Odds-based predictive improvement index for binary regression models

Boston–Keio–Tsinghua Workshop

Kenichi Hayashi
Keio University
Jun. 29, 2023

Outline

1. Background
2. Proposed index
3. Simulation studies
4. Summary

- This is joint work with Dr. Eguchi (ISM, Japan)

1. Background

Motivation

Comparison of two binary regression models wrt prediction

- Typical examples: Old model vs New model

  Old: established model, New: covariate(s) newly added

  1. Death by breast cancer

     Old: Age + ⋯ + Chemotherapy

     New: Age + ⋯ + Chemotherapy + Estrogen receptor

  2. Death by acute coronary syndromes

     Old: Age + ⋯ + peak CPK + e−GFR

     New: Age + ⋯ + peak CPK + e−GFR + hs−Tn

- Problems in existing indices

Setting

- Y: binary outcome (0 or 1)
- $\boldsymbol{X}$: covariate vector (patient background, biomarkers, etc.)
- $\boldsymbol{X}^{(y)} \sim \boldsymbol{X}|_{Y=y},\ y = 1,0$
- $p^{\star}(y|\boldsymbol{x}) = \mathrm{P}[Y = y|\boldsymbol{X} = \boldsymbol{x}],\ y = 1,0$
- $p_{\mathrm{new}}, p_{\mathrm{old}}$: models for the conditional probability $p^{\star}$

Example: For $\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_{\mathrm{old}} \\ \boldsymbol{x}_{\mathrm{new}} \end{pmatrix}$

$$p_{\mathrm{old}}(1|\boldsymbol{x}) = \mathrm{expit}\left(\boldsymbol{x}_{\mathrm{old}}^{\top}\boldsymbol{\beta}\right) = \frac{\exp(\boldsymbol{x}_{\mathrm{old}}^{\top}\boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_{\mathrm{old}}^{\top}\boldsymbol{\beta})}$$

$$p_{\mathrm{new}}(1|\boldsymbol{x}) = \frac{\exp(\boldsymbol{x}_{\mathrm{old}}^{\top}\boldsymbol{\beta} + \boldsymbol{x}_{\mathrm{new}}^{\top}\boldsymbol{\gamma})}{1 + \exp(\boldsymbol{x}_{\mathrm{old}}^{\top}\boldsymbol{\beta} + \boldsymbol{x}_{\mathrm{new}}^{\top}\boldsymbol{\gamma})}$$

$$\boldsymbol{X}^{(y)} \sim \boldsymbol{X}\Big|_{Y=y}, \; y = 1,0$$

Existing indices

$p_{\text{new}}, p_{\text{old}}$: models for $p^\star = \mathrm{P}[Y = y | \boldsymbol{X} = \boldsymbol{x}]$

- Area under the ROC curve: with independent $\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(0)}$

$$\Delta\mathrm{AUC}(p_{\text{new}}, p_{\text{old}}) = \mathrm{P}\left[p_{\text{new}}\left(1|\boldsymbol{X}^{(1)}\right) > p_{\text{new}}\left(1|\boldsymbol{X}^{(0)}\right)\right] \\ -\mathrm{P}\left[p_{\text{old}}\left(1|\boldsymbol{X}^{(1)}\right) > p_{\text{old}}\left(1|\boldsymbol{X}^{(0)}\right)\right]$$

  - Remark. AUC for a model p: $\mathrm{P}[p(1|\boldsymbol{X}) > p(1|\boldsymbol{X}')|Y = 1, Y' = 0]$

- IDI (integrated discrimination improvement)

$$\mathrm{IDI}(p_{\text{new}}, p_{\text{old}}) = \mathrm{E}\left[p_{\text{new}}\left(1|\boldsymbol{X}^{(1)}\right) - p_{\text{old}}\left(1|\boldsymbol{X}^{(1)}\right)\right] \\ +\mathrm{E}\left[p_{\text{new}}\left(0|\boldsymbol{X}^{(0)}\right) - p_{\text{old}}\left(0|\boldsymbol{X}^{(0)}\right)\right]$$

  - Pencina et al., (2008)

IDI: integrated discrimination improvement

- An index based on FPR and TPR（Pencina et al., 2008）

$$\begin{aligned}&\mathrm{IDI}(p_{\mathrm{new}}, p_{\mathrm{old}})\\&= (\mathrm{ITPR}(p_{\mathrm{new}}) - \mathrm{ITPR}(p_{\mathrm{old}})) + (\mathrm{IFPR}(p_{\mathrm{old}}) - \mathrm{IFPR}(p_{\mathrm{new}}))\\&= \textstyle\sum_{y=0}^{1} \mathrm{E}[p_{\mathrm{new}}(y|\boldsymbol{X}^{(y)}) - p_{\mathrm{old}}(y|\boldsymbol{X}^{(y)})]\end{aligned}$$

  - IFPR: integrated FPR, ITPR: integrated TPR

$$\because\ \mathrm{ITPR}(p) = \int_0^1 \mathrm{TPR}(u;p)du = \mathrm{E}\left[p\left(1\middle|\boldsymbol{X}^{(1)}\right)\right],\ \ \mathrm{TPR}(u;p) = P\left[p\left(1\middle|\boldsymbol{X}^{(1)}\right) > u\right]$$

  - Estimate

$$\begin{aligned}\widehat{\mathrm{IDI}}(p_{\mathrm{new}}, p_{\mathrm{old}}) = &\frac{1}{n_1}\textstyle\sum_{i=1}^{n}\left(p_{\mathrm{new}}(1|\boldsymbol{x}_i) - p_{\mathrm{old}}(1|\boldsymbol{x}_i)\right)y_i\\&+\frac{1}{n_0}\textstyle\sum_{i=1}^{n}\left(p_{\mathrm{new}}(0|\boldsymbol{x}_i) - p(0|\boldsymbol{x}_i)\right)(1-y_i)\end{aligned}$$

