Mathematical Modelling and Applications
2016; 1(1): 8-12
http://www.sciencepublishinggroup.com/j/mma
doi: 10.11648/j.mma.20160101.12

On the Hausdorff Distance Between the Heaviside Function and Some Transmuted Activation Functions

Nikolay Kyurkchiev[1], Anton Iliev[1, 2]

[1]Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria
[2]Faculty of Mathematics and Informatics, Paisii Hilendarski University of Plovdiv, Plovdiv, Bulgaria

Email address:
nkyurk@math.bas.bg (N. Kyurkchiev), aii@uni-plovdiv.bg (A. Iliev)

To cite this article:
Nikolay Kyurkchiev, Anton Iliev. On the Hausdorff Distance Between the Heaviside Function and Some Transmuted Activation Functions. *Mathematical Modelling and Applications.* Vol. 1, No. 1, 2016, pp. 8-12. doi: 10.11648/j.mma.20160101.12

Received: August 18, 2016; **Accepted**: October 12, 2016; **Published**: October 14, 2016

Abstract: In this paper we study the one-sided Hausdorff distance between the Heaviside function and some transmuted activation functions. Precise upper and lower bounds for the Hausdorff distance have been obtained. Numerical examples are presented throughout the paper using the computer algebra system MATHEMATICA. The results can be successfully used in the field of applied insurance mathematics.

Keywords: Transmuted Activation Functions, Heaviside Function, Hausdorff Distance, Upper and Lower Bounds, Squashing Function

1. Introduction and Preliminaries

In this paper we discuss some computational and approximation issues related to several classes of transmuted activation functions.

Sigmoidal functions find numerous applications in various fields related to Life sciences, chemistry, physics, artificial intelligence, signal processes, pattern recognition, machine learning, demography, economics, probability, financial mathematics, statistics, fuzzy set theory, etc.

The following are common examples of activation functions [2]:

$$\sigma(t) = \frac{1}{1+e^{-t}}$$

$$(\textit{the squashing function})$$

$$\sigma(t) = \begin{cases} 0, & \text{if} \quad t \le -1, \\ \dfrac{t+1}{2}, & \text{if} \quad -1 \le t \le 1, \\ 1, & \text{if} \quad t \ge 1. \end{cases}$$

$$(\textit{the piecewise linear function})$$

$$\sigma(t) = \frac{1}{2} + \frac{1}{\pi}\arctan(t)$$

$$(\textit{the arctan sigmoidal function})$$

Sigmoidal functions (are also known as "activation functions") find multiple applications to neural networks [9], [10], [11], [12], [13], [2].

Definition 1. Define the Heaviside step function as:

$$h(t) = \begin{cases} 0, & \text{if} \quad t < 0, \\ [0,1], & \text{if} \quad t = 0, \\ 1, & \text{if} \quad t > 0. \end{cases} \tag{1}$$

About approximation of the Heaviside step function by some cumulative distribution functions, see [14].

Definition 2. The arctan activation function (sigmoidal Cauchy cumulative distribution function) $f(t)$ is defined for $b > 0$ by [1]:

$$f(t) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{t}{b}\right). \tag{2}$$

Definition 3. A random variable T is said to have a transmuted distribution if its cumulative distribution function (*cdf*) is given by [3], [4]:

$$G_1(t) = (1+\lambda)F_1(t) - \lambda F_1^{\,2}(t), \quad |\lambda| \le 1, \tag{3}$$

where $F_1(t)$ is the cdf of the base distribution.

Definition 4. The Hausdorff 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 $\Omega \times \mathbb{R}$ [5], [6], [7].

More precisely, we have

$$\rho(f,g) = \max\{ \sup_{A\in F(f)} \inf_{B\in F(g)} \| A-B \|, \sup_{B\in F(g)} \inf_{A\in F(f)} \| A-B \|\}, \tag{4}$$

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 $||A-B|| = \max(|t_A-t_B|,|x_A-x_B|)$.

In this work we prove estimates for the one–sided Hausdorff approximation of the Heaviside function by transmuted Cauchy function.

The results are relevant for applied insurance mathematics [14].

The transmuted Cauchy distribution function has been used also in the analysis of extreme values.

Let us point out that the Hausdorff distance is a natural measuring criteria for the approximation of bounded discontinuous functions [7], [8].

