. [8] considered the space $L_2[a, b]$ of functions defined on the bounded segment $[a, b] \in \mathbb{R}$, where $[a, b]$ represents the sample space. It is pointed out that continuity is the only assumption to be made for the unknown function. Every continuous unknown function on the segment $[a, b]$ is bounded and belongs to the Hilbert space $L_2[a, b]$.

The general polynomial $P(x)$ can be written in terms of any sequence of basic polynomials of increasing degree. It is written in terms of the monomials $x^j$ which are the natural form polynomials. Monomials make all polynomials look very similar when plotted; that is they are correlated (G.K. Smyth, 1998). This shows that, first, a small change in $P(x)$ may arise from relatively large changes in the coefficients. Secondly, when there is a measurement or round-off error, the coefficients are not well determined. Therefore, monomials are not a good basis for the space of continuous functions because they are far from being orthogonal.

## IV. The Central Limit Theorem and the Characteristic Function

The characteristic function of a probability measure $\mu$ on the line is defined for real $t$ by:

$$\begin{aligned}\psi(x) &= \int_{-\infty}^{\infty} e^{itx}\mu\, dx \\ &= \int_{-\infty}^{\infty} \cos tx\mu(dx) + i\int_{-\infty}^{\infty} \sin tx\mu\,(dx)\end{aligned} \tag{16}$$

A random variable $X$ with the distribution $\mu$ has the characteristic function:

$$\psi(x) = \mathbb{E}[e^{itx}] = \int_{-\infty}^{\infty} e^{itx}\mu\, dx \tag{17}$$

The characteristic function is thus defined as the moment generating function but with the real argument $s$ replaced by $it$. As seen earlier, it has the advantage that it always exists because $e^{itx}$ is bounded. The characteristic function in nonprobabilistic contexts is called the Fourier transform. The knowledge of the Fourier transform is used in the Central Limit Theorem to send smooth functions to bounded functions and bounded functions to smooth functions.

For instance, if $g: \mathbb{R} \to \mathbb{R}$, the Fourier transform of $g$ is defined by:

$$\hat{g}(y) = \int_{-\infty}^{\infty} e^{ixy} g(x)dx \tag{18}$$

where $g$ is a Schwartz function, it is $C^\infty$ and all of its derivatives decay at $\pm\infty$ faster than every polynomial. If $g$ is Schwartz, then $\hat{g}$ is Schwartz. The inversion formula:

$$g(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixy}\,\hat{g}\,(y)dy \tag{19}$$

Hold for Schwartz functions. A straightforward computation shows that:

$$e^{-x^2/2} = \sqrt{2\pi}e^{-y^2/2} \tag{20}$$

Suppose $X \sim G(x)$ and $\varphi_x(t)$ denotes the characteristic function of $X$. If $\int_{-\infty}^{\infty}|\varphi_x(t)|dt < \infty$ then $g(x) = G^{'}(x)$exists and:

$$g(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixy}\,\varphi_x(t)\,dt \tag{21}$$

This gives the inversion formula for characteristic functions. Also:

$$\int_{-\infty}^{\infty} f(x)\hat{g}(x)dx = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} g(y)e^{-ixy}\,dy\right]$$

$$= \int_{-\infty}^{\infty} g(y)\left[\int_{-\infty}^{\infty} f(x)e^{-ixy}\,dx\right]dy$$

$$= \int_{-\infty}^{\infty}\hat{f}(y)g(y)dy \tag{22}$$

This is valid provided that the interchange of integrals can be justified; in particular, it holds for Schwartz $f, g$.

If $X \sim N(0,1)$, then by completing the square in the exponential, the characteristic function for a standard normal distribution can be obtained:

$$\phi(x) = \int_{-\infty}^{\infty} e^{itx}\,\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = e^{-t^2/2} \tag{23}$$

Then for any positive integer $k$, we have:

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-itx}e^{-t^2/2}\,(it)^k dt &= \frac{-1^k}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\,e^{-t^2/2}dt\\ &= -1^k\,\frac{d^k}{dx^k}\phi(x)\\ &= H_x(x)\phi(x)\end{aligned} \tag{24}$$

where (24) follows from (21) since $e^{-t^2/2}\,dt$ is the characteristic function for a standard normal distribution.

