

# The Linear Probability Density Function of Continuous Random Variables in the Real Number Field and Its Existence Proof

Yiyan Chen<sup>1</sup> and Ye Li<sup>2</sup>

<sup>1</sup>College of Software, Tianjin University, Tianjin, China

<sup>2</sup>College of Mathematical Science, Capital Normal University, Beijing, China

**Abstract.** For a continuous random variable in real number field, there must be a distribution and also a probability density function of this random variable. If there is a known function with this random variable as independent variable, its image is a smooth or piecewise smooth line, there must be at least one function that takes this random variable as its independent variable, these functions are bounded on the image of the first function. Any one of these functions conduct line integral operation to the line segment or arc length of the certain image of the first known function is the cumulative probability of this continuous random variable interval corresponding to the section of the image for line integral operation. A general designation for these functions are linear probability density function of continuous random variables. Conduct line integral operation to the linear probability density function and conduct integral operation to the probability density function have same results of the cumulative probability of continuous random variable. By the way, Line integration including curve integration. According to the uniqueness of the probability, the existence and the number of linear probability density function can be proved and calculated.

## 1 Introduction

For a continuous random variable as  $X$ ,  $X \in R$ , which meet the condition  $X \sim \Delta_X$ , obviously, there is a probability density function  $f(x)$ , meet the condition that  $\int_{-\infty}^x f(t)dt = F(x)$ .  $F(x)$  is the cumulative probability function of  $X$ . With reference to examples of curvilinear integral<sup>[1,2]</sup>, If the integral domain of  $X$ , which equivalent to interval of the value of random variable  $X$  from real number axis, changes to the function image ( or a part of it ) of  $g(x)$  which is a continuous function or countable piecewise function, there should be another function  $h(x)$  in real number field, according to the uniqueness of probability, it meets  $\int_{g(x)} h(u)du = F(x)$ . The integration path of continuous random variable  $X$  is a moving path of the points  $(x, g(x))$ , composed of the value of  $X$  and its corresponding value of the value  $g(x)$ , on the  $g(x)$  moving in the direction of  $x$ -axis.  $h(x)$  are called the linear probability density function of the continuous random variable  $X$  in the real number field. To this end, its need to define the linear probability density function, and prove its existence and the number of the function in real number field.

## 2 Relevant definition

Here is a need to give definition and properties of linear integral, linear probability density function.

### 2.1 Linear integral

#### 2.1.1 Definition of Linear integral

Set  $X$ ,  $X \in R$ ,  $X$  is a continuous random variable in the real number field,  $X \sim \Delta_X$ . Its probability density function is  $f(x)$ , cumulative probability function is  $F(x)$ , set  $g(x)$  is a function of random variables with  $X$  as independent variable, its image  $L$  is smooth or piecewise smooth in plane  $xOy$ , there is another function  $h(x)$  with independent variable the same as  $X$ ,  $h(x)$  is bounded on the image of  $g(x)$ , intercept an image  $L^*$  from  $L$ , its endpoints are  $A(x_m, g(x_m))$  and  $B(x_n, g(x_n))$ , using unlimited division, insert a finite point range  $M_1, M_2, \dots, M_{n-1}$  in the  $L^*$ , divide  $L^*$  into  $n$  short sections, set the length of the  $i$ th section is  $\Delta_{S_i}$ , also  $(\xi_i, \eta_i)$  is any point on the  $i$ th section, make

product  $h(\xi_i, \eta_i)\Delta_{s_i}$ ,  $i=1,2,\dots,n$ , and also make product sums  $\sum_{i=1}^n h(\xi_i, \eta_i)\Delta_{s_i}$ ,  $i=1,2,\dots,n$ , in this case, if  $\max[h(\xi_i, \eta_i)\Delta_{s_i}] = \lambda \rightarrow 0$ ,  $\lim_{\lambda \rightarrow 0} \sum_{i=1}^n h(\xi_i, \eta_i)\Delta_{s_i}$  is always exist, then called this limit as a function  $h(x, g(x))$  with doing linear integral to  $L^*$  on image of  $g(x)$ . The image  $L$  of  $g(x)$  could be a straight line, could also be a curve, could be bounded but also could be unbounded, but must meeting the smooth or piecewise smooth.

