

Journal of Financial and Actuarial Mathematics and Management

A joint edition of the Institute of Information Sciences and Institute of Social Sciences of Perun Academy
Publishing Company: PSP Publishing
ISSN: 2367-5810
Issue 9, Year IX (2021)
Founding Editor and Editor-in-Chief: Pavel Stoynov

Editorial Board: Hiroyuki Ono (Japan), Yasukazu Yoshizawa (Japan), Naoyuki Ishimura (Japan), Boris Melnikov (Russia), Elena Melnikova (Russia), Goran Karanovic (Croatia), James Osondu (Nigeria)

Contents

Y. Algani, W. Hibi. Sequences of Topological Spaces:
New Results……..3
P. Stoynov. ST activation function in dynamic systems……..……5
P. Stoynov. Stopit regression model……..……....11

Sequences of Topological Spaces: New Results

Yousef Methkal Abd Algani,
Department of Mathematics, Sakhnin College, Israel.
Department of Mathematics,
The Arab Academic college for education in Israel
yosefabdalgani@gmail.com
https://orcid.org/0000-0003-2801-5880
Wafiq Hibi
Department of Mathematics, Sakhnin College, Israel

Abstract.

Within the concept of the series, we deal with the subject of calculus. When studying the series, we explore the limits of series. Also in topology we will refer to series of real numbers Abd Algani (2019); Dirac (1961); Kozov (2008) proved some results in the sequences of topological spaces. In the present study we will begin by defining the series in Group X, and present the following claims:

Lemma 1: (X, d) will be a metric space. The series (x_n) converges to the point $a \in X$ if and only if for every $\epsilon > 0$ there is a natural number N so that $d(x_n, a) < \epsilon$ for every $> N$.

Lemma 2: (X, T) will be a topological space and $A \subseteq X$. If a is a limit point for a series within group A, then a belongs to the closed group A.

Lemma 3: f will be a continuous function from the topological space (X, T) to the topological space (Y, S) and (x_n) will be a series in X that converge to a, then the series $(f(x_n))$ will converge to $f(a)$.

Key words: Topological space, Metric space, Distance.

Introduction

Let's start with defining the function of distance:

Definition 1. Let R be a set of real numbers and a, b will be two elements in group R. We will define the distance $d(a, b)$ as follows:

$$d(a, b) = |a - b|.$$

So, $d: R \to R$ meets the following four conditions for every $a, b, c \in R$:

1. Non-negativity. $d(a, b) \geq 0$
2. $d(a, b) = 0$ if and only if $a = b$
3. Symmetries $of\ d(a, b) = d(b, a)$.
4. $d(a, c) \leq d(a, b) + d(b, c)$ (triangle inequality)

Definition 2. Let A_1 be a partial group of the metric space (A, d), and b will be an element of group A. The point b is called the limit point of group A_1 if each environment N of the point b contains at least one point of the points A_1 which is different from b, that is,

$$N \cap (A_1 - \{b\}) \neq \emptyset$$

From the above definition it can be concluded:

If b is a limit point for group A_1 then there is a series of points from A_1 that aspires to b. But if $b \in A_1$ so that there is an environment N to the point b and $A_1 \cap N = \{b\}$, then we call the point b, the isolated point from A_1.

Example 1: Let R will be the set of real numbers, and the metric is the matric of the absolute value and $A_1 = (0,1]$.

So the 0 is one of the limit points for group $A_1$and $\left\{1, \frac{1}{2}, \frac{1}{3}, \dots\right\}$ is a series that converge to 0.

But if the group is $A_1 = \{1, 2, 3, \dots\}$ then all the points are isolated if we assumed that $n \in A_1$ then $\left(n - \frac{1}{2}, n + \frac{1}{2}\right)$ is a surrounding for the point n and

$$\left(n-\frac{1}{2},n+\frac{1}{2}\right)\cap A_1=\{n\}$$

Definition 3. The function f from the set of natural numbers N to the set X is called a series in X.

It is clear that the function f is a defined group. Here, we will focus on the range of the function and we will mark it in $f(n)=x_n$ so that x_n is an element of X and we will mark the group with (x_n).

Definition 4. Let (X,T) be a topological space and (x_n) a series of elements from X. Suppose that (x_n) is a function that aspires to a , so that $a\in X$ if and only if for each partial open set A of X contains a. There exists a natural number N so that $x_n\in A$ for each $n>N$. In other words, a contains all the elements of the series except for a finite number of elements.