## V. The Proposed Model

Suppose $F$ is a function to approximate $T(\theta)$ such that $F: [a, b] \to \mathbb{R}$ and $g(x)$ is a polynomial approximation of $F$ with respect to the weight function $\omega(x)$. Let $X_1, \ldots, X_n$ be independent normal random variables where $X_i \sim N(\theta_i, 1)$ from $F(x)$. Suppose $\mathbb{E}(X_1) = 0$ and $var(X_1) = 1$; for otherwise, each $X_i$ is replaced by $(X_j - \mathbb{E}(X_1)/\sqrt{var(X_1)}$. Let also $\gamma = \mathbb{E}(X_j^3)$ and $\tau = \mathbb{E}(X_j^4)$ and suppose $\tau < \infty$. In this paper, the estimate of the distribution of the standardized sum:

$$S_n = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} X_j \tag{25}$$

is constructed.

From the CLT, for every $x$, $P(S_n \le x) \to \Phi(x)$, where $\Phi(x)$, denotes the standard normal distribution function. In this paper, a better approximation to $P(S_n \le x$ is constructed by the characteristic function of $S_n$.

$$\varphi S_n(t) = \mathbb{E}\left[exp\left\{\left(it/\sqrt{n}\right)\textstyle\sum_j X_j\right\}\right] = \left[\phi_X\left(\tfrac{t}{\sqrt{n}}\right)^n\right]$$

Then the Taylor's expansion of $exp\{itx/\sqrt{n}\}$ is used:

As $n \to \infty$,

$$\begin{aligned}\psi_{X\left(\frac{t}{\sqrt{n}}\right)} &= \mathbb{E}\left\{1+\frac{itX}{\sqrt{n}}+\frac{(it)^2X^2}{2n}+\frac{(it)^3X^3}{6n\sqrt{n}}+\frac{(it)^4X^4}{24n^2}\right\}+\sigma\left(\frac{1}{n^2}\right)\\ &= \left(1-\frac{t^2}{2n}\right)+\frac{(it)^3\gamma}{6n\sqrt{n}}+\frac{(it)^4\tau}{24n^2}+\sigma\left(\frac{1}{n^2}\right)\end{aligned} \tag{26}$$

where $\sigma\left(\frac{1}{n^2}\right)$ is the error in the Taylor's expansion and $i = -1$.

Raising this tetra nominal to the nth power most terms are $\sigma\left(\frac{1}{n}\right)$:

$$\left[\psi_X\left(\frac{t}{\sqrt{n}}\right)\right]^n =$$

$$\left(1-\frac{t^2}{2n}\right)^n + \left(1-\frac{t^2}{2n}\right)^{n-1} + \left(\frac{(it)^3\gamma}{6\sqrt{n}} + \frac{(it)^4\tau}{24n}\right) + \left(1-\frac{t^2}{2n}\right)^{n-2}\frac{(n-1)(it)^6\gamma^2}{72n^2} + \sigma\left(\frac{1}{n}\right) \qquad (27)$$

Using (27) and the binomial theorem equation for a fixed nonnegative integer $k$:

$$\left(1+\frac{a}{n}\right)^{n-k} = e^a\left(1-\frac{a(a+k)}{2n}\right) + \sigma\left(\frac{1}{n}\right) \text{ as } n \to \infty$$

$$\psi_{S_n}(t) = e^{-\frac{t^2}{2}}\left[1 - \frac{t^4}{8n} + \frac{(it)^3\gamma}{6\sqrt{n}} + \frac{(it)^4\tau}{24n} + \frac{(it)^6\gamma}{72n}\right] + \sigma\left(\frac{1}{n}\right)$$

$$= e^{-\frac{t^2}{2}}\left[1 + \frac{(it)^3\gamma}{6\sqrt{n}} + \frac{(it)^4(\tau-3)}{72n} + \frac{(it)^6\gamma^2}{72n}\right] + \sigma\left(\frac{1}{n}\right) \qquad (28)$$

Combining equation (28) with (21) the density function below is obtained as an approximation to the distribution of $S_n$:

$g(x) =$

$\frac{1}{2\pi}\left(\int_{-\infty}^{\infty} e^{-itx} e^{\frac{-t^2}{2}} +\right.$

$\gamma 6n-\infty\infty e-itxe-t22(it)3dt+\gamma-324n-\infty\infty e-itxe-t22(it)4dt$

$+\gamma 272n-\infty\infty e-itxe-t22(it)6dt$ (29)

Putting (28) with (23) gives:

$$g(x) = \phi(x)\left(1 + \frac{\gamma H_3(x)}{6\sqrt{n}} + \frac{(\tau-3)H_4(x)}{24n} + \frac{\gamma^2 H_6(x)}{72n}\right) \qquad (30)$$

Using (3), the antiderivative of $g(x)$ equals:

$$G(x) = \Phi(x) - \phi(x)\left(\frac{\gamma H_2(x)}{6\sqrt{n}} + \frac{(\tau-3)H_3(x)}{24n} + \frac{\gamma^2 H_5(x)}{72n}\right)$$

$$= \Phi(x) - \phi(x)\left(\frac{\gamma(x^2-1)}{6\sqrt{n}} + \frac{(\tau-3)(x^3-3x)}{24n} + \frac{\gamma^2(x^5-10x^3+15x)}{72n}\right) \qquad (31)$$

The letter $\phi$ denotes the probability density function and the corresponding distribution function is denoted $\Phi$.

