

# Power-IDI : IDIに基づく予測能改善の定量化

慶應義塾大学 理工学部 林 賢一  
統計数理研究所 数理推論研究系 江口 真透

バイオマーカーなどの変数がある事象をよく予測できるかどうかを評価することは、医学研究における重要な統計的課題である。本研究の目標は、応答変数  $D$  が二値の変数である場合の回帰モデルの予測力を、解釈可能な量として把握することである。より具体的には、事象の予測に有用であると思しき新しい変数を追加したモデルが、既存のモデルに比べ、どの程度予測力が向上するかを調べることである。このような問題について、ROC 曲線の下側面積の差分 ( $\Delta\text{AUC}$ ) が指標として用いられることが多い。しかし、これは予測能の指標ではないこと、検出力が低いなどの理由から批判がある。また、IDI (integrated discrimination improvement; Pencina et al., 2008) も多くの臨床研究で用いられている。しかし、IDI はモデル間の僅かな差も捉えてしまうため、予測能の改善を誤検出する問題が指摘されている (Hilden and Gerds, 2014)。

$\mathbf{X}$  を説明変数が従う確率ベクトル、 $p_{\text{new}}(\mathbf{x})$  と  $p_{\text{old}}(\mathbf{x})$  をそれぞれ  $P[D = 1 | \mathbf{X} = \mathbf{x}]$  に対する新たなモデル、既存のモデルとする。本研究では、 $\Delta\text{AUC}$  と IDI のもつ (それぞれ異なる) 問題を解決するための指標として

$$\text{IDI}_{\beta}(p_{\text{new}}, p_{\text{old}}) = \frac{1}{\beta} E[\tilde{p}_{\text{new}}(\mathbf{X}_1)^{\beta} - \tilde{p}_{\text{old}}(\mathbf{X}_1)^{\beta}] + \frac{1}{\beta} E[\tilde{q}_{\text{new}}(\mathbf{X}_0)^{\beta} - \tilde{q}_{\text{old}}(\mathbf{X}_0)^{\beta}]. \quad (1)$$

を提案する。ここで、 $\beta < 1$  は任意の定数であり、 $\mathbf{X}_d$  は確率ベクトル  $\mathbf{X} |_{D=d}$  である ( $d = 0, 1$ )。さらに、 $\tilde{p}_{\text{new}}(\mathbf{x}) = \frac{(p_{\text{new}}(\mathbf{x})/\pi_1)^{1/(1-\beta)}}{(p_{\text{new}}(\mathbf{x})/\pi_1)^{1/(1-\beta)} + (q_{\text{new}}(\mathbf{x})/\pi_0)^{1/(1-\beta)}}$ 、 $\tilde{q}_{\text{new}}(\mathbf{x}) = 1 - \tilde{p}_{\text{new}}(\mathbf{x})$  であり、 $\tilde{p}_{\text{old}}(\mathbf{x})$ 、 $\tilde{q}_{\text{old}}(\mathbf{x})$  も同様に定義する。

指標 (1) は、 $\beta \rightarrow 0$  のとき条件付尤度比の和に、 $\beta \rightarrow 1$  のとき条件付正判別確率の増分の和になる。すなわち、モデルの評価指標として汎く用いられる 2 つの指標の一般化とみなすことも可能である。さらに、 $\text{IDI}_{\beta}(p_{\text{new}}, p_{\text{old}})$  は Bayes リスク一貫性のみならず、Fisher 一貫性をもつ：

定理 (Hayashi and Eguchi, 2019)

$p^*(\mathbf{x}) = P[D = 1 | \mathbf{X} = \mathbf{x}]$  とするとき、任意の  $\beta \in (0, 1)$ 、 $p_{\text{new}}$ 、 $p_{\text{old}}$  に対して次の不等式が成り立つ。

$$\text{IDI}_{\beta}(p_{\text{new}}, p_{\text{old}}) \leq \text{IDI}_{\beta}(p^*, p_{\text{old}}).$$

等号は、 $p_{\text{new}} = p^*$  の場合にのみ成立する。

また、(1) の  $\beta$  に対する恣意性を回避する指標として

$$\text{DIDI}(p_{\text{new}}, p_{\text{old}}) = \int_0^1 \text{IDI}_{\beta}(p_{\text{new}}, p_{\text{old}}) d\beta \quad (2)$$

を提案する。提案した指標 (1)、(2) の挙動はシミュレーションや実データにより確認した。これらの詳細は、当日報告する。

## 参考文献

- [1] Hayashi, K., Eguchi, S. (2019). *Statistics in Medicine*, **38**, 2589–2604.
- [2] Hilden, J., Gerds, T.A. (2014). *Statistics in Medicine*, **33**, 3405–3414.
- [3] Pencina, M.J., et al. (2008). *Statistics in Medicine*, **27**, 175–172.