Problems in ΔAUC and IDI

$$\Delta\text{AUC}(p_{\text{new}}, p_{\text{old}}) = \text{P}\left[p_{\text{new}}\left(1|\boldsymbol{X}^{(1)}\right) > p_{\text{new}}\left(1|\boldsymbol{X}^{(0)}\right)\right] - \text{P}\left[p_{\text{old}}\left(1|\boldsymbol{X}^{(1)}\right) > p_{\text{old}}\left(1|\boldsymbol{X}^{(0)}\right)\right]$$

$$\text{IDI}(p_{\text{new}}, p_{\text{old}}) = \sum_{y=0}^{1} \text{E}\left[p_{\text{new}}\left(y|\boldsymbol{X}^{(y)}\right) - p_{\text{old}}\left(y|\boldsymbol{X}^{(y)}\right)\right]$$

- ΔAUC: insensitive to difference
  - AUC is a rank-based statistic
  - not an index for prediction improvement
- IDI: too sensitive, not safe
  - return positive values even if p_{new} is far from $p^{\star}$ (true)
    - Kerr et al. (2011), Hilden and Gerds (2014)
- Common problem: lack of Fisher consistency
  - Fisher consistency: if an index $\Phi(p_{\text{new}}, p_{\text{old}})$ holds
    1. $\Phi(p^{\star}, p_{\text{old}}) \geq \Phi(p_{\text{new}}, p_{\text{old}})$ for any p_{old}
    2. the equality in 1 holds iff $p_{\text{new}} \equiv p^{\star}$

Test of ΔAUC

- Case where two models are nested

  e.g. $p_{\text{old}}(y|\boldsymbol{x}) = p(y|x_1;\beta_1)$ VS $p_{\text{new}}(y|\boldsymbol{x}) = p(y|(x_1, x_2);(\beta_1, \beta_2))$
  - When $\beta_2 = 0$, $p_{\text{old}} \equiv p_{\text{new}}$
  - Test statistics of DeLong et al. (1988) degenerates
    - Z-type test fails (Demler et al., 2012)

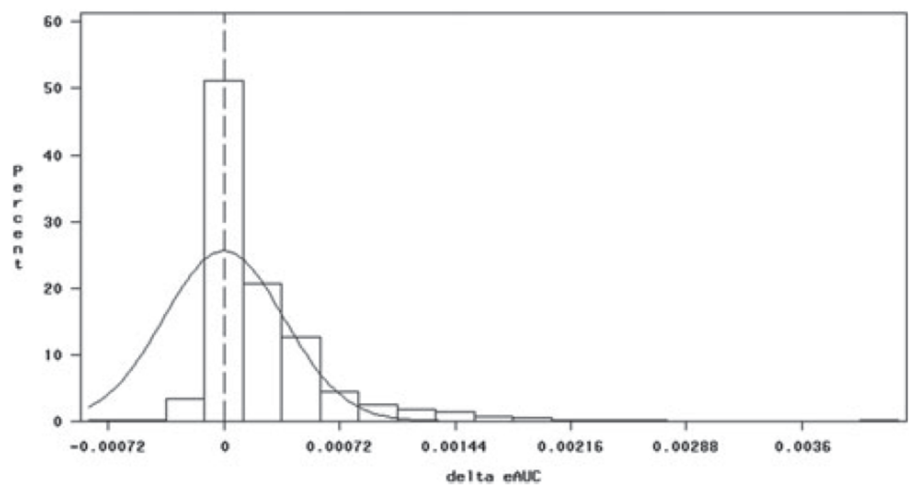
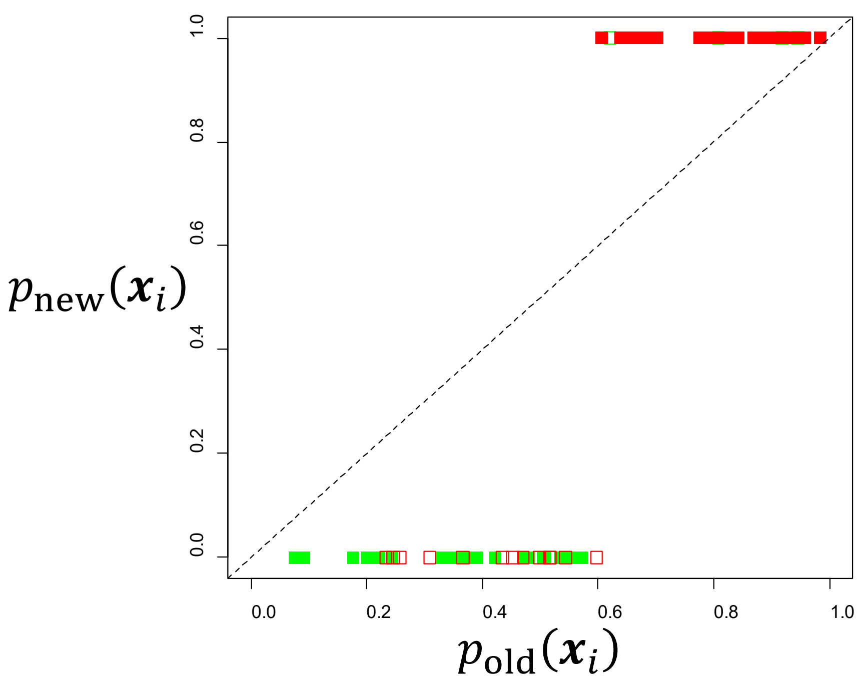


Figure 2. Histogram of change in eAUC under null hypothesis for multivariate normal data and sample size of 8365 with superimposed plot of corresponding distribution function used by DeLong test.

