Linear spline mapping in Normal Distribution

S. BOONTHIEM, S. BOONTA, W. KLONGDEE
Khon Kaen University, THAILAND
+669-1059-0942, kwatch@kku.ac.th

Abstract— This article presents a construction of a new distribution by using linear spline mapping based on probability density function of normal distribution where two end points have a value of probability density function as zero. In addition, we propose the cumulative distribution function and the inverse of cumulative distribution function of the distribution. Furthermore, we illustrate the parameter estimation of 77 data of the student's average intelligent quotient (IQ) for Grade 1 in Thailand by method of moments and propose minimum KS -statistics of the distribution by difference of the node width.
Keywords— Linear spline mapping, method of moments

1 Introduction

The inferential statistics can describe the behavior of data by parametric statistics and non-parametric statistics depending upon trend of the data. The trend of data indicates the shape of the distribution. Distribution theory is closely related to probability theory and important in order to calculate and approximate a probability distribution. Nowadays, many researchers are interested in the approximation of the standard cumulative normal distribution function, due to the integral of normal distribution does not have closed form theoretical expression. However, it give only approximation as the work of many researchers. For instances, Bowling, Khasawneh, Kaewkuekool, and Cho (2009) developed a logistic approximation to the cumulative normal distribution. The proposed algorithm has a simpler function form and gave higher accuracy with maximum error of less than 0.00014 for the entire range [1]. Vazquez-Leal et al. (2012) provided an approximate solution to the normal distribution integral by using the homotopy perturbation method (HPM). The HPM method have a high level of accuracy [2]. Choudhury (2014) proposed a very simple approximation formula to the standard cumulative normal distribution function. The formula could be implemented in any hand-held calculator. The maximum absolute error of the approximation was 0.00019 [3]. This article interested in the case study of the approximation of the cumulative normal distribution using linear spline mapping. We expect that this method can use to other symmetrical distributions. Linear spline is widely method for estimation in many fields such as statistics, mathematics, industrial engineering and so forth. In 1999, Chen studied the estimation of the probability density function and the cumulative distribution function of random variable. There were three spline density estimators consisting of a quantile regression spline estimator and two maximum likelihood spline estimators. These methods proposed for variable observed with measurement error [4]. In the same year, Lindstrom improved the computational and estimation properties of free-knot splines while retaining their adaptive smoothing properties and proposed estimator for the knots defined as the optimizer of a slightly penalized residual sum-of-squares [5]. In 2002, Zhang and Lin investigated the optimal models for building histograms based on linear spline techniques and presented efficient algorithms to achieve these proposed optimal models. The experimental results showed that these techniques could greatly improve the approximation accuracy comparing to the existing techniques [6]. In 2012, Holland employed penalized B-splines in the context of the partially linear model to estimate the non-parametric component, when both the number of knots and the penalty factor vary with the same size [7]. In 2013, Valenzuela, Pasadas, Ortuño, and Rojas presented a novel methodology for optimal placement and selected of knots, for approximating or fitting curves to data, using smoothing splines [8]. In 2015, Kang, Chen, Li, Deng, and Yang proposed a computationally efficient framework to calculate knots for splines fitting via sparse optimization [9]. In the same year, Tjahjowidodo, Dung, and Han proposed a fast method for knots calculation in a B-spline fitting based on the second derivative [10]. In 2017, Hussain, Abbas, and Irshad proposed a new quadratic trigonometric B-spline with control parameters was constructed to address the problems related to two dimentional digital image interpolation [11]. Our study is to construct a new distribution using linear spline mapping on probability density function of normal distribution and obtain the cumulative distribution function after that. Furthermore, this study proposes the inverse of cumulative distribution function that have useful for generating the random variable of some data.

2 Main Results

In this section, we interpret the principle of linear spline mapping through on the following definitions, propositions, and theorems. Throughout this article, we use the following notations,

$\Delta T_D\left(f\left(x_{i+1}\right)\right)=T_D\left(f\left(x_{i+1}\right)\right)-T_D\left(f\left(x_i\right)\right)$, $\Delta x_{i+1}=x_{i+1}-x_i$,

$\overline{x_i}=\dfrac{x_i+x_{i-1}}{2}$, $\Delta f\left(x_{i+1}\right)=f\left(x_{i+1}\right)-f\left(x_i\right)$,

$\overline{f\left(x_{i+1}\right)}=\dfrac{f\left(x_{i+1}\right)+f\left(x_i\right)}{2}$, $\overline{\phi\left(x_k\right)}=\dfrac{\phi\left(x_k\right)+\phi\left(x_{k-1}\right)}{2}$.

Definition 1. Given $C_B(\mathbb{R})=\left\{f:\mathbb{R}\to\mathbb{R} \mid f \text{ is continuous and bounded}\right\}$, $x_0,x_1,\ldots,x_n\in\mathbb{R}$, for each $x_0<x_1<\ldots<x_n$ and let $D=\left\{x_0,x_1,\ldots,x_n\right\}$ be a partition which corresponding to the function $T_D:C_B(\mathbb{R})\to C_B(\mathbb{R})$ which is given by

1. $T_D\left(f\left(x_0\right)\right)=T_D\left(f\left(x_n\right)\right)=0$,
2. $T_D\left(f\left(x_k\right)\right)=f\left(x_k\right)$, for all $k=1,2,\ldots,n-1$.