Definition 5. Consider the following transmuted Cauchy function

$$F(t) = \frac{1}{1+\frac{\lambda}{2}}\left((1+\lambda)\left(\frac{1}{2}+\frac{1}{\pi}\arctan\left(\frac{t}{b}\right)\right) - \lambda\left(\frac{1}{2}+\frac{1}{\pi}\arctan\left(\frac{t}{b}\right)\right)^2 \right), \tag{5}$$

$$F(0) = \frac{1}{2},$$

where $|\lambda| \leq 1$.

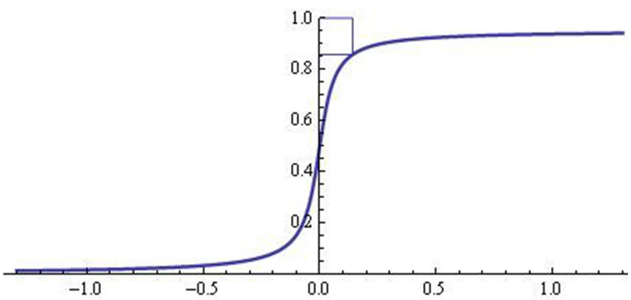

Figure 1. *Approximation of the Heaviside function by transmuted Cauchy function for the following data:* $\lambda = 0.1$, $b = 0.05$; *Hausdorff distance* $d = 0.141442$.

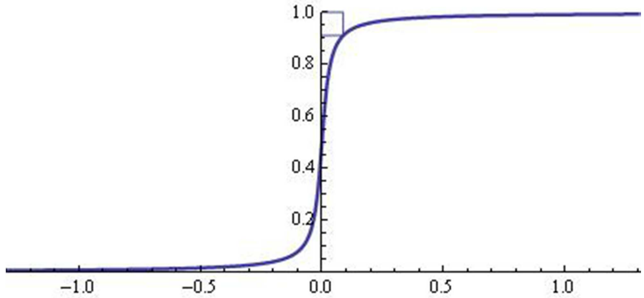

Figure 2. *Approximation of the Heaviside function by transmuted Cauchy function for the following data:* $\lambda = 0.001$, $b = 0.025$; *Hausdorff distance* $d = 0.0882481$.

Approximation of the Heaviside function by transmuted Cauchy function for specific values of λ, b and d can be seen on Fig. 1 and Fig. 2.

2. Main Results

We study the Hausdorff approximation d of the Heaviside function $h(t)$ by the transmuted Cauchy function (5) and look for an expression for the error of the best one–sided approximation.

The following Theorem gives upper and lower bounds for d:

Theorem 2.1 For the Hausdorff distance d between the function $h(t)$ and the transmuted Cauchy function (5) the following inequalities hold for $|\lambda| \leq 1$ and $b\pi(2+\lambda) < 4$:

$$d_l = \frac{1}{3\left(1+\frac{2}{b\pi(2+\lambda)}\right)} < d < \frac{\ln\left(3\left(1+\frac{2}{b\pi(2+\lambda)}\right)\right)}{3\left(1+\frac{2}{b\pi(2+\lambda)}\right)} = d_r. \tag{6}$$

Proof. We need to express d in terms of b and λ.

The Hausdorff distance d satisfies the relation

$$F(d) = \frac{1}{1+\frac{\lambda}{2}}\left((1+\lambda)\left(\frac{1}{2}+\frac{1}{\pi}\arctan\left(\frac{d}{b}\right)\right) - \lambda\left(\frac{1}{2}+\frac{1}{\pi}\arctan\left(\frac{d}{b}\right)\right)^2 \right) = 1-d. \tag{7}$$

Consider the function

$$G(d) = -\frac{1}{2} + \left(1+\frac{2}{b\pi(2+\lambda)}\right)d.$$

In addition $G' > 0$ and G is strictly monotonically increasing.

By means of Taylor expansion we obtain

$$G(d) - F(d) = O(d^2).$$

Hence $G(d)$ approximates $F(d)$ with $d \to 0$ as $O(d^2)$ (see Fig. 3).

Further, for $|\lambda| \le 1$ and $b\pi(2+\lambda) < 4$ we have

$$\begin{aligned} G(d_l) &= -\frac{1}{2} + \frac{1}{3} < 0, \\ G(d_r) &= -\frac{1}{2} + \frac{1}{3} \ln\left(3\left(1 + \frac{2}{b\pi(2+\lambda)}\right)\right) > 0. \end{aligned}$$

This completes the proof of the theorem.