New results

Lemma 1. (X,d) will be a metric space. The series (x_n) converges to the point $a\in X$ if and only if for every $\epsilon>0$ there is a natural number N so that $d(x_n,a)<\epsilon$ for every $n>N$.

Proof of Lemma 1: Suppose the series (x_n) aspires to point a, and suppose there exists $\epsilon>0$, then the open ball $B(a,\epsilon)$ represents an open environment group to the point a, then $B(a,\epsilon)$ contains the elements of the series (x_n). That is, there exists a natural number N so that x_n belongs to $B(a,\epsilon)$ per $n>N$. That is $d(x_n,a)<\epsilon$ per $n>N$. This means that the series converges to a.

Lemma 2. (X,T) will be a topological space and $A\subseteq X$. If a is a limit point for a series within group A, then a belongs to closed group A.

Proof of Lemma 2: x_n will be a series converging to a, so that $x_n\in A$ per N.

Suppose U is an open group containing a. That is U contains all the elements of the series (x_n) except for a finite number of elements, then $U\cap A\neq\emptyset$

This leads to a point in a closed group A.

Lemma 3. f will be a continuous function from the topological space (X,T) to the topological space (Y,S) and let (x_n) be a series in X that converge to a, then the series $(f(x_n))$ converge to $f(a)$.

Proof of Lemma 3: B will be an open group from Y, which contains the point $f(a)$ then $f^{-1}(B)$is an open group from X that contains the point a.

That is $f^{-1}(B)$ contains all the elements of the series (x_n) except for a finite number of elements. This implies that B contains all the elements of the series $(f(x_n))$ except a finite number of elements,

Then the series $f(f(x_n))$ converges to point $f(a)$.

References

Abd Algani, Y. (2019). The Topological Connectivity of the Independence Complex of Circular-Arc Graphs. *Universal Journal of Mathematics and Applications*, 2 (4), 159-169.

Dirac, G. A. (1961) On rigid circuit graphs, Math. sem. Univ. Hamburg, 25 (1961), 71-76.

Kozov, D. (2008) Combinatorial Algebraic Topology.

ST activation function in dynamic systems

Pavel Stoynov,
Technical University, Sofia, Bulgaria
pstoynov@tu-sofia.bg
Perun R&D Academy, Sofia, Bulgariua.
pavel_todorov_stoynov@yahoo.com

Abstract.
In the neural networks, the case of a single neuron arises from regression models. Logistic sigmoid is the function needed to interpret the output of a network as a probability when the hidden-units activation is governed by a member of the exponential family.

In this article a new type of non-linear activation functions is proposed based on Switch-Time (ST) distribution.

Key words: switch-time distribution, activation function, neural networks

Introduction

Artificial Neural Networks (ANN) are very fast developing branch of Artificial Intelligence (AI) research. Researchers study different kind of neural networks and apply them to different areas of applied science.

In neural networks, as well as in non-linear regression (which may be considered as a very simple case of neural network), different kinds of non-linear activation functions are used.

In this article a new type of non-linear activation function is proposed based on Switch-Time (ST) distribution.

Logistic activation function

The activation function is the function which is applied to the output of a neuron to obtain some specific, generally non-linear type of transformation of the input signals. We denote it by $g(u)$ and the neuron output is $y = g(u) = g(wx)$ where the vector x is the neuron input and w is the vector of weights. The function $g(u)$ should be monotonic with u.

Let us consider a two-class problem in which the class-conditional densities are given by Gaussian distribution with equal covariance matrices $\Sigma = \Sigma_1 = \Sigma_2$. So, for the conditional probability $p(x \mid C_k)$ for x given membership of class C_k, we have

$$p(x \mid C_k) = \frac{1}{(2\pi)^{\frac{d}{2}} \mid \Sigma \mid^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma^{-1}(x-\mu_k)}, k = 1,2,$$

where $C_k, k = 1,2,$ are the two classes into which we should classify the observations, d is the dimension of the distribution and $\mu_k, k = 1,2,$ is the vector of the mean of the Gaussian distribution, corresponding to the class C_k.

