Mathematics and Statistics 5(2): 74-77, 2017
DOI: 10.13189/ms.2017.050204 http://www.hrpub.org

On a Cumulative Distribution Function Related to the Bernoulli Process

Peter Kopanov[1,*], **Miroslav Marinov**[2]

[1]Department of Mathematics and Informatics, Plovdiv University "Paisii Hilendarski", 4000, Plovdiv, Bulgaria

[2]St Catherine's College, Oxford University, OX1 3UJ, Oxford, United Kingdom

Abstract We examine the properties of a cumulative distribution function which is related to the *Bernoulli process*. Results figuring in a paper[1] are shown and new ones are included. Most of them are connected to the behaviour of the probability density function (derivative) of the given distribution.

Keywords Bernoulli Process, Density Function, Criterion

1 Introduction

A **Bernoulli process** is a sequence (finite or infinite) of independent random variables a_1, a_2, a_3, ..., such that for each i:

- The value of a_i is either 0 or 1
- The probability of the event $a_i = 1$ is a real constant $p \in [0, 1]$ (respectively, the probability of $a_i = 0$ is $1 - p$).

Let Ω be the set of all infinite binary sequences. If we consider an infinite Bernoulli process, a specific outcome ω will belong to Ω. It is known that for all $\omega \in \Omega$ the probability ω occurs as a result of such a process is 0.

Now let $X \in [0, 1]$ be a random variable, such that for each $\omega = a_1, a_2, ...$ the value $X(\omega)$ is the fraction $0.a_1a_2....$ For example, the outcome $\omega = 1, 0, 0, 1, 0, 0, ...$ corresponds to the binary $X(\omega) = 0.1001$ (which in decimal system equals 0.5625). Obviously $X(\omega)$ depends on p.

For each $x \in \mathbb{R}$ by $F_p(x)$ we will represent the probability $\mathbb{P}[X(\omega) \leq x]$ (the **cumulative distribution function of** X). We have $F_p(x) = 0$ for all $x < 0$ and $F_p(x) = 1$ for all $x > 1$. The interesting domain is $[0, 1]$.

Note that, as any distribution function[3], $F_p(x)$ is non-decreasing and right-continuous in $[0, 1]$.

In the following section we shall cite properties of this function, some of which describe it as a singular function. As such it is natural to consider $F_p(x)$'s formal derivative - the **probability density function** $f_p(x) = F_p'(x)$ in $x \in [0, 1]$ (there are only partial results[1] on it). We will derive new results on it in the rest of the paper.

2 Basic properties

On *Figure 1* are shown different approximation (except $p = \frac{1}{2}$) graphs of $F_p(x)$, all of which depend on the value of p (as p increases, the $F_p(x)$-value for a given x decreases). We shall omit the trivial cases $p = 0$ and $p = 1$. If $p = \frac{1}{2}$ the graph lies on the line $y = x$. Indeed, here $\mathbb{P}[a_i = 1] = P[a_i = 0]$ thus $F_p(x)$ equals the probability a random number in $[0, 1]$ is less than or equal to a fixed number x (which is exactly x). The arc length when $p = \frac{1}{2}$ equals $\sqrt{2}$.

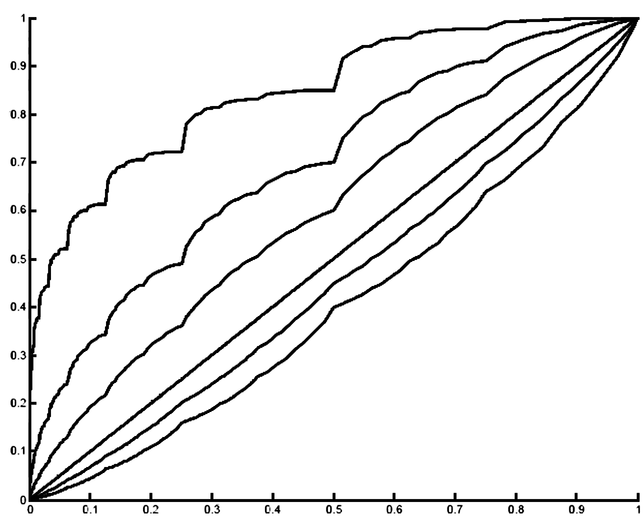

Figure 1. Different graphs of $F_p(x)$

For the non-trivial cases the graphs show that the function is **strictly increasing** and **continuous**. Its Riemann integral in $[0,1]$ equals $1-p$. For the arc length there is an interesting result - it always equals 2 regardless the value of p. The *Triangle inequality* implies that if we divide the graph into infinitely small partitions, each one is either parallel to the x or to the $y-$axis. All these statements have already been proven[1].