Figure from Demler et al. (2012)

Danger of IDI

- "Do not rely on IDI…"
  - Hilden and Gerds (2014)
  - IDI can be enlarged even when p_{new} is constructed by p_{old} only
  - $p_{\text{new}}(1|\boldsymbol{x}_i) = \begin{cases} 1 & \text{if } p_{\text{old}}(1|\boldsymbol{x}_i) \geq \bar{p} \\ 0 & \text{if } p_{\text{old}}(1|\boldsymbol{x}_i) < \bar{p} \end{cases}$ where $\bar{p} = \frac{1}{n}\sum_{i=1}^{n} p_{\text{old}}(1|\boldsymbol{x}_i)$



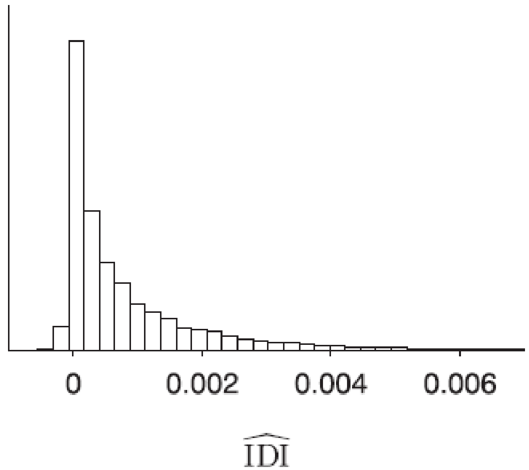
IDI=0.31
(p<0.001)

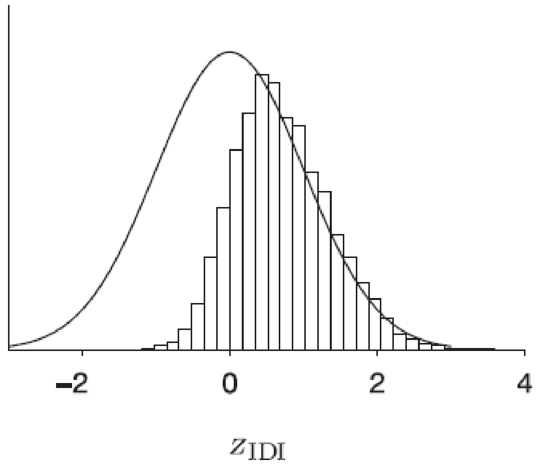
Problems in IDI

- False detection
  - "…*use is not always safe*" (Hilden and Gerds, 2014)
  - Model can be improved without adding measured information

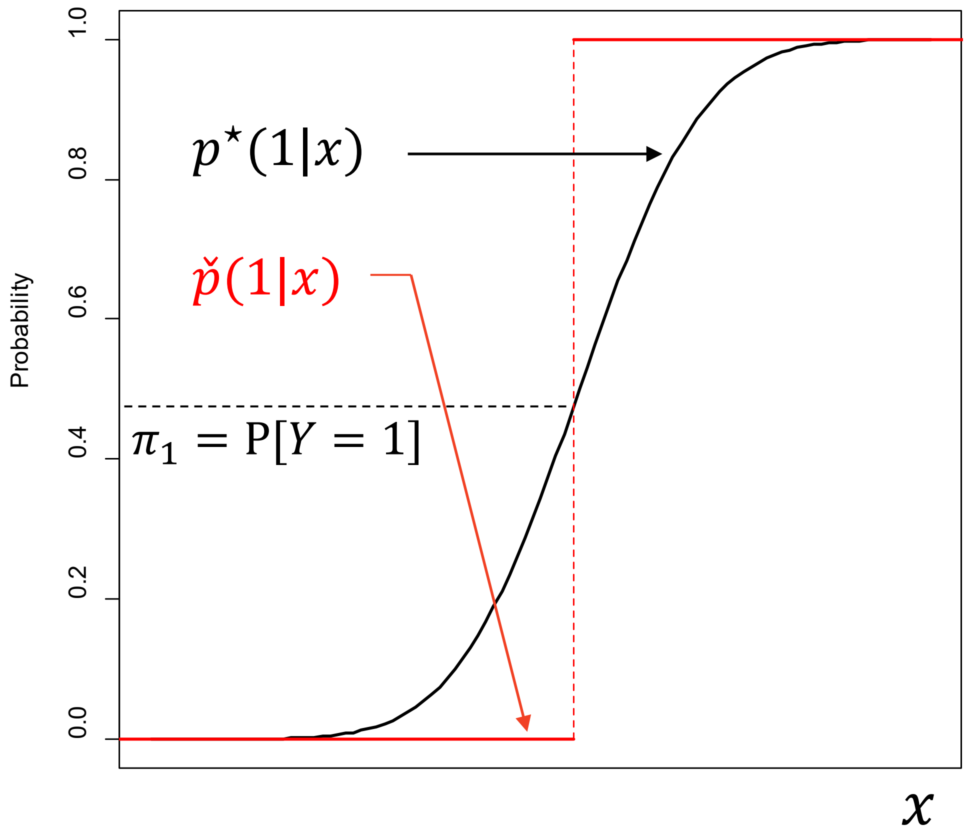
$$\mathrm{IDI}(p_{\mathrm{new}}, p_{\mathrm{old}}) = \sum_{y=0}^{1} \mathrm{E}[p_{\mathrm{new}}(y|\boldsymbol{X}^{(y)}) - p_{\mathrm{old}}(y|\boldsymbol{X}^{(y)})]$$

- Invalid test statistics
  - Kerr et al. (2011)





Maximizer of IDI



$$\check{p}(1|x) = \begin{cases} 1 & \text{if } \pi_0 p^\star(1|x) > \pi_1 p^\star(0|x) \\ 1/2 & \text{if } \pi_0 p^\star(1|x) = \pi_1 p^\star(0|x) \\ 0 & \text{if } \pi_0 p^\star(1|x) < \pi_1 p^\star(0|x) \end{cases}$$

➡ For any models p, p_{old}, $\mathrm{IDI}(\check{p}, p_{\mathrm{old}}) \geq \mathrm{IDI}(p, p_{\mathrm{old}})$

Properties for indices

1. Bayes risk consistency (BRC)

Index $\Phi(p_{\text{new}}, p_{\text{old}})$ has BRC

$\Leftrightarrow$ strictly increasing m: $[0,1] \to [0,1]$ exists and

$\Phi(m(p^\star), p_{\text{old}}) \geq \Phi(p_{\text{new}}, p_{\text{old}})$ for any p_{old}

- Remark: $p^\star(y|\boldsymbol{x}) = P[Y = y|\boldsymbol{X} = \boldsymbol{x}]$
- ΔAUC has Bayes risk consistency