Using the Bayes' theorem, the posterior probability of membership of class C_1 is given by Bishop (1994).

$$P(C_1 \mid x) = \frac{p(x \mid C_1)P(C_1)}{p(x \mid C_1)P(C_1) + p(x \mid C_2)P(C_2)} = \frac{1}{1+e^{-u}} = g(u)$$

where

$$u = \ln \frac{p(x \mid C_1)P(C_1)}{p(x \mid C_2)P(C_2)}.$$

So, we have that posterior probability is exactly the logistic sigmoid function

$$g(u) = \frac{1}{1+e^{-u}}.$$

Thus, the use of logistic sigmoid function provides interpretation of the output as posterior probabilities and we obtain not only a classification decision but also a probabilistic interpretation.

Switch-Time distributions

Switch-Time (ST) distributions are presented in details by Stoynov (2019). Here we just remember some basic facts about these distributions.

The random variable ξ with ST distribution has probability density function

$$f_{\xi}(x) = \begin{cases} C(n,\beta)e^{-\beta x}(1+x)^n, \ x \geq 0 \\ 0, x < 0, \end{cases}$$

where $C(n,\beta)$ are normalizing coefficients for which

$$C(n,\beta) = \frac{1}{I(n,\beta)}$$

and

$$I(n,\beta) = \frac{1}{\Gamma(1)}\int_0^{\infty} e^{-\beta t}(1+t)^n dt = \int_0^{\infty} e^{-\beta t}(1+t)^n dt.$$

In the special case when $n=0, \quad n=1$ and $n=2$ we have $C(0,\beta)=\beta, \quad C(1,\beta)=\frac{\beta^2}{\beta+1}$ and $C(2,\beta)=\frac{\beta^3}{\beta^2+2\beta+2}$.

For $I(n,\beta)$ we have

$$I(0,\beta) = \frac{1}{\beta},$$

$$I(n,\beta) = \frac{1}{\beta} + \frac{n}{\beta} I(n-1,\beta), n = 1,2,\ldots$$

The proof can be found in Stoynov (2019).

Below we can see some well known special cases of the distribution

1. Exponential distribution: $ST(0,\beta) \equiv Exp(\beta).$

2. Lindley distribution (Lindley): $ST(1,\beta) \equiv L(\beta).$ The random variable ξ with Lindley distribution has probability density

$$f_{\xi}(x) = \begin{cases} \frac{\beta^2}{\beta+1} e^{-\beta x}(1+x), \ x \geq 0 \\ 0, x < 0. \end{cases}$$

3. The distribution $ST(2,\beta).$ The random variable ξ with distribution $ST(2,\beta)$ has probability density

$$_{\xi}(x) = \begin{cases} \frac{\beta^3}{\beta^2+2\beta+2} e^{-\beta x}(1+x), \ x \geq 0 \\ 0, x < 0. \end{cases}$$

Theorem 1. The probability density function $f_{\xi}(x)$ for $x \geq 0$ of the random variable $\xi \in ST(n,\beta)$ can be presented as a mixture of Erlang distributions:

$$f_{\xi}(x)=\sum_{k=1}^{n+1}P(D^{n}=k)\,f_{G^{k}}(x),$$

where D^{n} are random variables with parameter $n,$ taking values $k=1,\ldots,(n+1)$ with probabilities $P(D^{n}=k)=\dfrac{C(n,\beta)n!}{\beta^{k}(n-k+1)!},\quad k=1,\ldots,(n+1),$ and G^{k} is a random variable with parameter $k,$ having distribution $G^{k}\in\Gamma\left(k,\dfrac{1}{\beta}\right)\equiv Erlang\left(k,\dfrac{1}{\beta}\right).$

Theorem 2. Let $\xi\in ST(2,\beta)$. Then for the moment generating function $M_{\xi}(s)=Ee^{s\xi}$ of ξ we have

$$M_{\xi}(s)=\frac{C(\beta)}{C(\beta-s)},$$

where $C(x)=\dfrac{x^{3}}{x^{2}+2x+2}$ and $\beta>s.$

Theorem 3. The characteristic function of the random variable $\xi\in ST(n,\beta)$ is $\psi_{\xi}(t)=C(n,\beta)\sum_{k=0}^{n}\binom{n}{k}\dfrac{k!}{(\beta-it)^{k+1}}.$

The proofs of Theorem 1, Theorem 2 and Theorem 3 can be seen in Stoynov (2019).

Probabilistic interpretation of Switch-Time activation function

Here we present the probabilistic interpretation of one specific type activation function which uses Switch-Time distribution.