T_D is called a linear spline mapping which corresponding to the partition $D=\left\{x_0,x_1,\ldots,x_n\right\}$.

Remark 1. From the Definition 1, we obtain that for all $x\in\mathbb{R}-D$,

$$T_D\left(f(x)\right)=\sum_{i=0}^{n-1}\left(\left(\frac{\Delta T_D\left(f\left(x_{i+1}\right)\right)}{\Delta x_{i+1}}\right)\left(x-x_i\right)+T_D\left(f\left(x_i\right)\right)\right)\chi_{\left[x_i,x_{i+1}\right)}(x),$$

$x_i<x<x_{i+1}$.

Proposition 2. Let $D=\left\{x_0,x_1,\ldots,x_n\right\}$ be a partition which corresponding to the function $T_D:C_B(\mathbb{R})\to C_B(\mathbb{R})$. Then T_D is a linear operator.

Proof. To show that T_D is a linear operator, we prove two properties as the following:

1. $T_D(f+g)(x)=T_D(f)(x)+T_D(g)(x)$, for all $f,g\in C_B(\mathbb{R})$.
2. $T_D(\alpha f)(x)=\alpha T_D(f)(x)$, for all $\alpha\in\mathbb{R}$ and $f\in C_B(\mathbb{R})$.

Let $f,g\in C_B(\mathbb{R})\, x\in\mathbb{R}$.

Consider

$$\begin{aligned}
&T_D(f)(x)+T_D(g)(x)\\
&=\sum_{i=0}^{n-1}\left[\left(\frac{\Delta T_D(f)\left(x_{i+1}\right)}{\Delta x_{i+1}}\right)\left(x-x_i\right)+T_D(f)\left(x_i\right)\right]\chi_{\left[x_i,x_{i+1}\right)}(x)\\
&\quad+\sum_{i=0}^{n-1}\left[\left(\frac{\Delta T_D(g)\left(x_{i+1}\right)}{\Delta x_{i+1}}\right)\left(x-x_i\right)+T_D(g)\left(x_i\right)\right]\chi_{\left[x_i,x_{i+1}\right)}(x)\\
&=\left[\left(\frac{\Delta T_D(f)\left(x_1\right)}{\Delta x_1}\right)\left(x-x_0\right)+T_D(f)\left(x_0\right)\right]\chi_{\left[x_0,x_1\right)}(x)+\cdots+\\
&\quad\left[\left(\frac{\Delta T_D(f)\left(x_n\right)}{\Delta x_n}\right)\left(x-x_{n-1}\right)+T_D(f)\left(x_{n-1}\right)\right]\chi_{\left[x_{n-1},x_n\right)}(x)\\
&\quad+\left[\left(\frac{\Delta T_D(g)\left(x_1\right)}{\Delta x_1}\right)\left(x-x_0\right)+T_D(g)\left(x_0\right)\right]\chi_{\left[x_0,x_1\right)}(x)+\cdots+\\
&\quad\left[\left(\frac{\Delta T_D(g)\left(x_n\right)}{\Delta x_n}\right)\left(x-x_{n-1}\right)+T_D(g)\left(x_{n-1}\right)\right]\chi_{\left[x_{n-1},x_n\right)}(x)\\
&=\sum_{i=0}^{n-1}\left[\left(\frac{\Delta T_D(f+g)\left(x_{i+1}\right)}{\Delta x_{i+1}}\right)\left(x-x_i\right)+T_D(f+g)\left(x_i\right)\right]\chi_{\left[x_i,x_{i+1}\right)}(x)\\
&=T_D(f+g)(x).
\end{aligned}$$

Let $\alpha\in\mathbb{R}$. Consider

$$\begin{aligned}
&\alpha T_D(f)(x)\\
&=\alpha\sum_{i=0}^{n-1}\left[\left(\frac{\Delta T_D(f)\left(x_{i+1}\right)}{\Delta x_{i+1}}\right)\left(x-x_i\right)+T_D(f)\left(x_i\right)\right]\chi_{\left[x_i,x_{i+1}\right)}(x)\\
&=\sum_{i=0}^{n-1}\left[\left(\frac{\Delta\alpha T_D(f)\left(x_{i+1}\right)}{\Delta x_{i+1}}\right)\left(x-x_i\right)+\alpha T_D(f)\left(x_i\right)\right]\chi_{\left[x_i,x_{i+1}\right)}(x)\\
&=\sum_{i=0}^{n-1}\left[\left(\frac{\Delta\alpha f\left(x_{i+1}\right)}{\Delta x_{i+1}}\right)\left(x-x_i\right)+\alpha f\left(x_i\right)\right]\chi_{\left[x_i,x_{i+1}\right)}(x)\\
&=\sum_{i=0}^{n-1}\left[\left(\frac{\Delta T_D(\alpha f)\left(x_{i+1}\right)}{\Delta x_{i+1}}\right)\left(x-x_i\right)+T_D(\alpha f)\left(x_i\right)\right]\chi_{\left[x_i,x_{i+1}\right)}(x)\\
&=T_D(\alpha f)(x).
\end{aligned}$$

Q.E.D.