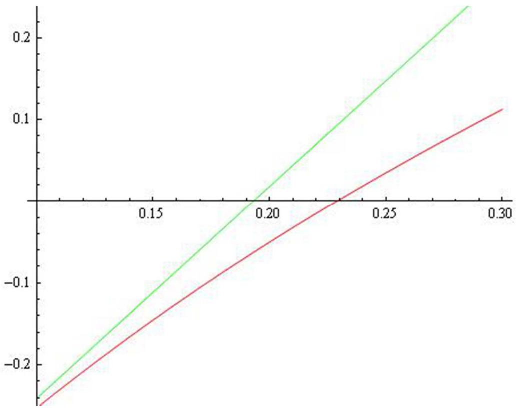

Figure 3. *The functions $F(d)$ and $G(d)$ for $b = 0.2$, $\lambda = 0.001$.*

On Fig. 3 can be seen a good approximation of the function $F(d)$.

Some computational examples using relations (6) are presented in Table 1.

Table 1 shows that at fixed b and decreasing values of parameter λ, Hausdorff distance d is reduced. The reader can visualize that declining trend.

Table 1. *Bounds for d computed by (6) for various λ, b.*

λ	b	d_l	d_r	d from (7)
0.18	0.2	0.135494	0.270829	0.251167
0.001	0.2	0.128663	0.263831	0.228837
0.1	0.1	0.0826817	0.20615	0.185118
0.18	0.1	0.0850281	0.209575	0.198145

The last column of Table 1 contains the values of d computed by solving the nonlinear equation (7).

The resulting "fork" to the root $d_l < d < d_r$ and particularly the assessment above are satisfactory.

3. Conclusion

New estimates for the Hausdorff distance between an interval Heviside step function and its best approximating transmuted Cauchy function are obtained.

The assessment of the value of the best Hausdorff approximation is in close contact with many interesting problems in the field of mathematics in insurance - an assessment of the value of the achieved liability insurance - unfortunately at a slow pace, which is in unison with the law of diminishing marginal returns.

In the present paper we do not consider transmuted cumulative distribution functions, such as the

Raised–Cosine transmuted function:

$$H(t) = (1+\lambda)\left(0.5\left(1 + \frac{t}{b} + \frac{1}{\pi}\sin\left(\frac{t\pi}{b}\right)\right)\right) - \lambda\left(0.5\left(1 + \frac{t}{b} + \frac{1}{\pi}\sin\left(\frac{t\pi}{b}\right)\right)\right)^2$$

Half–Cauchy transmuted function:

$$H_1(t) = (1+\lambda)\left(\frac{2}{\pi}\arctan\left(\frac{t}{b}\right)\right) - \lambda\left(\frac{2}{\pi}\arctan\left(\frac{t}{b}\right)\right)^2$$

Kumaraswamy–Half–Cauchy transmuted function:

$$H_2(t) = (1+\lambda)\left(1 - \left(1 - \left(\frac{2}{\pi}\arctan\left(\frac{t}{b}\right)\right)^{\alpha}\right)^{\beta}\right) - \lambda\left(1 - \left(1 - \left(\frac{2}{\pi}\arctan\left(\frac{t}{b}\right)\right)^{\alpha}\right)^{\beta}\right)^2$$

Hyperbolic–Secant transmuted function:

$$H_3(t) = (1+\lambda)\left(\frac{2}{\pi}\arctan\left(e^{\frac{t}{b}}\right)\right) - \lambda\left(\frac{2}{\pi}\arctan\left(e^{\frac{t}{b}}\right)\right)^2$$

where $|\lambda| \le 1$.

Based on the methodology proposed in the present note, the reader may formulate the corresponding approximation problems on his/her own.

The Hausdorff approximation of the interval step function by the logistic and other sigmoid functions is discussed from various approximation, computational and modelling aspects in [15]–[25].

4. Remarks

Definition 6. The cut function $c_{\gamma,\delta}$ on $\Delta = [\gamma - \delta, \gamma + \delta]$ is defined for $t \in \mathbb{R}$ by

$$c_{\gamma,\delta}(t) = \begin{cases} 0, & \text{if} \quad t < \Delta, \\ \dfrac{t - \gamma + \delta}{2\delta}, & \text{if} \quad t \in \Delta, \\ 1, & \text{if} \quad \Delta < t. \end{cases}$$