Let us again consider a two-class problem in which the class-conditional densities are given by Gaussian distribution with equal covariance matrices $\Sigma=\Sigma_{1}=\Sigma_{2}.$ So, for the conditional probability $p(x\mid C_{k})$ for x given membership of class $C_{k},$ we have

$$p(x\mid C_{k})=\frac{1}{(2\pi)^{\frac{d}{2}}\mid\Sigma\mid^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu_{k})^{T}\Sigma^{-1}(x-\mu_{k})},$$

where $C_{k},\ k=1,2,$ are the two classes into which we should classify the observations, d is the dimension of the distribution and $\mu_{k},\ k=1,2,$ is the vector of the mean of the Gaussian distribution, corresponding to the class $C_{k}.$

Using the Bayes' theorem, the posterior probability of membership of class C_{1} can be presented as

$$P(C_{1}\mid x)=\frac{p(x\mid C_{1})P(C_{1})}{p(x\mid C_{1})P(C_{1})+p(x\mid C_{2})P(C_{2})}=$$

$$=\frac{1}{1+C(n,\beta)e^{-\beta u}(1+u)^{n}}=g(u;n,\beta)$$

where

$$C(n,\beta)e^{-\beta u}(1+u)^{n}=\frac{p(x\mid C_{2})P(C_{2})}{p(x\mid C_{1})P(C_{1})}.$$

So, we have that posterior probability is exactly the Switch-Time sigmoid function which is a generalization of the logistic sigmoid function because if we take the parameters $n=0$ and $\beta=1,$ we obtain exactly the logistic sigmoid function.

Thus, the use of Switch-Time sigmoid function also provides interpretation of the output as posterior probabilities.

ST activation function

Here we present different types of activation functions based on ST-distribution.
Using ST-function as activation function was for first time proposed in Stoynov (2021).
Here we present different kinds of ST-functions as activation functions.
ST activation function with threshold zero which uses one-side ST distribution is

$$N_{kt} = \begin{cases} \int_0^{n_{kt}} C(n,b)e^{-bt}(1+t)^n dt, \, n_{kt} \geq 0, \\ 0, \, n_{kt} < 0. \end{cases}$$

This function depends on two parameters - n and b. Standard ST-activation functions can be obtained by different choice of these two parameters.

With $n = 0$ and $b = 1$ we obtain standard exponential activation function with threshold zero:

$$N_{kt} = \begin{cases} \int_0^{n_{kt}} e^{-t} dt = 1 - e^{-n_{kt}}, \, n_{kt} \geq 0, \\ 0, \, n_{kt} < 0. \end{cases}$$

The graphic of this function is presented on Figure 1.

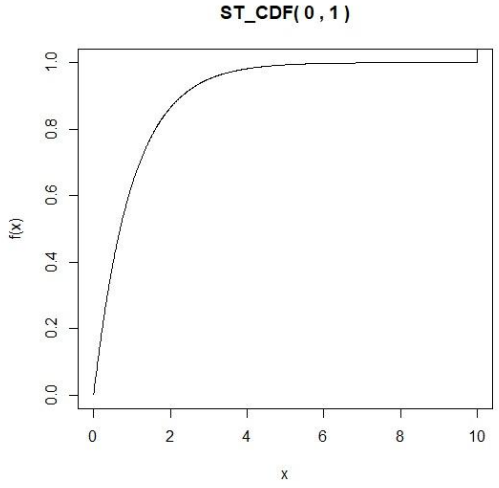


Fig1. Standard exponential activation function with zero threshold and parameters $n = 0$ and $b = 1$.

The double-sided ST-distribution has a density function

$$s_{\xi}(x) = \begin{cases} \frac{1}{2} C(n,b)e^{-bx}(1+x)^n, \, x \geq 0, \\ \frac{1}{2} C(n,b)e^{bx}(1-x)^n, x < 0. \end{cases}$$

This distribution permits to define activation functions without thresholds with general formula

$$N_{kt} = \int_{-\infty}^{n_{kt}} s(x)dx = \begin{cases} \frac{1}{2} + \frac{1}{2} \int_0^{n_{kt}} C(n,b)e^{-bx}(1+x)^n dx, \, n_{kt} \geq 0, \\ \frac{1}{2} \int_{-\infty}^{n_{kt}} C(n,b)e^{bx}(1-x)^n dx, \, n_{kt} < 0. \end{cases}.$$

The function depends on two parameters – n and b. Standard ST-activation functions without thresholds are obtained by selecting different values for these parameters.

The graphics of the density function with $n = 0$ and $b = 1$ is presented in Figure 2.