2. Fisher consistency (FC)

Index $\Phi(p_{\text{new}}, p_{\text{old}})$ has FC

$\Leftrightarrow \Phi(p^\star, p_{\text{old}}) \geq \Phi(p_{\text{new}}, p_{\text{old}})$ for any p_{old}

and the equality holds iff $p_{\text{new}} \equiv p^\star$

- Remark: FC implies BRC

Power-IDI 1/2

- Power-IDI (Hayashi and Eguchi, 2019): for $\beta \in (0,1)$

$$\mathrm{IDI}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}}) = \frac{1}{\beta}\sum_{y=0}^{1} \mathrm{E}\left[\tilde{p}_{\mathrm{new}}\left(y|\boldsymbol{X}^{(y)}\right)^{\beta} - \tilde{p}_{\mathrm{old}}\left(y|\boldsymbol{X}^{(y)}\right)^{\beta}\right]$$

$$\text{where } \tilde{p}_{\mathrm{new}}(y|\boldsymbol{x}) = \frac{\left(p_{\mathrm{new}}(y|\boldsymbol{x})/\pi_y\right)^{1/(1-\beta)}}{\left(p_{\mathrm{new}}(1|\boldsymbol{x})/\pi_1\right)^{1/(1-\beta)} + \left(p_{\mathrm{new}}(0|\boldsymbol{x})/\pi_0\right)^{1/(1-\beta)}}$$

  - $\pi_y = \mathrm{P}[Y = y]$ and $\tilde{p}_{\mathrm{old}}$ is defined as well as $\tilde{p}_{\mathrm{new}}$
  - Power-IDI has Fisher consistency
- Relationship to power divergence: with $\gamma = \beta/(1-\beta)$

$$C_\beta(f, g) \propto -\frac{\int f(\boldsymbol{x})^{\beta/(1-\beta)} g(\boldsymbol{x}) d\boldsymbol{x}}{\left\{\int f(\boldsymbol{x})^{1/(1-\beta)} d\boldsymbol{x}\right\}^{\beta}}$$

: power cross entropy between f & g (Eguchi et al., 2011)

Power-IDI 2/2

- Power-IDI (Hayashi and Eguchi, 2019): for $\beta \in (0,1)$

$$\mathrm{IDI}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}}) = \frac{1}{\beta}\sum_{y=0}^{1} \mathrm{E}\left[\tilde{p}_{\mathrm{new}}\left(y|\boldsymbol{X}^{(y)}\right)^\beta - \tilde{p}_{\mathrm{old}}\left(y|\boldsymbol{X}^{(y)}\right)^\beta\right]$$

$$\text{where } \tilde{p}_{\mathrm{new}}(y|\boldsymbol{x}) = \frac{\left(p_{\mathrm{new}}(y|\boldsymbol{x})/\pi_y\right)^{1/(1-\beta)}}{(p_{\mathrm{new}}(1|\boldsymbol{x})/\pi_1)^{1/(1-\beta)} + (p_{\mathrm{new}}(0|\boldsymbol{x})/\pi_0)^{1/(1-\beta)}}$$

  - $\pi_y = \mathrm{P}[Y = y]$ and $\tilde{p}_{\mathrm{old}}$ is defined as well as $\tilde{p}_{\mathrm{new}}$
  - Power-IDI has Fisher consistency

Remaining problems

① Interpretation of values of β

  - $\mathrm{DIDI}(p_{\mathrm{new}}, p_{\mathrm{old}}) = \int_0^1 \mathrm{IDI}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}})\mathrm{d}\beta$ is also proposed, but…

② Interpretation of two-step transformation

2. Proposed index

① Meaning of β?
② Meaning of transformation?

Odds-based index

- With $\mathrm{O}(p) = p/(1-p)$,

$$\mathrm{IDI}_{\mathrm{odds}}(p_{\mathrm{new}}, p_{\mathrm{old}})$$
$$= -\sum_{y=0}^{1} \pi_y \mathrm{E}\left[\mathrm{O}\left(p_{\mathrm{new}}(y|\boldsymbol{X}^{(y)})\right)^{-1/2} - \mathrm{O}\left(p_{\mathrm{old}}(y|\boldsymbol{X}^{(y)})\right)^{-1/2}\right]$$
$$= -\mathrm{E}\left[\mathrm{O}\big(p_{\mathrm{new}}(Y|\boldsymbol{X})\big)^{-1/2} - \mathrm{O}\big(p_{\mathrm{old}}(Y|\boldsymbol{X})\big)^{-1/2}\right]$$

- holds Fisher consistency
- No hyperparameters ➡ ① solved
- No transformation ➡ ② solved (partially)
- Estimation: with sample $\{(y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)\}$

$$\widehat{\mathrm{IDI}}_{\mathrm{odds}}(p_{\mathrm{new}}, p_{\mathrm{old}}) = -\frac{1}{n}\sum_{i=1}^{n}\left\{\left(\frac{p_{\mathrm{new}}(y_i|\boldsymbol{x}_i)}{p_{\mathrm{new}}(1-y_i|\boldsymbol{x}_i)}\right)^{-y_i+\frac{1}{2}} - \left(\frac{p_{\mathrm{old}}(y_i|\boldsymbol{x}_i)}{p_{\mathrm{old}}(1-y_i|\boldsymbol{x}_i)}\right)^{-y_i+\frac{1}{2}}\right\}$$

Possible interpretation 1

- p_{old} & p_{new} are logistic regression model: with $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_{\mathrm{old}} \\ \boldsymbol{X}_{\mathrm{new}} \end{pmatrix}$

$$p_{\mathrm{old}}(1|\boldsymbol{X}) = \frac{\exp(\boldsymbol{X}_{\mathrm{old}}^{\top}\boldsymbol{\beta}')}{\exp(\boldsymbol{X}_{\mathrm{old}}^{\top}\boldsymbol{\beta}') + \exp(-\boldsymbol{X}_{\mathrm{old}}^{\top}\boldsymbol{\beta}')},$$

$$p_{\mathrm{new}}(1|\boldsymbol{X}) = \frac{\exp(\boldsymbol{X}_{\mathrm{old}}^{\top}\boldsymbol{\beta} + \boldsymbol{X}_{\mathrm{new}}^{\top}\boldsymbol{\gamma})}{\exp(\boldsymbol{X}_{\mathrm{old}}^{\top}\boldsymbol{\beta} + \boldsymbol{X}_{\mathrm{new}}^{\top}\boldsymbol{\gamma}) + \exp(-\boldsymbol{X}_{\mathrm{old}}^{\top}\boldsymbol{\beta} - \boldsymbol{X}_{\mathrm{new}}^{\top}\boldsymbol{\gamma})}$$

