

# 信用リスクスコアリングにおける AUC 最大化法

総合研究大学院大学 三浦翔  
統計数理研究所 山下智志, 江口真透

## 1 従来の手法と提案する新手法について

企業の信用リスクスコアリングの分野において広く用いられている従来からのモデル作成の手法は、2 値回帰モデル、特にロジットモデルを仮定し、最尤推定法により線形スコアリングのパラメータの推定値を求め、モデルを作成するものである。また、作成されたモデルの判別力を測定する指標として、ROC(Receiver Operating Characteristic) 曲線の下側面積で定義される AUC(Area under Curve) が用いられることが多い。デフォルト企業群、非デフォルト企業群の各データ数を  $n_D, n_{ND}$  とし、それぞれの企業群に属する企業の説明変量ベクトルをそれぞれ  $\mathbf{X}_i^D, \mathbf{X}_j^{ND} (i = 1, 2, \dots, n_D, j = 1, 2, \dots, n_{ND})$  とすると、データから得られる AUC の値  $\overline{\text{AUC}}(\beta)$  は、ヘビサイド関数  $I$  を用いて、

$$\overline{\text{AUC}}(\beta) = \frac{1}{n_D n_{ND}} \sum_{i=1}^{n_D} \sum_{j=1}^{n_{ND}} I(\beta^T \mathbf{X}_i^D - \beta^T \mathbf{X}_j^{ND}) \quad (1)$$

で表され、値が大きいほど作成されたモデルは判別力を有すると判断する。しかし、従来からの手法で作成されたモデルは、一般的には AUC における最適性は有していない。そこで、本研究で提案する手法では、 $\overline{\text{AUC}}(\beta)$  を最大化する線形スコアリングのパラメータの推定値  $\hat{\beta}_A$  を求め、モデルを作成する。

$$\begin{aligned} \hat{\beta}_A &= \arg \max_{\beta} \overline{\text{AUC}}(\beta) \\ &= \arg \max_{\beta} \frac{1}{n_D n_{ND}} \sum_{i=1}^{n_D} \sum_{j=1}^{n_{ND}} I(\beta^T \mathbf{X}_i^D - \beta^T \mathbf{X}_j^{ND}) \end{aligned}$$

このとき、 $\overline{\text{AUC}}(\beta)$  は微分できないヘビサイド関数で定義されるため、(1) 式中のヘビサイド関数を微分可能なシグモイド関数  $s_{\sigma}(x) = 1/\{1 + \exp(-x/\sigma)\}$  で近似することにより、微分可能な近似 AUC を最大化するようなパラメータの推定値  $\hat{\beta}_{\sigma}$  を求める。 $\overline{\text{AUC}}(\beta)$  のヘビサイド関数をシグモイド関数で近似した関数を  $\text{sAUC}_{\sigma}(\beta)$  とすると、

$$\begin{aligned} \hat{\beta}_{\sigma} &= \arg \max_{\beta} \text{sAUC}_{\sigma}(\beta) \\ &= \arg \max_{\beta} \frac{1}{n_D n_{ND}} \sum_{i=1}^{n_D} \sum_{j=1}^{n_{ND}} s_{\sigma}(\beta^T \mathbf{X}_i^D - \beta^T \mathbf{X}_j^{ND}) \end{aligned}$$

と表される。 $\text{sAUC}_{\sigma}(\beta)$  を最大化する推定量  $\hat{\beta}_{\sigma}$  を求め、モデルを作成する。

従来のロジットモデル・最尤推定法により得られる推定量は、モデルを仮定することにより推定量を求めている。一方で、 $\text{sAUC}_{\sigma}(\beta)$  は線形スコアリングの順位にしかよらず、 $\hat{\beta}_{\sigma}$  を求める際にモデルの仮定は用いていない。したがって、提案する手法は従来の手法よりも異常値に対してロバストであると考えた。以下に、実際のデータを用いた両手法の結果を記す。

## 2 実データを用いた分析結果

今回用いたデータは、1999~ 2003 年の上場企業において観測されたデフォルト企業 25 社を用いた。また、同程度の規模の非デフォルト企業を 1:2 の割合のなるように、50 社を選び出し、合計 75 社のデータを用いた。説明変量は、自己資本比率、売上高営業利益率、流動比率の 3 つを用いた。

