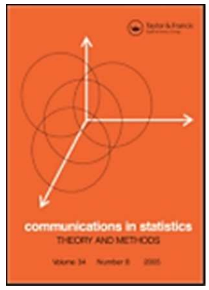

On the Confusion Matrix in Credit Scoring and Its Analytical Properties

Journal:	*Communications in Statistics – Theory and Methods*
Manuscript ID	LSTA-2018-0441
Manuscript Type:	Original Paper
Date Submitted by the Author:	25-Apr-2018
Complete List of Authors:	Zeng, Guoping; Elevate, Risk Management
Keywords:	Confusion Matrix, Sensitivity, Specificity, ROC, KS, Accuracy
Abstract:	Confusion Matrix is an important measure to evaluate the accuracy of credit scoring models. However, the literature about Confusion Matrix is limited. The analytical properties of Confusion Matrix are ignored. Moreover, the concept of Confusion Matrix is confusing. In this paper, we systematically study Confusion Matrix and its analytical properties. We enumerate 16 possible variants of Confusion Matrix and show that only 8 are reasonable. We study the relationship between Confusion Matrix and 2 other performance measures: the receiver operating characteristic curve (ROC) and Kolmogorov-Smirnov statistic (KS). We show that an optimal cutoff score can be attained by KS.

On the Confusion Matrix in Credit Scoring and Its Analytical Properties

Guoping Zeng

4522 Oak Shores Dr, Plano TX 75024

guopingtx@yahoo.com

Abstract. Confusion Matrix is an important measure to evaluate the accuracy of credit scoring models. However, the literature about Confusion Matrix is limited. The analytical properties of Confusion Matrix are ignored. Moreover, the concept of Confusion Matrix is confusing. In this paper, we systematically study Confusion Matrix and its analytical properties. We enumerate 16 possible variants of Confusion Matrix and show that only 8 are reasonable. We study the relationship between Confusion Matrix and 2 other performance measures: the receiver operating characteristic curve (ROC) and Kolmogorov-Smirnov statistic (KS). We show that an optimal cutoff score can be attained by KS.

Keywords: Confusion Matrix; Sensitivity; Specificity; ROC; KS; False Positive; False Negative; True Positive; True Negative

1. Introduction

Recently, credit scoring models have been extensively used by financial institutions to make credit granting decisions (Siddiqi 2006; Refaat 2011; Zeng 2013; Zeng 2014a; Zeng, 2014b; Zeng 2015; Zeng 2016a; Zeng 2017a; Zeng 2017b; Zeng 2018).

ROC is a two-dimensional performance measure of credit scoring models (Siddiqi 2006; Refaat 2011). To construct the ROC curve, for each score, the cumulative proportion of goods is plotted on the x-axis and the cumulative proportion of bads is plotted on the y-axis. In general, the higher the ROC curve above the diagonal line $y = x$, the better the model is. The area under the ROC curve (AUC), known as the AUC statistic, or the c-statistic, is a quantitative measure of ROC. AUC measures the predictive accuracy of a credit scoring model. The larger AUC, the more accurate the model is. As illustrated in Figure 1, AUC is the sum of areas of triangles, rectangles and trapezoids (Zeng 2018).

KS is another major measure to evaluate credit scoring models (Siddiqi 2006; Refaat 2011). KS is the maximum difference between the cumulative proportion of bads and the cumulative proportion of goods among all scores. It measures the predictive ability of the model to separate goods from bads. The KS statistic measures the maximum, across all scores, of the difference in the cumulative percentages of goods and bads.

Figure 2 is a typical KS graph for a credit scorecard model. The X axis shows scores and the Y axis denotes the cumulative proportions of observations in each outcome class (*Goods* vs. *Bads)*. The top curve represents the cumulative percentages of bads, whereas bottom curve represents the cumulative percentages of goods. Then, the KS is the maximum vertical distance between the two curves for that system. The further apart are the two curves, the greater separation the *goods and bads*, and thus, the better is the model.

Confusion (or misclassification) Matrix is a fundamental term in machine learning (Kohavi and Provost 1998). It is used in machine learning to measure the accuracy of a model by comparing predicted values with actual values (Laniz 2015; Mailund 2017). Confusion Matrix has been used in credit scoring (Siddiqi 2006; Refaat 2011; Thomas 2009) to measure the accuracy of a model by comparing the number of true goods and bads with the number of predicted goods and bads for a certain cutoff score.

