2 項モデルの予測によるリスク総額最小化

赤司健太郎　リスク解析戦略研究センター　特任研究員

● はじめに

例えば、保険・与信契約といった 2 項モデルのリスク総額を企業側の総利潤で捉えてみよう。このとき、総利潤を確率収束の意味で最大化する簡潔な最適 2 項予測を与えることができる。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する手続きがある。こうした際でも、次期に実現する最大化総利潤の信頼区間、ないし最小信頼区間が構成されることを示す。数値実験では、漸近論の結果は有限標本における近似も良好を示している。

● 2 項予測による総利潤最大化と最小信頼区間

標準的な 2 項モデルを考える。

$$y_i = \mathbb{I}\{\ \boldsymbol{\beta}'\mathbf{x}_i + \epsilon_i \geq 0\ \}$$

ここで、y_i は借り手 i $(i = 1, ..., N)$ の事故発生時に 1、無事故で 0 となる指示関数、$(\mathbf{x}_i, \boldsymbol{\beta})$ はそれぞれ説明変数と係数ベクトルである。F で $-\epsilon_i$ の累積分布関数を表せば、事故率は $\Pr(y_i = 1|\mathbf{x}_i) = F(\boldsymbol{\beta}'\mathbf{x}_i)$ となる。

次に、返済時の利潤を $r_i(> 0)$、事故発生時の損失額を指示関数 $d_i(> 0)$ とする。また z_i を貸し出すか否かの指示関数とする (貸す時は 1, その他で 0)。まず期待利潤の最大化問題を考えると、

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

仮に最適解を z_i^* として貸出数 N が大きいとき、リスク総額は次の意味で最小化されている。

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

標準的には $z_i = \mathbb{I}\{\ F(\boldsymbol{\beta}'\mathbf{x}_i) \leq c_i\ \}$ の形が用いられるが、

結果 1： 最大解となる $(r_i, d_i, \mathbf{x}_i')$ に関し可測な指示関数 z_i^* $(i = 1, ..., N)$ は、次となる。

$$z_i^* = \mathbb{I}\{\ F(\boldsymbol{\beta}'\mathbf{x}_i) \leq \frac{r_i}{r_i + d_i}\ \}$$

つまり、最適解は事故率に対してを簡潔な閾値としてもつ形となり、個体間で異なる確率変数 (random threshold) である。

事故の有無によって期末に損益が確定する以前に、総利潤の信頼区間を構成したい。このとき、他より得た最尤推定量を代入して、$\hat{z}_i^* = \mathbb{I}\{\ F(\hat{\boldsymbol{\beta}}'\mathbf{x}_i) \leq \frac{r_i}{r_i+d_i}\ \}$ に依る。

結果 2： $\hat{R}_{N_2}^* = [\ \hat{l}_{N_2}^*, \hat{u}_{N_2}^*\ ]$ は、最大化総利潤の漸近的 $100(1-\alpha)\%$ 信頼区間である。特に $N_2/N_1 \to 0$ のとき最小区間。

$$\Pr\Big(\ \hat{l}_{N_2}^* \leq \sum_{i=1}^{N_2} \pi(\hat{z}_i^*) \leq \hat{u}_{N_2}^*\ \Big) \underset{N_1,N_2\to\infty}{\longrightarrow} (1-\alpha)$$

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

$$\begin{aligned}
\hat{R}_{N_2}^* &= \sum_{i=1}^{N_2}[\ r_i(1 - F(\hat{\boldsymbol{\beta}}'\mathbf{x}_i))\hat{z}_i^* - d_i F(\hat{\boldsymbol{\beta}}'\mathbf{x}_i)\hat{z}_i^*\ ] \pm \hat{\sigma}_{N_2}^{**} t_{\alpha/2} \\
\hat{\sigma}_{N_2}^{*2} &= \sum_{i=1}^{N_2}(r_i + d_i)^2 F(\hat{\boldsymbol{\beta}}'\mathbf{x}_i)(1 - F(\hat{\boldsymbol{\beta}}'\mathbf{x}_i))\hat{z}_i^* \\
\hat{\sigma}_{N_2}^{**} &= \Big[\ \hat{\sigma}_{N_2}^{*2} + \hat{\mathbf{f}}_{N_2}' \hat{\mathbf{V}}_{N_1}^{-1} \hat{\mathbf{f}}_{N_2}\ \Big]^{\frac{1}{2}}
\end{aligned}$$

● 数値実験：保険の例

設定は Logit model とし、平均事故率が 3% の場合を見てみよう。例えば、保険料率 $r_i = 0.1$ に対して、医療費支払い $d_i = 2.0$ と想定する。最適な判断は $c_i^* \simeq 0.05$ だが、比較の為 (i)$c = 0.01$ と (ii)$c = 0.5$ を考える。結果は 100%で利潤が上がり、平均上昇率は (i)5.6%(ii)46.2%である。最大化利潤の名目信頼係数 95%に対し、実際には 94.5%と、漸近論は好い近似を与える。