### 2.1.2 Classification of solving linear integral

If the image  $L$  of  $g(x)$  is bounded, set  $L = L^*$ , according to the methods just now, mark the linear integral as  $\int_L h(x, g(x))ds$ , there are:

$$\int_L h(x, g(x))ds = \lim_{\lambda \rightarrow 0} \sum_{i=1}^n h(\xi_i, \eta_i)\Delta_{s_i} \quad (1)$$

If  $L$  is unbounded, then  $L^* \subset L$ , two endpoints of  $L^*$  are  $A(x_m, g(x_m))$  and  $B(x_n, g(x_n))$ , if it is need to required out linear integral of  $L^*$ , there are:

$$\begin{aligned} & \int_{L^*} h(x, g(x))ds \\ &= \int_{L(A \rightarrow B)} h(x, g(x))ds = \lim_{\lambda \rightarrow 0} \sum_{i=1}^n h(\xi_i, \eta_i)\Delta_{s_i} \end{aligned} \quad (2)$$

If  $L$  is unbounded, then  $L^* \subset L$ , if it is need to required out linear integral of  $L$ , insert an infinite point range  $M_1, M_2, \dots, M_{n-1}, \dots$  in  $L$ , divide  $L$  into  $r$  short sections,  $r \rightarrow +\infty$ , set the length of  $i$ th section is  $\Delta_{s_i}$ , also  $(\xi_i, \eta_i)$  is any point on the  $i$ th section, make product  $h(\xi_i, \eta_i)\Delta_{s_i}$ ,  $i=1,2,\dots,r$ , in this case, also

make  $\sum_{i=1}^r h(\xi_i, \eta_i)\Delta_{s_i}$ , if  $\max[h(\xi_i, \eta_i)\Delta_{s_i}] = \lambda \rightarrow 0$ ,  $\lim_{\lambda \rightarrow 0} \sum_{i=1}^r h(\xi_i, \eta_i)\Delta_{s_i}$  is always exist, then called this

limit as a function  $h(x, g(x))$  with doing linear integral to  $L$  on image of  $g(x)$ . Still mark the linear integral of  $L$  as  $\int_L h(x, g(x))ds$ , there are:

$$\int_L h(x, g(x))ds = \lim_{\lambda \rightarrow 0} \sum_{i=1}^r h(\xi_i, \eta_i)\Delta_{s_i} \quad (3)$$

In short, if it is a need to find out the entire or a part of image of  $g(x)$  for operating linear integral, the method of infinite division could be used to realize to solve the linear integral calculation in traditional.

## 2.2 Linear probability density function

### 2.2.1 Definition of linear probability density function

According to the last section, reference to the definition and properties of probability density function and cumulative probability function<sup>[3,4,5]</sup>,  $h(x)$  is a function of  $X$ ,  $g(x)$  is also a function of  $X$ , set  $F(x)$  as the cumulative probability function of  $X$ ,  $\int_{g(x)} h(u)du$  is operating linear integral by x-axis direction along the image of  $g(x)$ , if:

$$\int_{g(x)} h(u)du = F(x) \quad (4)$$

It is said  $g(x)$  is path of linear integral of  $X$ ,  $h(x)$  is the linear density function of  $X$ . Another note that  $h(x) = h(x, g(x))$ ,  $g(x)$  is not a variable of two-dimensional random but a parameter of  $h(x)$ .

### 2.2.2 Properties of linear probability density function

It is obvious that linear integral of the linear probability density function is a change form of the curve integral, and the linear probability density function is also a change form of probability density function, so the linear probability density function has the characteristics of both of them.