However, the literature about Confusion Matrix is limited. The analytical properties of Confusion Matrix are ignored. Moreover, the concept of Confusion Matrix is confusing. There is no unified definition for Confusion Matrix. In this paper, for the first time we systematically study Confusion Matrix and its analytical properties. The rest of the paper is organized as follows. In Section 2, we enumerate 16 possible variants of Confusion Matrix and show that only 8 are reasonable. In Section 3, we first study the properties of measures for prediction accuracy. We then study the invariant properties of 4 sum-kind measures for the 8 reasonable Confusion Matrix. Next, we study the relationship between Confusion Matrix and ROC, and between Confusion Matrix and KS. Finally, we prove the optimal cutoff score of Confusion Matrix can be attained by KS. The paper is concluded in Section 4.

Throughout the paper**,** we consider a sample of N observations (also called records or accounts) with binary dependent variable y. We assume that the sample has n_B bads (with $y = 1$) and n_G

goods (with $y = 0$) so that $n_B + n_G = N$. We also assume this sample has been scored by a credit scoring model and has m different scores $S_1 < S_2 < \cdots < S_m$.

2. Confusion Matrix

Assume S is the cutoff score. Let us first define a predicted good and bad. For convenience, we simply call an actual bad or good a bad or good.

Definition 2.1. A sample is a predicted good if its score is larger than S.

Definition 2.2. A sample is a predicted bad if its score is less than or equal to S.

Combining Predicted Good and Bad with Actual Good or Bad, we are ready to define Confusion Matrix.

Definition 2.3. Confusion (or Misclassification) Matrix is a two-way frequency table $C = (c_{ij})_{2\times2}$ with two binary variables Actual (Good or Bad) and Predicted (Good or Bad). The values of the 4 elements $c_{11}, c_{12}, c_{21}, c_{22}$ are called **True Negative (TN), True Positive (TP), False Negative (FN) and False Positive (FP),** not necessarily in order.

For convenience, a third row or a third column is often added into Confusion Matrix for the aggregate purpose.

Given a cutoff score, the 4 values of the two-way frequency table can be uniquely determined. Let's denote the 4 values by $n_{gg}, n_{gb}, n_{bg}, n_{bb}$, where the first subscript represent Actual status g (Good) or b (Bad) and the second subscript represents Predcted status g (Good) or b (Bad).

Naturally,

$$n_{gg} + n_{gb} + n_{bg} + n_{bb} = TN + TP + FN + FP = n_G + n_B = N.$$

However, the concept of Confusion Matrix is confusing. It involves 4 pairs of binary concepts: Good or Bad, Positive or Negative, True or False, Actual or Predicted. It depends upon the position of Actual or Predicted: horizontal or vertical. It also depends upon the positions of Good and Bad in the 2×2 table: Good before Bad or Bad before Good.

Some (Siddiqi 2006; Lantz 2015; Louzada et al. 2016) make Actual horizontal and Predicted vertical. Others (Refaat 2011; Thomas, Edelman, and crook 2002; Thomas 2009) make Actual vertical and Predicted horizontal. Some explicitly or implicitly define good as Negative and bad as Positive (Refaat 2011; Louzada et al. 2016). Others implicitly define good as Positive and bad as Negative (Siddiqi 2006). Some even do not mention Positive or Negative at all (Thomas, Edelman, and Crook 2002; Thomas 2009).

Indeed, it is just a matter of thinking. For instance, if you think Good as Negative as in medical diagnosis, then Bad is Positive. Conversely, if you think Good as Positive as in financial reports, then Bad is Negative. Once Positive and Negative have been determined, Confusion Matrix will be easy to determine:

- True Negative is the number of (Actual) negatives that are correctly classified as negatives.
- False Negative is the number of (Actual) positives that are incorrectly classified as negatives.
- True Positive is the number of (Actual) positives that are correctly classified as positive.
- False Positive is the number of (Actual) negatives that are incorrectly classified as positives.

From the point view of linguistics, we may interpret False as "not actual". In the following we shall enumerate all variants of Confusion Matrix.

Note that in credit scoring, good and bad are concepts of a pair of complementary opposites. The definition of "bad" means bad performance of an account such as bankruptcy, charge-off, fraud, and conceptual-delinquency based definitions like 30+ days s past due in 180 days. Therefore, the concepts of good and bad cannot be switched. By the time the model is developed, the status (Good or Bad) of an account is known.

Therefore, Confusion Matrix will depend on the following 5 conditions

(1) whether good is before bad in rows,
(2) whether good is before bad in columns,
(3) whether Actual or Predicted is placed horizontally or vertically,
(4) whether good or bad is defined as Positive or Negative,

(5) which off-diagonal element is False Positive or False Negative.

So, there are a total of $2 \times 2 \times 2 \times 2 \times 2 = 32$ possible variants of Confusion matrix. Once Conditions (1) and (2) are determined, the values of the 4 elements $c_{11,}c_{12}, c_{21}, c_{22}$ are determined. Then, it is a matter to call who is True Positive, who is True Negative, who is False Positive and who is False Negative.

