

2項モデルの予測による金融リスク最小化： 理論と応用

赤司 健太郎¹・川崎 能典²

(受付 2010年9月29日；改訂 2011年3月14日；採択 3月17日)

要 旨

本稿では、2項モデルで表現される保険・与信契約といった事象にかかわるリスク総額最小化問題を総利潤の側から捉え、総利潤を確率収束の意味で最大化する最適2項予測を与え、その形が簡潔な閾値モデルに帰着することを示す。現実には、不連続関数である指示関数に最尤推定値を代入して予測を行う。こうした際でも、次期に実現する最大化総利潤の予想区間ないし最小予想区間が、通常の漸近論の結果から構成されることを示す。数値実験では、我々の漸近論の結果は有限標本における近似も良好である。実証分析では、南ドイツ銀行の個人向けローンのデータを用い、総利潤の実績値を数%上回る結果が2項予測の閾値の最適化によって得られることを示す。

キーワード：最適2項予測，利潤最大化，最小予想区間，指示関数の近似。

1. はじめに

近年2項予測に関わる問題においては、経済学における消費者の選好あるいは効用最大化問題を考慮した議論がある。社会厚生関数に関わる最適2項予測の問題については、Granger and Pesaran (2000)やLieli and Nieto-Barthaburu (2010)を挙げることができる。また一方で、消費者と並んで生産者あるいは企業側にとっても2項予測は身近な問題であろう。

例えば銀行の企業への貸付や、信販会社の個人への信用供与を考えると、貸出側はまず借り手の返済能力から返済可能か否かを予測して、貸し出すか否かの意思決定を下している。あるいは医療・損害保険では、被保険者が事故に見舞われる確率と保険料率を勘案して契約が行われている。これは実務的に古くから行われているクレジットスコアリング問題である。すなわち、統計的手法を用い、新規顧客や既存顧客に関する情報を重み付けして何らかのスコアに変換し、それを組み合わせて合計スコアを算出し、このスコアの高低を支払い能力の高低を測る指標と見なすのである。この点は例えば、Crouhy et al. (2006)の第9章およびそこに紹介されている文献に詳しい。また、信用リスクにおけるデフォルト確率推定の文脈では、財務比率を用いてデフォルトを予測する手法(AltmanのZスコア法)がしばしば出発点として参照される(例えばHull, 2007, 第11章の簡潔な記述が参考になる)。

以上に鑑みれば、現実には観察される収損益のデータは、外生的な収損データを元に何らかの意味で貸し手側の最適な判断を下した結果が含まれていると解釈できよう。また、経済原理か

¹ 学習院大学 経済学部：〒171-8588 東京都豊島区目白 1-5-1

² 統計数理研究所：〒190-8562 東京都立川市緑町 10-3

らは、貸し手は利潤を最大化するように行動していると考えられ、その結果である最適な判断は内生変数である。

審査情報に重み付けを行う統計手法には、線形判別分析、ロジスティック回帰、樹形モデル、ニューラルネットワークモデルなどがあるが、現実には2項予測にしばしば用いられる最も一般的なモデルは、ロジットモデル等の確率自体をモデル化する方法である（再びCrouhy et al., 2006, の9.2節を参照）。このとき、教科書的取り扱いでは50%を判別の閾値とすることが殆どである。本研究では、企業側の利潤最大化に照らせば、閾値が利潤や損失の設定に依存して変わりうることを示すが、その形は非常に簡潔である。

2項予測に関わる利潤最大化と効用最大化問題の違いに関して述べれば、効用関数は一般に非線形でありかつ効用の値自体は意味を持たないが、利潤関数は損益に関し線形で、かつ金額で表される目的関数の値自体も意味を持つ。以下で、こうした貸出や保険モデルを標準的な計量経済モデルで記述し、顧客が十分多く大数法則が成り立つ状況を考える。換言すれば、個々の顧客のエクスポージャーが小粒であるため、一顧客のデフォルトが貸出側を揺るがすほどにはならない、という状況を考える。このとき現実問題として、未実現(例えば次期)の最大化総利潤に対して予想区間の構成を考える。特に、スコアを構成する母数の一致推定量をトレーニングサンプルあるいは前期の営業データから構成し、それを指示関数に代入することの影響を解析的に検討する。

本稿の構成は以下の通りである。第2節で2項予測による利潤最大化を定式化し、総利潤の最小予想区間を考える。第3節で総利潤関数に関する予想区間の標本版を与え、第4節で数値実験の結果を示す。第5節で実際のデータを扱う場合の注意点に言及し、南ドイツ銀行の貸出データを用い最適化の実証分析を行う。最後に第6節で結論を述べる。

2. 2項予測による総利潤最大化と最小予想区間

2.1 最適2項予測

標準的な2項モデルを考える。

$$(2.1) \quad y_i = \mathbb{1}\{\beta' \mathbf{x}_i + \epsilon_i \geq 0\}$$

ここで、 y_i は借り手あるいは被保険者 $i (i = 1, \dots, N)$ の事故発生時に1、無事故で0となる指示関数(indicator function)、 (\mathbf{x}_i, β) はそれぞれ説明変数と係数ベクトルである。 \mathbf{x}_i と誤差項 ϵ_i が独立なとき、 F で $-\epsilon_i$ の累積分布関数を表せば事故率は、

$$(2.2) \quad \Pr(y_i = 1 | \mathbf{x}_i) = F(\beta' \mathbf{x}_i)$$

となる。特に、 ϵ_i に正規分布またはロジスティック分布を仮定すると、それぞれプロビット、ロジットモデルと呼ばれ、未知係数 β の推定には最尤法が用いられる。以下では、 $(\mathbf{x}_i, \epsilon_i)$ は互いに独立 ($i = 1, \dots, N$) とするが、 \mathbf{x}_i は同一分布に従わなくてもよいとする。同時に、貸し手にとっては F などは既知で2項モデルの定式化が正しいとする。

次に、貸し手である企業側が、無事故の際に受け取る元本と利子、あるいは保険料などの利益を $r_i (> 0)$ とする。また、事故発生時の貸し倒れや事故補償などの損失額を $d_i (> 0)$ とする。そして、 z_i を貸し手が貸し出しや契約を結ぶか否かの指示関数とする(貸す時は1、その他で0)。不確実性があるので、利潤 $\pi(z_i)$ に対して、まず期待利潤の最大化問題を考え、

$$(2.3) \quad \pi(z_i) = r_i(1 - y_i)z_i - d_i y_i z_i$$

$$(2.4) \quad \max_{z_i} \mathcal{E}[\pi(z_i)] = \max_{z_i} \mathcal{E}[r_i(1 - y_i)z_i - d_i y_i z_i]$$

と定式化する。仮に0,1の値をとる指示関数としての最適解 z_i^* が存在するとしよう。このと

き、利潤総額 $\sum_{i=1}^N \pi(z_i)$ あるいは期待損失としてのリスク総額は、貸出数 N が大きいとき大数法則から次の意味で最小化されていると云える。

$$(2.5) \quad \frac{\sum_{i=1}^N \pi(z_i)}{\sum_{i=1}^N \pi(z_i^*)} \xrightarrow{p} \lim_{N \rightarrow \infty} \frac{\frac{1}{N} \sum_{i=1}^N \mathcal{E}[\pi(z_i)]}{\frac{1}{N} \sum_{i=1}^N \mathcal{E}[\pi(z_i^*)]} \leq 1$$

つまり、期待利潤の総和を最大にするように判断を下せば、実現する総利潤が最大化されている確率は 1 に収束し、最大化の意味では不確実性は無いと考えられる。その為には、各々の借り手に対して期待利潤の最適解を求める必要がある。