3 Useful functional equations

Set $x=\sum_{i=1}^{\infty}\frac{a_i}{2^i}$. We have the following equalities[1]:
Proposition 1.1. $F_p(\frac{x}{2})=(1-p)F_p(x)$
Proposition 1.2. $F_p(\frac{x+1}{2})=1-p+pF_p(x)$
Proposition 2. $F_p(x)=1-F_{1-p}(1-x)$

From the first two propositions we can directly deduce that
Proposition 3. $F_p(0.a_1a_2...)=(1-p)a_1+(1-p+(2p-1)a_1)F_p(0.a_2a_3...)$
This equation is suitable for creating a recursive algorithm for calculating $F_p(x)$ for an arbitrary $x\in[0,1]$ (the process may be an approximation with a negligible error if x is non-dyadic).

4 The derivative (density) of the cumulative function

We shall omit the case $p=\frac{1}{2}$, where clearly $f_p(x)=1$ in $x\in(0,1)$.

4.1 Basic facts and useful lemmas

In **Section 1** we mentioned the *infinite staircase behaviour* of $F_p(x)'$s graph. It directly implies that in each point in $[0,1]$ the derivative $f_p(x)=F_p'(x)$ of our function is either 0, $+\infty$ or does not exist. Moreover, since $F_p(x)$ is a monotonic (strictly increasing) function, due to **Lebesgue's theorem for monotonic functions**[4] its derivative exists **almost everywhere** in $[0,1]$. So how does the density function exactly behave $-$ how can we describe, depending on p, which are the points in which it is defined? We shall derive a criterion which answers the previous question.

Firstly, we have the following:
Lemma 4.1. If $f_p(x)$ exists then $f_p(x)=f_{1-p}(1-x)$.
Proof. Differentiating both sides of the equation in **Proposition 2** with respect to x gives

$$f_p(x)=\frac{dF_p(x)}{dx}=-\frac{dF_{1-p}(1-x)}{dx}=$$

$$=-\frac{dF_{1-p}(1-x)}{d(1-x)}\frac{d(1-x)}{dx}=\frac{dF_{1-p}(1-x)}{d(1-x)}=f_{1-p}(1-x)$$

as desired.
Remark. In particular, the upper equality also holds for the left and right-hand derivatives of $F_p(x)$ and $F_{1-p}(1-x)$.

It would be suitable to have an expression of the form $F_p(x+h)-F_p(x)$ for an appropriate h (in this case: $h=\frac{1}{2^n}$). Then using that $f_p(x)=0$ or $+\infty$ when it exists, a squeeze-type technique (using this expression) will be sufficient to describe its behaviour in a given point.

Set $s_n(x)=\sum_{i=1}^{n}a_i$ (i.e. $s_n(x)$ represents the number of 1's among the first n bits in x's binary representation). Then:
Lemma 4.2. For every $n\in\mathbb{N}$ we have:

$$F_p(0.a_1...a_{n-1}1a_{n+1}...)-F_p(0.a_1...a_{n-1}0a_{n+1}...)=$$
$$=p^{s_n(x)}(1-p)^{n-s_n(x)}\left(1+\frac{2p-1}{1-p}F_p(y_n)\right)$$

where $x=0.a_1...a_{n-1}0a_{n+1}...$ and $y_n=0.a_{n+1}a_{n+2}....$
Proof. Apply **Proposition 3.** $n-1$ times (the $i-$th for $F_p(0.a_ia_{i+1}...)$), then:
$F_p(0.a_1a_2...1...)-F_p(0.a_1a_2...0...)=$
$=(1-p)a_1+(1-p+(2p-1)a_1)((1-p)a_2+(1-p+(2p-1)a_2)((1-p)a_3+(1-p+(2p-1)a_3)(...((1-p)a_{n-1}+(1-p+(2p-1)a_{n-1})F_p(0.1a_{n+1}...))...)$
$-(1-p)a_1-(1-p+(2p-1)a_1)((1-p)a_2+(1-p+(2p-1)a_2)((1-p)a_3+(1-p+(2p-1)a_3)(...((1-p)a_{n-1}+(1-p+(2p-1)a_{n-1})F_p(0.0a_{n+1}...))...)$