Proposition 3. Let $\left(C_B(\mathbb{R}),\|.\|_\infty\right)$ be a normed space for all $x\in\mathbb{R}$, $D=\{x_0,x_1,\dots,x_n\}$ be a partition which corresponding to the function $T_D:C_B(\mathbb{R})\to C_B(\mathbb{R})$ and T_D be a linear spline mapping. Then T_D is bounded linear operator.

Proof. Let $f\in C_B(\mathbb{R})$. Consider

$$\begin{aligned}\|T_D(f)\|_\infty &\le \max\left\{|f(x_0)|,|f(x_1)|,\dots,|f(x_{n-1})|,|f(x_n)|\right\}\\ &\le \sup_{x\in\mathbb{R}}\left\{|f(x)|\right\}\\ &=\|f\|_\infty.\end{aligned}$$

Q.E.D.

Remark 4. Let $L_B^1=\left\{f:\int_{\mathbb{R}}|f|\,dx<\infty\right\}, f\in L_B^1(\mathbb{R})$, $D=\{x_0,x_1,\dots,x_n\}$ be a partition which corresponding to the function $T_D:C_B(\mathbb{R})\to C_B(\mathbb{R})$ and $IT_D:L_B^1(\mathbb{R})\to\mathbb{R}$ such that $IT_D(f)=\int_{-\infty}^{\infty}(T_Df)(x)dx.$ Then IT_D is a linear operator.

Proposition 5. Let $D=\{x_0,x_1,\dots,x_n\}$ be a partition which corresponding to the function $T_D:C_B(\mathbb{R})\to C_B(\mathbb{R})$, $\left(L_B^1(\mathbb{R}),\|.\|_\infty\right)$ be a normed space for all $x\in\mathbb{R}$, and $IT_D:L_B^1(\mathbb{R})\to\mathbb{R}$ be a linear spline mapping. Then IT_D is bounded.

Proof. Let $f\in L_B^1(\mathbb{R})$. Consider

$$\begin{aligned}&\|IT_D(f)\|_\infty\\ &\le \max\left\{n|f(x_0)|,n|f(x_1)|,\dots,n|f(x_{n-1})|,n|f(x_n)|\right\}\\ &=n\max\left\{|f(x_0)|,|f(x_1)|,\dots,|f(x_{n-1})|,|f(x_n)|\right\}\\ &\le n\sup_{x\in\mathbb{R}}|f(x)|\\ &=n\|f\|_\infty.\end{aligned}$$

Q.E.D.

Theorem 6. Let $D=\{x_0,x_1,\dots,x_n\}$ be a partition corresponding to a linear spline mapping $T_D:C_B(\mathbb{R})\to C_B(\mathbb{R})$. Then

$$\int_{x_0}^{x_n}T_D(f(x))dx=\sum_{i=0}^{n-1}\overline{f(x_{i+1})}\Delta x_{i+1}.$$

Proof. Consider

$$\begin{aligned}&\int_{x_0}^{x_n}T_D(f(x))dx\\ &=\sum_{i=0}^{n-1}\left(\frac{\Delta f(x_{i+1})}{\Delta x_{i+1}}\frac{(x-x_i)^2}{2}+f(x_i)x\right)\chi_{[x_i,x_{i+1})}(x)\Bigg|_{x=x_i}^{x=x_{i+1}}\\ &=\sum_{i=0}^{n-1}\left(\frac{\Delta f(x_{i+1})}{\Delta x_{i+1}}\frac{(x_{i+1}-x_i)^2}{2}+f(x_i)x_{i+1}\right)\\ &\quad-\sum_{i=0}^{n-1}\left(\frac{\Delta f(x_{i+1})}{\Delta x_{i+1}}\frac{(x_i-x_i)^2}{2}+f(x_i)x_i\right)\\ &=\sum_{i=0}^{n-1}\left(\frac{\Delta f(x_{i+1})}{2}\Delta x_{i+1}+f(x_i)\Delta x_{i+1}\right)\\ &=\sum_{i=0}^{n-1}\overline{f(x_{i+1})}\Delta x_{i+1}.\end{aligned}$$

Q.E.D.

Definition 2. Let $D=\{x_0,x_1,\dots,x_n\}$ be a partition corresponding to a linear spline mapping $T_D:C_B(\mathbb{R})\to C_B(\mathbb{R})$. The partition D is called a $P-$partition corresponding to T_D for a probability density function $f\in C_B(\mathbb{R})$, if

$$\int_{-\infty}^{\infty}T_D(f(x))dx=1.$$

In addition, T_Df is a probability density function.

Theorem 7. Let $x\in\mathbb{R}$, $D=\{x_0,x_1,\dots,x_n\}$, $x_0<x_1<\cdots<x_n$ where

$$\Delta x_i=\begin{cases}d, & i=2,3,\dots,n-1\\ \ell, & i=1 \text{ or } i=n,\end{cases}$$

and $d>0$, $\ell=\dfrac{1}{\phi(x_1)}\left[1-\sum_{i=2}^{n-1}\overline{\phi(x_i)}\,d\right]$, $T_D:C_B(\mathbb{R})\to C_B(\mathbb{R})$ be a linear spline mapping and $\phi(x)$ be any probability density function. Then D is $P-$partition for the probability density function ϕ.