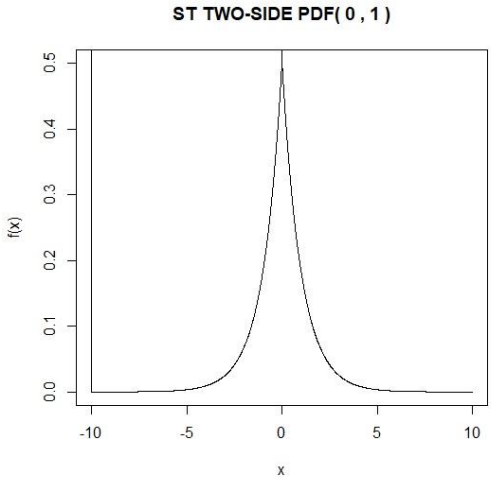


Fig 2. Two-sided ST-density function parameters $n = 0$ and $b = 1.$

With $n = 0$ and $b = 1$ we obtain standard two-sided exponential activation function, called also Laplace activation function:

$$N_{kt} = \begin{cases} \frac{1}{2} + \frac{1}{2}\int_0^{n_{kt}} e^{-x}dx = 1 - \frac{1}{2}e^{-n_{kt}}, \; n_{kt} \geq 0, \\ \frac{1}{2}\int_{-\infty}^{n_{kt}} e^{x}dx = \frac{1}{2}e^{n_{kt}}, n_{kt} < 0. \end{cases}$$

The graphics of the cumulative distribution function of two-sided ST distribution with parameters $n = 0$ and $b = 1$ is presented in Figure 3.

Conclusion

ST distributions proposed for activation function are alternative of the currently used functions and distributions. It may be used in two forms - with and without threshold – based on one-sided and two-sided ST distributions. Also, different values of the parameter n may be chosen with values 0, 1, and 2 giving three standard types ST activation functions for every of the two forms.

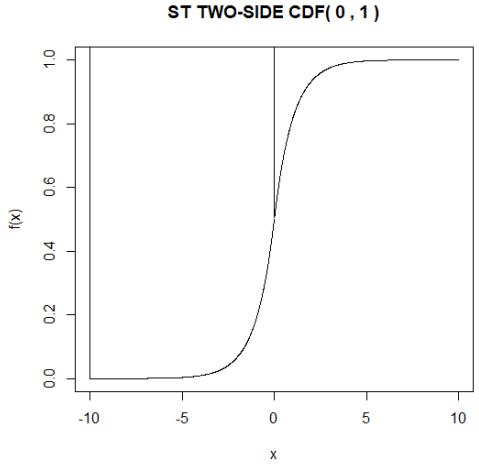


Fig 3. Two-sided ST cumulative distribution function parameters $n = 0$ and $b = 1.$

References

Bishop, P. (1994) Neural Networks for Pattern Recognition.1994. Oxford University Press. ISBN 0 19 853864.

Stoynov, P. (2019) Switch-time distributions and processes and their applications in finance and insurance. Pliska Studia Mathematica 30, (2019), pp. 201-214.

Stoynov, P. (2021) Applications of ST Distributions to Neural Networks and Regression Models. Proceedings of Eighth International Conference on New Trends in the Applications of Differential Equations in Science (NTADES'21), 6-9 September, St. Constantine and Helena, Bulgaria. (accepted for publishing in AIP Conference Proceedings)

Stopit regression model

Pavel Stoynov,
Technical University, Sofia, Bulgaria,
pstoynov@tu-sofia.bg
Perun R&D Academy, Sofia, Bulgaria
pavel_todorov_stoynov@yahoo.com

Introduction

In recent decades neural computing was successfully applied in many fields and became object of deep scientific research. Many concepts here have been inspired by biological networks. But the perspective of statistical analysis in many cases provides more direct link between the basic concepts.

For example, the case of a single neuron in Artificial Neural Networks (ANN) arises from regression models.

In this article a new type of non-linear regression model is proposed based on Switch-Time (ST) distribution and the term stopit regression is defined.

Switch-Time distributions

Switch-Time (ST) distributions are presented in details by Stoynov (2019). Here we just remember some basic facts about these distributions.

