ON THE "SUPERSATURATION" OF THE GENERALIZED LOG–BURR–III CUMULATIVE FUNCTION

Anna Malinova¹, Olga Rahneva², Todorka Terzieva³, Evgenia Angelova⁴

¹,³,⁴Faculty of Mathematics and Informatics
University of Plovdiv Paisii Hilendarski
24, Tzar Asen Str., 4000 Plovdiv, BULGARIA

²Faculty of Economy and Social Sciences
University of Plovdiv Paisii Hilendarski
24, Tzar Asen Str., 4000 Plovdiv, BULGARIA

ABSTRACT: In this paper we study the characteristic - "supersaturation" of the cumulative distribution function of the generalized Log–Burr–III distribution to the horizontal asymptote in the Hausdorff sense.

We also analyze some experimental data.

The experiments show that in some cases the use of the model proposed in [1] and analyzed in this article with "respect to the Hausdorff distance" is satisfactory.

Obviously, such studies are a must for the experimenter in the search for dialectical unity "data–model".

Numerical examples, illustrating our results are presented using programming environment CAS Mathematica.

AMS Subject Classification: 41A46
Key Words: cumulative distribution function of the generalized Log–Burr–III distribution, Heaviside function, Hausdorff approximation

Received: May 9, 2019
Revised: October 3, 2019
Published: October 18, 2019
doi: 10.12732/ijdea.v18i1.9

1. INTRODUCTION AND PRELIMINARIES

It is well known that many datasets from finance, reliability, biochemical sciences and
other fields do not follow known continuous distributions.

In this regard, in [1], the authors examined a generalized Log–Burr–III (GLBIII) distribution developed on the basis of a Log Pearson differential equation.

**Definition 1.** The cdf of the generalized Log–Burr–III (GLBIII) distribution is defined by [1]

\[ M(t) = \left( 1 + \left( \frac{\ln t}{b} \right)^{-2a} \right)^{-p}, \tag{1} \]

where \( t \geq 1, a > 0, b > 0, p > 0. \)

The Burr III distribution is well known and widely used in many problems related to forestry, reliability quality control, risk analysis and many other areas of research.

Some modifications, properties and applications can be found in [2]–[8].

**Definition 2.** The *shifted Heaviside step function* is defined by

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

**Definition 3.** The Hausdorff distance [9] (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 \( \| \cdot \| \) 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 \( ||A - B|| = \max(|t_A - t_B|, |x_A - x_B|) \).

We study the Hausdorff approximation of the *shifted Heaviside step function* by the family of type (1).

### 2. MAIN RESULTS AND NUMERICAL EXAMPLES

We consider the following class of this family:

\[ M(t) = \left( 1 + \left( \frac{\ln t}{b} \right)^{-2a} \right)^{-p}, \]

\[ M(t_0) = \frac{1}{2}; \quad t_0 = e^{b(2p-1)^{-\frac{1}{2a}}}. \tag{2} \]
The one-sided Hausdorff distance $d$ between the function $h_{t_0}(t)$ and the sigmoid (2) satisfies the relation

$$M(t_0 + d) = 1 - d.$$  \hspace{1cm} (3)

For given $a > 0$, $b > 0$, $p > 0$ and $t_0$, the nonlinear equation $M(t_0 + d) - 1 + d = 0$ has unique positive root $-d$.

The model (2) for $a = 2.3$, $b = 0.7$, $p = 0.5$ and $t_0 = 1.73548$ is visualized on Fig. 1.

From the nonlinear equation (2) we have: $d = 0.287576$.

The model (2) for $a = 1.3$, $b = 0.1$, $p = 0.3$ and $t_0 = 1.04374$ is visualized on Fig. 2.

From the nonlinear equation (3) we have: $d = 0.103401$.

Some computational examples are presented in Table 1.

<table>
<thead>
<tr>
<th>$a$</th>
<th>$b$</th>
<th>$p$</th>
<th>$t_0$</th>
<th>$H$–distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3</td>
<td>0.7</td>
<td>0.5</td>
<td>1.73548</td>
<td>0.287576</td>
</tr>
<tr>
<td>1.3</td>
<td>0.1</td>
<td>0.3</td>
<td>1.04374</td>
<td>0.103401</td>
</tr>
<tr>
<td>3.3</td>
<td>0.05</td>
<td>0.25</td>
<td>1.03373</td>
<td>0.0349754</td>
</tr>
<tr>
<td>4.3</td>
<td>0.01</td>
<td>0.2</td>
<td>1.00673</td>
<td>0.00789708</td>
</tr>
<tr>
<td>5.3</td>
<td>0.001</td>
<td>0.1</td>
<td>1.00052</td>
<td>0.0010214</td>
</tr>
</tbody>
</table>

Table 1: The Hausdorff distance $d$ computed by nonlinear equation (3)

From the above examples, it can be seen that the ”supersaturation” is faster.
3. APPLICATIONS

1. We will now analyze a sample of experimental data obtained by the biologist T. Carlson in 1913 about the development of Saccharomyces culture in nutrient medium (see, for example [13], [12]):

   We consider the following data:

   \[
   \text{data\_Carlson2} := \{\{2,29\},\{3,47.2\},\{4,71.1\},\{5,119.1\},\{6,174.6\},\{7,257.3\},\{8,350.7\},\{9,441\},\{10,513.3\},\{11,559.7\},\{12,594.8\},\{13,629.4\},\{14,640.8\},\{15,651.1\},\{16,655.9\},\{17,659.6\}\};
   \]

   After that using the model \( M^*(t) = \omega M(t) \) for \( a = 10.5195, \ b = 2.4171, \ p = 0.198804 \) and \( \omega = 659.6 \) we obtain the fitted model (see, Fig. 3).

