COMMENTS ON SOME MODIFICATION OF SUJA CUMULATIVE FUNCTIONS WITH APPLICATIONS TO THE THEORY OF COMPUTER VIRUSES PROPAGATION. VII

TODORKA TERZIEVA¹, ANTON ILIEV², ASEN RAHNEV³, NIKOLAY KYURKCHIEV⁴
¹,²,³,⁴Faculty of Mathematics and Informatics
University of Plovdiv Paisii Hilendarski
24, Tzar Asen Str., 4000 Plovdiv, BULGARIA

ABSTRACT: In this note we study the Hausdorff approximation of the Heaviside function \( h_{t_0}(t) \) (\( t_0 \) is the "median") by the cdf of some modification of Suja distribution [1]

\[
M_1(t) = 1 - \frac{24 + 24t\alpha + 12t^2\alpha^2 + 4t^3\alpha^3 + \alpha^4 + t^4\alpha^4}{24 + \alpha^4} e^{-\alpha t}
\]
\[
M_2(t) = 1 - \frac{120 + \alpha(120t + 60t^2\alpha + 20t^3\alpha^2 + (1 + 5t^4)\alpha^3 + t\alpha^4(1 + t^4))}{120 + \alpha^4} e^{-\alpha t}
\]
\[
M_3(t) = 1 - \frac{\alpha^4 + \alpha^5t^5\beta + 5\alpha^4t^4\beta + 20\alpha^3t^3\beta + 60\alpha^2t^2\beta + \alpha(\alpha^4 + 120)t\beta + 120}{120 + \alpha^4} e^{-\alpha t^\beta}.
\]

We will show that the proposed models can be successfully used with success (of course, after extensive research) in the field of analysis of Computer Viruses Propagation and Debugging Theory.

We also analyze some experimental data. Numerical examples, illustrating our results are presented using programming environment CAS Mathematica.

AMS Subject Classification: 41A46
Key Words: Suja distribution, length–biased Suja distribution (LBSD), power length–biased Suja distribution (PLBSD), Heaviside step–function \( h_{t_0}(t) \), Hausdorff distance

Received: March 19, 2020
Revised: August 10, 2020
Published: August 24, 2020
doi: 10.12732/ijdea.v19i1.6

Academic Publications, Ltd.
1. INTRODUCTION AND PRELIMINARIES

In this note we study the Hausdorff approximation of the Heaviside function \( h_{t_0}(t) \) by the cdf of some modification of Suja distribution [1] with applications to the theory of Computer Viruses Propagation and Debugging.

**Definition 1.** Shanker [1] developed a new continuous distribution known as Suja distribution (SD) with cumulative distribution function (cdf) for \( t \geq 0 \):

\[
M_1(t) = 1 - \frac{24 + 24t\alpha + 12t^2\alpha^2 + 4t^3\alpha^3 + \alpha^4 + t^4\alpha^4}{24 + \alpha^4}e^{-\alpha t}
\]  

(1)

where \( \alpha > 0 \).

**Definition 2.** Al–Omari and Alsmairan [2] suggested length–biased Suja distribution (LBSD) with cdf for \( t \geq 0 \):

\[
M_2(t) = 1 - \frac{120 + \alpha(120t + 60t^2\alpha + 20t^3\alpha^2 + (1 + 5t^4)\alpha^3 + t\alpha^4(1 + t^4))}{120 + \alpha^4}e^{-\alpha t}
\]  

(2)

where \( \alpha > 0 \).

**Definition 3.** Al–Omari, Alhyasat, Ibrahin and Bakar [3] proposed power length–biased Suja distribution (PLBSD) with cdf for \( t \geq 0 \):

\[
M_3(t) = 1 - \frac{\alpha^4 + \alpha^5t^5\beta + 5\alpha^4t^4\beta + 20\alpha^3t^3\beta + 60\alpha^2t^2\beta + \alpha(\alpha^4 + 120)t^2\beta + 120}{120 + \alpha^4}e^{-\alpha t^\beta}
\]  

(3)

where \( \alpha, \beta > 0 \).

**Definition 4.** The shifted Heaviside step function is defined by

\[
h_{t_0}(t) = \begin{cases} 
0, & \text{if } t < t_0, \\
[0,1], & \text{if } t = t_0, \\
1, & \text{if } t > t_0
\end{cases}
\]  

(4)

**Definition 5.** [11] The Hausdorff distance (the H–distance) \( \rho(f,g) \) between two interval functions \( f, g \) on \( \Omega \subseteq \mathbb{R} \), is the distance between their completed graphs \( F(f) \) and \( F(g) \) considered as closed subsets of \( \Omega \times \mathbb{R} \). More precisely,

\[
\rho(f,g) = \max\{ \sup_{A \in F(f)} \inf_{B \in F(g)} ||A - B||, \sup_{B \in F(g)} \inf_{A \in F(f)} ||A - B|| \},
\]
wherein $||.||$ is any norm in $\mathbb{R}^2$, e. g. the maximum norm $||(t, x)|| = \max\{|t|, |x|\};$ hence the distance between the points $A = (t_A, x_A), B = (t_B, x_B)$ in $\mathbb{R}^2$ is $||A - B|| = \max(|t_A - t_B|, |x_A - x_B|)$.

