HOW TO MODEL ROC CURVES – A CREDIT SCORING PERSPECTIVE

Błażej Kochański

Abstract

ROC curves, which derive from signal detection theory, are widely used to assess binary classifiers in various domains. The AUROC (area under the ROC curve) ratio or its transformations (the Gini coefficient) belong to the most widely used synthetic measures of the separation power of classification models, such as medical diagnostic tests or credit scoring.

Frequently a need arises to model an ROC curve. In the biostatistical context, modelling ROC curves was discussed mainly in the context of scarcity of available data and estimation, but in case of credit scoring the modelling may be required for other reasons.

When a model for an ROC curve is needed, several options are available. In the article binormal, bilogistic, bigamma and bibeta models are defined, along with the novel approach: a bifractal ROC curve model.

The models are tested against publicly presented empirical ROC data. As it turns out, taking into account goodness of fit, all presented models, except for the bilogistic curve, are comparable and fit the data quite well. The choice of the model should therefore be driven by other features of the curves under consideration.

Key words: ROC curves, credit scoring, Gini coefficient, AUC

JEL Code: G21, C15, C38, C53, D81

Introduction

ROC curves derive from signal detection theory, but currently they are popular in many domains as a method to graphically present separation power of binary classifiers (Fawcett, 2005). Those binary classifiers may be known for example as diagnostic tests in biostatistics, biomarkers in clinical epidemiology, detectors in signal processing, credit scores in banking and, generally, in machine learning and data mining research and applications.

An ROC curve plots true positive fraction (in biostatistics: sensitivity, in credit scoring: cumulative bad proportion) against false positive fraction (in biostatistics:1-specificity, in credit scoring: cumulative good proportion).

An ROC curve may be summarised with its “area under curve” (AUC or AUROC) index. Generally, with some reservations regarding uncertainty and the shape of the curve, the higher the AUROC, the better. In credit scoring domain the Gini coefficient is more popular, which, in fact, is a function of the area under the curve:

$$Gini = 2 \cdot AUROC - 1 \quad (1)$$

Frequently, a need to model a ROC curve may arise. In biostatistical contexts the scarcity of data drives this need – the question is then how much uncertainty there is about the real path of the curve. The ROC curve is estimated and confidence intervals for AUROC may be computed. Some of the models described below were derived in this context and serve as a basis for such estimation.

In the credit scoring domain scarcity of the data may not be the key problem. If there is uncertainty regarding the ROC curve, it is not driven by limited number of observations – the uncertainty may be driven by the macroeconomic risk or by the fact that the data is censored (a lender does not have a good/bad information on rejected applicants – a reject inference problem). Still, the models for ROC curves may turn out to be useful in such situations.

Another example, more often encountered by the author, when the need for modelling ROC curves arises, is the situation when the financial institution does not have a model yet. For example the lender wants to know how much better should the model be (in other words, how much should its Gini coefficient improve) in order to achieve business goals. Or, the other way round, the lender knows that the quant team is able to produce a new scoring model (based on new available data or with the help of a new classification methodology) and is able to achieve a Gini coefficient higher by 15 percentage points – a question arises, to what extent it will reduce credit losses or increase loan portfolio profitability. Again, models for ROC curves described below may prove useful.

1 ROC curve models

Generally, a ROC curve may be viewed as a function [0; 1]→[0; 1] which is built using two cumulative distribution functions. In the context of credit scoring, those two CDFs are those for good and bad customers based on their scores. The general formula for a ROC curve is then:

$$y = F_B\left(F_G^{-1}(x)\right), \tag{2}$$

where F_B is a CDF of scores of bad customers and F_G^{-1} is inverse distribution function of scores of good customers. In other word, the ROC curve may be defined using a parameter:

$$\begin{aligned} y &= F_B(s) \\ x &= F_G(s), \end{aligned} \tag{3}$$

where *s* denotes the value of the test, that is – in the credit scoring context – the score. Several ROC curve models proposed below are based on this simple observation – two CDFs may be assumed to follow specific probability distributions.