The random variable ξ with ST distribution has probability density function

$$f_{\xi}(x) = \begin{cases} C(n,\beta)e^{-\beta x}(1+x)^n, & x \geq 0 \\ 0, & x < 0, \end{cases}$$

where $C(n,\beta)$ are normalizing coefficients for which

$$C(n,\beta) = \frac{1}{I(n,\beta)}$$

and

$$I(n,\beta) = \frac{1}{\Gamma(1)}\int_0^{\infty} e^{-\beta t}(1+t)^n dt = \int_0^{\infty} e^{-\beta t}(1+t)^n dt.$$

In the special case when $n=0,\quad n=1$ and $n=2$ we have $C(0,\beta)=\beta,\quad C(1,\beta)=\frac{\beta^2}{\beta+1}$ and $C(2,\beta)=\frac{\beta^3}{\beta^2+2\beta+2}$.

For $I(n,\beta)$ we have

$$I(0,\beta) = \frac{1}{\beta},$$

$$I(n,\beta) = \frac{1}{\beta} + \frac{n}{\beta} I(n-1,\beta), n = 1,2,\ldots$$

The proof can be found in Stoynov (2019).

Below we can see some well known special cases of the distribution

4. Exponential distribution: $ST(0,\beta) \equiv Exp(\beta).$

5. Lindley distribution (Lindley): $ST(1,\beta) \equiv L(\beta).$ The random variable ξ with Lindley distribution has probability density

$$f_{\xi}(x)=\begin{cases}\dfrac{\beta^2}{\beta+1}e^{-\beta x}(1+x),\ x\geq 0\\ 0, x<0.\end{cases}$$

6. The distribution $ST(2,\beta)$. The random variable ξ with distribution $ST(2,\beta)$ has probability density

$$_{\xi}(x)=\begin{cases}\dfrac{\beta^3}{\beta^2+2\beta+2}e^{-\beta x}(1+x),\ x\geq 0\\ 0, x<0.\end{cases}$$

Theorem 1. The probability density function $f_{\xi}(x)$ for $x\geq 0$ of the random variable $\xi\in ST(n,\beta)$ can be presented as a mixture of Erlang distributions:

$$f_{\xi}(x)=\sum_{k=1}^{n+1}P(D^n=k)\,f_{G^k}(x),$$

where D^n are random variables with parameter n, taking values $k=1,\ldots,(n+1)$ with probabilities $P(D^n=k)=\dfrac{C(n,\beta)n!}{\beta^k(n-k+1)!}$, $k=1,\ldots,(n+1)$, and G^k is a random variable with parameter k, having distribution $G^k\in\Gamma\left(k,\dfrac{1}{\beta}\right)\equiv Erlang\left(k,\dfrac{1}{\beta}\right)$.

Theorem 2. Let $\xi\in ST(2,\beta)$. Then for the moment generating function $M_{\xi}(s)=Ee^{s\xi}$ of ξ we have

$$M_{\xi}(s)=\frac{C(\beta)}{C(\beta-s)},$$

where $C(x)=\dfrac{x^3}{x^2+2x+2}$ and $\beta>s.$

Theorem 3. The characteristic function of the random variable $\xi\in ST(n,\beta)$ is $\psi_{\xi}(t)=C(n,\beta)\sum_{k=0}^{n}\binom{n}{k}\dfrac{k!}{(\beta-it)^{k+1}}.$

The proofs of Theorem 1, Theorem 2 and Theorem 3 can be seen in Stoynov (2019).

The double-sided ST-distribution (DST distribution) has a density function

$$f_{\xi}(x)=\begin{cases}\dfrac{1}{2}C(n,b)e^{-bx}(1+x)^n,\ x\geq 0,\\ \dfrac{1}{2}C(n,b)e^{bx}(1-x)^n, x<0.\end{cases}$$

One popular special case is the distribution $DST(0,\beta)$ known also as two-sided expomnential distribution or Laplace distribution.

Regression based on Switch-Time distribution – stopit regression

The regression models can be seen as simplest cases of neural networks. The logistic regression assumes that a binary dependent variables takes two possible values – one or zero – and the probability for taking value one is

$$p_i=\frac{1}{1+e^{-(x_i\beta+\beta_0)}}$$

where we have that $y_i = \beta_0 + \beta x_i + \varepsilon_i$ is a linear regression model and the function $y = \dfrac{1}{1+e^{-x}}$ is the logistic function. This is the so called logit model which is popular for classification in two classes.

An example for another similar model is the Gompit model which assumes that the binary dependent variable takes values zero and one and the probability for value one is

$$p_i = 1 - e^{-e^{-(x_i\beta+\beta_0)}}$$

where the distribution of Weibull is used which is also called Gompertz distribution. From this name is derived the short name gompit model.