Special case 1. For $\gamma = 0$ we obtain a cut function on the interval $\Delta = [-\delta, \delta]$:

$$c_{0,\delta}(t)=\begin{cases}0, & \text{if } t<-\delta,\\ \dfrac{t+\delta}{2\delta}, & \text{if } -\delta\le t\le\delta,\\ 1, & \text{if } \delta<t.\end{cases}$$

Special case 2. For $\gamma=\delta$ we obtain the cut function on $\Delta=[0,2\ \delta]$:

$$c_{\delta,\delta}(t)=\begin{cases}0, & \text{if } t<0,\\ \dfrac{t}{2\delta}, & \text{if } 0\le t\le 2\delta,\\ 1, & \text{if } 2\delta<t.\end{cases}$$

Definition 7. The smooth sigmoid raised-cosine cumulative distribution function (SRCCDF.cdf) $f_{\gamma,b}(t)$ is defined for $\gamma-b<t<\gamma+b$, $\gamma\in R$, $b>0$ by:

$$f_{\gamma,b}(t)=\frac{1}{2}\left(1+\frac{t-\gamma}{b}+\frac{1}{\pi}\sin\left(\frac{(t-\gamma)\pi}{b}\right)\right). \qquad (8)$$

Approximation of the cut function by (SRCCDF.cdf) $f_{\gamma,b}(t)$.

Note that the function (8) has an inflection at its ”centre” $(\gamma,\frac{1}{2})$ and its slope at γ is equal to $\frac{1}{b}$.

On Fig. 4 can be seen the cut and (SRCCDF.cdf) for specific concrete values of γ,δ and b.

The following Theorem is valid.

Theorem. The function $f_{0,b}(t)$ with $\frac{1}{b}=\frac{1}{2\delta}$ is the SRCCDF.cdf function of best Hausdorff one-side approximation to function $c_{\delta,\delta}(t)$ and for H-distance h the following holds for $0<b<2.75$:

$$\frac{1}{2.1\left(1+\frac{1}{b}\right)}<h<\frac{\ln\left(2.1\left(1+\frac{1}{b}\right)\right)}{2.1\left(1+\frac{1}{b}\right)}.$$

The proof of the Theorem follows the ideas given in this paper and will be omitted.

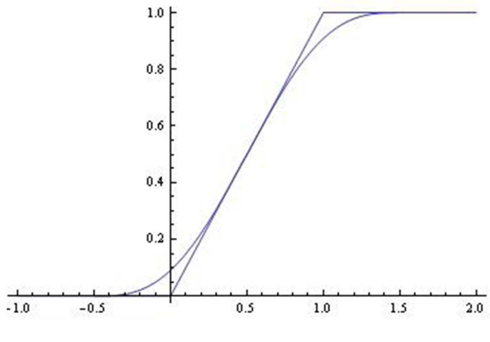

Figure 4. *The cut and (SRCCDF.cdf) for* $\gamma=\delta=1/2$, $b=1$.

Considering this interesting result, the reader may formulate and explore the generalized transmuted cumulative distribution functions of the above functions - $H(t), H_1(t), H_2(t)$ and $H_3(t)$.

As an example,

$$H^*(t)=(1+\lambda)\left(0.5\left(1+\frac{t-\gamma}{b}+\frac{1}{\pi}\sin\left(\frac{(t-\gamma)\pi}{b}\right)\right)\right)-$$
$$\lambda\left(0.5\left(1+\frac{t-\gamma}{b}+\frac{1}{\pi}\sin\left(\frac{(t-\gamma)\pi}{b}\right)\right)\right)^2.$$

References

[1] N. Johnson, S. Kotz, and N. Balakrishnan, Continuous Univariate Distributions, Willey, NY, Vol. 1, 1994.

[2] N. Guliyev and V. Ismailov, A single hidden layer feedforward network with only one neuron in the hidden layer san approximate any univariate function, Neural Computation, Vol. 28, 2016, pp. 1289–1304.

[3] W. Shaw and I. Buckley, The alchemy of probability distributions: beyond Gram-Charlier expansion and a skew-kurtotic-normal distribution from rank transmutation map, Research report, 2007.

[4] R. C. Gupta, O. Akman, and S. Lvin, A study of log–logistic model in survival analysis, Biometrical Journal, Vol. 41 (4), 1999, pp. 431–443.

[5] F. Hausdorff, Set Theory (2 ed.) (Chelsea Publ., New York, (1962 [1957]) (Republished by AMS-Chelsea 2005), ISBN: 978–0–821–83835–8.