Formula (2) shows that a ROC curve is invariant to monotone transformations of underlying scores – the score does not go directly to the ROC equation, only the CDFs go.

If the scores for bad borrowers are distributed according to a beta distribution with parameters α_B and β_B and the scores for good customers follow *Beta*(α_G, β_G), then the formula for the ROC curve is

$$y = F_{\alpha_B,\beta_B}\left(F_{\alpha_G,\beta_G}^{-1}(x)\right), \tag{4}$$

where $F_{\alpha_B,\beta_B}$and F_{α_G,β_G} are CDFs of the two beta distributions. Such a model is called a "bibeta" ROC curve (Chen & Hu, 2016).

A "bigamma" model (Dorfman et al., 1997), by analogy, assumes that the scores, or some monotone transformation of them, follow two gamma distributions:

$$y = G_{\alpha_B,\beta_B}\left(G_{\alpha_G,\beta_G}^{-1}(x)\right). \tag{5}$$

The same analogy could be used to build a binormal model:

$$y = F_{\mu_B,\sigma_B}\left(F_{\mu_G,\sigma_G}^{-1}(x)\right). \tag{6}$$

In this case, equation (6) can be reformulated to arrive at the following equation:

$$y = \Phi\big(a + b\Phi^{-1}(x)\big), \tag{7}$$

where Φ is a CDF of a standard normal variable, Φ^{-1} is its inverse,

$$a = \frac{\mu_G - \mu_B}{\sigma_G}$$

and

$$b = \frac{\sigma_B}{\sigma_G}.$$

Φ may be viewed as a (probit) link function in the more general equation for a ROC curve:

$$y = g\big(\alpha_0 + \alpha_1 g^{-1}(x)\big), \tag{8}$$

where g( ) is a link function. If a logit function is taken as a link function, namely:

$$g(\cdot) = \exp(\cdot)/(1 + \exp(\cdot)), \qquad (9)$$

a bilogistic ROC curve model is derived. After several transformations the formula for a bilogistic curve looks as follows:

$$y = \left(1 + \exp(\alpha_1 \ln\left(\frac{1}{x} - 1\right) - \alpha_0)\right)^{-1} \qquad (10)$$

With the bilogistic curve we move from parametric ROC curve models, where one starts with distributions and then arrives at the formula, to "algebraic" ROC curve models, where the underlying distribution is somewhat "secondary" to the formula itself.

A simple example of an algebraic model is a "power function":

$$y = x^{\theta}, \theta < 1 \qquad (11)$$

The power function may be also derived as a "Lehmann ROC curve", based on proportional hazards specification (Gönen & Heller, 2010). From the credit scoring perspective the power ROC curve has an interesting property, which could be called „fractal". If the shape of an ROC curve follows equation (11), then if we take any fraction of the lowest-scored customers and graph an ROC curve for this group, the shape of the ROC curve remains the same and the *AUROC*=1/(1+θ). Also, the Gini coefficient (=$(1-\theta)/(1+\theta)$) remains constant. If one would like to make the Gini coefficient an explicit function parameter, the equation can be reformulated in the following way:

$$y = x^{\frac{1-\gamma}{1+\gamma}},\ 0 < \gamma < 1 \qquad (12)$$

where γ is a parameter for the Gini coefficient.

Analogically, one can derive a function which keeps the shape (and the AUROC/Gini) when plotted for any fraction of the highest-scored customers. It could be named a "right-hand" fractal curve, as opposed of the left-hand curve described above. The formula for the right-hand fractal curve is as follows:

$$y = 1 - (1-x)^{\frac{1+\gamma}{1-\gamma}} \qquad (13)$$

As it turns out (see Kochański (2017) for the details), the empirical ROC curves lie somewhere in-between those two extremes, therefore a „bifractal" model can be proposed:

$$y = \beta\left(1 - (1-x)^{\frac{1+\gamma}{1-\gamma}}\right) + (1-\beta)x^{\frac{1-\gamma}{1+\gamma}}, \qquad (14)$$

which is a linear combination of the two „fractal" curves.