ところで z_i の構成についてであるが、Greene (2000) の 19 章の議論にあるように、標準的な 2 項予測は次で与えられる。

$$(2.6) \quad z_i = \mathbb{1}\{F(\beta' \mathbf{x}_i) \leq c_i\}$$

ここで、 c_i は $0 < c_i < 1$ となる閾値 (cut-off point) であり、事故率が閾値以下であれば貸し出しや契約を行うことを意味する。「どのような c_i を選べば z_i を含む損失関数が最適化されるか」という点について、ファイナンスや保険におけるリスク解析の文脈で論じた先行研究としては、Lee and Urrutia (1996) が挙げられる。ただし、そこでの研究は具体的なデータ解析問題の中で最適な閾値を実証的 (empirical) に探索するものであり、本稿のように解析的な結論を与えるものではない。以降では、最初は応用上の視点からは成立が難しいと思われるような強い仮定を置きながら解析的に導出される解を示し、その後段階的に現実的な設定下での議論へと移行してゆく。

定理 1. (A1-I) F と $(-\epsilon_i, \beta' \mathbf{x}_i)$ の密度関数 (f, g) はそれぞれ微分可能. (A1-II) $f > 0$ とする. このとき、最大解となる (2.6) 式の定数閾値 c_i^* ($i = 1, \dots, N$) は、

(i) (r_i, d_i) が定数 (r_{0i}, d_{0i}) のとき、

$$(2.7) \quad c_i^* = \frac{r_{0i}}{r_{0i} + d_{0i}}$$

(ii) (r_i, d_i) と $(\mathbf{x}'_i, \epsilon_i)$ が独立であるとき、

$$(2.8) \quad c_i^* = \frac{\mathcal{E}[r_i]}{\mathcal{E}[r_i] + \mathcal{E}[d_i]}$$

で与えられる (証明は付録を参照)。

最適な閾値は、分布関数 F の形状や平均的な事故率 $(1/N) \sum_{i=1}^N \Pr(y_i = 1)$ の大小などに依存せず簡潔な形をとる。しかしながら、解析解を得るため c_i を定数とみなすと、 (r_i, d_i) と \mathbf{x}'_i との独立性が必要で、 (r_i, d_i) を説明変数に含めることができず現実的でない。以下では、説明変数 \mathbf{x}_i の一部に (r_i, d_i) の関数が含まれることを認める。

$$(2.9) \quad (r_i, d_i) \subset \mathbf{x}'_i$$

このとき、貸し手が利用可能なすべての情報 $\mathbf{w}_i = (r_i, d_i, \mathbf{x}'_i)'$ を用いて期待利潤を最大化することを考える。定理 1 で特別な場合をまず考察したが、一般の最適解は次で与えられる。

定理 2. (A2) $\mathbf{w}_i = (r_i, d_i, \mathbf{x}'_i)'$ と ϵ_i は独立とする. このとき、最大解となる \mathbf{w}_i に関し可測な指示関数 z_i^* ($i = 1, \dots, N$) は、

$$(2.10) \quad \begin{aligned} z_i^* &= \mathbb{1}\{r_i[1 - F(\beta' \mathbf{x}_i)] \geq d_i F(\beta' \mathbf{x}_i)\} \\ &= \mathbb{1}\left\{F(\beta' \mathbf{x}_i) \leq \frac{r_i}{r_i + d_i}\right\} \end{aligned}$$

で与えられる(証明は付録を参照).

つまり, 最適解は事故率に対して $0 < r_i/(r_i + d_i) < 1$ を閾値としてもつ形に帰着する. ここで, (2.6)式の c_i を定数に制約すると定理1(ii)などの閾値と異なる場合があるが, (2.10)式の無制約の解がより大きい最大値を与える.

また, 現実には (r_i, d_i) は保険の契約や貸出などを実施する時点では所与であると考えられるが, 最適な閾値は個体間で異なる確率変数(random threshold)であり, 個別のデータに依存する. こうした最適解の導出の考え方は, Rao(1965)の7章などにみられる.

2.2 最大化総利潤の予想区間

続く論点として総利潤の予想区間の構成が考えられる. 例えば事故の有無によって期末に損益が確定する以前に, 総利潤の実現しうる何らかの範囲を予測することはリスク管理の上で重要である. 一般に総利潤は増大列となるので, 総利潤を挟み込んでいる増大列であり, y_i に依存せずに観察可能あるいは推定可能な確率変数列を考える. 具体的には最大化利潤の条件付き期待値

$$(2.11) \quad \mathcal{E}[\pi(z_i^*)|\mathbf{w}_i] = r_i(1 - F(\beta' \mathbf{x}_i))z_i^* - d_i F(\beta' \mathbf{x}_i)z_i^*$$

を用いる. このとき次の定理が従う.

定理 3. (A3-I) $0 < F < 1$ かつ $\Pr(z_i^* = 1) > \Delta_* > 0 (i = 1, \dots, N)$.

(A3-II) $\|\mathbf{x}_i\| < \Delta_x, 0 < \Delta_r < r_i < \Delta_r$ かつ $0 < \Delta_d < d_i < \Delta_d (i = 1, \dots, N)$ とする. このとき,

(i) $[l_N^*, u_N^*]$ は, 最大化総利潤の漸近的 $100(1 - \alpha)\%$ 予想区間である.

$$(2.12) \quad \Pr\left(l_N^* \leq \sum_{i=1}^N \pi(z_i^*) \leq u_N^*\right) \xrightarrow{N \rightarrow \infty} (1 - \alpha)$$

ここで, $t_{\alpha/2}$ を標準正規分布の上側 $(\alpha/2)\%$ 点として,

$$\begin{aligned} u_N^* &= \sum_{i=1}^N \mathcal{E}[\pi(z_i^*)|\mathbf{w}_i] + \sigma_N^* t_{\alpha/2} \\ &= \sum_{i=1}^N [r_i(1 - F(\beta' \mathbf{x}_i))z_i^* - d_i F(\beta' \mathbf{x}_i)z_i^*] + \sigma_N^* t_{\alpha/2} \\ l_N^* &= \sum_{i=1}^N \mathcal{E}[\pi(z_i^*)|\mathbf{w}_i] - \sigma_N^* t_{\alpha/2} \\ \sigma_N^* &= \left[\sum_{i=1}^N \text{Var}[\pi(z_i^*) - \mathcal{E}[\pi(z_i^*)|\mathbf{w}_i]] \right]^{\frac{1}{2}} \\ &= \left[\sum_{i=1}^N \mathcal{E}[(r_i + d_i)^2 F(\beta' \mathbf{x}_i)(1 - F(\beta' \mathbf{x}_i))z_i^*] \right]^{\frac{1}{2}}. \end{aligned}$$

(ii) (A3-III) h_N を $\{\mathbf{w}_1, \dots, \mathbf{w}_N\}$ の可測関数として, 次の漸近正規性

$$\Pr\left(t_l \leq \frac{1}{\sigma_N} \left[\sum_{i=1}^N (\pi(z_i^*) - \mathcal{E}[\pi(z_i^*)|\mathbf{w}_i]) - (h_N - \mathcal{E}[h_N]) \right] \leq t_u\right) \xrightarrow{N \rightarrow \infty} \Phi(t_u) - \Phi(t_l)$$

を満たす予想区間を考える. ここで, $\sigma_N^2 = \text{Var}[\sum_{i=1}^N \pi(z_i^*) - h_N]$, Φ は標準正規累積分布関数を表す. このとき, $[l_N^*, u_N^*]$ と漸近的等価な区間が, 任意の被覆係数 $(1 - \alpha)$ と h_N に対して, 最大化総利潤の最小予想区間となる(証明は付録を参照).