Property 1. If  $h(x)$  is the linear probability density function of a continuous random variable  $X$ , then:

$$h(x) \geq 0 \quad (5)$$

Property 2. If  $h(x)$  is the linear probability density function of a continuous random variable  $X$ ,  $g(x)$  is the path of linear integral,  $g(x|-\infty < x < +\infty)$  express the whole image of  $g(x)$  on x-axis, then:

$$\int_{g(x|-\infty < x < +\infty)} h(u)du = 1 \quad (6)$$

Property 3. Set  $\alpha, \beta$  as constants,  $h(x)$  is the linear probability density function of a continuous random variable  $X$ ,  $g(x)$  is the path of linear integral, if  $h(x) = \alpha h_1(x, g(x)) + \beta h_2(x, g(x))$ , then:

$$\begin{aligned} & \int_{g(x)} [\alpha h_1(u, g(u)) + \beta h_2(u, g(u))] du \\ &= \alpha \int_{g(x)} h_1(u, g(u)) du + \beta \int_{g(x)} h_2(u, g(u)) du \end{aligned} \quad (7)$$

Property 4.  $h(x)$  is the linear probability density function of a continuous random variable  $X$ ,  $g(x)$  is the path of linear integral, set the whole image of  $g(x)$  on the x-axis is  $g(x|-\infty < x < +\infty)$ , in this case, make  $g(x|-\infty < x < +\infty) = L$ , if  $L$  could be divided into a finite number of short sections as  $L_i$ ,  $i=1,2,\dots,n$ ,

$\sum_{i=1}^n L_i = L$  , then:

$$\int_L h(u)du = \sum_{i=1}^n \int_{L_i} h(u)du \quad (8)$$

Property 5.  $h(x)$  is the linear probability density function of a continuous random variable  $X$  ,  $g(x)$  is the path of linear integral, set the whole image of  $g(x)$  on the x-axis is  $g(x|-\infty < x < +\infty)$  , in this case, make  $g(x|-\infty < x < +\infty) = L$  , if  $h(x) \leq t(x)$  on the  $L$  , then:

$$\int_L h(u)du \leq \int_L t(u)du \quad (9)$$

Property 6.  $h(x)$  is the linear probability density function of a continuous random variable  $X$  ,  $g(x)$  is the path of linear integral,  $f(x)$  is the probability density function of  $X$  ,  $F(x)$  is cumulative probability of  $X$  , then:

$$\int_{g(x)} h(u)du = \int_{-\infty}^x f(t)dt = F(x) \quad (10)$$

Property 7 is the uniqueness of the probability, for a continuous random variable, it is bound to be the only existence of the cumulative probability in a given region, no matter what form of expression of probability. The following is a proof of it.

**Proof:**  $X$  is a continuous random variable, its probability density function is  $f(x)$  , its cumulative probability is  $F(x)$  , linear probability density function is  $h(x)$  , path of linear integral is  $g(x)$  ,  $g(x|a \leq x \leq b)$  express the image about  $x \in [a, b]$  on the  $g(x)$  , if there is a specified integral region which like  $a \leq x \leq b$  ,  $\int_a^b f(x)dx \neq \int_{g(x|a \leq x \leq b)} h(x)dx$  , set  $\int_a^b f(x)dx = P_1$  , set  $\int_{g(x|a \leq x \leq b)} h(x)dx = P_2$  , then  $P(a \leq x \leq b)$  there are two values  $P_1$  and  $P_2$  , and meanwhile, the rest of the integral region of random variables satisfied  $\int_{g(x)} h(u)du = \int_{-\infty}^x f(t)dt = F(x)$  , set the cumulative probability of the rest of integral region of random variables as  $P_s$  , according to the regularity of  $f(x)$  , it is a need to meet  $\int_{-\infty}^{+\infty} f(t)dt = 1$  , but  $P_1 \neq P_2$  , obviously,  $P_1 + P_s$  and  $P_2 + P_s$  must have one not equal to 1, it is unsatisfied to the regularity of probability density function, therefore, for any  $x$  ,  $x \in X$  , satisfied to  $\int_{g(x)} h(u)du = \int_{-\infty}^x f(t)dt = F(x)$ . That is to say, no matter the cumulative probability express by the integral of probability density function or linear integral of linear probability density function, for any value region of random variable  $X$  must be equal everywhere.