The apparent advantage of the bifractal function is that it contains Gini coefficient (or AUROC) as its explicit parameter. Also, the second parameter of the bifractal ROC curve has a meaningful and intuitive interpretation as a distance between to extremes – left- and right-hand fractal curves.

This could be a decisive advantage for the bifractal against the binormal model. However, as it turns out, the binormal model has a quite simple formula analytic formula for area under the ROC curve (Bandos et al., 2017):

$$AUROC = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right), \tag{15}$$

which means that it can be transformed in such a way that its formula has the Gini coefficient as its parameter (γ):

$$y = \Phi\left(\Phi^{-1}\left(\frac{\gamma+1}{2}\right)\sqrt{1+b^2} + b\Phi^{-1}(x)\right) \tag{16}$$

Because of the fact that having explicit Gini coefficient seems to be quite important in modelling, equation (16) will be used as the standard binormal model in the next section.

2 Fitting ROC curve models to empirical data

Models for an ROC curve presented in the previous section may be fitted to the empirical ROC data of real-life financial institutions' scoring models. In this paper empirical ROC curves are taken from publicly available papers (Řezáč & Řezáč, 2011, Wójcicki & Migut, 2010, Tobback & Martens, 2017) and presentations (Conolly, 2015, Jennings, 2017). In all but one case the curves reflect the separation characteristic of some form of a credit scoring model, the only exception is the anti-fraud model from Wójcicki & Migut (2010). In order to obtain the numbers (x and y coordinates of the points constituting the empirical ROC), it was almost always necessary to read the data from the graph itself, therefore an internet-based tool for transforming graphs into numbers with some pointing and clicking was used.

Once the data is available, the question emerges what is the correct procedure for fitting the curve. The binormal and bilogistic curve may be fitted quite intuitively through probit/logit transformation and simple linear regression fitting. However, such a procedure is not available for other models. The aim of the author was to use the same fitting procedure for all ROC curve models. Therefore numerical optimization has been used: bobyqa function from the R package minqa (Powell, 2009, Bates et al., 2014). The "objective function" to be minimized was the following:

$$f_{obj} = \sum_i |y_i - r(x_i)| \cdot w_i, \tag{17}$$

where $|a|$ denotes absolute value of a, x_i and y_i are coordinates of the points from the empirical ROC curve, $r(x_i)$ is a vertical coordinate of x_i point obtained from a modelled ROC curve and w_i are weights set as illustrated in the Figure 1. In other words, weighted average absolute vertical distance between empirical points and the theoretical curve was minimized. Optimisation procedure used in this exercise is analogical to the procedure used for fitting the bifractal curve in Kochański (2017).