上述の定理では、利潤の総和を考えているのだから、まず最大化総利潤としてのリスク量を正規分布で近似することができるということを述べて、正規分布の性質から利潤の条件付き期待値が良い予測量となることを考えている。

3. β の推定と漸近論

以上の議論は、未知係数ベクトル β の真値を所与として進めてきた。むしろ現実には、最尤法などで推定しなければならない。ただし、 β の推定には $(y_i)_{i=1}^{N_1}$ の情報も必要であるが、これら $(y_i)_{i=1}^{N_1}$ が実現し事故の有無が判明する前に予測や総利潤の予想区間を構成するのが実際の関心事であるから、異なるデータから β を事前に推定する設定下で考える必要がある。例えば $(y_j, \mathbf{x}_j)_{j=1}^{N_1}$ を前期のデータとして係数の推定に用い、今期の総利潤 $\sum_{i=1}^{N_2} \pi(z_i^*)$ の予測などに用いることを考える。

Amemiya (1985) の 4 章にある正則条件 (regularity condition) のもとで、最尤推定量

$$(3.1) \quad \hat{\beta} = \operatorname{argmax}_{\beta} \sum_{j=1}^{N_1} y_j \log F(\beta' \mathbf{x}_j) + (1 - y_j) \log (1 - F(\beta' \mathbf{x}_j))$$

は、 $N_1 \rightarrow \infty$ として次の性質をもつ。

$$\sqrt{N_1}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}_1^{-1}), \quad \mathbf{V}_1^{-1} = \left[\operatorname{plim}_{N_1 \rightarrow \infty} \frac{1}{N_1} \sum_{j=1}^{N_1} \frac{[f(\beta' \mathbf{x}_j)]^2}{F(\beta' \mathbf{x}_j)[1 - F(\beta' \mathbf{x}_j)]} \mathbf{x}_j \mathbf{x}_j' \right]^{-1}$$

そして、実際の判断は推定量 $\hat{\beta}$ を用いた次の 2 項予測に基づくことになる。

$$(3.2) \quad \hat{z}_i = \mathbb{I}\{F(\hat{\beta}' \mathbf{x}_i) \leq c_i\}$$

$$(3.3) \quad \pi(\hat{z}_i) = r_i(1 - y_i)\hat{z}_i - d_i y_i \hat{z}_i$$

特に、(3.2) 式右辺の c_i を $c_i^* = r_i / (r_i + d_i)$ で与えるとき、左辺を \hat{z}_i^* と表す。

推定値で置き換えても今までの議論が成立することが望ましいが、2 値不連続関数である指示関数に推定量 $\hat{\beta}$ を代入しているので、標準的な議論と異なる点に注意する必要がある。目的は異なるが、指示関数を連続関数で近似する漸近論は Horowitz (1992) を挙げることができる。以下の定理の証明では、次の指示関数の表現を利用し漸近論を考えている。

$$(3.4) \quad \mathbb{I}\{v \leq 0\} = \lim_{b_N \rightarrow 0} K\left(\frac{v}{b_N}\right)$$

ここで、 $v \neq 0$ 及び $b_N > 0$ として、 $K(v)$ は微分可能な有界連続関数で $\lim_{v \rightarrow \infty} K(v) = 0$ 、 $\lim_{v \rightarrow -\infty} K(v) = 1$ である。注意すべきこととして、実際の推定の際にカーネル関数 K やバンド幅 b_N の指定をする必要はない。また簡単化のために、モデルのパラメータを推定するための標本と予測評価用の標本では個体の重複は無いものと仮定し、 N_1 個と N_2 個の標本は独立としよう。

$$(3.5) \quad (y_j, \mathbf{x}'_j)_{j=1}^{N_1} \perp (y_i, \mathbf{x}'_i, r_i, d_i)_{i=1}^{N_2}$$

このとき次の定理が成り立つ。

定理 4. (A4-I) 定理 3 の (A3-I) と (A3-II) を仮定する。

(A4-II) ある $\epsilon > 0$ に対し、すべての i で $\Pr(|s_{i\beta}| < \epsilon) = 0$ とする。ここで、 $s_{i\beta} = \beta' \mathbf{x}_i - F^{-1}(c_i)$ である。

(A4-III) (c_1, \dots, c_{N_2}) に対して $\lim_{N_2 \rightarrow \infty} (1/N_2) \sum_{i=1}^{N_2} \mathcal{E}[\pi(z_i)]$ が存在する. 特に,

$$(3.6) \quad \lim_{N_2 \rightarrow \infty} \frac{1}{N_2} \sum_{i=1}^{N_2} \mathcal{E}[\pi(z_i^*)] > 0.$$

(i) このとき, (A4-III) を満たす $(z_i, z_i^*)_{i=1}^{N_2}$ に対して $(N_1, N_2) \rightarrow \infty$ とすると,

$$(3.7) \quad \frac{\sum_{i=1}^{N_2} \pi(\hat{z}_i)}{\sum_{i=1}^{N_2} \pi(\hat{z}_i^*)} \xrightarrow{p} \lim_{N_2 \rightarrow \infty} \frac{\sum_{i=1}^{N_2} \mathcal{E}[\pi(z_i)]}{\sum_{i=1}^{N_2} \mathcal{E}[\pi(z_i^*)]} \leq 1.$$

さらに次を仮定する.

(A4-IV) 定理3の仮定に加えて, $N_2/N_1 \rightarrow \eta$ ($0 \leq \eta < \infty$) とする.

(ii) このとき $[\hat{l}_{N_2}^*, \hat{u}_{N_2}^*]$ は, $\hat{\beta}$ に基づく最大化総利潤の漸近的 $100(1 - \alpha)\%$ 予想区間である.

$$(3.8) \quad \Pr\left(\hat{l}_{N_2}^* \leq \sum_{i=1}^{N_2} \pi(\hat{z}_i^*) \leq \hat{u}_{N_2}^*\right) \xrightarrow{N_1, N_2 \rightarrow \infty} (1 - \alpha)$$

ここで, $t_{\alpha/2}$ を標準正規分布の上側 $(\alpha/2)\%$ 点として,

$$(3.9) \quad \hat{u}_{N_2}^* = \sum_{i=1}^{N_2} [r_i(1 - F(\hat{\beta}' \mathbf{x}_i)) \hat{z}_i^* - d_i F(\hat{\beta}' \mathbf{x}_i) \hat{z}_i^*] + \hat{\sigma}_{N_2}^{**} t_{\alpha/2}$$

$$(3.10) \quad \hat{l}_{N_2}^* = \sum_{i=1}^{N_2} [r_i(1 - F(\hat{\beta}' \mathbf{x}_i)) \hat{z}_i^* - d_i F(\hat{\beta}' \mathbf{x}_i) \hat{z}_i^*] - \hat{\sigma}_{N_2}^{**} t_{\alpha/2}$$

$$(3.11) \quad (\hat{\sigma}_{N_2}^*)^2 = \sum_{i=1}^{N_2} (r_i + d_i)^2 F(\hat{\beta}' \mathbf{x}_i) (1 - F(\hat{\beta}' \mathbf{x}_i)) \hat{z}_i^*$$

$$(3.12) \quad \hat{\sigma}_{N_2}^{**} = [(\hat{\sigma}_{N_2}^*)^2 + \hat{\mathbf{f}}'_{N_2} \hat{\mathbf{V}}_{N_1}^{-1} \hat{\mathbf{f}}_{N_2}]^{\frac{1}{2}}$$

そして,

$$(3.13) \quad \hat{\mathbf{V}}_{N_1}^{-1} = \left[\sum_{j=1}^{N_1} \frac{[f(\hat{\beta}' \mathbf{x}_j)]^2}{F(\hat{\beta}' \mathbf{x}_j) [1 - F(\hat{\beta}' \mathbf{x}_j)]} \mathbf{x}_j \mathbf{x}_j' \right]^{-1}$$

$$(3.14) \quad \hat{\mathbf{f}}_{N_2} = \sum_{i=1}^{N_2} (r_i + d_i) \hat{z}_i^* f(\hat{\beta}' \mathbf{x}_i) \mathbf{x}_i$$

である (証明は付録を参照).