2. MAIN RESULTS

2.1. A NOTE ON THE CDF (1)

The investigation of the characteristic ”supersaturation” of the cdf (1) to the horizontal asymptote is important.

Let $t_0$ is the value for which $M_1(t_0) = \frac{1}{2}$.

The one–sided Hausdorff distance $d$ between the function $h_{t_0}(t)$ and the cdf (1) satisfies the relation

$$M_1(t_0 + d) = 1 - d. \quad (5)$$

For given $\alpha$ and $t_0$, the nonlinear equation $M_1(t_0 + d) - 1 + d = 0$ has a positive root $-d$.

The cdf (1) for $\alpha = 6$ and $t_0 = 0.118635$ is visualized on Fig. 1 a.

From the nonlinear equation (5) we have: $d = 0.180583$.

The cdf (1) for $\alpha = 14$ and $t_0 = 0.049551$ is visualized on Fig. 1 b.

From the nonlinear equation (5) we have: $d = 0.109057$.

From the above examples, it can be seen that the ”supersaturation” by the (cdf) $M_1(t)$ is faster.

2.2. A NOTE ON THE CDF (2)

Let $t_0$ is the value for which $M_2(t_0) = \frac{1}{2}$.

The one–sided Hausdorff distance $d$ between the function $h_{t_0}(t)$ and the cdf (2) satisfies the relation

$$M_2(t_0 + d) = 1 - d. \quad (6)$$

For given $\alpha$ and $t_0$, the nonlinear equation $M_2(t_0 + d) - 1 + d = 0$ has a positive root $-d$.

The cdf (2) for $\alpha = 26$ and $t_0 = 0.0645677$ is visualized on Fig. 2 a.

From the nonlinear equation (6) we have: $d = 0.0901374$. 
Figure 1: a) The cdf (1) for $\alpha = 6$ and $t_0 = 0.118635$; H–distance $d = 0.180583$; b) The cdf (1) for $\alpha = 14$ and $t_0 = 0.0495551$; H–distance $d = 0.109057$.

Figure 2: a) The cdf (2) for $\alpha = 26$ and $t_0 = 0.0645677$; H–distance $d = 0.0901374$; b) The cdf (2) for $\alpha = 50$ and $t_0 = 0.0335675$; H–distance $d = 0.0577949$.

The cdf (2) for $\alpha = 50$ and $t_0 = 0.0335675$ is visualized on Fig. 2 b.

From the nonlinear equation (6) we have: $d = 0.0577949$. 
2.3. A NOTE ON THE CDF (3)

Let \( t_0 \) is the value for which \( M_3(t_0) = \frac{1}{7} \).

The one–sided Hausdorff distance \( d \) between the function \( h_{t_0}(t) \) and the cdf (3) satisfies the relation

\[
M_3(t_0 + d) = 1 - d. \tag{7}
\]

For given \( \alpha \), \( \beta \) and \( t_0 \), the nonlinear equation \( M_3(t_0 + d) - 1 + d = 0 \) has a positive root \(-d\).

The cdf (3) for \( \alpha = 12 \), \( \beta = 4 \) and \( t_0 = 0.612368 \) is visualized on Fig. 3 a.

From the nonlinear equation (7) we have: \( d = 0.128439 \).

The cdf (3) for \( \alpha = 20 \), \( \beta = 10 \) and \( t_0 = 0.780577 \) is visualized on Fig. 3 b.

From the nonlinear equation (7) we have: \( d = 0.0758028 \).

Obviously, this "advantage" can actually be used to approximate some specific data from the field of analysis of Computer Viruses Propagation and Debugging Theory.

In the next Section, we will support what is said by analyzing real datasets.
2.4. APPLICATIONS

Example 1. Analysis of Witty worm infection behavior.
Here we will give an application of model $M_3^*(t)$ when provide analysis of this real "data" [4]

\[\text{data}_{\text{Witty World}} = \]
\[
\{(0.1, 150), (5, 869), (10, 2141), (15, 3637), (20, 5312),
\{26, 6602\}, \{31, 7562\}, \{36, 8340\}, \{41, 8941\}, \{46, 9389\}, \{51, 9734\},
\{56, 10060\}, \{61, 10349\}, \{66, 10586\}, \{71, 10800\}, \{76, 11169\},
\{86, 11362\}, \{91, 11532\}, \{96, 11684\}, \{101, 11823\}, \{106, 11972\},
\{111, 12118\}, \{116, 12256\}, \{121, 12372\}\]

The fitted model $M_3^*(t) = \omega M_3(t)$ for $\omega = 12372; \alpha = 1.30886; \beta = 0.453772$ is visualized on Fig. 4.