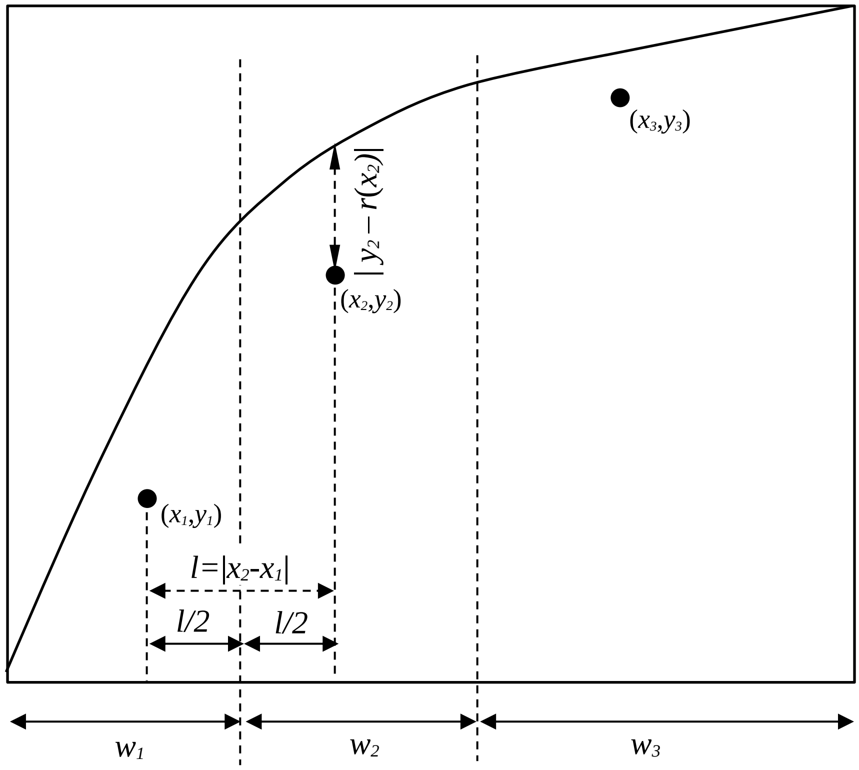
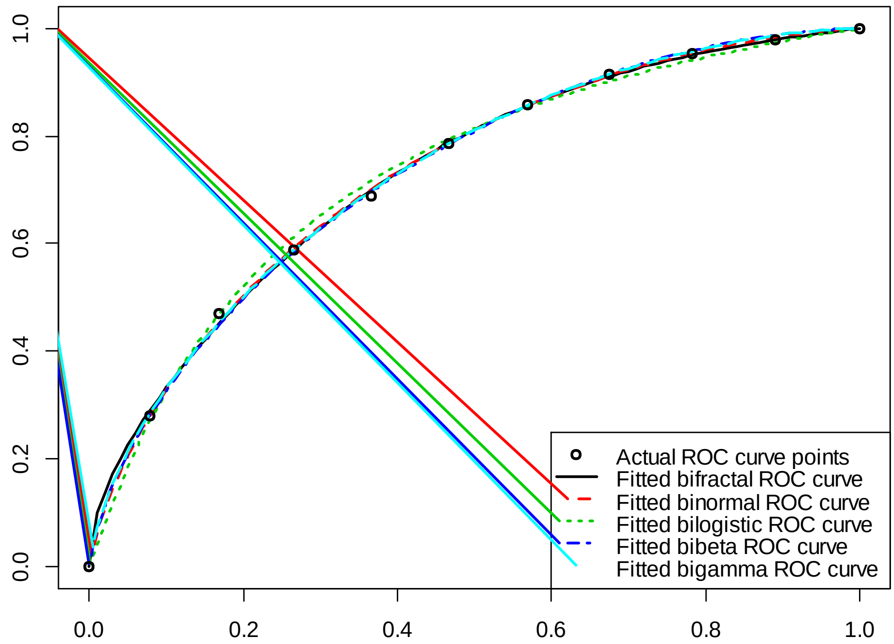


Fig. 1: Setting weights for the bobyqa objective function.

Source: Kochański, 2017.

Fig. 2: Results of fitting the models to the Řezáč & Řezáč (2011) data.



Source: Own calculations.

Example results of fitting the five models (bifractal, binormal, bilogistic, bibeta and bigamma) are presented in figure 2. As it can be seen, in case of data from Řezáč & Řezáč (2011), basically 4 of 5 models fit the data quite good, the most deviation is observed in the case of the bilogistic curve.

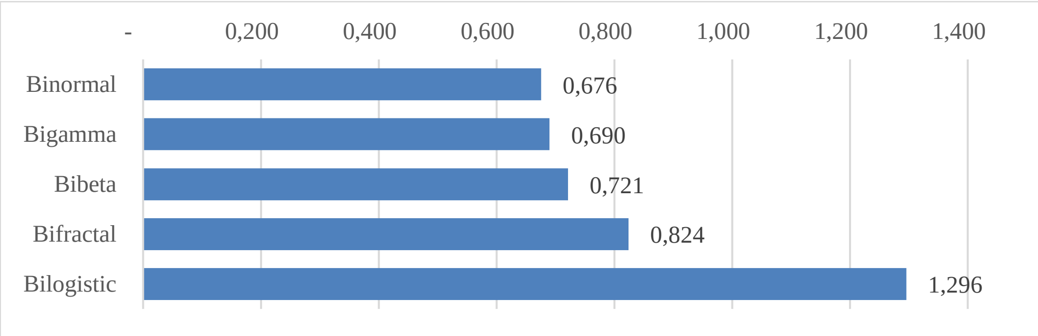
Table 1 presents the results of the fitting process in terms of the goodness of fit for all gathered data sets. Objective (f_{obj}) is multiplied by 100 for clarity reasons – please note it can be interpreted as (weighted) average difference between the empirical ROC points and the fitted ROC curve, expressed in percentage points. As it turns out, in all the cases the bilogistic model shows the worst fit. Binormal model was the best in terms of goodness of fit in 4 cases, bigamma twice and bibeta once. Bifractal model also shows the good fit, but it is worse than binormal in all but one cases.