Among the possible 32 variants, we may reduce to 16 possible variantes by combining Conditions (1) and (2) into one if we naturally let Good and Bad have the same order in rows and in columns:

- whether good is before bad in both rows and columns
- whether Actual or Predicted is horizontal or vertical
- whether good or bad is defined as Positive or Negative
- which off-diagonal element is False Positive or False Negative.

2.1. **Good before Bad**

Let's first assume Good is placed before Bad in both rows and columns in Confusion Matrix. In this case, the 2 "True" cases will always lie in the main diagonal and the 2 "False" cases in the off-diagonal.

2.1.1. **Actual Vertical and Predicted Horizontal**

(I) Variant 1

This variant has the following conditions as in Refaat (2011):

- Good as Negative and Bad as Positive,
- False Positive as c_{21} and False Negative as c_{12}.

Confusion Matrix is illustrated in Table 1. In this case, False Positive means "not actual bad". Since element c_{21} represents actual good but predicted bad, it is "not actual bad". Thus, False Positive as c_{21} makes sense. Similarly, Negative as c_{12} makes sense too.

(II) Variant 2

This variant results from Variant 1 by switching the off-diagonal elements:

- Good as Negative and Bad as Positive,

- False Negative as c_{21} and False Positive as c_{12}.

Confusion Matrix is illustrated in Table 2. In this case, False Negative means “not actual good”. Since element c_{21} represents actual good but predicted bad, it is “not actual bad”. Therefore, False Negative as c_{21}does not make sense.

(III) Variant 3

Variant 3 results from Variant 2 by switching the main diagonal elements:

- Good as Positive and Bad as Negative,
- False Negative as c_{21} and False Positive as c_{12}.

Confusion Matrix is illustrated in Table 3. In this case, False Negative means “not actual bad”. Since element c_{21} represents actual good but predicted bad, it is “not actual bad”. Thus, False Negative as c_{21}makes sense. Similarly, False Positive as c_{12} makes sense too.

(IV) Variant 4

Variant 4 results from Variant 3 by switching the off-diagonal elements:

- Good as Positive and Bad as Negative,
- False Positive as c_{21} and False Negative as c_{12}.

Confusion Matrix is illustrated in Table 4. In this case, False Positive means “not actual good”. Since element c_{21} represents actual good but predicted bad, it is “not actual bad”. False Positive as c_{21} does not makes sense.

2.1.2. **Actual Horizontal and Predicted Vertical**

(V) Variant 5

This variant has the following conditions as in Siddiqi (2006):

- Good as Positive and Bad as Negative,
- False Positive as c_{21} and False Negative as c_{12}.

Confusion Matrix is illustrated in Table 5. In this case, False Positive means “not actual good”. Since element c_{21} represents actual bad but predicted good, it is “not actual good”. False Positive as c_{21}thus makes sense. Similarly, False Negative as c_{12} makes sense too.

(VI) Variant 6

This variant results from Variant 5 by switching the off-diagonal elements:

- Good as Positive and Bad as Negative,
- False Negative as c_{21} and False Positive as c_{12}.

Confusion Matrix is illustrated in Table 6. In this case, False Negative means "not actual bad". Since element c_{21} represents actual bad but predicted good, it is "not actual good". Therefore, False Negative as c_{21} does not make sense.

(VII) Variant 7

Variant 7 results from Variant 6 by switching the main diagonal elements:

- Good as Negative and Bad as Positive,
- False Negative as c_{21} and False Positive as c_{12}.

Confusion Matrix is illustrated in Table 7. In this case, False Negative means "not actual good". Since element c_{21} represents actual bad but predicted good, it is "not actual good". False Negative as c_{21} thus makes sense. Similarly, False Positive as c_{12} makes sense too.

(VIII) Variant 8

Variant 8 results from Variant 7 by switching the off-diagonal elements:

- Good as Negative and Bad as Positive,
- False Positive as c_{21} and False Negative as c_{12}.

Confusion Matrix is illustrated in Table 8. In this case, False Positive means "not actual bad". Since element c_{21} represents actual bad but predicted good, it is "not actual good". False Positive as c_{21} does not makes sense.

2.2. Bad before Good

Next, let's assume Bad is placed before Good in both rows and columns in Confusion Matrix. In this case, the 2 "True" cases will still lie in the main diagonal and the 2 "False" cases in the off-diagonal. Following arguments similar to those in section 2.1, we find 4 reasonable Variants (9 – 12) as illustrated in Tables 9 – 12.

3. Analytical Properties of Confusion Matrix

From Section 2, we see that 8 variants (Variants 1, 3, 5, 7, 9 – 12) are reasonable definitions of Confusion Matrix. Among the 8 reasonable definitions, Variant 1 is the most frequently used one, followed by variants 5 and 7.