仮定 (A4-II) については若干のコメントが必要であろう. これは真値 β を所与としたとき, 確率 $F(\beta' \mathbf{x}_i)$ が閾値 c_i の近傍に入らないことを要請しており, これによって定理の証明で $K(v)$ の原点での近似に関する議論が簡略化されている. この仮定が現実的に妥当性を持つことは, 例えば説明変数 \mathbf{x}_i と (r_i, d_i) がすべて離散型確率変数である状況を想定するとよい. このとき $\beta' \mathbf{x}_i$ は, 取り得る有限個の値の上に確率質点に乗っているという意味で離散変数となり, $F(\beta' \mathbf{x}_i)$ の各点が $r_i/(r_i + d_i)$ の近傍に集中する可能性は, 実用上無視できると考える.

もちろん, 実際の統計的モデリングで連続型変数が混入することは十分考えられるが, そのような説明変数にも測定の最小単位はあり, それが事実上の bin 幅であるような離散変数と見なすことが可能である. 仮にこの見方に無理があると感じられるほど bin が非常に小さな値であったとしても, 応答変数の予測の観点からは, 隣接する bin の併合を繰り返して少数のカテゴリ値しか持たない離散変数を生成した方が, 効率の良いモデリングにつながる場合が少なくない (そのような方法論と実例は, 坂元, 1985 を参照されたい).

定理 4 の結果は、まず \hat{z}_i^* に基づく判断でも確率収束の意味で総利潤は最大化されることを述べている。次に予想区間に関しては、 β の推定に伴う影響は指示関数への代入ではなく、 β の連続関数 $-(r_i + d_i)F(\beta'x)$ を経由する推定誤差としてのみ(3.12)式の第 2 項に現われ、予想区間の幅がその分広がる。これは、指示関数の傾きが平らであることに起因すると考えられる。しかしながら、

$$(3.15) \quad \frac{1}{N_2}(\hat{\sigma}_{N_2}^{**})^2 = \frac{1}{N_2}(\hat{\sigma}_{N_2}^*)^2 + \left(\frac{N_2}{N_1}\right) \frac{1}{N_2} \hat{f}'_{N_2} \left[\frac{1}{N_1} \hat{V}_{N_1} \right]^{-1} \frac{1}{N_2} \hat{f}_{N_2}$$

の関係から、 $N_2/N_1 \rightarrow 0$ と十分はやく β が推定されるならば推定誤差を無視できて、定理 3 (ii) と同じく最大化総利潤の条件付き期待値による予想区間が最小である、と云うことができる。

4. 数値実験

この節では、簡単なモンテカルロ実験の結果を述べる。疑似乱数の生成過程は、 $N_1 = N_2 = 10000$ としてロジットモデルを考えた。

$$(4.1) \quad F(\beta'x_i) = \frac{\exp(\beta_1 + \beta_2 x_i)}{1 + \exp(\beta_1 + \beta_2 x_i)}$$

ここで、 $\beta_2 = -1$ であり、 x_i は自由度 15 の χ^2 分布から生成した。一般に、債務不履行や保険の支払いは稀な事象と考えられるので、事故率 $(1/N_2) \sum_{i=1}^{N_2} y_i$ もそれに対応して低くなるように β_1 を調整し、1% ($\beta_1 = 4.2$) と 3% ($\beta_1 = 5.8$) の 2 通りを考えた。

利得と損失は、 i ($i=1, \dots, N_2$) に対して共通で $r_i = 0.1$ と $d_i = 2.0$ とした。 $r_i < d_i$ なので、保険料率が r_i で医療費支払いが d_i であるような保険の事例などが想定される。このとき最適な閾値は、 $c_i^* = 0.1/(0.1 + 2.0) \simeq 0.05$ である。比較のため閾値 $c_i = 0.5$ と、より慎重に保険契約を結ぶ $c_i = 0.01$ を考えた。

繰り返し数 3000 に対する平均値と割合を表 1 にまとめた。ここで丸括弧内の数値は真値 $\beta = (\beta_1, \beta_2)'$ を所与とした結果を示す。表 1 の上段には、(3.7) 式の逆数に基づく

$$(4.2) \quad \left[\frac{\sum_{i=1}^{N_2} \pi(\hat{z}_i^*)}{\sum_{i=1}^{N_2} \pi(\hat{z}_i)} - 1 \right] \times 100$$

を報告しており、これにより最適化による総利潤上昇率(%)を観察している。 $c_i = 0.5$ の場合は c_i^* からの乖離が大きいため総利潤の上昇率が大きい傾向がある。同時に事故率 $(1/N_2) \sum_{i=1}^{N_2} y_i$ が高いほどその上昇率が大きい。表の最大化確率とは、繰り返し数 3000 のうちで(4.2)式が正となる割合で、殆どの場合で 100%となっている。

表 1. 最適な閾値 c_i^* との比較(上段)と最大化総利潤の予想区間(下段)。推定値 $\hat{\beta}$ に基づく結果。このとき $N_2/N_1 = 1$ 。丸括弧内は真値 β を所与とした結果を表す。

| | 事故率 1% | | 事故率 3% | |
|------------------|-------------|--------------|-------------|--------------|
| 閾値 c_i | 0.01 | 0.5 | 0.01 | 0.5 |
| 総利潤上昇率 [%] | 3.4 (3.8) | 12.4 (12.0) | 5.6 (6.1) | 45.2 (43.7) |
| 最大化確率 [%] | 99.8 (100) | 100 (100) | 100 (100) | 100 (100) |
| | 事故率 1% | | 事故率 3% | |
| 最大化総利潤の 95% 予想区間 | [885, 934] | ([890, 927]) | [799, 855] | ([805, 849]) |
| 最大化総利潤の 95% 被覆係数 | 94.9 (94.7) | | 94.5 (94.5) | |

次に表1下段は、定理3(i)と定理4(ii)による最大化総利潤(c_i^* の場合)の予想区間に関する結果である。95%予想区間は事故率が低いほど狭く、事故率が低い方がその水準自体も高くなっている。括弧内は真値 β を所与とした結果なので、定理4に従い推定値を用いるといずれの事故率の下でも約30%予想区間の幅を広げなければならない。繰り返し数3000のうち予想区間に入った割合(%)を示す被覆係数については、(3.12)式の第2項を無視すると予想区間が狭くなり、実際には80%台(表中未報告)になってしまう。一方その第2項を追加することで、名目の95%被覆係数に対して表1の実現値は95%に近くなっており、この場合漸近論の近似は有限標本の下でも良好と云えよう。

5. 実証分析

5.1 実データでの留意点

データ解析に先立ち、前節で観察した数値実験の結果と現実のデータを扱う場合との関連を以下で考察しよう。現実を観察されるデータに対する注意点としては、ある会社側の判断 z_i^+ で契約を結んでもらえた個人($z_i^+ = 1$)の標本しか収集されていない可能性が高い、という点が挙げられる。これは一見すると、ランダムに標本を抽出しているのではなく、むしろ債務不履行に陥りにくい人ばかりを集めて契約していることに起因する標本の偏り(いわゆる選択バイアスの問題)があると感じられるかもしれないが、我々のモデル設定では分析の障害とはならない。以下そのことを確認しよう。まず、このとき最尤推定量は、

$$(5.1) \quad \hat{\beta}^+ = \operatorname{argmax}_{\beta} \sum_{j=1}^{N_1} z_j^+ y_j \log F(\beta' \mathbf{x}_j) + z_j^+ (1 - y_j) \log (1 - F(\beta' \mathbf{x}_j))$$