Tab. 1: ROC model curve fitting – goodness of fit ($f_{obj} \cdot 100$)

	Bifractal	Binormal	Bilogistic	Bibeta	Bigamma
Jennings, 2017	0.45	**0.35**	0.74	0.70	0.59
Conolly, 2015, curve I	0.74	**0.67**	0.77	0.75	0.72
Conolly, 2015, curve II	1.03	0.65	1.25	0.66	**0.61**
Tobback & Martens, 2017	1.44	1.59	2.51	**1.32**	1.35
Řezáč & Řezáč, 2011	0.48	**0.43**	1.10	0.51	0.49
Řezáč & Řezáč, 2011 - additional data points read from the graph	0.53	0.52	1.28	0.51	**0.47**
Wójcicki & Migut, 2010	1.10	**0.52**	1.42	0.60	0.60

Source: own calculations.

Fig. 3: Average goodness of fit ($f_{obj} \cdot 100$) for particular models



Source: own calculations, corresponding to Table 1.

Figure 3 summarises finding from table 1 – average goodness of fit for each model is presented. Once again, the summarised data shows that all presented models, except for the bilogistic curve, are comparable in terms of the average goodness of fit. The choice of the model should therefore be driven by other features of the curves under consideration.

3 Choice of the model – other aspects

When assessing the appropriateness of a particular ROC curve model, apart from goodness of fit, number and interpretability of parameters should be considered. Binormal, bifractal and bilogistic models require just two parameters. Bifractal model has an explicit Gini parameter, which is important in the context of credit scoring modelling. When binormal curve is built with equation (16), the explicit Gini parameter for the binormal curve is also available. Other models do not allow simple reformulation aimed at obtaining Gini coefficient as input.

Second parameter of the bifractal model, responsible for the shape of the curve, has also apparent meaning, while the interpretation of the shape parameter in the binormal model is less obvious. The bifractal model, however, when compared with the dual normal distribution assumption of the binormal model, lacks a theoretical foundation, which may be considered a substantial disadvantage of this approach.

One potential shortcomings of the binormal model is presence of "hooks" – non-concave regions, which are irrational for an ROC curve (a portion of the curve is below the 45º diagonal, so random guessing in those regions would be better than ROC-based decision making). Due to this fact "proper" ROC curves (as opposed to "improper" conventional binormal one) are suggested (Chen & Hu, 2016, Dorfman et al., 1997). In practice however, "hook" regions of fitted binormal curves, especially in the credit scoring context, seem to be relatively small, so this shortcoming is not necessarily the crucial one.

Another argument against the binormal curve (as well as against the bibeta and the bigamma one) is that those models require quite complex mathematical operations. This argument would support bilogistic and bifractal approach, however due to availability of specialized computer software and cheap computing power it is not as important as it probably would be half a century ago.

Table 2 brings together the advantages and disadvantages of the ROC curve models discussed above.