### 3 Proof of existence of linear probability density and its number

According to the uniqueness of the probability, it could prove the existence and number of linear probability density function. There are two kinds of situations, first one,  $f(x)$  is known,  $h(x)$  is unknown,  $g(x) \neq h(x)$  ; Second one,  $f(x)$  is known,  $h(x)$  is unknown,  $g(x) = h(x)$  . The meaning of  $f(x)$  ,  $g(x)$  and  $h(x)$  are the same as above section. If to prove the existence of linear probability density function, that is to prove the existence of  $h(x)$  , it could be based on the equation of the probability like  $\int_{g(x)} h(u)du = \int_{-\infty}^x f(t)dt = F(x)$  to solve  $h(x)$  , if the equation has real roots, it is proved the existence of  $h(x)$  in real number field, the number of real roots is the number of  $h(x)$  in real number field.

#### 3.1 First situation

**Proof:**  $f(x)$  is known,  $h(x)$  is unknown,  $g(x) \neq h(x)$  , The meaning of  $f(x)$  ,  $g(x)$  and  $h(x)$  are the same as above section. To reference the content of calculation method of curve integral of arc length in mathematical analysis, it could provide that :

$$\int_{g(x)} h(u)du = \int_{-\infty}^x f(t)dt = F(x)$$

$$\Leftrightarrow \int_{-\infty}^x h(u) \sqrt{1 + [g'(u)]^2} du = F(x)$$

Seek the first derivative on both sides, then:

$$\Leftrightarrow h(x) \sqrt{1 + [g'(x)]^2} = f(x)$$

$$\Leftrightarrow h^2(x) \left\{ 1 + [g'(x)]^2 \right\} = f^2(x)$$

$$\Leftrightarrow h^2(x) = f^2(x) / \left\{ 1 + [g'(x)]^2 \right\}$$

$$\Leftrightarrow h(x) = \pm \sqrt{f^2(x) / \left\{ 1 + [g'(x)]^2 \right\}}$$

$f(x)$  and  $g(x)$  are both real functions, obviously,  $h(x)$  are also real functions, therefore, in real number field, the existence of  $h(x)$  is proved. The number of  $h(x)$  is 2.

#### 3.2 Second situation

**Proof:**  $f(x)$  is known,  $h(x)$  is unknown,  $g(x) = h(x)$  , The meaning of  $f(x)$  ,  $g(x)$  and  $h(x)$  are the same as above section. To reference the content of calculation

method of curve integral of arc length in mathematical analysis, it could provide that :

$$\int_{g(x)} h(u) du = \int_{-\infty}^x f(t) dt = F(x)$$

$$\Leftrightarrow \int_{-\infty}^x g(u) \sqrt{1 + [g'(u)]^2} du = F(x)$$

Seek the first derivative on both sides, then:

$$\Leftrightarrow g(x) \sqrt{1 + [g'(x)]^2} = f(x)$$

$$\Leftrightarrow g^2(x) \left\{ 1 + [g'(x)]^2 \right\} = f^2(x)$$

$$\Leftrightarrow g^2(x) + g^2(x) \cdot [g'(x)]^2 = f^2(x)$$

$$\Leftrightarrow [g'(x)]^2 = [f^2(x) - g^2(x)] / g^2(x)$$

$$\Leftrightarrow [g'(x)]^2 = [1/g^2(x)] \cdot f^2(x) - 1$$

Set  $g(x) = 1/k$ , then  $g'(x) = -1/k^2$ ,  $k \neq 0$ , substitute them to the equation, then :

$$\Leftrightarrow \left(-1/k^2\right)^2 = k^2 f^2(x) - 1$$

$$\Leftrightarrow 1 = k^6 f^2(x) - k^4$$

Set  $f^2(x) = C$ , substitute it to the equation, then :

$$\Leftrightarrow \left(-1/k^2\right)^2 = k^2 C - 1$$

$$\Leftrightarrow 1 = k^6 C - k^4$$

$$\Leftrightarrow k^6 C - k^4 - 1 = 0$$

$$\Leftrightarrow k x^6 - x^4 - 1 = 0$$

Solving higher order equation, as long as seeking out  $x$ , thus to seek out  $h(x)$ , with the aid of MATLAB, input code : `syms x k;x=solve(k*x^6-x^4-1)`, it could get:

$x =$

$$1/6/k/(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} * 6^{1/2} * (k*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} * ((108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{2/3} + 4 + 2*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3}))^{1/2}$$

$$-1/6/k/(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} * 6^{1/2} * (k*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} * ((108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{2/3} + 4 + 2*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3}))^{1/2}$$

$$1/6/k/(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} * 3^{1/2} * (k*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} * (i*3^{1/2}*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{2/3} - (108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{2/3} - 4 + 4*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} - 4*i*3^{1/2}))^{1/2}$$

$$-1/6/k/(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} * 3^{1/2} * (k*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} * (i*3^{1/2}*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{2/3} - (108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{2/3} - 4 + 4*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} - 4*i*3^{1/2}))^{1/2}$$

$$1/6/k/(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} * (-3*k*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} * (i*3^{1/2}*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{2/3} + (108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{2/3} + 4 - 4*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} - 4*i*3^{1/2}))^{1/2}$$

$$-1/6/k/(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} * (-3*k*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} * (i*3^{1/2}*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{2/3} + (108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{2/3} + 4 - 4*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} - 4*i*3^{1/2}))^{1/2}$$

There are 2 real roots and 4 imaginary roots. Seek out  $h(x)$  in real number field, seeking reciprocal of real roots of  $x$  by MATLAB, then:

$x =$

$$k*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} * 6^{1/2} / (k*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} * ((108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{2/3} + 4 + 2*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3}))^{1/2}$$

$$-k*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} * 6^{1/2} / (k*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3} * ((108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{2/3} + 4 + 2*(108*k^2+8+12*3^{1/2}*(27*k^2+4)^{1/2}*k)^{1/3}))^{1/2}$$

Replace  $k$  with  $f(x)$ , then it could get :

$$h(x) = f(x) * (108 * f(x)^2 + 8 + 12 * 3^{1/2} * (27 * f(x)^2 + 4)^{1/2} * f(x))^{1/3} * 6^{1/2} / (f(x) * (108 * f(x)^2 + 8 + 12 * 3^{1/2} * (27 * f(x)^2 + 4)^{1/2} * f(x))^{1/3} * ((108 * f(x)^2 + 8 + 12 * 3^{1/2} * (27 * f(x)^2 + 4)^{1/2} * f(x))^{2/3} + 4 + 2 * (108 * f(x)^2 + 8 + 12 * 3^{1/2} * (27 * f(x)^2 + 4)^{1/2} * f(x))^{1/3}))^{1/2}$$

$$h(x) = -f(x) * (108 * f(x)^2 + 8 + 12 * 3^{1/2} * (27 * f(x)^2 + 4)^{1/2} * f(x))^{1/3} * 6^{1/2} / (f(x) * (108 * f(x)^2 + 8 + 12 * 3^{1/2} * (27 * f(x)^2 + 4)^{1/2} * f(x))^{1/3} * ((108 * f(x)^2 + 8 + 12 * 3^{1/2} * (27 * f(x)^2 + 4)^{1/2} * f(x))^{2/3} + 4 + 2 * (108 * f(x)^2 + 8 + 12 * 3^{1/2} * (27 * f(x)^2 + 4)^{1/2} * f(x))^{1/3}))^{1/2}$$

$f(x)$  is real functions, obviously,  $h(x)$  are also real functions, therefore, in real number field, the existence of  $h(x)$  is proved. The number of  $h(x)$  is 2.

In short, for two kinds of situation above, there are 4  $h(x)$  in real number field.

#### 4 Brief summary

Linear probability density function is a kind of probability density function, operating linear integral for linear probability density function is another expression of seeking cumulative probability. It could according to the known specified path to obtain different linear probability density function.

#### References

1. Li Zhong, Fang Liping, *A course in mathematical analysis*, Higher Education Press, Beijing, **2** (2008): 70-75.
2. Department of Applied Mathematics of Tongji University, *Higher mathematics*, Higher Education Press, Beijing, **2** (2002): 126-130.
3. Mao Shisong, Pu Xiaolong, Cheng Yiming, *A concise on probability theory and mathematical statistics*, Higher Education Press, Beijing, (2012): 42-44.
4. Mao Shisong, Lv Xiaoling, *Mathematical Statistics*(China Renmin University Press, Beijing, (2011), 35-45.
5. Wu Xizhi, *Statistics: from data to conclusions*, China Statistics Press, Beijing, (2011), 57-64