## VI. Conclusion

Our observation led us to the conclusion that estimating non-smooth functional exhibits some features that are significantly different from those in estimating smooth functional. Lack of these properties highlights the reason why standard techniques fail to give sharp results. Therefore, the best polynomial approximation and the Hermite polynomial are used in the derivation of lower bounds and construction of optimal estimators.

Saddle point approximation is an important tool in asymptotic analysis and statistical analysis and it provides accurate approximation of the distribution even in small sample. This statistical procedure can be used on all types of data which are nominally, ordinal, interval and/or ratio scaled. It is also easy to apply the procedure on a small sample size, which would demand the distributions to be known precisely in order for parametric tests to be applied. Therefore, the procedure is used in many applications, for example, hierarchical clustering, data ranking, anomaly detection, detecting regions of brain activity, mapping environmental contamination, estimating contour levels in digital elevation mapping and identification of genes with significant expression levels in micro array analysis. It is quite useful in finding excess mass in tinned food stuff.

## Acknowledgements

First and foremost, I am grateful to Almighty God who through His infinite mercy and love gave me life and guided me throughout the writing of this paper. My special gratitude goes to my supervisors Prof. George O. Orwa, Dr. Joseph K. Munga'tu and Prof. Romanus O. Odhiambo whose continuous support, guidance and assistance made it possible to complete this paper. You directly and indirectly gave ideas which were incorporated towards the development of this paper. I was deeply touched by your invaluable input, continued stewardship and scholarly contributions throughout the period of this study. God bless you so much. I thank you all.

## Reference

**Journal Papers:**

[1] Ibragimov, I. A. and Khasminski, R., *Asymptotic normal families of distributions and effective estimation., Ann. Statist 19*, 1991, 1681-1724.
[2] Lepski, O., Nemirovski, A. and Spokoiny, V. (1999), *On estimation of the $L_r$ norm of a regression function, Probab. Theory Relat. Fields 113*, 1999, 221-253.
[3] Donoho, D. L. and Liu, R. C. (1991), *Geometrizing rates of convergence II, Ann. Statist,19*,1999, 633-667.
[4] Cai, T. and Low, M. , *Testing composite, Hermite polynomials, and optimal estimation of a non-smooth functional, Ann statist, 39(2)* ,2011,1012-1041.
[5] Lugannani, R., and Rice, S., *Saddle point approximation for the distribution of the sum of independent random variables, Advances in Applied Probability, 12*,1980, 475-490.
[6] Daniels, H.E., *Tail probability approximations. International Statistical Review 55*,1987, 37-48.
[7] Chichilnisky, G., *The limits of econometrics: Nonparametric functions in Hilbert spaces, Economic Theory, 25*,2009, 1-17.
[8] Bergstrom, R., *The estimation of nonparametric functions in a Hilbert space, Economic Theory, 1*, 1985, 7-26.

**Books:**

[9] W. Larry, *All of nonparametric statistics,* (New York: Springer Science + Business Media, Inc, 2006).

**Chapters in Books:**

[10] Terry Rockafellar, Mathematical Programming: State of the Art , ( University of Michigan Press: Ann Arbor, 1994), 248-248.

**Proceedings Papers:**

[11] Korostelev, A. P. and Tsybakov, A. B, Minimax Theory of Image Reconstruction, Lecture Notes in Statist.Springer, 1994
[12] Richard, M., Functional analysis lecture notes, Department of Mathematics, MIT, 2013
[13] Herman, J.B., Introduction to Hilbert spaces, Pennsylvania State University, 2007.
[14] Ivan, A., Notes on Hilbert spaces. New Mexico Tech Socorro, NM 87801, 2007
[15] Timothy Sauer, Numerical analysis, Pearson Education Inc. (C) 2006-2012.

Kololi Moses. "Nonparametric Estimator for the Standardized sum using Edgeworth Expansions." IOSR Journal of Mathematics (IOSR-JM) 14.1 (2018): 37-44.