Example 2. Storm worm one of the most biggest cyber threats of 2008.
We analyze the following data [5]
COMMENTS ON SOME MODIFICATION OF SUJA CUMULATIVE...

Figure 5: The fitted model $M_3^*(t)$. 

\[ \text{data}_{\text{Storm	extunderscore IDs}} := \{\{1.8,0.843\}, \{4,0.926\}, \{5,0.954\}, \{6,0.967\}, \{7,0.976\}, \{8,0.981\}, \{9,0.985\}, \{10,0.991\}, \{22,0.995\}, \{38,0.997\}, \{51,0.998\}, \{64,0.9985\}, \{74,0.999\}, \{83,1\}, \{100,1\}\} \]

The cdf $M_3^*(t) = \omega M_3(t)$ for $\alpha = 4.31451, \beta = 0.435348, \omega = 0.99$ is visualized on Fig. 5.

**Example 3.** Approximating cdf of the number of Bitcoin received per address [6]

We consider the following data (see, [6]):

\[ \text{data}_{\text{CDF of Bitcoin received \textunderscore inransoms} \textunderscore per \text underscore address \textunderscore in \textunderscore CCL} := \{\{1,0.0857\}, \{2,0.1238\}, \{3,0.6571\}, \{4,0.6854\}, \{5,0.8381\}, \{6,0.8476\}, \{7,0.8810\}, \{8,0.9095\}, \{9,0.9143\}, \{10,0.9333\}, \{12,0.9429\}, \{14,0.9571\}, \{18,0.9667\}, \{20,0.9762\}, \{23,0.9810\}, \{27,0.9857\}, \{40,0.9905\}, \{46,0.9952\}, \{59,0.9981\}\}. \]

The cdf $M_3^*(t) = \omega M_3(t)$ for $\alpha = 2.04944, \beta = 0.862203, \omega = 0.98$ is visualized on Fig. 6.

**Example 4.** On July 26, 2004 a variant of MyDoom attacks Google, AltaVista and Lycos, completely stopping the function of the popular Google search engine for the larger portion of the workday, and creating noticeable slow-downs in the AltaVista and Lycos engines for hours [7].

We analyze the following data
Figure 6: The fitted model $M_3^*(t)$.

Figure 7: The fitted model $M_3^*(t)$. 

\begin{verbatim}
data_MyDoom :=
{[1.01, 800], [2, 3000], [3, 9610], [4, 23274], [5, 38846], [6, 50000],
[7, 53846], [8, 57300]}
\end{verbatim}

The fitted model $M_3^*(t) = \omega M_3(t)$ for $\alpha = 0.917566$, $\beta = 1.24861$, $\omega = 57300$ is visualized on Fig. 7.

**Example 5.** Software Failure Data – Release #1. These data derive from one major release of software products at Tandem Computers [8], [9]
Figure 8: The fitted model $M^*_3(t)$.

$$data_{PhamNordmanZhang} := \{(1,16), (2,24), (3,27), (4,33),$$
$$ (5,41), (6,49), (7,54), (8,58), (9,69), (10,75), (11,81), (12,86),$$
$$ (13,90), (14,93), (15,96), (16,98), (17,99), (18,100), (19,100),$$
$$ (20,100)\};$$

For this data the fitted model for estimated parameters: $\omega = 110; \alpha = 2.00345; \beta = 0.518826$ is plotted on Fig. 8.

**Example 6.** We analyze the following data for Welchia worm [10]

$$data_{Welchia} := \{(0.1,0.1), (1,7250), (2,15714), (3,24107),$$
$$ (4,29464), (5,29643), (6,29821)\};$$

For this data the fitted model for estimated parameters: $\omega = 29821; \alpha = 2.55859; \beta = 0.980523$ is plotted on Fig. 9.

### 3. CONCLUDING REMARKS

Finally, we note that the studied model produces extremely good results, generally when approximating specific "cumulative data" from Computer Viruses Propagation. For other approximation and modelling results, see [12]–[34].

We hope that the results will be useful for specialists in this scientific area.
Figure 9: The fitted model $M^*_3(t)$.

ACKNOWLEDGMENTS

This paper is supported by the National Scientific Program ”Information and Communication Technologies for a Single Digital Market in Science, Education and Security (ICTinSES)”, financed by the Ministry of Education and Science.

REFERENCES


COMMENTS ON SOME MODIFICATION OF SUJA CUMULATIVE...