⇨
$$\mathrm{IDI}_{\mathrm{odds}}(p_{\mathrm{new}}, p_{\mathrm{old}}) = -\mathrm{E}\left[\exp\left(-\tilde{Y}\left(\boldsymbol{X}_{\mathrm{new}}^{\top}\boldsymbol{\gamma} + \boldsymbol{X}_{\mathrm{old}}^{\top}\boldsymbol{\beta}\right)\right) - \exp\left(-\tilde{Y}\boldsymbol{X}_{\mathrm{old}}^{\top}\boldsymbol{\beta}'\right)\right]$$

where $\tilde{Y} = \begin{cases} +1 & \text{if } Y = 1 \\ -1 & \text{if } Y = 0 \end{cases}$

⇨ Increment of negative exponential risk

Exponential loss and AdaBoost

- AdaBoost: ensemble method for binary classification
  - Freund and Schapire (1997); Friedman et al. (2001)
- Construct $F_T(\boldsymbol{x}) = \sum_{j=1}^{T} \gamma_t f_t(\boldsymbol{x})$ in a sequential manner
  - $f_t: \boldsymbol{x} \mapsto \{+1, -1\}$ (weak learner)

Algotirhm: $\{(\tilde{y}_i, \boldsymbol{x}_i); i = 1, \dots, n\}$ and a set of weak learner $\mathcal{F}$

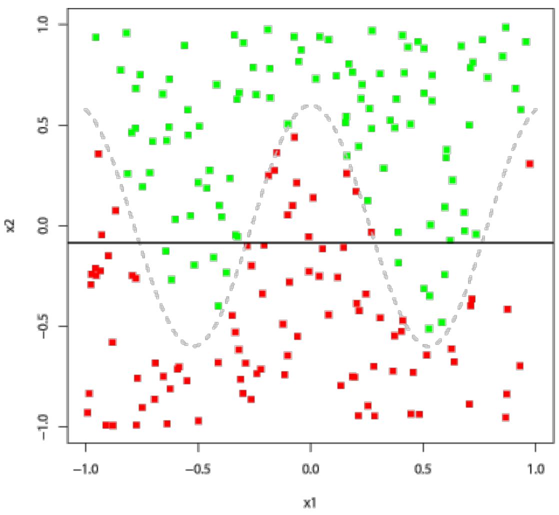
  - $\tilde{y}_i \in \{+1, -1\}$

- Set $F_0(\boldsymbol{x}) \equiv 0$ and for $t = 1, \dots, T$

① $$f_t = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \frac{\partial}{\partial \gamma} \sum_{i=1}^{n} \exp\left(-\tilde{y}_i\left(F_{t-1} + \gamma f(\boldsymbol{x}_i)\right)\right)\Big|_{\gamma=0}$$

② $$\gamma_t = \underset{\gamma \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} \exp\left(-\tilde{y}_i\left(F_{t-1} + \gamma f_t(\boldsymbol{x}_i)\right)\right)$$

③ $$F_t = F_{t-1} + \gamma_t f_t$$



Multicategory case

- $p(y|\boldsymbol{x})$: a model for $\mathrm{P}[Y = y|\boldsymbol{x}],\ y = 1, \dots, M$

$$\mathrm{O}_M(p(y|\boldsymbol{x})) = \prod_{y'=1}^{M} \frac{p(y|\boldsymbol{x})}{p(y'|\boldsymbol{x})},\ \pi_y = \mathrm{P}[D = y]$$

$$\mathrm{IDI}_{\mathrm{odds}}(p_{\mathrm{new}}, p_{\mathrm{old}})$$
$$= -\sum_{y=1}^{M} \pi_y \,\mathrm{E}\left[\mathrm{O}_M\left(p_{\mathrm{new}}\left(y|\boldsymbol{X}^{(y)}\right)\right)^{-\frac{1}{M}} - \mathrm{O}_M\left(p_{\mathrm{old}}\left(y|\boldsymbol{X}^{(y)}\right)\right)^{-\frac{1}{M}}\right]$$
$$= -\mathrm{E}\left[\mathrm{O}_M\left(p_{\mathrm{new}}(Y|\boldsymbol{X})\right)^{-\frac{1}{M}} - \mathrm{O}_M\left(p_{\mathrm{old}}(Y|\boldsymbol{X})\right)^{-\frac{1}{M}}\right]$$

- Multicategory version also has Fisher consistency
- Corresponds to multicategory AdaBoost (Zhu et al., 2009)
  - Exponential type loss function
  - When two models are multinomial logit regression models

Possible interpretation 2

Reference value of each term in Odds-IDI is 1

cf. reference value of AUC is 0.5 (with random choice)

- Binary: $\mathrm{E}\left[\mathrm{O}\left(p(Y|\boldsymbol{X})\right)^{-\frac{1}{2}}\right] \geq 2\mathrm{E}\left[\sqrt{p^{\star}(1|\boldsymbol{X})p^{\star}(0|\boldsymbol{X})}\right] \geq 1$
  - Lower bound is proportional to the variance of binomial distribution
  - Random choice models attain the smallest lower bound
- Multicategory: $\mathrm{E}\left[\mathrm{O}_M\left(p(Y|\boldsymbol{X})\right)^{-\frac{1}{M}}\right] \geq M\mathrm{E}\left[\prod_{y=1}^{M} p^{\star}(y|\boldsymbol{X})^{\frac{1}{M}}\right] \geq 1$
  - Random choice models attain the lower bound
  - Lower bound: prop. to the generalized variance of multinom. dist.