Proof. Consider

$$\begin{aligned}&\int_{-\infty}^{\infty}T_D(\phi(x))\,dx\\ &=\int_{x_0}^{x_n}T_D(\phi(x))\,dx\\ &=\int_{x_0}^{x_1}T_D(\phi(x))\,dx+\int_{x_1}^{x_{n-1}}T_D(\phi(x))\,dx+\int_{x_{n-1}}^{x_n}T_D(\phi(x))\,dx\end{aligned}$$

$$= \overline{\phi(x_1)}\,\ell + \sum_{i=2}^{n-1} \overline{\phi(x_i)}\, d + \overline{\phi(x_n)}\,\ell$$
$$= \phi(x_1)\ell + \sum_{i=2}^{n-1} \overline{\phi(x_i)}\, d$$
$$= \phi(x_1)\frac{1}{\phi(x_1)}\left[1 - \sum_{i=2}^{n-1} \overline{\phi(x_i)}\, d\right] + \sum_{i=2}^{n-1} \overline{\phi(x_i)}\, d$$
$$= 1.$$

Thus D is $P-$partition for a probability density function ϕ.

Q.E.D.

Corollary 8. Let $D = \{x_0, x_1, \ldots, x_n\}$ be a $P-$partition corresponding to a linear spline mapping $T_D : C_B(\mathbb{R}) \to C_B(\mathbb{R})$ where $x_0 < x_1 < \cdots < x_n$,

$$\Delta x_i = \begin{cases} d, & i = 2,3,\ldots,n-1 \\ \ell, & i = 1 \text{ or } i = n, \end{cases}$$

$d > 0$, $\ell = \frac{1}{\phi(x_1)}\left[1 - \sum_{i=2}^{n-1} \overline{\phi(x_i)}\, d\right]$ and $\phi(x)$ be any probability density function. Then $\sigma D + \mu$ is a partition corresponding to a $\sigma T_{\sigma D+\mu}$ for a probability density function of normal distribution $N(\mu, \sigma^2)$.

Proof. For each $x_i \in D$, we consider

$$f(\sigma x_i + \mu; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\sigma x_i + \mu - \mu}{\sigma}\right)^2}$$
$$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}x_i^2}$$
$$= \frac{1}{\sigma}\phi(x_i)$$

such that ϕ is a probability density function of a standard normal distribution. Consider

$$\int_{-\infty}^{\infty} \sigma T_{\sigma D+\mu}\left(f(\sigma x_i + \mu; \mu, \sigma)\right) dx$$
$$= \sigma \int_{-\infty}^{\infty} T_{\sigma D+\mu}\left(\frac{1}{\sigma}\phi(x)\right) dx$$
$$= \sigma \frac{1}{\sigma} \int_{-\infty}^{\infty} T_{\sigma D+\mu}\left(\phi(x)\right) dx$$
$$= \sum_{i=0}^{n-1}\left(\frac{\Delta\phi(x_{i+1})}{\Delta x_{i+1}} \frac{(x - x_i)^2}{2} + \phi(x_i)x\right)\chi_{[x_i, x_{i+1})}(x)\Bigg|_{x=x_i}^{x=x_{i+1}}$$
$$= \sum_{i=0}^{n-1}\left(\frac{\Delta\phi(x_{i+1})}{\Delta x_{i+1}} \frac{(\Delta x_{i+1})^2}{2} + \phi(x_i)x_{i+1} - \frac{\Delta\phi(x_{i+1})}{\Delta x_{i+1}}\frac{(x_i - x_i)^2}{2} - \phi(x_i)x_i\right)$$
$$= \sum_{i=0}^{n-1}\left(\frac{\Delta\phi(x_{i+1})}{2}\Delta x_{i+1} + \phi(x_i)\Delta x_{i+1}\right)$$
$$= \sum_{i=0}^{n-1} \overline{\phi(x_{i+1})}\Delta x_{i+1}$$
$$= 1.$$

Q.E.D.

Next, we illustrate to support our theorems that help reader to understand the approach. Firstly, we consider

$$f(x) = \begin{cases} 0 & ; x \le x_0 \\ 5 + \frac{5}{\ell}(x - x_1) & ; x_0 < x \le x_1 \\ 5 & ; x_1 < x \le x_2 \\ 5 - \frac{5}{\ell}(x - x_2) & ; x_2 < x \le x_3 \\ 0 & ; x \ge x_3, \end{cases}$$

and set $d = \frac{1}{6}$, we calculate ℓ value as the following equation:

$$\ell = \frac{1}{f(x_1)}\left[1 - \frac{f(x_1) + f(x_2)}{2} d\right] = \frac{1}{5}\left[1 - \left(\frac{5+5}{2}\right)\frac{1}{6}\right] = \frac{1}{30},$$

and then we get

$$\int_{x_0}^{x_3} f(x)\, dx = \frac{1}{2}f(x_1)\ell + \frac{f(x_1) + f(x_2)}{2}d + \frac{1}{2}f(x_2)\ell$$
$$= \frac{1}{2}(5)\left(\frac{1}{30}\right) + \left(\frac{5+5}{2}\right)\left(\frac{1}{6}\right) + \frac{1}{2}(5)\left(\frac{1}{30}\right)$$
$$= 1.$$