With Confusion Matrix, we can formulize the following measures for prediction accuracy in a standard way by just using Positive and Negative (Refaat 2011; Siddiqi 2006; Louzada et al. 2016).

Definition 3.1. The Accuracy (or **Overall Accuracy)** is the ratio of True Positive and True Negative to the total number of accounts:

$$Accuracy = \frac{TP + TN}{n_B + n_G}. \tag{3.1}$$

Definition 3.2. Error Rate (ERR) is the ratio of False Positive and False Negative to the total number of accounts:

$$ERR = \frac{FN + FP}{n_B + n_G}. \tag{3.2}$$

Definition 3.3. Sensitivity is the accuracy of positives or the true positive rate, defined as

$$Sensitivity = \frac{TP}{Total\ Actual\ Positives}. \tag{3.3}$$

Definition 3.4. Specificity is the accuracy of negatives or the true negative rate, defined as

$$Specificity = \frac{TN}{Total\ Actual\ Negatives}. \tag{3.4}$$

Definition 3.5. Type I Error is the false positive rate or the proportion of Negatives predicted as Positives, defined as

$$FPR = \frac{FP}{Total\ Actual\ Negatives}. \tag{3.5}$$

Definition 3.6. Type II Error is the false negative rate or the proportion of Positives predicted as Negatives, defined as

$$FNR = \frac{FN}{Total\ Actual\ Positives}. \tag{3.6}$$

Remark 3.7. It follows from the definitions of Specificity and Sensitivity in Thomas, Edelman, and Crook (2002) and Thomas (2009) that they implicitly treat Good as Negative and Bad as Positive. Hence, their version of Confusion Matrix is essentially Variant 1.

3.1. **Properties of Measures for Prediction Accuracy**

The first 3 of the following properties are implied in literature (Thomas 2009; Huang 2014) without proofs. We explicitly state them here and provide analytic proofs for all the 8 reasonable Variants 1, 3, 5, 7, 9 - 12 of Confusion Matrix.

Theorem 3.8. The measures (3.1) – (3.6) have the following properties for all the 8 reasonable Variants 1, 3, 5, 7, 9 – 12:

(1) $Accuracy < Sensitivity + Specificity.$

(2) $Accuracy + ERR = 1.$

(3) Sensitivity + Type II Error = 1.

(4) Specificity + Type I Error = 1.

Proof. (1) Consider 2 mutual exclusive cases: (i) Good as Positive and Bad as Negative; (ii) Good as Negative and Bad as Positive. For (i), $Total\ Actual\ Positives = n_G$ and $Total\ Actual\ Negatives = n_B$. Hence,

$$Accuracy = \frac{TP + TN}{n_B + n_G}$$

$$= \frac{TP}{n_B + n_G} + \frac{TN}{n_B + n_G}$$

$$< \frac{TP}{n_G} + \frac{TN}{n_B}$$

$$= \frac{TP}{Total\ Actual\ Positives} + \frac{TN}{Total\ Actual\ Negatives}$$

$$= Sensitivity + Specificity.$$

Case (ii) can be proved similarly.

(3) It follows from Tables 1, 3, 5, 7, 9 – 12 that $TP + FN = n_B, n_G, n_G, n_B, n_G, n_B, n_G, n_B$ in Variants 1, 3, 5, 7, 9 – 12, respectively**,** which are all actual positives. Hence,

$$
\begin{aligned}
&Sensitivity\ +\ Type\ II\ Error \\
&= \frac{TP}{Total\ Actual\ Positives} + \frac{FN}{Total\ Actual\ Positives} \\
&= \frac{TP\ +\ FN}{All\ Actual\ Positive} \\
&= \frac{All\ Actual\ Positive}{All\ Actual\ Positive} \\
&= 1.
\end{aligned}
$$

(4) Can be proved in an analogous way to (3). ■

3.2. Invariant Properties of Confusion Matrix

While individual measures such as False Positive, False Negative, True Positive, Truce negative, Sensitivity and Specificity depend on the definition of Confusion Matrix, the following 6 sum-kind measures are invariant for all the 8 variants of Confusion Matrix.

Theorem 3.9. The following 6 measures are all invariant for all the 8 reasonable Variants 1, 3, 5, 7, 9 – 12 of Confusion Matrix:

(1) TN + TP.
(2) FN + FP.
(3) Accuracy.
(4) Error Rate.
(5) Sensitivity + Specificity.
(6) Type I Error + Type II Error.

Proof. (1) TN and TP are the 2 diagonal elements, whose sum is $n_{gg} + n_{bb}$ for all the 8 variants of Confusion Matrix.

(2) FN and FP are the 2 off-diagonal elements, whose sum is $n_{gb} + n_{bg}$ for all the 8 variants of Confusion Matrix.