となることに注意する。ここで、 z_j^+ は \mathbf{x}_j に依存していると想定されるが、例えば事故の有無以前に契約を結ぶといった実態に照らして考えれば、被説明変数 y_j に依存するとは考えられないので、 β の推定に偏りは生じない。Amemiya(1985)の9章5節に従えば最尤法における外生的標本抽出(exogenous sampling)の一例である。ただし、標本の観測の制限により次のような問題は存在する。本来の最大化利潤と、ある判断 z_i^+ による利潤の無条件期待値は次のように分解される。

$$(5.2) \quad \mathcal{E}[\pi(z_i^*)] = \mathcal{E}[\pi(z_i^*)|z_i^+ = 1] \Pr(z_i^+ = 1) + \mathcal{E}[\pi(z_i^*)|z_i^+ = 0][1 - \Pr(z_i^+ = 1)]$$

$$(5.3) \quad \mathcal{E}[\pi(z_i^+)] = \mathcal{E}[\pi(z_i^+)|z_i^+ = 1] \Pr(z_i^+ = 1) + \mathcal{E}[\pi(z_i^+)|z_i^+ = 0][1 - \Pr(z_i^+ = 1)]$$

契約したデータしかないとすれば第2項同士の差分、つまり、「(i)もし契約していたら利潤に貢献していたかもしれない分」は、実際には契約していないことから個人の属性データを追跡していないため、測ることはできない。しかし、「(ii)現実には契約したがもし契約しなければ利潤を上げていたかもしれない分」については計測できる。より正確に述べると、($z_i^+ = 1, y_i = 1$)に対し $z_i^* = 0$ と判断することによる利潤上昇分が、($z_i^+ = 1, y_i = 0$)に対し $z_i^* = 0$ と判断したことによる利潤機会の逸失を上回る分が、利潤の上昇や、あるいは損失を未然に防いだことに対応していると解釈できる。これはある契約を所与 $\Pr(z_i^+ = 1) = 1$ とした条件付き期待値の第一項同士の差分で、

$$(5.4) \quad \mathcal{E}[\pi(z_i^*)|z_i^+ = 1] - \mathcal{E}[\pi(z_i^+)|z_i^+ = 1] \geq 0$$

の成立を意味する。ここで(5.4)式が非負であるのは、判断 z_i^+ も w_i の関数とすると、

$$(5.5) \quad z_i^* = \operatorname{argmax}_{z_i} \mathcal{E}[\pi(z_i)] = \operatorname{argmax}_{z_i} \mathcal{E}[\pi(z_i)|z_i^+ = 1]$$

となることによる。つまり、 $z_i^+ = 1$ を条件付けても利潤最大化のための 2 項予測 z_i^* の形は変わらない。

前節の数値実験との関連で述べれば、本小節でいうケース (ii) の可能性は実験での $c_i = 0.5$ の場合に相当し、すべての i において定数閾値で常に $0.05 \simeq c_i^* < c_i = 0.5$ である。よって、最適化により利潤が上昇した理由は、 $c_i^* < \Pr(\hat{\beta}' x_i) \leq c_i$ の個人 i に対しては契約せずに損失を抑えたことにある。

5.2 南ドイツ銀行の貸出データ

以下では、Fahrmeir and Tutz (1994) で分析されているドイツ南部のある銀行(特定化を避けている)の個人向けローンのデータを用いて、貸出モデルの利潤最大化に関する検証結果を示す。ここで、当行を「南ドイツ銀行」と仮に呼ぶことにしよう。なお、用いたデータはミュンヘン大学統計学部のウェブサイト (http://www.stat.uni-muenchen.de/service/datenarchiv/kredit/kredit_e.html) から入手可能である。南ドイツ銀行が実際に貸し出した 1000 人 ($\sum_{i=1}^N z_i^+ = 1000$) の内で 300 人が返済できていない。やや注意を要するのは、この標本が Amemiya (1985, 9 章) や Fahrmeir and Tutz (1994, p. 34) にあるように内生的標本抽出 (endogenous sampling) から得られたと考えられる点である。しかし、真の比率 $\Pr(y_i = 1)$ を推定するための情報が無いことと、議論が複雑になるので、本稿ではそのまま無作為標本として分析を行った。まず、定理 4 の想定通りに、 β^+ の事前の推定のため無作為に標本分割した ($N_1 = 498, N_2 = 502$)。ここでは、全データに一様乱数を付し、その値が 0.5 以下なら標本 N_1 に含めた。2 項モデルに関しては、近年セミパラメトリックな手法が発展を見せているが、ここでは定式化の誤りの問題には踏み込まずに代表的で扱いやすいロジットモデルを仮定し、説明変数の選択については同じデータを用いた分析 Fahrmeir and Tutz (1994, p. 31) に依拠した。

- $x_{i1[1]}$: 0 medium running account, 1 good running account
- $x_{i1[2]}$: 0 medium running account, 1 no account
- x_{i3} : duration of loan in months
- x_{i4} : the size of loan in DM/1000
- x_{i5} : 0 previously bad payer, 1 previously good payer
- x_{i6} : 0 professional, 1 private
- x_{i7} : 0 male, 1 female
- x_{i8} : 0 living alone, 1 not living alone

ここで、 x_{i3} と x_{i4} 以外はダミー変数である。分析に必要な変数として $d_i = x_{i4}$ とし、全額貸倒が起きる場合を考えた。 r_i に関する情報は得られないので、各個人で共通のローン利率を幾通りか仮定した。つまり、返済額は利率 $\rho = (.05, \dots, .30)$ として $r_i = (1 + \rho)d_i$ である。このとき、 (d_i, r_i) の単位はドイツマルク/1000 で、最適な閾値は、

$$(5.6) \quad c_i^*(\rho) = \frac{r_i}{r_i + d_i} = \frac{1 + \rho}{2 + \rho}$$

となる。標本数 N_1 ($\sum_{j=1}^{N_1} y_j = 152$) を用いて β^+ を推定した結果が表 2 である。

これより各 ρ の想定に応じて最適 2 項予測 $\hat{z}_i^* = \mathbb{1}\{F(\mathbf{x}_i' \hat{\beta}^+) \leq c_i^*(\rho)\}$ を構成し、最大化総利潤と南ドイツ銀行の実際の総利潤

$$(5.7) \quad \sum_{i=1}^{N_2} [r_i(1 - y_i)\hat{z}_i^* - d_i y_i \hat{z}_i^*] z_i^+, \quad \sum_{i=1}^{N_2} [r_i(1 - y_i)z_i^+ - d_i y_i z_i^+] z_i^+$$

を、標本数 N_2 ($\sum_{i=1}^{N_2} y_i = 148$) を用いて比較した。ここで、(5.7) 式において $z_i^+ = 1, r_i = (1 + \rho)d_i$ である。

表 2. 最尤推定量 $\hat{\beta}^+$ の結果.

| | 定数 | $x_{1[1]}$ | $x_{1[2]}$ | x_3 | x_4 |
|----------------|--------|------------|------------|--------|--------|
| 推定値 | -0.45 | -0.62 | -2.1 | 0.06 | -0.04 |
| (<i>t</i> -値) | (0.89) | (2.44) | (4.86) | (6.92) | (0.69) |
| | | x_5 | x_6 | x_7 | x_8 |
| 推定値 | | -0.63 | -0.53 | 0.64 | -0.22 |
| (<i>t</i> -値) | | (1.74) | (2.30) | (2.15) | (0.75) |

表 3. 最大化総利潤と南ドイツ銀行の総利潤との比較.

| 想定ローン利率 ρ | .05 | .10 | .15 | .20 | .25 | .30 |
|----------------|------|-----|-----|------|------|------|
| 総利潤上昇率 [%] | 10.6 | 6.6 | 6.3 | 4.4 | 1.7 | 3.4 |
| 最大化総利潤 | 569 | 605 | 660 | 704 | 739 | 807 |
| 95%予想区間下限 | 515 | 553 | 589 | 629 | 670 | 709 |
| 95%予想区間上限 | 860 | 913 | 969 | 1021 | 1074 | 1130 |

各ローン利率 ρ に対して, (4.2) 式に対応する利潤上昇率と最大化総利潤, ならびに定理 4 に基づく最大化総利潤の 95% 予想区間を表 3 にまとめた.