Theorem 9. Let $D = \{x_0, x_1, \ldots, x_n\}$ be a $P-$partition corresponding to a linear spline mapping $T_D : C_B(\mathbb{R}) \to C_B(\mathbb{R})$, $x_0 < x_1 < \cdots < x_n$ where

$$\Delta x_i = \begin{cases} d, & i = 2,3,\ldots,n-1 \\ \ell, & i = 1 \text{ or } i = n, \end{cases}$$

$d > 0$, $\ell = \frac{1}{\phi(x_1)}\left[1 - \sum_{i=2}^{n-1} \overline{\phi(x_i)}\, d\right]$, ϕ be probability density function of normal distribution and $\phi(x_i) = f(x_i)$. If $0 < \sum_{i=2}^{n-1} \overline{\phi(x_i)}\, d < 1$, $N(x) = \min\{k : x \le x_k\}$, $CT_D f(x) = \int_{-\infty}^{x} T_D(f(s))\chi_{[x_i, x_{i+1})}\, ds$, and given

$$CT_D\left(f(x)\right)=\begin{cases}0 & ;x\le x_0\\ \dfrac{f(x_1)}{\ell}\left(\dfrac{x^2}{2}-x_0x+\dfrac{x_0^2}{2}\right) & ;x_0<x\le x_1\\ \dfrac{f(x_1)\ell}{2}+\sum\limits_{i=2}^{N(x)-1}\overline{f(x_i)}d+\dfrac{\Delta f\left(x_{N(x)}\right)}{2d}x^2- & \\ \left[\dfrac{\Delta f\left(x_{N(x)}\right)}{d}x_{N(x)-1}-f\left(x_{N(x)-1}\right)\right]x+ & \\ \dfrac{\Delta f\left(x_{N(x)}\right)}{2d}x_{N(x)-1}^2-f\left(x_{N(x)-1}\right)x_{N(x)-1} & ;x_1<x\le x_{N(x)}\\ CT_D f(x_{n-1})+\dfrac{\Delta f(x_n)}{2\ell}x^2- & \\ \left[\dfrac{\Delta f(x_n)}{\ell}x_{n-1}-f(x_{n-1})\right]x+ & \\ \dfrac{\Delta f(x_n)}{2\ell}x_{n-1}^2-f(x_{n-1})x_{n-1} & ;x_{n-1}<x\le x_n\\ 1 & ;x\ge x_n\end{cases}$$

Then $CT_D f$ is the cumulative distribution function of T_D.

Proof. Since $T_D\left(f(x_0)\right)=T_D\left(f(x_n)\right)=0,$ we get obviously

$$\int_{-\infty}^{x_0}T_D\left(f(x)\right)dx=0=\int_{x_n}^{\infty}T_D\left(f(x)\right)dx.$$

We consider in case of $x_0<x\le x_1$,

$$CT_D f(x_1)=\int_{-\infty}^{\infty}T_D\left(f(x)\right)\chi_{[x_0,x)}dx$$
$$=\int_{x_0}^{x}\left[\frac{\Delta T_D\left(f(x_1)\right)}{\Delta x_1}(s-x_0)+T_D\left(f(x_0)\right)\right]ds$$
$$=\left[\frac{\Delta T_D\left(f(x_1)\right)}{\Delta x_1}\left(\frac{s^2}{2}-x_0s\right)+T_D\left(f(x_0)\right)s\right]_{s=x_0}^{s=x}$$

$CT_D f(x_1)$

$$=\frac{\Delta T_D f(x_1)}{\Delta x_1}\left(\frac{x^2-x_0^2}{2}-x_0(x-x_0)\right)+T_D\left(f(x_0)\right)(x-x_0)$$
$$=\frac{f(x_1)}{\ell}\left(\frac{x^2-x_0^2}{2}-x_0(x-x_0)\right)$$
$$=\frac{f(x_1)}{\ell}\left(\frac{x^2}{2}-x_0x+\frac{x_0^2}{2}\right).$$

For $x=x_1$, we obtain $CT_D f(x_1)=\dfrac{T_D\left(f(x_1)\right)}{2}\ell.$

Next, we consider in case of $x_{i-1}<x\le x_i$, $i=1,2,\ldots,N(x)$ where $N(x)=\min\{k:x\le x_k\}$ so

$CT_D f\left(x_{N(x)}\right)$

$$=\int_{-\infty}^{\infty}T_D\left(f(x)\right)\chi_{[x_{i-1},x_i)}dx$$
$$=\sum_{i=1}^{N(x)}\int_{x_{i-1}}^{x}\left[\frac{\Delta T_D\left(f(x_i)\right)}{\Delta x_i}(s-x_{i-1})+T_D\left(f(x_{i-1})\right)\right]ds$$
$$=\frac{\Delta T_D\left(f(x_1)\right)}{\Delta x_1}\left(\frac{x^2-x_0^2}{2}-x_0(x-x_0)\right)+T_D\left(f(x_{i-1})\right)(x-x_0)+$$
$$\frac{\Delta T_D\left(f(x_2)\right)}{\Delta x_2}\left(\frac{x^2-x_1^2}{2}-x_1(x-x_1)\right)+T_D\left(f(x_1)\right)(x-x_1)+\cdots+$$
$$\frac{\Delta T_D\left(f\left(x_{N(x)-1}\right)\right)}{\Delta x_{N(x)-1}}\left(\frac{x^2-x_{N(x)-2}^2}{2}-x_{N(x)-2}\left(x-x_{N(x)-2}\right)\right)+$$
$$T_D\left(f\left(x_{N(x)-2}\right)\right)\left(x-x_{N(x)-2}\right)+\int_{x_{N(x)-1}}^{x}T_D\left(f(s)\right)ds$$
$$=\frac{\left(f(x_1)\right)\ell}{2}+\sum_{i=2}^{N(x)-1}\overline{f(x_i)}d+\frac{\Delta f\left(x_{N(x)}\right)}{2d}x^2-$$
$$\left[\frac{\Delta f\left(x_{N(x)}\right)}{d}x_{N(x)-1}-f\left(x_{N(x)-1}\right)\right]x+$$
$$\frac{\Delta f\left(x_{N(x)}\right)}{2d}x_{N(x)-1}^2-f\left(x_{N(x)-1}\right)x_{N(x)-1}.$$

