Fourteenth International Conference
Financial and Actuarial Mathematics – FAM-2021

Stopit regression model

Pavel Stoynov,
Technical University, Sofia.
pstoynov@tu-sofia.bg

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

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

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

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

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

Theorem 1. The probability density function $f_\xi(x)$ for $x \ge 0$ 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 $k = 1,\dots,(n+1)$ with probabilities $P(D^n = k) = \dfrac{C(n,\beta)n!}{\beta^k (n-k+1)!}$, $k = 1,\dots,(n+1),$ 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

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

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 $y = \frac{1}{1+e^{-x}}$ 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).