Generalized variance

- Product of nonnegative eigenvalues for covariance matrix
  - Diaconis and Efron (1983)

<u>Examples</u>

1. Multivariate normal distribution $N_M(\boldsymbol{\mu}, \Sigma)$

   ➡ Generalized variance is $|\Sigma|$

2. Multinomial distribution with $\pi_1, \dots, \pi_M$ $(\sum_{m=1}^{M} \pi_m = 1)$

   ➡ Generalized variance is proportional to $\prod_{m=1}^{M} \pi_m$

3. Simulation studies

Settings

- Objective: comparison of $\mathrm{IDI}_{\mathrm{odds}}$ with existing indices
- Data generation
  - $p^{\star}(1|\boldsymbol{x}) = \mathrm{expit}(\lambda_0 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3)$
    - $\lambda_0 = 0$ and $\lambda_1 = \lambda_2 = \lambda_3 = 0.05$
  - (x_1, x_2, x_3) are generated from trivariate standard normal distribution
- Scenarios
  - The old model is fixed as $p_{\mathrm{old}}(1|\boldsymbol{x}) = \mathrm{expit}(\lambda_0 + \lambda_1 x_1)$

(A) new model that is (nearly) $p^{\star}$

- Interest: how correctly indices detect better model p_{new}

(B) new model that is far away from $p^{\star}$

- Interest: how correctly indices ignore “wrong” model p_{new}

- 10^5 datasets are generated
- Parameters are estimated based on ML method

$p^{\star}(1|\boldsymbol{x}) = \mathrm{expit}(\lambda_0 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3)$

Results 1/3

- Scenario (A)
  - $p_{\mathrm{old}}(1|\boldsymbol{x}) = \mathrm{expit}(\lambda_0 + \lambda_1 x_1)$, $p_{\mathrm{new}}(1|\boldsymbol{x}) = \mathrm{expit}(\lambda_0 + \lambda_1 x_1 + \lambda_2 x_2)$
- Averages of 10^5 values

n	$\mathrm{IDI}_{\mathrm{odds}}$	Power-IDI			DIDI	IDI	2ΔAUC
		$\beta = 0.2$	$\beta = 0.3$	$\beta = 0.4$			
100	0.0060	0.0127	0.0134	0.1431	0.0192	0.0114	0.0556
400	0.0019	0.0041	0.0044	0.0047	0.0077	0.0038	0.0323
1600	0.0009	0.0020	0.0021	0.0023	0.0042	0.0018	0.0227

- Proportion of positive values (%)

n	$\mathrm{IDI}_{\mathrm{odds}}$	Power-IDI			DIDI	IDI	2ΔAUC
		$\beta = 0.2$	$\beta = 0.3$	$\beta = 0.4$			
100	99.25	99.38	99.18	98.75	82.95	99.25	78.18
400	99.79	99.83	99.79	99.65	85.88	99.86	82.63
1600	99.93	99.97	99.96	99.89	90.15	99.98	89.96

$p^{\star}(1|\boldsymbol{x}) = \mathrm{expit}(\lambda_0 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3)$

Results 2/3

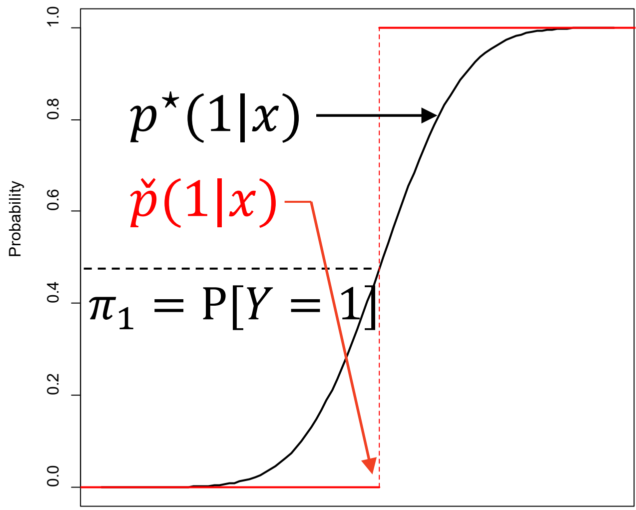
- Scenario (A')
  - $p_{\mathrm{old}}(1|\boldsymbol{x}) = \mathrm{expit}(\lambda_0 + \lambda_1 x_1)$, $p_{\mathrm{new}}(1|\boldsymbol{x}) = p^{\star}(1|\boldsymbol{x})$
- Averages of 10^5 values

n	$\mathrm{IDI}_{\mathrm{odds}}$	Power-IDI			DIDI	IDI	2ΔAUC
		$\beta = 0.2$	$\beta = 0.3$	$\beta = 0.4$			
100	0.0118	0.0251	0.0264	0.0282	0.0367	0.0224	0.0989
400	0.0038	0.0082	0.0867	0.0094	0.0147	0.0074	0.0570
1600	0.0018	0.0041	0.0044	0.0048	0.0838	0.0038	0.0403

- Proportion of positive values (%)

n	$\mathrm{IDI}_{\mathrm{odds}}$	Power-IDI			DIDI	IDI	2ΔAUC
		$\beta = 0.2$	$\beta = 0.3$	$\beta = 0.4$			
100	99.97	99.98	99.98	99.93	93.66	99.97	90.12
400	100.00	100.00	100.00	99.99	94.96	100.00	93.06
1600	100.00	100.00	100.00	100.00	98.02	100.00	97.72

Results 3/3



- Scenario (B)
  - $p_{\text{old}}(1|\boldsymbol{x}) = \text{expit}(\lambda_0 + \lambda_1 x_1)$, $p_{\text{new}} = \check{p}$
- Averages of 10^5 values

n	IDI_{odds}	Power-IDI			DIDI	IDI	2ΔAUC
		$\beta = 0.2$	$\beta = 0.3$	$\beta = 0.4$			
100	-46.497	-2.991	-1.866	-1.173	-1.780	0.0388	-0.011
400	-46.592	-2.992	-1.864	-1.168	-1.772	0.0444	0.001
1600	-46.557	-2.987	-1.859	-1.164	-1.767	0.0471	0.009