For $N(x)=n-1$, we obtain

$$CT_D f\left(x_{N(x)}\right)=\frac{\left(f(x_1)\right)}{2}\ell+\sum_{i=2}^{n-1}\overline{f(x_i)}d.$$

Next, we combine in case of $x_{i-1}<x\le x_i$, $i=1,2,\ldots,n-1$ and $x_{n-1}<x\le x_n$, we obtain

$CT_D\left(f(x)\right)$

$$=\int_{-\infty}^{\infty}T_D\left(f(x)\right)\chi_{[x_0,x_{n-1})}dx+\int_{-\infty}^{\infty}T_D\left(f(x)\right)\chi_{[x_{n-1},x_n)}dx$$
$$=CT_D\left(f\left(x_{N(x)}\right)\right)+\int_{n-1}^{x}\left[\frac{\Delta T_D\left(f(x_n)\right)}{\Delta x_n}(s-x_{n-1})+T_D\left(f(x_{n-1})\right)\right]ds$$
$$=CT_D\left(f\left(x_{N(x)}\right)\right)+\frac{\Delta T_D\left(f(x_n)\right)}{2\ell}x^2-$$
$$\left[\Delta T_D\left(f(x_n)\right)x_{n-1}-T_D\left(f(x_{n-1})\right)\right]x+$$
$$\frac{\Delta T_D\left(f(x_n)\right)}{\ell}x_{n-1}^2-T_D\left(f(x_{n-1})\right)x_{n-1}$$
$$=CT_D f\left(x_{N(x)}\right)+\frac{\Delta f(x_n)}{2\ell}x^2-\left(\frac{\Delta f(x_n)}{\ell}x_{n-1}-f(x_{n-1})\right)x+$$
$$\frac{\Delta f(x_n)}{2\ell}x_{n-1}^2-f(x_{n-1})x_{n-1}.$$

For $N(x)=n$, we obtain

$$CT_D f(x_n)=\frac{\left(f(x_1)\right)}{2}\ell+\sum_{i=2}^{n-1}\overline{f(x_i)}d+\frac{\left(f(x_{n-1})\right)}{2}\ell$$

$$= \left(f\left(x_1\right)\right)\ell + \sum_{i=2}^{n-1} \overline{f\left(x_i\right)} d$$
$$= f\left(x_1\right)\frac{1}{f\left(x_1\right)}\left[1-\sum_{i=2}^{n-1}\overline{f\left(x_{i-1}\right)}\, d\right]+\sum_{i=2}^{n-1}\overline{f\left(x_i\right)} d$$
$$=1.$$

Q.E.D.

Next, we propose the parameter estimation and the inverse of the cumulative distribution function of $T_D f\left(x\right)$ that it has useful for generating a random variable of some data as the Remark 10. We estimate parameter of $T_D f\left(x\right)$ based on normal distribution using method of moments (MM) and use the 77 data of the student's average intelligent quotient (IQ) for Grade 1 that survey between November 2015 and February 2016 in all provinces of Thailand as shown the frequency of data as Figure 1.

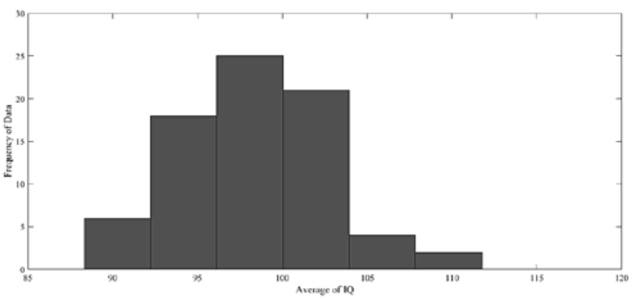

Fig. 1 Frequency of average IQ of student Grad 1 in Thailand

Table 1. Test of normality of the interesting data

Kolmogorov-Smirnov			Shapiro-Wilk		
Statistic	df	Sig.	Statistic	df	Sig.
0.056855	77	0.200*	0.990617	77	0.848593

Table 1 shows that the interesting data do not reject normal distribution at the 0.05 significance level.

Next, we propose the algorithm of parameter estimation as the following steps:

Step 1. Input data and sort data by ascending.