Tab. 2: ROC curve models – comparison

	Pros	Cons
Bigamma/bibeta	Good fit. "Proper" ROC curve obtained.	4 parameters, no explicit AUROC parameter, complicated implementation (requires Beta and Gamma distribution functions).
Bilogistic	2 parameters, simple mathematical operations.	Worst fit, presence of non-concave regions.
Bifractal	2 parameters, explicit Gini parameter, interpretable shape parameter, good fit, only the simplest mathematical operations needed, monotone in the whole domain.	Lack of theoretical background.
Binormal	Best fit, 2 parameters, the model may be reformulated to produce an explicit Gini parameter, lack of apparent interpretability of the second parameter.	Presence of "hooks": non-concave regions of the curve. More complicated implementation where standard normal CDF and quantile function are not available.

Source: prepared by the author.

Conclusion

Empirical ROC curves for credit scoring data may be modelled with the theoretical ROC curve models. Taking into account goodness of fit – measured on publicly available empirical ROC curve data for credit scorecards – the binormal, bibeta, bigamma and bifractal models have comparable performance, while the bilogistic one performs slightly but consistently worse. As the four models fit the data quite well, the choice of the model should therefore be driven by other features of the curves under consideration, including number of parameters and their interpretation (an explicit Gini coefficient in the binormal and bifractal formula is clearly a great advantage), as well as computational perspective and specific properties of particular curves.

References

Bandos, A. I., Guo, B., & Gur, D. (2017). Estimating the Area Under ROC Curve When the Fitted Binormal Curves Demonstrate Improper Shape. *Academic Radiology, 24*(2), 209–219.

Bates, D., Mullen, K. M., Nash, J. C. and Varadhan, R. (2014). minqa: Derivativefree optimization algorithms by quadratic approximation. R package version 1.2.4. (Available from https://CRAN.R-project.org/package=minqa.)

Chen, W., & Hu, N. (2016). Proper bibeta ROC model: algorithm, software, and performance evaluation. In C. K. Abbey & M. A. Kupinski (Eds.), *Medical Imaging 2016: Image Perception, Observer Performance, and Technology Assessment* (Vol. 9787, p. 0E). San Diego, CA.

Conolly, S. (2017). Personality and risk: a new chapter for credit assessment. In *Credit Scoring and Credit Control XV Conference - Presented Papers*. Retrieved April 27, 2018, from https://www.business-school.ed.ac.uk/crc-conference/accepted-papers/

Dorfman, D. D., Berbaum, K. S., Metz, C. E., Lenth, R. V., Hanley, J. A., & Abu Dagga, H. (1997). Proper receiver operating characteristic analysis: the bigamma model. *Academic Radiology, 4*(2), 138-149.

Gönen, M., Heller, G. (2010). Lehmann family of ROC curves. *Medical Decision Making, 30*, 509–17.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874.

Jennings, A. (2015). Expanding the credit eligible population in the USA: A case study. In *Credit Scoring and Credit Control XIV Conference - Conference Papers*. Retrieved April 27, 2018, from https://www.business-school.ed.ac.uk/crc/category/conference-papers/2015/

Kochański, B. (2017), Fractal ROC curves – a simple model for the impact of the Gini coefficient's improvement on credit losses, Manuscript submitted for publication.

Powell, M. J. D. (2009), The BOBYQA algorithm for bound constrained optimization without derivatives. Technical Report DAMTP 2009/NA06, Cambridge: Centre for Mathematical Sciences, University of Cambridge. Retrieved April 27, 2018, from http://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf

Řezáč, M., & Řezáč, F. (2011). How to Measure the Quality of Credit Scoring Models. *Czech Journal of Economics and Finance (Finance a Uver), 61*(5), 486-507.

Tobback, E., & Martens, D. (2017). Retail credit scoring using fine-grained payment data. In *Credit Scoring and Credit Control XV Conference - Presented Papers*. Retrieved April 27, 2018, from https://www.business-school.ed.ac.uk/crc-conference/accepted-papers/

Wójcicki, B., & Migut, G. (2010). Wykorzystanie skoringu do przewidywania wyludzen w Invest Banku. In *Skoring w Zarządzaniu Ryzykiem* (pp. 47-57). Kraków: Statsoft.

Contact

Błażej Kochański
Faculty of Management and Economics
Gdańsk University of Technology
bkochanski@zie.pg.gda.pl