(3) By (1) TN + TP is invariant for all the 8 reasonable Variants. Hence, $Accuracy = \frac{TP+TN}{n_B+n_G}$ is invariant for all the 8 reasonable Variants.

(4) can be proved similarly to (3).

(5) $Sensitivity\ +\ Specificity = \frac{TP}{Total\ Actual\ Positives} + \frac{TN}{Total\ Actual\ Negatives}$.

The right of the above is equal to

$$\frac{n_{bb}}{n_B} + \frac{n_{gg}}{n_G}, \frac{n_{gg}}{n_G} + \frac{n_{bb}}{n_B}, \frac{n_{gg}}{n_G} + \frac{n_{bb}}{n_B}, \frac{n_{bb}}{n_B} + \frac{n_{gg}}{n_G}, \frac{n_{gg}}{n_G} + \frac{n_{bb}}{n_B}, \frac{n_{bb}}{n_B} + \frac{n_{gg}}{n_G}, \frac{n_{gg}}{n_G} + \frac{n_{bb}}{n_B},$$

$\frac{n_{bb}}{n_B} + \frac{n_{gg}}{n_G}$ for variants 1, 3, 5, 7, 9, 10, 11, 12, respectively, which are all equal

$\frac{n_{bb}}{n_B} + \frac{n_{gg}}{n_G}$.

(6) By (3) and (4) of Theorem 3.8, we have

$$Type\ I\ Error\ +\ Type\ II\ Error\ =\ 2 - (Sensitivity\ +\ Specificity).$$

So, $Type\ I\ Error\ +\ Type\ II\ Error$ is invariant for all the 8 reasonable Variants of Confusion Matrix by (5). ■

3.3. Confusion Matrix and ROC

Unlike Confusion Matrix, ROC does not use a cutoff score. Nevertheless, ROC can be still related to Confusion Matrix. According to Siddiqi (2006), Refaat (2011), and Thomas, Edelman, and Crook (2009), ROC can be generated using the concept of Confusion Matrix. However, no analytical proofs have been given. In the following, we generalize and summarize into a theorem for all the 8 reasonable variants of Confusion Matrix and then provide a mathematical proof.

Theorem 3.10. ROC can be generated by plotting Sensitivity and (1 – Specificity) in the xy-plane by varying cutoff with all the scores $S_1, S_2, \dots, S_m$.

(1) If Bad is defined as Positive and Good as Positive, then Sensitivity is on the y-axis and (1 – Specificity) on the x-axis.

(2) If otherwise Bad is defined as Negative and Good as Positive, then Specificity is on the y-axis and (1 – Sensitive) on the x-axis.

Proof. (1) For any i from 1 to m, if we take S_i as a cutoff score, then by Definition 2.2, all the bads with scores less than or equal to S_i are also predicted bads. Denote the number of bads with scores less than or equal to S_i by n_{b,S_i}. Then $n_{b,S_i} = TP$ and so

$$\frac{n_{b,S_i}}{n_B} = \frac{TP}{n_B},$$

where the left is the cumulative bads for score S_i and the right is Sensitivity with regard to cutoff score S_i. Therefore, Sensitivity with regard to cutoff score S_i can be put on the y-axis.

By Definition 2.2, all the goods with scores less than or equal to S_i are also predicted bads. Denote the number of goods with scores less than or equal to S_i by n_{g,S_i}. Then $n_{g,S_i} = FP$ and so

$$\frac{n_{g,S_i}}{n_G} = \frac{FP}{n_G},$$

where the left is the cumulative goods for score S_i and the right is $Type\ I\ Error\ = 1 - Specificity$ by Property (4) in Theorem 3.8. Therefore, (1 – Specificity) with regard to cutoff score S_i can be put on the x-axis.

(2) For any i from 1 to m, if we take S_i as a cutoff score, then by Definition 2.2, all the bads with scores less than or equal to S_i are also predicted bads. Using notation n_{b,S_i} in (1), we obtain $n_{b,S_i} = TN$ and so

$$\frac{n_{b,S_i}}{n_B} = \frac{TN}{n_B},$$

where the left is the cumulative bads for score S_i and the right is Specificity. Therefore, Specificity with regard to cutoff score S_i can be put on the y-axis.

By Definition 2.2, all the goods with scores less than or equal to S_i are predicted bads. Using notation n_{g,S_i} from (1), we obtain $n_{g,S_i} = FN$ and so

$$\frac{n_{g,S_i}}{n_G} = \frac{FN}{n_G},$$

where the left is the cumulative goods for score S_i and the right is $Type\ II\ Error\ =$ $1 - Sensitivity$ by Property (3) in Theorem 3.8. Therefore, (1 – Sensitivity) with regard to cutoff score S_i can be put on the x-axis. ■

Remark 3.11. Bad is implicitly defined **as** Negative in Siddiqi (2006). Hence, Specificity is on the y-axis and (1 – Sensitive) on the x-axis for ROC. Yet, Sensitivity is mistakenly put on the y-axis and (1 – Specificity) on the x-axis in Siddiqi (2016).