- Proportion of positive values (%)

n	IDI_{odds}	Power-IDI			DIDI	IDI	2ΔAUC
		$\beta = 0.2$	$\beta = 0.3$	$\beta = 0.4$			
100	0.00	0.00	0.00	0.00	0.00	65.46	45.94
400	0.00	0.00	0.00	0.00	0.00	81.65	50.20
1600	0.00	0.00	0.00	0.00	0.00	97.59	61.58

4. Summary

Summary and Future tasks

Summary

- Proposed an index for prediction improvement

  ➡ $\mathrm{IDI}_{\mathrm{odds}}$: based on odds $\frac{p}{1-p}$
- Odds-IDI has Fisher consistency
  - No hyperparameters, no transformations
- Simulation suggests superiority of odds-IDI to existing indices

Future tasks

- Extension to ordered multicategory: $1 \prec 2 \prec \cdots \prec M$
- Extension to survival outcomes

Thank you for your attention

Appendix

References (selected)

- Diaconis and Efron (1983) *Tech Rep Stanford Univ*
- Freund and Schapire (1997) *J Comput Syst Sci*
- Friedman et al. (2001) *Ann Statist*
- Gneiting and Raftery (2007) *JASA*
- Hayashi and Eguchi (2019) *Stat Med*
- Hilden and Gerds (2014) *Stat Med*
- Kerr et al. (2011) *Am J Epidemiol*
- Pencina et al. (2008) *Stat Med*

Odds-IDI: a sketch of derivation

- $\Psi_\alpha(p_{\text{new}}, p_{\text{old}}) = -\pi_1 \mathrm{E}\left[\left(\frac{p_{\text{new}}(1|\boldsymbol{X}^{(1)})}{p_{\text{new}}(0|\boldsymbol{X}^{(1)})}\right)^\alpha - \left(\frac{p_{\text{old}}(1|\boldsymbol{X}^{(1)})}{p_{\text{old}}(0|\boldsymbol{X}^{(1)})}\right)^\alpha\right]$

  $-\pi_0 \mathrm{E}\left[\left(\frac{p_{\text{new}}(0|\boldsymbol{X}^{(0)})}{p_{\text{new}}(1|\boldsymbol{X}^{(0)})}\right)^\alpha - \left(\frac{p_{\text{old}}(0|\boldsymbol{X}^{(0)})}{p_{\text{old}}(1|\boldsymbol{X}^{(0)})}\right)^\alpha\right]$

  Stationary condition of the variational function

  $$\frac{\partial}{\partial \varepsilon} \Psi_\alpha\big((1-\varepsilon)p^* + \varepsilon r, p_{\text{old}}\big)\Big|_{\varepsilon=0} = 0$$

  - p^* : maximizer (function) of Ψ_α with respect to the first argument
  - r : arbitrary regression function

  ➡ $\alpha = -1/2$ iff $p^*(y|\boldsymbol{x}) = p^\star(y|\boldsymbol{x}) = \mathrm{P}[Y = y|\boldsymbol{X} = \boldsymbol{x}]$
- Uniqueness of p^* can be established
- Remark: log odds does not have Fisher consistency

Power-IDI

$$\mathrm{IDI}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}}) = \frac{1}{\beta}\sum_{y=0}^{1} \mathrm{E}\left[\tilde{p}_{\mathrm{new}}\left(y|\boldsymbol{X}^{(y)}\right)^\beta - \tilde{p}_{\mathrm{old}}\left(y|\boldsymbol{X}^{(y)}\right)^\beta\right]$$

where $\tilde{p}_\square(y|\boldsymbol{x}) = \dfrac{\left(p_\square(y|\boldsymbol{x})/\pi_y\right)^{1/(1-\beta)}}{(p_\square(1|\boldsymbol{x})/\pi_1)^{1/(1-\beta)} + (p_\square(0|\boldsymbol{x})/\pi_0)^{1/(1-\beta)}}$

and $\beta \in (0,1)$

- IDI_β has Fisher consistency
  - Equality holds iff $p_{\mathrm{new}} \equiv p^\star$
- Relationship to power divergence: with $\gamma = \beta/(1-\beta)$

$C_\beta(f,g) \propto -\dfrac{\int f(\boldsymbol{x})^{\beta/(1-\beta)} g(\boldsymbol{x}) d\boldsymbol{x}}{\{\int f(\boldsymbol{x})^{1/(1-\beta)} d\boldsymbol{x}\}^\beta}$: power cross entropy between f & g

(Eguchi et al., 2011)

Derivation of Power-IDI

$$F_\beta(p_{\mathrm{new}}, p_{\mathrm{old}}) = \frac{1}{\beta} \sum_{y=0}^{1} \mathrm{E}\left[p_{\mathrm{new}}\left(y|\boldsymbol{X}^{(y)}\right)^\beta - p_{\mathrm{old}}\left(y|\boldsymbol{X}^{(y)}\right)^\beta \right]$$

- p_β^*: maximizer of $F_\beta(p, p_{\mathrm{old}})$ wrt the first argument

$$p_\beta^*(y|\boldsymbol{x}) = \frac{\left(p^\star(y|\boldsymbol{x})/\pi_y\right)^{1/(1-\beta)}}{(p^\star(1|\boldsymbol{x})/\pi_1)^{1/(1-\beta)} + (p^\star(0|\boldsymbol{x})/\pi_0)^{1/(1-\beta)}}$$

- F_β has Bayes risk consistency
  - Maximizer is proportional to $\Lambda(\boldsymbol{x}) = p^\star(1|\boldsymbol{x})/p^\star(0|\boldsymbol{x})$
- But still $F_\beta(p^\star, p_{\mathrm{old}}) \leq F_\beta\left(p_\beta^*, p_{\mathrm{old}}\right)$

Form of $\tilde{p}_{\mathrm{new}}$

$$\tilde{p}_{\mathrm{new}}(1|\boldsymbol{x}) = \frac{(p_{\mathrm{new}}(1|\boldsymbol{x})/\pi_1)^{1/(1-\beta)}}{(p_{\mathrm{new}}(1|\boldsymbol{x})/\pi_1)^{1/(1-\beta)} + (p_{\mathrm{new}}(0|\boldsymbol{x})/\pi_0)^{1/(1-\beta)}}$$

$$= \frac{1}{1+\exp(-g(\boldsymbol{x}))}$$

where $g(\boldsymbol{x}) = \frac{1}{1-\beta}\left(\log\frac{\pi_0}{\pi_1} + \log\frac{p_{\mathrm{new}}(1|\boldsymbol{x})}{p_{\mathrm{new}}(0|\boldsymbol{x})}\right)$

- If p_{new} is the logistic linear models, $\tilde{p}_{\mathrm{new}}$ has
    - coefficients = original $\times\ 1/(1-\beta)$
  - intercept = original $\times\ 1/(1-\beta)$ + const.