## 2.1 実データを用いた両手法の $\overline{AUC}$ の比較

実データを用いて両手法により得られるモデルの  $\overline{AUC}$  の比較を行う。下図はそれぞれ、(左図:a) ロジットモデル・最尤推定法により得られる ROC 曲線と  $\overline{AUC}$  の値と (右図:b) 提案する新手法により得られる ROC 曲線と  $\overline{AUC}$  の値である。

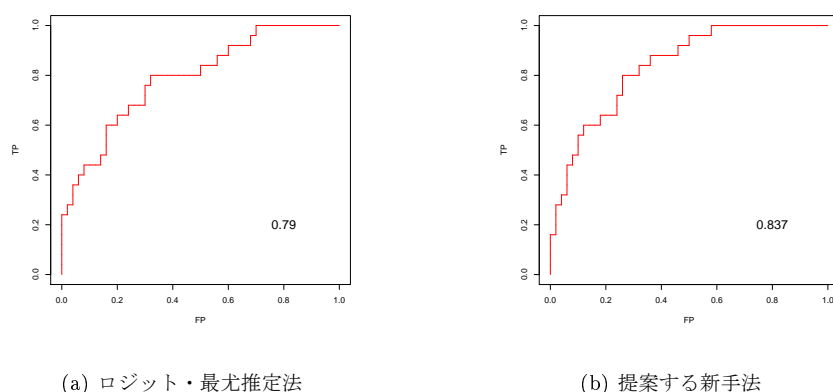


図 1: 実データを用いた分析結果

それぞれのモデルの  $\overline{AUC}$  の値は、0.790 と 0.837 となり大きな差がみられることがわかった。

## 2.2 異常値を発生させたデータを用いた両手法の $\overline{AUC}$ の比較

データ中に異常値があるときの両手法の比較を行う。今回用いたデータ中には異常値とみなせるデータが存在しなかったため、上の分析において用いたデータのデフォルト企業で最もスコアリングの悪かった企業のラベルを、非デフォルトを表わすラベルに張り替えることによって異常値を発生させた。以下では、このデータを用いて両手法のモデルによる  $\overline{AUC}$  の比較を行った。下図はそれぞれ、(左図) ロジットモデル・最尤推定法により得られる ROC 曲線と  $\overline{AUC}$  の値と (右図) 提案する新手法により得られる ROC 曲線と  $\overline{AUC}$  の値である。

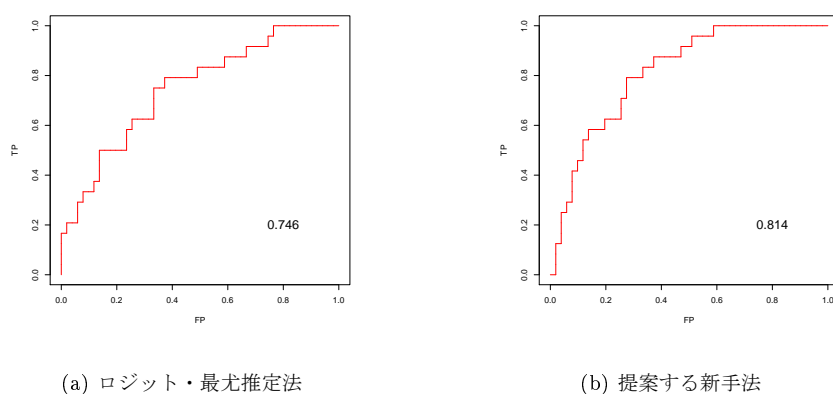


図 2: 異常値を発生させたデータを用いた分析結果

従来の手法においては  $\overline{AUC}$  の値が 0.790 から 0.746 へと 94.4%減少したのに対し、提案する新手法においては 0.837 から 0.814 へと 97.3%減少するにとどまった。このことは、提案する新手法が異常値に対してよりロバストであることを示している。また、各分析結果の (右図:b) を比較すると ROC 曲線の形状が非常によく似ていることがわかる。これは、得られるパラメータの推定値  $\hat{\beta}_\sigma$  が異常値によって変化していないことが理由である。

以上から、本研究で提案する手法により作成されるモデルは、AUC に関して最適性を有するだけでなく、従来の手法で得られるモデルよりも異常値に対してロバストであるという結果が得られた。金融の分野においては、異常値が含まれるデータが多くみられるため、異常値に対してロバストなモデルが得られる手法は、金融データにおいて非常に有効な手法である。