[6] B. Sendov, Hausdorff Approximations (Kluwer, Boston, 1990), doi: 10.1007/978-94-009-0673-0.

[7] R. Anguelov and S. Markov, Hausdorff Continuous Interval Functions and Approximations, In: M. Nehmeier et al. (Eds), Scientific Computing, Computer Arithmetic, and Validated Numerics, 16th International Symposium, SCAN 2014, LNCS 9553, pp. 3–13, 2016, Springer, doi: 10.1007/978-3-319-31769-4

[8] N. Kyurkchiev and A. Andreev, Approximation and antenna and filter synthesis: Some moduli in programming environment Mathematica, LAP LAMBERT Academic Publishing, Saarbrucken, 2014, ISBN: 978-3-659-53322-8.

[9] D. Costarelli and R. Spigler, Approximation results for neural network operators activated by sigmoidal functions, Neural Networks, Vol. 44, 2013, pp. 101–106.

[10] D. Costarelli and G. Vinti, Pointwise and uniform approximation by multivariate neural network operators of the max-product type, Neural Networks, 2016, doi: 10.1016/j.neunet.2016.06.002.

[11] D. Costarelli and R. Spigler, Solving numerically nonlinear systems of balance laws by multivariate sigmoidal functions approximation, Computational and Applied Mathematics, 2016, doi: 10.1007/s40314-016-0334-8.

[12] D. Costarelli and G. Vinti, Convergence for a family of neural network operators in Orlicz spaces, Mathematische Nachrichten 2016, doi: 10.1002/mana.20160006.

[13] J. Dombi and Z. Gera, The Approximation of Piecewise Linear Membership Functions and Lukasiewicz Operators, Fuzzy Sets and Systems, Vol. 154 (2), 2005, pp. 275–286.

[14] N. Kyurkchiev, On the Approximation of the step function by some cumulative distribution functions, Compt. rend. Acad. bulg. Sci., Vol. 68 (12), 2015, pp. 1475–1482.

[15] N. Kyurkchiev and S. Markov, Sigmoid functions: Some Approximation and Modelling Aspects, LAP LAMBERT Academic Publishing, Saarbrucken, 2015, ISBN: 978-3-659-76045-7.

[16] N. Kyurkchiev and S. Markov, On the Hausdorff distance between the Heaviside step function and Verhulst logistic function. J. Math. Chem., Vol. 54 (1), 2016, pp. 109–119, doi: 10.1007/S10910-015-0552-0.

[17] N. Kyurkchiev and S. Markov, Sigmoidal functions: some computational and modelling aspects. Biomath Communications, Vol. 1 (2), 2014, pp. 30–48, doi: 10.11145/j.bmc.2015.03.081.

[18] N. Kyurkchiev and S. Markov, On the approximation of the generalized cut function of degree $p+1$ by smooth sigmoid functions, Serdica J. Computing, Vol. 9 (1), 2015, pp. 101–112.

[19] N. Kyurkchiev, A note on the new geometric representation for the parameters in the fibril elongation process, Compt. rend. Acad. bulg. Sci., Vol. 69 (8), 2016, pp. 963–972.

[20] N. Kyurkchiev, S. Markov, and A. Iliev, A note on the Schnute growth model, Int. J. of Engineering Research and Development, vol. 12 (6), 2016, 47-54, ISSN: 2278-067X.

[21] V. Kyurkchiev and N. Kyurkchiev, On the Approximation of the Step function by Raised-Cosine and Laplace Cumulative Distribution Functions. European International Journal of Science and Technology, Vol. 4 (9), 2016, pp. 75–84.

[22] A. Iliev, N. Kyurkchiev, and S. Markov, On the Approximation of the Cut and Step Functions by Logistic and Gompertz Functions, BIOMATH, Vol. 4 (2), 2015, 1510101, doi: 10.11145/j.biomath.2015.10.101.

[23] A. Iliev, N. Kyurkchiev, and S. Markov, On the Approximation of the step function by some sigmoid functions, Mathematics and Computers in Simulation, 2015, doi: 10.1016/j.matcom.2015.11.005.

[24] N. Kyurkchiev and A. Iliev, A note on some growth curves arising from Box-Cox transformation, Int. J. of Engineering Works, Vol. 3 (6), 2016, pp. 47–51, ISSN: 2409-2770.

[25] N. Kyurkchiev and A. Iliev, On some growth curve modeling: approximation theory and applications, Int. J. of Trends in Research and Development, Vol. 3 (3), 2016, pp. 466–471.