3.4. Confusion Matrix and KS

Since the KS graph uses the cumulative percentages of goods and bads like in ROC, we immediately obtain the following result from the proof of Theorem 3.10.

Theorem 3.12. The KS graph can be generated by Confusion Matrix by putting scores in the x-axis and by making each score as a cutoff score.

(1) If Bad is defined as Positive and Good as Negative, then the top curve has Sensitivity on the y-axis and the bottom curve has (1 – Specificity) on the y-axis.

(2) If Bad is defined as Negative and Good as Positive, then the top curve has Specificity on the y-axis and the bottom curve has (1 – Sensitivity) on the y-axis.

3.5. Optimal Cutoff Score by KS

As pointed out in (SIDDIQI 2006), a good credit scoring model would be the one where the “true” cases are maximized and “false” cases are minimized. That is, the Accuracy is maximized or Error Rate is minimized, since they add up to 1, or the sum of Sensitivity and Specificity is maximized.

While KS does not guarantee the best accuracy (Thomas, 2009) in terms of the (Overall) Accuracy, it gives an optimal cutoff score in terms of the sum of Sensitivity and Specificity.

Theorem 3.13. For any cutoff score, we have

$$Sensitivity + Specificity \leq 1 + KS.$$

The equality holds if the cutoff score is the score S^* attaining the maximum difference in KS.

Proof. KS in the maximum difference between the cumulative percentages of goods and bads. By (5) of Theorem 3.9, we can use Variant 1 to represent all the 8 reasonable Variants of Confusion Matrix. To begin with, let us take any score S_i as the cutoff score for $1 \leq i \leq m$.

From the proof of Theorem 3.10, we see that

$$Sensitivity = \frac{n_{b,S_i}}{n_B},$$

and

$$Specificity = 1 - \frac{n_{g,S_i}}{n_G}.$$

Adding the above 2 equations, we have

$$Sensitivity + Specificity$$

$$= 1 + \frac{n_{b,S_i}}{n_B} - \frac{n_{g,S_i}}{n_G}.$$

Since $\frac{n_{b,S_i}}{n_B} - \frac{n_{g,S_i}}{n_G}$ is the difference between cumulative proportions of Bads and Goods, by the definition of KS, we have

$$\frac{n_{b,S_i}}{n_B} - \frac{n_{g,S_i}}{n_G} \leq 1 + KS$$

with equality holding at $S_i = S^*$. Hence,

$$Sensitivity + Specificity \leq 1 + KS$$

with equality holding at $S_i = S^*$. ■

4. Conclusions

In this paper, we have systematically studied Confusion Matrix and its analytical properties. We clarify the confusing concept of Confusion Matrix by showing that only 8 are reasonable among 16 possible Variants of Confusion Matrix. We study the properties of measures for prediction accuracy, the invariant properties of 4 sum-kind measures for the 8 reasonable variants of Confusion Matrix, the relationship between Confusion Matrix and ROC, and the relationship between Confusion Matrix and KS. We also show that an optimal cutoff score of Confusion Matrix can be attained by KS.

Although all the 8 reasonable Variants of Confusion Matrix are good, Variants 1, 5 and 7 are recommended to use. In addition, our methodology and results can be easily extended to machine learning.

References

Huang, J. 2014. Feature Selection in Credit Scoring – A quadratic Programming Approach Solving with Bisection Method Based on TABU Search. Ph.D. Dissertation. Texas A & M International University.

Kohavi, R., and Provost, F. 1998. On Applied Research in Machine Learning. In *Editorial for the Special Issue on Applications of Machine Learning and the Knowledge Discovery Process, Columbia University, New York*, volume 30.

Lantz, B. 2015. *Machine Learning with R*. 2nd ed. Birmingham: PACKT Publishing.

Louzada, F., Ara, A., and Fernandes, G. B. 2016. Classification methods applied to credit scoring: A systematic review and overall comparison. *Surveys in Operations Research and Management Science* 21(2):117–134.

Mailund, T. 2017. *Beginning Data Science in R - Data Analysis, Visualization, and Modelling for the Data Scientist*. Apress.

Refaat, M. 2011. *Credit Risk Scorecards: Development and Implementation Using SAS*. Raleigh, North Carolina, USA: LULU.COM.

Siddiqi, N. 2006. *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Thomas, L. C. 2009. *Consumer Credit Models: Pricing, Profit and Portfolios*. Oxford University Press, 1st Edition.