Now we can firstly eliminate $(1-p)a_1$ and $-(1-p)a_1$. After that we take out in front of the brackets $(1-p+(2p-1)a_1)$. In the brackets we eliminate $(1-p)a_2$ and $-(1-p)a_2$, then we take out $(1-p+(2p-1)a_2)$ to the front and so on (this occurs $n-1$ times). We reach:

$$F_p(0.a_1a_2...1...)-F_p(0.a_1a_2...0...)=$$
$$=\prod_{i=1}^{n-1}(1-p+(2p-1)a_i).(F_p(0.1a_{n+1}...)-F_p(0.0a_{n+1}...))$$

Now if $a_i=0$, the i-th multiplier in the upper expression equals $1-p$, else it equals p. Therefore:

$$\prod_{i=1}^{n-1}(1-p+(2p-1)a_i)=p^s(1-p)^{n-s-1}$$

On the other hand, **Propositions 1.1. and 1.2.** give:

$$F_p(0.1a_{n+1}...)-F_p(0.0a_{n+1}...)=$$
$$1-p+pF_p(y_n)-(1-p)F_p(y_n)=$$
$$=1-p+(2p-1)F_p(y_n)=(1-p)\left(1+\frac{2p-1}{1-p}F_p(y_n)\right)$$

Multiplying the last two equalities gives the result.

Remark 4.3. For simplicity, let $A(y_n)=\left(1+\frac{2p-1}{1-p}F_p(y_n)\right)$.
Note that $A(y_n)$ is bounded and always positive. The first is because $p=const$ and $f_p(y)\in[0,1]$ - for $p\in(\frac{1}{2},1)$ we have $1\le A(y_n)\le 1+\frac{2p-1}{1-p}$, and if $p\in(0,\frac{1}{2})$ then $1+\frac{2p-1}{1-p}\le A(y_n)\le 1$. The second follows from $1+\frac{2p-1}{1-p}=\frac{p}{1-p}>0$ (referring to the previous double inequalities).

4.2 The special dyadic case

An important characteristic property of the dyadic numbers is that the value $s_n(x)$ is finite for all n - that is due to the fact that from a bit onwards there are only 0's. Set $\lim_{n\to\infty} s_n(x) = s_\infty(x)$. We shall prove that:

Theorem 5. Let $x \in [0,1]$ be a dyadic number. Then $f_p(x)$ does not exist.

Proof. It suffices to show that as $h \to 0$ the limit of $\frac{F_p(x+h)-F_p(x)}{h}$ is different in the cases $h = \pm\frac{1}{2^n}$. Consider the limit in the case $h = \frac{1}{2^n}$ (denote it by $C_p(x)$). We have:

$$C_p(x) = \lim_{n\to\infty}\left[2^n p^{s_n(x)}(1-p)^{n-s_n(x)}A(y_n)\right] =$$

$$= \left(\frac{p}{1-p}\right)^{s_\infty(x)} \lim_{n\to\infty}\left[\big(2(1-p)\big)^n A(y_n)\right]$$

Neglecting the positive $\left(\frac{p}{1-p}\right)^{s_\infty(x)}$ and the bounded multiplier $A(y_n)$, depending on whether $2(1-p)$ is less or greater than 1 (recall that $p \neq \frac{1}{2}$), we deduce that:

$$C_p(x) = \begin{cases} 0 & \text{if } p > \frac{1}{2} \\ +\infty & \text{if } p < \frac{1}{2} \end{cases}$$

However, if $D_p(x)$ is the same limit, but with $h = -\frac{1}{2^n}$, then by the remark in **Lemma 4.1** $D_p(x) = C_{1-p}(1-x)$, so since $1-x$ is also a dyadic number, we have:

$$D_p(x) = \begin{cases} 0 & \text{if } p < \frac{1}{2} \\ +\infty & \text{if } p > \frac{1}{2} \end{cases}$$

We conclude that $C_p(x) \neq D_p(x)$ and the proof is completed.

4.3 The criterion and examples

The criterion will mainly rely on **Lemma 4.2**. Since the expression there contains a combination of $s_n(x)$ and n, it would be convenient to use the notation $k_n(x) = \frac{s_n(x)}{n}$ (i.e. $k_n(x)$ is the density of the 1-bits among the first n). Also, set $k(x) = \lim_{n\to\infty} k_n(x)$ (if this limit exists) and $t = p - 0.5$ - the oriented distance from p to 0.5 on the real line.