表 3 からは, 利率 ρ の設定によらず, 最適な 2 項判断によって総利潤の上昇が数% あったことが観察できる. また, すべての場合で最大化利潤は 95% 予想区間に入っている. ここで, (5.6) 式の $c_i^*(\rho)$ は ρ の増加関数なので, ρ を上げると $z_i^* = 1$ の数も増えて南ドイツ銀行の判断 (すべてで $z_i^* = 1$) に近づき, 両者の総利潤の値も近づく. しかし, (5.7) 式において $r_i = (1 + d_i)\rho$ も ρ の増加関数なので, 総利潤最大化のための最適予測が当たると利潤の上げ幅も大きくなり, 結果として表 3 の利潤上昇率は単調には減少しないと解釈できる. また, (5.6) 式での最適な閾値の 1% や 2% の若干の変化で, 利潤上昇率が変動する点も興味深い.

6. 結論

本稿では, 保険・与信契約といった事象への 2 項モデリングを想定し, リスクと表裏一体の目的関数として総利潤を考えた. これを最大化する最適 2 項予測は非常に簡潔であり, 現実問題へ応用することも容易である. 技術的な問題として, 実際には不連続関数である指示関数に最尤推定値を代入しなければならないが, この際も次期に実現する最大化総利潤の予想区間ないし最小予想区間が構成されることを示した. 数値実験では我々の漸近論の結果は有限標本における近似も良好であった. 実証分析では, 南ドイツ銀行の個人向けローンのデータを用い, 最適化により同行の実績値を数% 上回る結果を示した. 我々の議論が妥当性を持つ範囲では, 最適な 2 項予測による最大化総利潤と現実の総利潤との乖離を分析することが可能となる.

付 録

定理 1 の証明. はじめに (r_i, d_i) が定数 (r_{0i}, d_{0i}) であるときを考える. $s_i = \beta' x_i$, F の逆関数を F^{-1} として $c_i^- = F^{-1}(c_i)$ と表すと,

$$(A.1) \quad \mathcal{E}[(1 - y_i)z_i] = \Pr(s_i + \epsilon_i < 0, F(s_i) \leq c_i)$$

$$\begin{aligned}
&= \int_{-\infty}^{c_i^-} \int_s^{\infty} f(\epsilon)g(s)d\epsilon ds \\
&= \int_{-\infty}^{c_i^-} [1 - F(s)]g(s)ds
\end{aligned}$$

$$\begin{aligned}
(A.2) \quad \mathcal{E}[y_i z_i] &= \Pr(s_i + \epsilon_i \geq 0, F(s_i) \leq c_i) \\
&= \int_{-\infty}^{c_i^-} \int_{-\infty}^s f(\epsilon)g(s)d\epsilon ds \\
&= \int_{-\infty}^{c_i^-} F(s)g(s)ds.
\end{aligned}$$

したがって 1 階条件は、 $dc_i^-/dc_i = 1/f(c_i)$ を用いて、

$$(A.3) \quad \frac{d}{dc_i} \mathcal{E}[r_{0i}(1 - y_i)z_i - d_{0i}y_i z_i] = r_{0i}(1 - F(c_i^-)) \frac{g(c_i^-)}{f(c_i)} - d_{0i}F(c_i^-) \frac{g(c_i^-)}{f(c_i)} = 0$$

$F(c_i^-) = c_i$ なので、 $c_i^* = r_{0i}/(r_{0i} + d_{0i})$ を得る。2 階微分は、

$$\begin{aligned}
(A.4) \quad \frac{d^2}{(dc_i)^2} \mathcal{E}[r_{0i}(1 - y_i)z_i - d_{0i}y_i z_i] \\
&= \frac{d}{dc_i} [r_{0i}(1 - F(c_i^-)) - d_{0i}F(c_i^-)] \left[\frac{g(c_i^-)}{f(c_i)} \right] \\
&= -(r_{0i} + d_{0i}) \left[\frac{g(c_i^-)}{f(c_i)} \right] + [r_{0i}F(c_i^-) - d_{0i}(1 - F(c_i^-))] \left[\frac{d}{dc_i} \frac{g(c_i^-)}{f(c_i)} \right].
\end{aligned}$$

c_i^* で評価すると第 2 項は 0、第 1 項は常に負なので、 c_i^* は最大解である。

(r_i, d_i) が確率変数のときも条件(ii)の下では、 $\mathcal{E}[r_i(1 - y_i)z_i - d_i y_i z_i] = \mathcal{E}[\mathcal{E}[r_i(1 - y_i)z_i - \mathcal{E}[d_i]y_i z_i]]$ なので、同じ議論が成り立つ。□

定理 2 の証明. 次の分解に注意する。

$$(A.5) \quad \mathcal{E}[r_i(1 - y_i)z_i - d_i y_i z_i] = \mathcal{E}[r_i(1 - y_i)z_i + d_i y_i(1 - z_i)] - \mathcal{E}[d_i y_i]$$

よって、第 1 項を最大にするように z_i を選ばばよい。第 1 項は、 $\mathbf{w}_i = (r_i, d_i, \mathbf{x}_i)'$ として次で表される。

$$\begin{aligned}
(A.6) \quad \mathcal{E}[\mathcal{E}[r_i(1 - y_i)z_i + d_i y_i(1 - z_i)|\mathbf{w}_i]] \\
= \mathcal{E}[r_i[1 - \Pr(y_i = 1|\mathbf{w}_i)]z_i + d_i \Pr(y_i = 1|\mathbf{w}_i)(1 - z_i)]
\end{aligned}$$

ここで、 $\mathcal{E}[y_i|\mathbf{w}_i] = \Pr(y_i = 1|\mathbf{w}_i) = F(\beta' \mathbf{x}_i)$ である。 z_i と $(1 - z_i)$ は排反なので、 $\{r_i[1 - \Pr(y_i = 1|\mathbf{w}_i)] \geq d_i \Pr(y_i = 1|\mathbf{w}_i)\}$ のときに $z_i^* = 1$ とすれば条件付き期待値は最大値をとる、つまり任意の z_i に対して、

$$(A.7) \quad \mathcal{E}[r_i(1 - y_i)z_i + d_i y_i(1 - z_i)|\mathbf{w}_i] \leq \mathcal{E}[r_i(1 - y_i)z_i^* + d_i y_i(1 - z_i^*)|\mathbf{w}_i].$$

したがって期待値の単調性より、無条件期待値も最大値をとる。□

定理 3 の証明. はじめに $[l_N^*, u_N^*]$ が漸近的な $100(1 - \alpha)\%$ 予想区間であることを示す。以下では $\pi_i^* = \pi(z_i^*)$ と表記する。

まず、 $\sigma_N^{*2} = \sum_{i=1}^N \text{Var}[\pi_i^* - \mathcal{E}[\pi_i^*|\mathbf{w}_i]]$ であるが、これを評価する。 $\mathcal{E}[\pi_i^*] = \mathcal{E}[\mathcal{E}[\pi_i^*|\mathbf{w}_i]]$ なので、 $\text{Var}[\pi_i^* - \mathcal{E}[\pi_i^*|\mathbf{w}_i]] = \mathcal{E}[(\pi_i^* - \mathcal{E}[\pi_i^*|\mathbf{w}_i])^2] = \mathcal{E}[\text{Var}[\pi_i^*|\mathbf{w}_i]]$ より、

$$\begin{aligned} & \mathcal{E}[\text{Var}[\pi_i^*|\mathbf{w}_i]] \\ &= \mathcal{E}[r_i^2 z_i^{*2} \text{Var}[y_i|\mathbf{w}_i] + d_i^2 z_i^{*2} \text{Var}[(1-y_i)|\mathbf{w}_i] - 2r_i d_i z_i^{*2} \text{Cov}[y_i, (1-y_i)|\mathbf{w}_i]]. \end{aligned}$$

ここで, $z_i^{*2} = z_i^*$, $\text{Var}[y_i|\mathbf{w}_i] = F(\beta' \mathbf{x}_i)(1 - F(\beta' \mathbf{x}_i))$, $\text{Cov}[y_i, (1-y_i)|\mathbf{w}_i] = -\text{Var}[y_i|\mathbf{w}_i]$ を用いて, 定理の表現となる.