Thomas, L. C., Edelman, D. B., and Crook, J. N. 2002. *Credit Scoring and Its Applications*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Zeng, G. 2013. Metric divergence measures and information value in credit scoring. *Journal of Mathematics*, vol. 2013, Article ID 848271. doi: 10.1155/2013/848271.

Zeng, G. 2014a. A rule of thumb for reject inference in credit scoring. *Mathematical Finance Letters*, vol. 2014, article 2, 1-13.

Zeng, G. 2014b. A necessary condition for a good binning algorithm in credit scoring. *Applied Mathematical Sciences Applied Mathematical Sciences* 8(65): 3229–3242. doi: 10.12988/ams.2014.44300.

Zeng, G. 2015. A unified definition of mutual information with applications in machine learning. *Mathematical Problems in Engineering*, vol. 2015, Article ID 201874. doi: 10.1155/2015/201874.

Zeng, G. 2017a. A comparison Study of Computational Methods of Kolmogorov–Smirnov Statistic in Credit Scoring. *Communications in Statistics: Simulation and Computation* 46(10): 7744-7760. doi: 10.1080/03610918.2016.1249883.

Zeng, G. 2017b. Invariant Properties of Logistic Regression Model in Credit Scoring under Monotonic Transformations. *Communications in Statistics: Theory and Methods* 46 (17), 8791-8807. doi: 10.1080/03610926.2016.1193200.

Zeng, G. 2017c. On the Existence of Maximum Likelihood Estimates for Weighted Logistic Regression. *Communications in Statistics: Theory and Methods* 46 (22) : 11194-11203. doi: 10.1080/03610926.2016.1260742.

Zeng, G., and Zeng, E. 2018. On the Three-way Equivalence of AUC in Credit Scoring with Tied Scores. *Communications in Statistics: Theory and Methods.* doi: 10.1080/03610926.2018.1435814.

Figure 1. ROC Curve

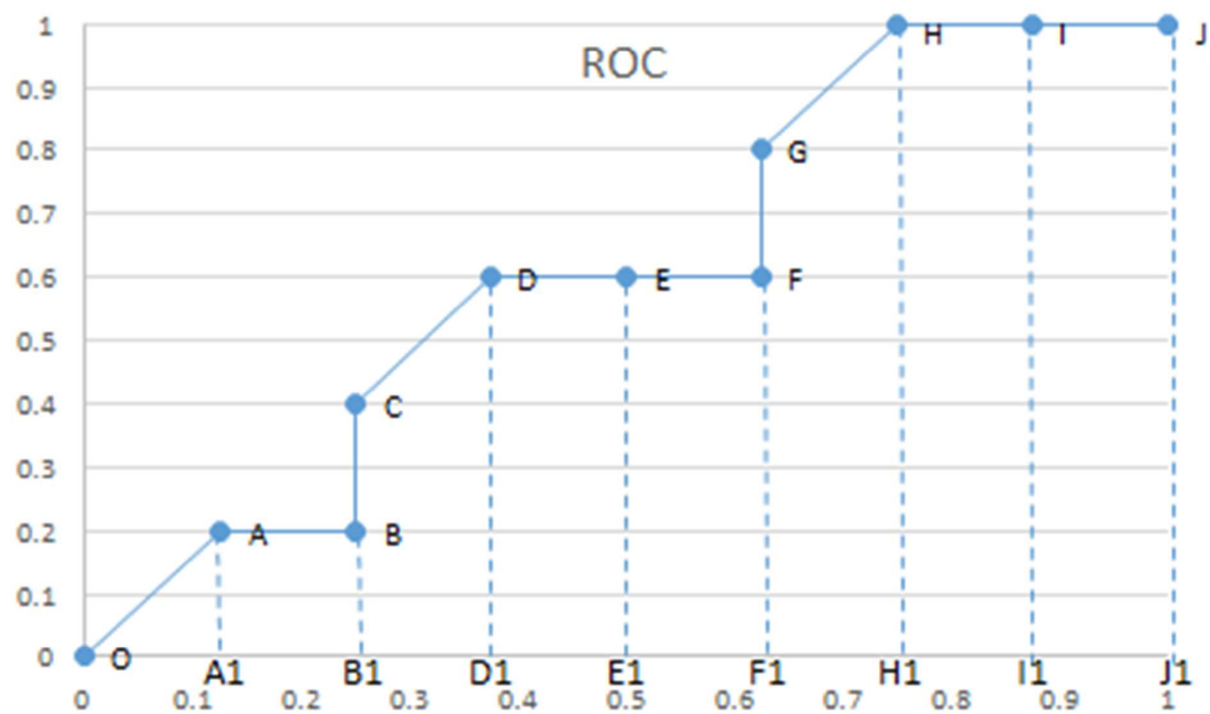

Figure 2.