Theorem 6. *A Criterion for the behaviour of $f_p(x)$ in almost all real $x \in [0,1]$*

For all real $x \in [0,1]$ for which $k(x)$ exists and $k(x) \neq 0, 1$ we have

$$f_p(x) = \begin{cases} 0, & \text{if } (1+2t)^{k(x)}(1-2t)^{1-k(x)} < 1 \\ +\infty, & \text{if } (1+2t)^{k(x)}(1-2t)^{1-k(x)} > 1 \\ \text{does not exist}, & \text{if } (1+2t)^{k(x)}(1-2t)^{1-k(x)} = 1 \end{cases}$$

Proof. Consider $\frac{1}{2^{n+1}} \leq h \leq \frac{1}{2^n}$. Then since $F_p(x)$ is strictly increasing, we have:

$$\frac{F_p(x+\frac{1}{2^{n+1}}) - F_p(x)}{\frac{1}{2^n}} \leq \frac{F_p(x+h) - F_p(x)}{h} \leq \frac{F_p(x+\frac{1}{2^n}) - F_p(x)}{\frac{1}{2^{n+1}}}$$

hence

$$\lim_{n\to\infty}\left[\frac{F_p(x+\frac{1}{2^{n+1}}) - F_p(x)}{\frac{1}{2^n}}\right] \leq f_p(x)$$

$$\leq \lim_{n\to\infty}\left[\frac{F_p(x+\frac{1}{2^n}) - F_p(x)}{\frac{1}{2^{n+1}}}\right]$$

The left-hand and right-hand sides are respectively equal to (note that $k(x)(1-k(x)) \neq 0$)

$$L_p(x) = \lim_{n\to\infty}\left[2^n p^{s_{n+1}(x)}(1-p)^{n+1-s_{n+1}(x)}A(y_{n+1})\right] =$$

$$= \lim_{n\to\infty}\left[\frac{1}{2}\big((1+2t)^{k(x)}(1-2t)^{1-k(x)}\big)^{n+1}A(y_{n+1})\right]$$

and (analogously)

$$R_p(x) = \lim_{n\to\infty}\left[2\big((1+2t)^{k(x)}(1-2t)^{1-k(x)}\big)^n A(y_n)\right]$$

As t and x vary, we have the following cases for the multiplier $(1+2t)^{k(x)}(1-2t)^{1-k(x)}$:

1. If it equals 1, then consider $f_p(x)$ again with $h = \frac{1}{2^n}$. We analogously have:

   $$C_p(x) = \lim_{n\to\infty}\left[\big((1+2t)^{k(x)}(1-2t)^{1-k(x)}\big)^n A(y_n)\right] =$$

   $$= \lim_{n\to\infty} A(y_n)$$

   Here everything depends on $A(y_n)$, in particular - on $F_p(y_n)$. We shall prove that $F_p(y_n)$ oscillates around $F_p(\frac{1}{2}) = 1-p$ but does not tend to it. If from a bit onwards in y's binary expansion there are only 1s, then y is actually a dyadic, but this case **has already been examined**. Hence we can assume that y is not dyadic.
   Depending on n, the $(n+1)$-th bit (after the decimal point) will be either 1 or 0. We have an infinite alternation (not necessarily periodic) of these two possibilities. The two cases respectively give $y > \frac{1}{2}$ and $y < \frac{1}{2}$. Therefore, to have an existing limit, we need y to tend (from left and right) to $\frac{1}{2}$. In order to happen from the right side, we need from a bit onwards to have: if it is 1, the next one is 0 (else $0.11... > \frac{3}{4}$). Conversely, from the left side we need from a bit onwards: if this bit is 0, then the next one is 1. Combining the two necessities, we have (from a bit onwards) either $y = 0.(01)$ or $y = 0.(10)$, respectively $\frac{1}{3} < \frac{1}{2}$ and $\frac{2}{3} > \frac{1}{2}$, a contradiction.
   Hence the limit $C_p(x)$ does not exist and so does the derivative.

2. If it is strictly less than 1, then the bounded multiplier $A(y_n) > 0$ affects nothing - we have $L_p(x) = R_p(x) = 0$, therefore $f_p(x) = 0$.

3. If it is strictly greater than 1, then the bounded multiplier $A(y_n) > 0$ affects nothing - we have $L_p(x) = R_p(x) = +\infty$, therefore $f_p(x) = +\infty$.