Step 2. Estimate parameter of normal distribution using the MM. If we have the dataset $x_1, x_2, \ldots, x_m$, we get the MM of normal distribution that have the location parameter μ and the scale parameter σ approximate by

$$\hat{\mu} = \frac{1}{m}\sum_{i=1}^{m} x_i = \overline{x},\ \ \hat{\sigma} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(x_i - \overline{x}\right)^2}.$$

Step 3. Set N (is node) for determining the intervals of distribution with the width of the intervals is equal to d for each interval $\left[x_i, x_{i+1}\right], i = 1, 2, \cdots, n-2$. The other intervals are $\left[x_0, x_1\right] = \ell = \left[x_{n-1}, x_n\right]$.

Step 4. Compute KS -statistics,

$$KS = \max_{\theta}\left|CT_D f\left(x_i;\theta\right) - \frac{i}{m}\right|,$$

where θ is the parameter of distribution and m is the size of data.

Remark 10. Let $u \in \left(0,1\right)$ be a uniform distribution function.

If $CT_D f\left(x_0\right) < u \le CT_D f\left(x_1\right)$ then

$$x = x_0 \pm \sqrt{\frac{2u\ell}{f\left(x_1\right)}}.$$

If $CT_D f\left(x_{i-1}\right) < u \le CT_D f\left(x_i\right),\ i = 1, 2, 3, \ldots, N\left(x\right)$ where $N\left(x\right) = \min\left\{k : x \le x_k\right\}$ then we set

$$A = \frac{\Delta f\left(x_{N(x)}\right)}{d},\ B = \frac{\Delta f\left(x_{N(x)}\right)}{d} x_{N(x)-1} - f\left(x_{N(x)-1}\right), \text{ and}$$
$$C = \frac{\Delta f\left(x_{N(x)}\right)}{d} x_{N(x)-1}^2 - 2 f\left(x_{N(x)-1}\right) x_{N(x)-1} + 2CT_D f\left(x_{N(x)-1}\right) - 2u,$$

i.e., $x = \frac{1}{A}\left(B \pm \sqrt{B^2 - AC}\right).$

If $CT_D f\left(x_{n-1}\right) < u \le CT_D f\left(x_n\right)$ then we set

$$H = \frac{f\left(x_n\right)}{\ell},\ J = \frac{\Delta f\left(x_n\right)}{\ell} x_{n-1} - f\left(x_{n-1}\right),$$

and $K = \frac{\Delta f\left(x_n\right)}{\ell} x_{n-1}^2 - 2 f\left(x_{n-1}\right) x_{n-1} + 2CT_D f\left(x_{N(x)}\right) - 2u,$

i.e., $x = \frac{1}{H}\left(J \pm \sqrt{J^2 - HK}\right).$

In the rest of this section, we propose the KS -statistics of the interesting data through MM as shown in Table 1 and Fig. 2-4. The experimental situations consider $N = 1, 2, \ldots, 300$ for $d = 0.1$, $N = 1, 2, \ldots, 2000$ for $d = 0.01$, and $N = 1, 2, \ldots, 3000$ for $d = 0.001$. Table 2 shows the minimum KS -statistics of the interesting data for $d = 0.1, 0.01$ and 0.001 which found that at $N = 26, 259$ and 2591 obtain the minimum KS -statistics as 0.0497. Moreover, we see that the new distribution has KS -statistics less than of the normal distribution that show in Table 1 and Table 2.

Table 2. The minimum KS -statistics of the interesting data through MM

N	d	ℓ	$\hat{\mu}$	$\hat{\sigma}$	KS -statistics
26	0.1	6.9098	98.1642	4.2340	0.0497
259	0.01	6.9195	98.1642	4.2340	0.0497
2591	0.001	6.9185	98.1642	4.2340	0.0497