Figure 2. KS Graph

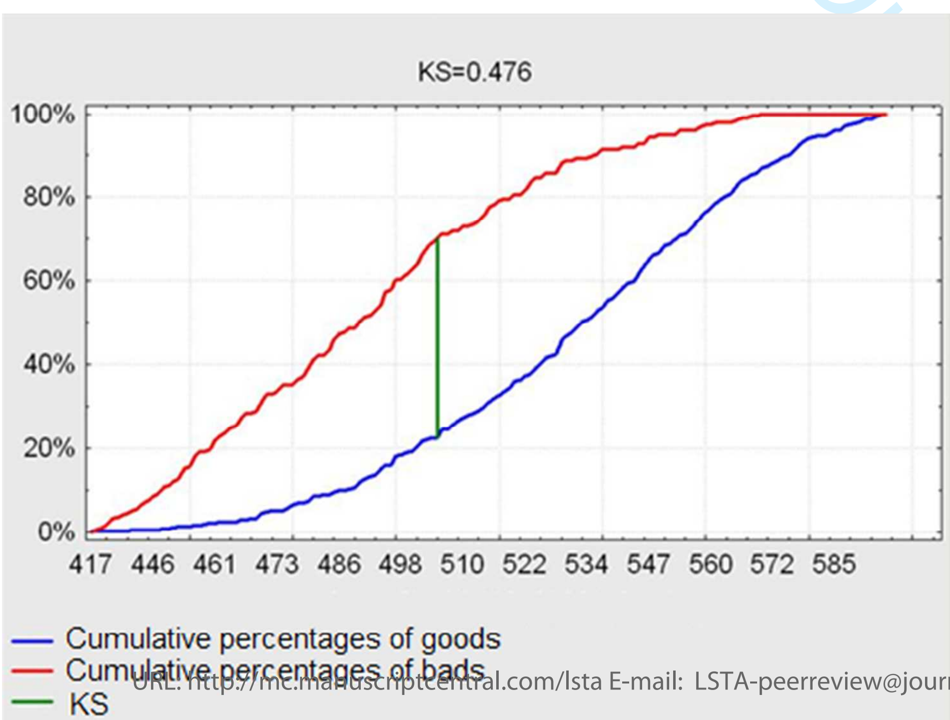

Table 1. Confusion Matrix Variant 1

		Actual	
		Good	Bad
Predicted	Good	True Negative	False Negative
	Bad	False Positive	True Positive
Total		n_G	n_B

Table 2. Confusion Matrix variant 2

		Actual	
		Good	Bad
Predicted	Good	True Negative	False Positive
	Bad	False Negative	True Positive
Total		n_G	n_B

Table 3. Confusion Matrix Variant 3

		Actual	
		Good	Bad
Predicted	Good	True Positive	False Positive
	Bad	False Negative	True Negative
Total		n_G	n_B

Table 4. Confusion Matrix Variant 4

		Actual	
		Good	Bad
Predicted	Good	True Positive	False Negative
	Bad	False Positive	True Negative
Total		n_G	n_B

Table 5. Confusion Matrix Variant 5

		Predicted		
		Good	Bad	**Total**
Actual	Good	True Positive	False Negative	n_G
	Bad	False Positive	True Negative	n_B

Table 6. Confusion Matrix Variant 6

		Predicted		
		Good	Bad	**Total**
Actual	Good	True Positive	False Positive	n_G
	Bad	False Negative	True Negative	n_B

Table 7. Confusion Matrix Variant 7

		Predicted		
		Good	Bad	**Total**
Actual	Good	True Negative	False Positive	n_G
	Bad	False Negative	True Positive	n_B

Table 8. Confusion Matrix Variant 8

		Predicted		
		Good	Bad	**Total**
Actual	Good	True Negative	False Negative	n_G
	Bad	False Positive	True Positive	n_B

Table 9. Confusion Matrix Variant 9

		Actual	
		Bad	Good
Predicted	Bad	True Negative	False Negative
	Good	False Positive	True Positive
Total		n_B	n_G

Table 10. Confusion Matrix Variant 10

		Actual	
		Bad	Good
Predicted	Bad	True Positive	False Positive
	Good	False Negative	True Negative
Total		n_B	n_G

Table 11. Confusion Matrix Variant 11

		Predicted		
		Bad	Good	**Total**
Actual	Bad	True Negative	False Positive	n_B

	Good	False Negative	True Positive	n_G

Table 12. Confusion Matrix Variant 12

		Predicted		
		Bad	Good	**Total**
Actual	Bad	True Positive	False Negative	n_B
	Good	False Positive	True Negative	n_G