総利潤の漸近正規性を考える. 仮定 (A3-I) と (A3-II) より, 任意の i に対しある Δ_1 が存在して,

$$(A.8) \quad \text{Var}[\pi_i^* - \mathcal{E}[\pi_i^*|\mathbf{w}_i]] = \mathcal{E}[(r_i + d_i)^2 F(\beta' \mathbf{x}_i)(1 - F(\beta' \mathbf{x}_i)) |z_i^* = 1] \text{Pr}(z_i^* = 1) \geq \Delta_1 > 0.$$

また, 仮定 (A3-II) よりある Δ_2 に対し,

$$(A.9) \quad \begin{aligned} \mathcal{E}[|\pi_i^* - \mathcal{E}[\pi_i^*|\mathbf{x}_i]|^3] &= \mathcal{E}[|-(d_i + r_i)|^3 |y_i - F(\beta' \mathbf{w}_i)|^3 z_i^{*3}] \\ &\leq \mathcal{E}[(r_i + d_i)^3] < \Delta_2. \end{aligned}$$

$i (i=1, \dots, N)$ は互いに独立で, 上の議論から次のリアプノフ条件

$$(A.10) \quad \sum_{i=1}^N \mathcal{E}[|\pi_i^* - \mathcal{E}[\pi_i^*|\mathbf{w}_i]|^3] = o\left(\left[\sum_{i=1}^N \text{Var}[\pi_i^* - \mathcal{E}[\pi_i^*|\mathbf{w}_i]]\right]^{\frac{3}{2}}\right)$$

が成り立つ. よって中心極限定理から,

$$(A.11) \quad \frac{1}{\sigma_N^*} \sum_{i=1}^N (\pi_i^* - \mathcal{E}[\pi_i^*|\mathbf{w}_i]) \xrightarrow{d} \mathcal{N}(0, 1).$$

したがって,

$$(A.12) \quad \begin{aligned} & \text{Pr}\left(l_N^* \leq \sum_{i=1}^N \pi_i^* \leq u_N^*\right) \\ &= \text{Pr}\left(-t_{\alpha/2} \leq \frac{1}{\sigma_N^*} \sum_{i=1}^N (\pi_i^* - \mathcal{E}[\pi_i^*|\mathbf{w}_i]) \leq t_{\alpha/2}\right) \xrightarrow{N \rightarrow \infty} (1 - \alpha). \end{aligned}$$

最小予想区間であることについて,

$$\begin{aligned} & \lim_{N \rightarrow \infty} \text{Pr}\left((h_N - \mathcal{E}[h_N]) + \sum_{i=1}^N \mathcal{E}[\pi_i^*] + \sigma_N t_l \leq \sum_{i=1}^N \pi_i^* \leq (h_N - \mathcal{E}[h_N]) + \sum_{i=1}^N \mathcal{E}[\pi_i^*] + \sigma_N t_u\right) \\ &= \Phi(t_u) - \Phi(t_l) = (1 - \alpha). \end{aligned}$$

このとき予想区間の幅は $(t_u - t_l)\sigma_N$ である. 標準正規密度関数は原点对称で単峰形なので, 任意の α に対して $(t_u - t_l) = t_{\alpha/2} - (-t_{\alpha/2})$ で最小となる. σ_N の最小値に関しては,

$$(A.13) \quad \sigma_N^2 = \mathcal{E}\left[\left(\sum_{i=1}^N (\pi_i^* - \mathcal{E}[\pi_i^*]) - (h_N - \mathcal{E}[h_N])\right)^2\right].$$

この平均 2 乗誤差を最小にする $\{\mathbf{w}_1, \dots, \mathbf{w}_N\}$ の可測関数 $(h_N - \mathcal{E}[h_N])$ は, 条件付期待値 $\mathcal{E}[\sum_{i=1}^N (\pi_i^* - \mathcal{E}[\pi_i^*])|\mathbf{w}_1, \dots, \mathbf{w}_N] = \sum_{i=1}^N (\mathcal{E}[\pi_i^*|\mathbf{w}_i] - \mathcal{E}[\pi_i^*])$ であり, 任意の N で $\sigma_N^* \leq \sigma_N$. よって, $[l_N^*, u_N^*]$ と漸近的に等価な予想区間が最小区間をもつ. \square

定理 4 の証明. はじめに次の補題を与える.

補題. $s_{i\beta} = \beta' \mathbf{x}_i - F^{-1}(c_i)$, $K(v) = 1/[1 + \exp(v)]$, $b_{N_1} = 1/\sqrt{N_1}$ とする. このとき仮定 (A4-I) と (A4-II) の下, すべての $i \in N_2$ で N_1 が十分大きいとき,

$$(A.14) \quad \mathbb{1}\{s_{i\hat{\beta}} \leq 0\} = K\left(\frac{s_{i\hat{\beta}}}{b_{N_1}}\right) + o_p\left(\frac{1}{\sqrt{N_1}}\right)$$

$$(A.15) \quad = K\left(\frac{s_{i\beta}}{b_{N_1}}\right) + o_p\left(\frac{1}{\sqrt{N_1}}\right)$$

$$(A.16) \quad = \mathbb{1}\{s_{i\beta} \leq 0\} + o_p\left(\frac{1}{\sqrt{N_1}}\right).$$

補題の証明. まず (A.14) 式について, K の対称性から, 任意の ϵ_1 に対して次を考える.

$$(A.17) \quad \Pr\left(\sqrt{N_1} \left| \mathbb{1}\{s_{i\hat{\beta}} \leq 0\} - K\left(\frac{s_{i\hat{\beta}}}{b_{N_1}}\right) \right| \leq \epsilon_1\right) = \Pr\left(\sqrt{N_1} K\left(\frac{|s_{i\hat{\beta}}|}{b_{N_1}}\right) \leq \epsilon_1\right) \xrightarrow{N_1 \rightarrow \infty} 1$$

つまり, 次を云いたい.

$$(A.18) \quad \sqrt{N_1} K\left(\frac{|s_{i\hat{\beta}}|}{b_{N_1}}\right) = \frac{1}{1/\sqrt{N_1} + \exp\{\sqrt{N_1}(|s_{i\hat{\beta}}| - \log \sqrt{N_1}/\sqrt{N_1})\}} \xrightarrow{p} 0$$

N_1 が十分大きいとき $\log \sqrt{N_1}/\sqrt{N_1} \rightarrow 0$ で, また $\hat{\beta}$ の一貫性と仮定 (A4-II) より, i について一様に $|s_{i\hat{\beta}}| \xrightarrow{p} |s_{i\beta}| > \epsilon > 0$ となる. よって, (A.18) 式が成り立つ.