2. We also analyze the following experimental data: "The growth of population of Oryzaephilus in renewed wheat" obtained by the Crombie in 1945 [11]:
After that using the model $M^*(t) = \omega M(t)$ for $a = 7.26127$, $b = 4.08014$, $p = 1.75789$ and $\omega = 480$ we obtain the fitted model (see, Fig. 4).

3. Analysis of MyDoom worm propagation [14], [15]
We consider the following data:

\[
\text{data\_MyDoom1} := \{\{1.01, 800\}, \{2, 3000\}, \{3.9610, 4, 23274\}, \{5, 38846\}, \{6, 50000\}, \{7, 53846\}, \{8, 57300\}\}.
\]

After that using the model \(M^*(t) = \omega M(t)\) for \(a = 7.92607, b = 1.77069, p = 0.229937\) and \(\omega = 57300\) we obtain the fitted model (see, Fig. 5).

4. Analysis of data ”growth of the cumulative number of TREZ publications” \cite{38, 39}

\[
\text{data\_Journal} := \{\{1.1, 5\}, \{2, 37\}, \{3, 107\}, \{4, 201\}, \{5, 298\}, \{6, 439\}, \{7, 617\}, \{8, 773\}, \{9, 936\}, \{10, 1121\}, \{11, 1316\}, \{12, 1451\}, \{13, 1563\}, \{14, 1629\}, \{15, 1722\}, \{16, 1788\}\}.
\]

After that using the model \(M^*(t) = \omega M(t)\) for \(a = 14.5595, b = 2.61943, p = 0.123592\) and \(\omega = 1788\) we obtain the fitted model (see, Fig. 6).

We analyze the \text{data\_Journal} with the Verhulst logistic model \(V(t)\)

\[
V(t) = \frac{\omega}{1 + me^{-kt}}
\]

The logistic model for \(\omega = 1788, m = 49.2367\) and \(k = 0.449469\) is visualized on Fig. 6.
Figure 6: The fitted models $M^*(t)$—blue and $V(t)$—dashed.

5. Analysis of data "Mobile" [40]

\[
data_{\text{Mobile}} := \{(3,20.29), (4,25.08), (5,30.81), (6,38.75), (7,45), (8,49.16), (9,55.15), (10,62.852), (11,68.63), (12,76.64), (13,82.47), (14,85.68), (15,89.14), (16,91.86), (17,95.28), (18,98.17)\};
\]

After that using the model $M^*(t) = \omega M(t)$ for $a = 35.0878$, $b = 2.84793$, $p = 0.02848$ and $\omega = 98.17$ we obtain the fitted model (see, Fig. 7).

The logistic model $V(t)$ for $\omega = 98.17$, $m = 10.0279$ and $k = 0.297664$ is visualized on Fig. 7.

### 3.1. CONCLUSIONS

The proposed growth model can be successfully used with success (of course, after extensive research) in the field of analysis of Computer Viruses Propagation, Biochemical sciences and Debugging and Test Theory.

For some approximation, computational and modelling aspects, see [17]–[37].

Finally, we consider the following data "cdf of the number of Bitcoin received per
address” (see, [10]):

\[
data_{CDF\_of\_Bitcoin\_received\_(inransoms)\_per\_address\_in\_CCL} := \{(1.1, 0.0857), (2, 0.1238), (3, 0.6571), (4, 0.6854), (5, 0.8381),
\{(6, 0.8476), (7, 0.8810), (8, 0.9095), (9, 0.9143), (10, 0.9333), \}
\{(12, 0.9429), (14, 0.9571), (18, 0.9667), (20, 0.9762), (23, 0.9810),
\{(27, 0.9857), (40, 0.9905), (46, 0.9952), (59, 0.9981)\}.
\]

The model (2) for \(a = 1.37337, b = 0.0853445, p = 600.612, t_0 = 2.72224\) and \(d \approx 0.38\) is visualized on Fig. 8.

In this particular case, this is the saturation \(d\) that can be expected when approximating the specific data for example \(t = 1, 7(1)\) (see, Fig. 8) with our fixed model.

Obviously, such studies are a must for the experimenter in the search for dialectical unity "data–model”.

The experiments show that in some cases the use of the model proposed in [1] and analyzed in this article with ”respect to the Hausdorff distance” is satisfactory.

Specialists working in this scientific field have a say.

It will be noted that interval estimates of the Hausdorff distance can be obtained using one of the techniques described in [16].

From the experiments conducted (Examples 1-6) it can be concluded that for some specific datasets, the generalized Log–Burr–III cumulative function produces satisfactory results.
Figure 8: The fitted model (2) for approximation of the data: "cdf of the number of Bitcoin received per address" [10].

Very qualitative comparisons between the logistics model and the Gompertz model when approximating the data from Examples 4–5 can be found in the original article by Satoh [39].

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