The proof is completed.

Remark. We have considered $k(x) \neq 0, 1$ because $\lim_{n\to\infty}((1+2t)^{k_n(x)})^{n+1} = \lim_{n\to\infty}((1+2t)^{k(x)})^{n+1}$ is not necessarily true if $k(x) = 0$ (similarly for $1-k(x) = 0$).

Our criterion is ready and we can test it. The following example illustrates the **measure one** subset of $[0,1]$, in which the derivative is 0 for all non-trivial p. Indeed, it is proven[2] that the subset of real numbers with $k(x) = \frac{1}{2}$ (the so-called **Regular numbers**) has Lebesgue measure 1 in $[0,1]$. Now:

Theorem 7. Let $k \in (0,1)$ and consider the set of all real numbers x with limit density value $k(x) = k$. Then for all non-trivial p, it holds that $f_p(x) = 0$ at all numbers in this set if and only if $k = \frac{1}{2}$.

Proof. Due to **Theorem 6.** we actually have to prove that: *The inequality* $(1+2t)^k(1-2t)^{1-k} < 1$ *holds for all* $t \in (-0.5, 0) \cup (0, 0.5)$ *if and only if* $k = \frac{1}{2}$.

($\Leftarrow$) If $k = \frac{1}{2}$, then we have $\sqrt{1-4t^2} < 1$. The last holds for all given t.

($\Rightarrow$) The given inequality can be rewritten as $(1-4t^2)^{\frac{1}{2}} < \left(\frac{1-2t}{1+2t}\right)^{k-\frac{1}{2}}$.

Denote the left-hand and right-hand sides by $f(t)$ and $g(t)$, respectively. Note that $f(0) = g(0) = 1$ and that $f(t)$ attains its maximum at $t = 0$. Hence if $f(t)$ and $g(t)$ intersect at $t = 0$, then there will be a region where $g(t) < f(t)$, a contradiction. Therefore, a necessary condition for the inequality to hold is $-$ f and g are tangent at $t = 0$. But then $f'(0) = g'(0)$ and since $f'(0) = 0$, we finally have:

$$g'(t) = \left(k - \frac{1}{2}\right)\left(\frac{1-2t}{1+2t}\right)^{-k-\frac{1}{2}} * \left(\frac{-4}{(1+2t)^2}\right) \Leftrightarrow$$

$$0 = g'(0) = 2 - 4k \Leftrightarrow k = \frac{1}{2}$$

which completes the proof.

The last two theorems show that for all non-trivial p there exist sets (of measure 0 over $\mathbb{R}$) in which $f_p(x)$ does not exist. All these are interesting only from the constructive point of view (but not from the probabilistic because their measure is 0). Let us show an example.

Proposition 8. For all $l \in \mathbb{N}$:

$$f_p\left(\frac{6}{7.2^l}\right) \begin{cases} = 0, & p \in (0, \frac{1}{2}) \cup (\frac{\sqrt{5}+1}{4}, 1) \\ +\infty, & p \in (\frac{1}{2}, \frac{\sqrt{5}+1}{4}) \\ \text{does not exist}, & p = \frac{\sqrt{5}+1}{4} \end{cases}$$

Proof. Note that all such numbers have a binary expansion $0.00..0(110)$, hence they have limit density $k = \frac{2}{3}$ and the derivative depends on the expression $(1+2t)^{\frac{2}{3}}(1-2t)^{\frac{1}{3}}$. With Wolfram Mathematica we can check that the last is greater than or equal to 1 when $t \in (0, \frac{\sqrt{5}-1}{4}]$ and less than 1 in the other cases. Finally, $p = 0.5 + t$ gives the result.

4.4 The derivative when $k_n(x)$ diverges

In this case both of the non-constant multipliers in the expressions for $L_p(x)$ and $R_p(x)$ diverge, too. This, in general, does not exclude the possibility that their product converges. The author still does not have any results about such cases.

REFERENCES

[1] Cohen M., Nater J., Wear P., *Tossing a coin, 18.821 Project Laboratory in Mathematics (2013)*

[2] Ross S., *A First Course in Probability, California, Pearson, 8-th edition (2010), Chapter 4, p.168*

[3] Lamperti J., *Probability, New York, A survey of the mathematica theory, (1966), p.39, Theorem 2.*

[4] Riss F., *Lecons D'Analyse Fonctionnelle, Budapest, Academiai Kiado, Sixteme edition (1972), p.15*