次に (A.15) 式について, β に関するテーラー展開から次を云えば良い.

$$(A.19) \quad \sqrt{N_1} \left[K\left(\frac{s_{i\hat{\beta}}}{b_{N_1}}\right) - K\left(\frac{s_{i\beta}}{b_{N_1}}\right) \right] = \left[\sqrt{N_1} K'\left(\frac{s_{i\hat{\beta}}}{b_{N_1}}\right) \right] \sqrt{N_1} (\hat{\beta} - \beta)' \mathbf{x}_i \\ = o_p(1) \times O_p(1)$$

ここで, $K'(v) = -\exp(v)/[1 + \exp(v)]^2$, $\bar{\beta} = \beta + \lambda_{N_1}(\hat{\beta} - \beta)$ ($0 < \lambda_{N_1} < 1$), (A4-I) より $\|\sqrt{N_1}(\hat{\beta} - \beta)\| \max_{1 \leq i \leq N_2} \|\mathbf{x}_i\| = O_p(1)$ である. ここで, K' も裾がはやく 0 に減衰し原点对称であるから, (A.18) 式と同様の議論で, (A.19) 式が成り立つ.

最後に (A.16) 式は, (A.17) 式で $\hat{\beta}$ を真値に置き換えて成立する. \square

定理 4 の最大化に関する確率収束については, 補題の結果から,

$$(A.20) \quad \frac{1}{N_2} \sum_{i=1}^{N_2} \pi(\hat{z}_i) - \frac{1}{N_2} \sum_{i=1}^{N_2} \pi(z_i) = o_p\left(\frac{1}{\sqrt{N_1}}\right).$$

続いて, 予想区間に関して漸近正規性を考える.

$$(A.21) \quad \frac{1}{\sqrt{N_2}} \sum_{i=1}^{N_2} \pi(\hat{z}_i^*) - [r_i(1 - F(\hat{\beta}' \mathbf{x}_i))\hat{z}_i^* - d_i F(\hat{\beta}' \mathbf{x}_i)\hat{z}_i^*] \\ = \frac{1}{\sqrt{N_2}} \sum_{i=1}^{N_2} \pi(z_i^*) - [r_i(1 - F(\hat{\beta}' \mathbf{x}_i))z_i^* - d_i F(\hat{\beta}' \mathbf{x}_i)z_i^*] + o_p(1) \\ = \frac{1}{\sqrt{N_2}} \sum_{i=1}^{N_2} \pi(z_i^*) - \mathcal{E}[\pi(z_i^*) | \mathbf{w}_i] \\ + \left[\frac{\sqrt{\eta}}{N_2} \sum_{i=1}^{N_2} (r_i + d_i) z_i^* f(\beta' \mathbf{x}_i) \mathbf{x}_i' \right] \sqrt{N_1} (\hat{\beta} - \beta) + o_p(1)$$

1 番目の等号については、補題の結果から次が成り立つことによる。

$$\frac{1}{\sqrt{N_2}} \sum_{i=1}^{N_2} o_p \left(\frac{(r_i + d_i) |F(\hat{\beta}' \mathbf{x}_i) - y_i|}{\sqrt{N_1}} \right) = o_p(1)$$

(A.21) 式の 2 番目の等号に関しては、 $F(\hat{\beta}' \mathbf{x}_i)$ に対するテーラー展開を用いている。

最後に、以上と同様な議論から、

$$\begin{aligned} \frac{1}{N_1} \sum_{j=1}^{N_1} \frac{[f(\hat{\beta}' \mathbf{x}_j)]^2}{F(\hat{\beta}' \mathbf{x}_j)[1 - F(\hat{\beta}' \mathbf{x}_j)]} \mathbf{x}_j \mathbf{x}_j' &= \text{plim}_{N_1 \rightarrow \infty} \frac{1}{N_1} \sum_{j=1}^{N_1} \frac{[f(\beta' \mathbf{x}_j)]^2}{F(\beta' \mathbf{x}_j)[1 - F(\beta' \mathbf{x}_j)]} \mathbf{x}_j \mathbf{x}_j' + o_p(1) \\ \frac{1}{N_2} \sum_{i=1}^{N_2} (r_i + d_i) \hat{z}_i^* f(\hat{\beta}' \mathbf{x}_i) \mathbf{x}_i' &= \text{plim}_{N_2 \rightarrow \infty} \frac{1}{N_2} \sum_{i=1}^{N_2} (r_i + d_i) z_i^* f(\beta' \mathbf{x}_i) \mathbf{x}_i' + o_p(1) \end{aligned}$$

が云えるので、定理 3 の結果と合わせて (A.21) 式の漸近正規性が成り立つ。□

謝 辞

平成 21 年度研究報告会(統計数理研究所)に於いて国友直人氏(東京大学大学院経済学研究科)、山下智志氏(統計数理研究所)から、国友教授のご還暦記念研究発表会に於いては加藤賢悟氏(広島大学大学院理学研究科)から、ご助言を頂戴した。関西計量経済学研究会の参加者の方々から貴重なコメントを頂いた。また、匿名の 2 名のレフェリーの方々から有益なコメントと初稿における証明の誤りをご指摘頂いたことをここに記して感謝申し上げる。なお、この研究は統計数理研究所 共同研究課題(22-共研-2043)に基づくと同時に、科研費(09005898, 09019321, 10011616, 10001711)からの助成および情報・システム研究機構 新領域融合研究プロジェクト「コミュニケーション情報学」からの助成によっている。

参 考 文 献

- Amemiya, T. (1985). *Advanced Econometrics*, Basil Blackwell, Oxford and New York.
- Crouhy, M., Gala, D. and Mark, R. (2006). *The Essentials of Risk Management*, McGraw-Hill, New York. (三浦良造 訳者代表(2008). 『リスクマネジメントの本質』, 共立出版, 東京.)
- Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer, New York.
- Granger, C. W. J. and Pesaran, M. H. (2000). Economic and statistical measures of forecast accuracy, *Journal of Forecasting*, **19**, 537–560.
- Greene, W. H. (2000). *Econometric Analysis*, 4th ed., Prentice Hall, Upper Saddle River, New Jersey.
- Horowitz, J. L. (1992). A smoothed maximum score estimator for the binary response model, *Econometrica*, **60**, 505–531.
- Hull, J. C. (2007). *Risk Management and Financial Institutions*, Pearson Education International, Upper Saddle River, New Jersey. (竹谷仁宏 訳(2008). 『フィナンシャルリスクマネジメント』, ピアソンエデュケーション, 東京.)
- Lee, S. H. and Urrutia, J. L. (1996). Analysis and prediction of insolvency in the property-liability insurance industry: A comparison of logit and hazard models, *The Journal of Risk and Insurance*, **63**, 121–130.
- Lieli, R. P. and Nieto-Barthaburu, A. (2010). Optimal binary prediction for group decision making, *Journal of Business & Economic Statistics*, **28**(2), 308–319.

Rao, C. R. (1965). *Linear Statistical Inference and Its Applications*, John Wiley and Sons, Inc., New York, London and Sydney.

坂元慶行(1985). 『カテゴリカルデータのモデル分析』, 共立出版, 東京.

Binary Prediction for Minimization of Financial Risk: Theory and Applications

Kentaro Akashi¹ and Yoshinori Kawasaki²

¹Faculty of Economics, Gakushuin University

²The Institute of Statistical Mathematics

Risk involved with financial contracts can often be viewed as uncertainty of binary outcomes. This paper treats risk minimization as a problem of profit maximization, and gives an optimal solution of the cut-off point for binary prediction. This optimality or profit maximization will be asymptotically attained in the sense of convergence in probability. In practice, we have to replace the true parameters inside an indicator function by estimates. Because indicator functions are discontinuous, apparently it looks like a non-standard argument. We show, in spite of this, that we can construct an interval for maximized profit, and even minimize it based on asymptotic theories where the MLE is simply plugged in. Simulation results suggest that the finite sample properties of our asymptotic theories are satisfactory. In an empirical analysis using personal loan data of a south German bank, we show that the total profit realized by our optimal prediction exceeds the actually observed profit regardless of the settings of loan interest.