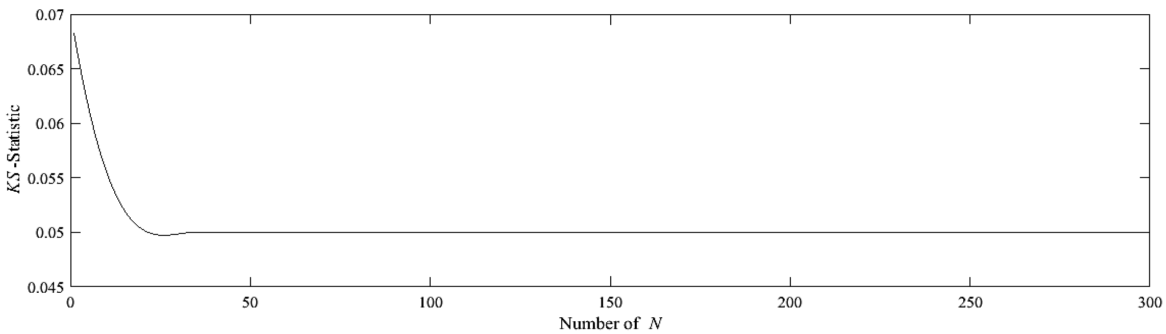

Fig. 2 KS -statistics of the interesting data for $d = 0.1$

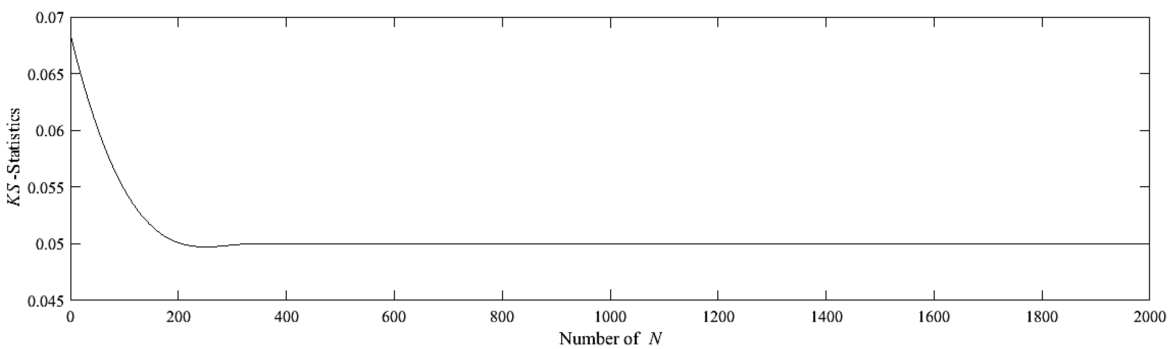

Fig. 3 KS -statistics of the interesting data for $d = 0.01$

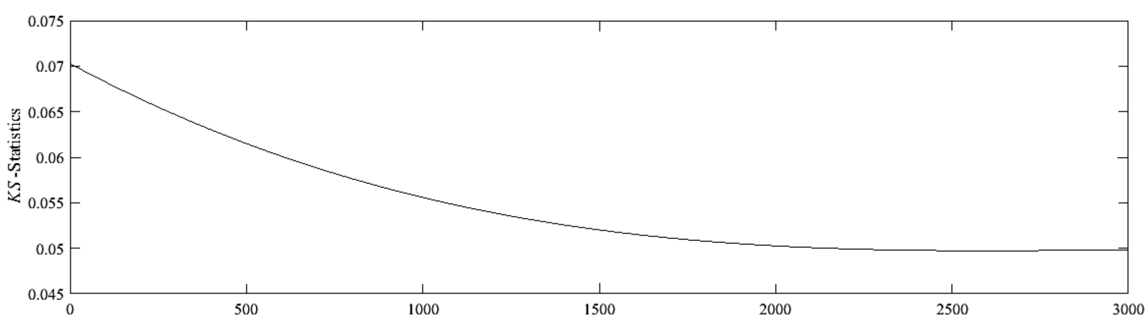

Fig. 4 KS -statistics of the interesting data for $d = 0.001$

3 Conclusion

The linear spline mapping on the probability density function of normal distribution can be represent the normal distribution. Moreover, the new distribution has the inverse of cumulative distribution function that can generate the random variable for determining some of data which give the value of parameter as of normal distribution.

Acknowledgment I

The authors would like to thank both reviewers for their insightful comments on the paper, as these comments led us to an improvement of the work.

Acknowledgment II

This research is supported by Mathematics and Statistics Program, Faculty of Science and Technology, Sakon Nakhon Rajabhat University, Thailand.

.

REFERENCES

[1] S.R. Bowling, M.T. Khasawneh, S. Kaewkuekool and B.R. Cho. "A logistic approximation to the cumulative normal distribution". Journal of Industrial Engineering and Management, Vol. 2, pp. 114-127, Mar. 2009.
[2] H. Vazquez-Leal, R. Castaned-Sheissa, U. Filobello-Nino, A. Sarmient-Reyes and J.S. Orea. "High accurate simple approximation of normal distribution integral". Mathematical Problem in Engineering, Vol. 12, pp. 1-22, Oct. 2012.
[3] A. Choudhury. "A simple approximation to the area under standard normal curve". Mathematics and Statistics, Vol. 2, pp. 147-149, 2014.
[4] C. Chen. "Spline estimators of the distribution function of a variable measured with error", Ph.D. dissertation, Statistics, Iowa State Univ., 1999.
[5] M.J. Lindstrom. "Penalized estimation of free-knot splines". Journal of Computational and Graphical Statistics, Vol. 8, pp. 333-352, Jun. 1999.
[6] Q. Zhang and X. Lin. "On linear spline based histograms". WAIM, pp. 354-366, 2002.
[7] A.D. Holland. "Penalized spline estimation in the partially linear model", Ph.D. dissertation, Statistics, Applied and Interdisciplinary Mathematics, The University of Michigan. 2012.
[8] O.Valenzuela, M.Pasadas, F. Ortuño, and I.Rojas. "Optimal Knots Allocation in Smoothing Splines using intelligent system. Application in bio-medical signal processing". In *Proc. 4th IWBBIO*, Spain, 2013, pp. 289-295.
[9] H. Kang, F. Chen, Y. Li, J. Deng, and Z. Yang. "Knot calculation for spline fitting via sparse optimization". ADC, Vol. 58, pp. 179-188 Jan. 2015.
[10] T. Tjahjowidodo, VT. Dung, and ML. Han. "A Fast Non-Uniform Knots Placement Method for B-Spline Fitting". IEEE Int. Conf. on AIM, Busan, Korea, pp. 1490-1495, July 2015.
[11] M. Z. Hussain, S. Abbas, and M. Irshad. "Quadratic trigonometric B-spline for image interpolation using GA". PLoS ONE 12(6): e0179721. Available online: https://doi.org/10.1371/journal.pone.0179721