As a third example of binary classification we may consider the probit model. In the framework of this model, the dependent binary variable takes possible values zero and one and the probability for value one is

$$p_i = \Phi(x_i\beta + \beta_0) = \int_{-\infty}^{x_i\beta+\beta_0} f(t)dt,$$

where Φ is the standard cumulative distribution function (CDF) of the normal distribution and f is the probability density function (PDF) of the normal distribution.

By analogy here a new regression model is proposed (Stoynov, 2021). Here we precise two kinds of the stopit regression – one-side stopit regression and two-side stopit regression model.

With one-side stopit model it is assumed that the binary dependent variables accepts value one with probability

$$p_i = S_1(x_i\beta + \beta_0) = \begin{cases} \int_0^{x_i\beta+\beta_0} C(m,b)e^{-bt}(1-t)^n\,dt, & x_i\beta + \beta_0 \geq 0, \\ 0, & x_i\beta + \beta_0 < 0. \end{cases}$$

where S_1 is the cumulative distribution function (CDF) of the one-sided Switch-time (ST) with probability density function $f(x) = C(m,b)e^{-bx}(1-x)^n,\ x \geq 0.$ Here $C(m,b)$ is the normalizing constant.

At the two-sided stopit it is assumed that the binary dependent variable receives value one with probability

$$p_i = S_2(x_i\beta + \beta_0) = \begin{cases} \frac{1}{2} + \frac{1}{2}\int_0^{n_{kt}} C(m,b)e^{-bx}(1+x)^n\,dx, & n_{kt} \geq 0, \\ \frac{1}{2}\int_{-\infty}^{n_{kt}} C(m,b)e^{bx}(1-x)^n\,dx, & n_{kt} < 0. \end{cases}$$

where S_2 is the cumulative distribution function of the two-sided switch-time (ST) distribution.

To estimate the regression coefficient the maximum likelihood method is used. The likelihood function is

$$L(y \mid \beta) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i}.$$

Here $y_i \in \{0,1\}$ and $P(y_i = 1) = p_i,$ correspondingly $P(y_i = 0) = 1 - p_i.$ Also, we have $y_i \approx \beta x_i + \beta_0.$

The natural logarithm of the likelihood function is

$$\ln L(y \mid \beta) = \sum_{i=1}^{n} y_i \ln p_i + (1-y_i)\ln(1-p_i).$$

To estimate the parameters β and β_0 using the method of the maximal likelihood, we look for

$$\arg\max_{\beta,\beta_0} L(y \mid \beta) = \arg\max_{\beta,\beta_0} \ln L(y \mid \beta).$$

When β is a scalar, we have only one explanatory variable. When β is a vector, we have several explanatory variables. In this case we have $y_i = \beta_0 + \sum_{k=1}^{d} \beta_k x_{ik} + \varepsilon_i$. For the derivative of the probability p_i with respect to the explanatory variables x_{ik} in the stopit model we have

$$\frac{\partial p_i}{\partial x_{ik}} = C(m,b)e^{-b(x_i\beta+\beta_0)}(1-x_i\beta+\beta_0)^m\beta_k,$$
$$i = 1,2,\ldots,n,\ k = 1,2,\ldots,d.$$

This partial derivative may be used to identify the characteristics which lead to considerable increase or decrease of the probability for the corresponding state.

If the estimated $p_i \geq 0.5,$ it is assumed that $y_i = 1.$ If the estimated $p_i < 0.5,$ it is assumed that $y_i = 0.$

The regression model for binary choice can be considered as a special case of a neural network for binary choice with one neuron.

Conclusion

Stopit regression model based on ST and DST distribution can be successfully used as a non-linear regression model in many application.

References

McNelis, P. (2005) Neural Networks in Finance: Gaining Predictive Edge in the Market (Elsevier Inc., 2005).

Stoynov, P. (2019) Switch-time distributions and processes and their applications in finance and insurance. Pliska Studia Mathematica 30, (2019), pp. 201-214.

Stoynov, P. (2021) Applications of ST Distributions to Neural Networks and Regression Models. Proceedings of Eighth International Conference on New Trends in the Applications of Differential Equations in Science (NTADES'21), 6-9 September, St. Constantine and Helena, Bulgaria (accepted for publishing in AIP Conference Proceedings).