Power-IDI: two special cases

$$\mathrm{IDI}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}}) = \frac{1}{\beta}\sum_{y=0}^{1} \mathrm{E}\left[\tilde{p}_{\mathrm{new}}\left(y|\boldsymbol{X}^{(y)}\right)^\beta - \tilde{p}_{\mathrm{old}}\left(y|\boldsymbol{X}^{(y)}\right)^\beta\right]$$

where $\tilde{p}_\square(y|\boldsymbol{x}) = \dfrac{\left(p_\square(y|\boldsymbol{x})/\pi_y\right)^{1/(1-\beta)}}{(p_\square(1|\boldsymbol{x})/\pi_1)^{1/(1-\beta)} + (p_\square(0|\boldsymbol{x})/\pi_0)^{1/(1-\beta)}}$

and $\beta \in (0,1)$

- $\frac{1}{\beta}\mathrm{E}\left[\tilde{p}_{\mathrm{new}}(1|\boldsymbol{X}_1)^\beta\right] \to \mathrm{P}[\pi_0 p_{\mathrm{new}}(1|\boldsymbol{X}_1) > \pi_1 p_{\mathrm{new}}(0|\boldsymbol{X}_1)]$ as $\beta \to 1$
  - Hit rate of p_{new} or NRI at event rate
    - Pencina et al. (2017)
- $\frac{1}{\beta}\mathrm{E}\left[\tilde{p}_{\mathrm{new}}(1|\boldsymbol{X}_1)^\beta\right] \to \mathrm{E}\left[\log\frac{p_{\mathrm{new}}(1|\boldsymbol{X}_1)}{q_{\mathrm{new}}(0|\boldsymbol{X}_1)}\right]$ as $\beta \to 0$
  - Conditional log likelihood ratio of p_{new}

Estimation of IDI_β

- Sample: $\{(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_n, y_n)\}$

$$\widehat{\mathrm{IDI}}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}}) = \frac{1}{n_1\beta}\sum_{i=1}^{n}\left(\tilde{p}_{\mathrm{new}}(1|\boldsymbol{x}_i)^\beta - \tilde{p}_{\mathrm{old}}(1|\boldsymbol{x}_i)^\beta\right)y_i$$
$$+\frac{1}{n_0\beta}\sum_{i=1}^{n}\left(\tilde{p}_{\mathrm{new}}(0|\boldsymbol{x}_i)^\beta - \tilde{p}_{\mathrm{old}}(0|\boldsymbol{x}_i)^\beta\right)(1-y_i)$$

  - $n_1 = \sum_{i=1}^{n} y_i$, $n_0 = n - n_1$
  - Estimate π_y by $\hat{\pi}_y = n_y/n$ $(d = 0,1)$

DIDI: doubly-integrated DI

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

- Estimation: with $\beta_1 < \beta_2 < \cdots < \beta_B$

$$\widehat{\mathrm{DIDI}}(p_{\mathrm{new}}, p_{\mathrm{old}}) = \frac{1}{B} \sum_{\ell=1}^{B} \widehat{\mathrm{IDI}}_{\beta_\ell}(p_{\mathrm{new}}, p_{\mathrm{old}})$$

Setting

$$p_{\text{ext}}(1|\boldsymbol{x}) = \begin{cases} \min(p_{\text{old}}(1|\boldsymbol{x}) + \varepsilon, 1) & \text{if } p_{\text{old}}(\boldsymbol{x}) \geq \bar{p} \\ \max(p_{\text{old}}(1|\boldsymbol{x}) - \varepsilon, 0) & \text{if } p_{\text{old}}(\boldsymbol{x}) < \bar{p} \end{cases}$$

- Objective: comparison of $\mathrm{IDI}_{\text{odds}}$ with existing indices
- Data generation
  - $p^\star(1|\boldsymbol{x}) = \operatorname{expit}(\lambda_0 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3)$
  - (x_1, x_2, x_3) are generated from 3 variate standard normal distribution
- Scenarios

(A) compare $p_{\text{old}}(1|\boldsymbol{x}) = \operatorname{expit}(\lambda_0 + \lambda_1 x_1)$ with

$$p_{\text{new}}(1|\boldsymbol{x}) = \operatorname{expit}(\lambda_0 + \lambda_1 x_1 + \lambda_2 x_2)$$

(A') compare $p_{\text{old}}(1|\boldsymbol{x}) = \operatorname{expit}(\lambda_0 + \lambda_1 x_1)$ with $p_{\text{new}} = p^\star$

- Interest: how correctly indices detect better model p_{new}

(B) compare $p_{\text{old}}(1|\boldsymbol{x}) = \operatorname{expit}(\lambda_0 + \lambda_1 x_1)$ with pnew is

$$\check{p}(1|x) = \begin{cases} 1 & \text{if } \pi_0 p^\star(1|x) > \pi_1 p^\star(0|x) \\ 1/2 & \text{if } \pi_0 p^\star(1|x) = \pi_1 p^\star(0|x) \\ 0 & \text{if } \pi_0 p^\star(1|x) < \pi_1 p^\star(0|x) \end{cases}$$

Power-IDI

$$\mathrm{IDI}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}}) = \frac{1}{\beta}\sum_{y=0}^{1} \mathrm{E}\left[\tilde{p}_{\mathrm{new}}\left(y|\boldsymbol{X}^{(y)}\right)^\beta - \tilde{p}_{\mathrm{old}}\left(y|\boldsymbol{X}^{(y)}\right)^\beta\right]$$

where $\tilde{p}_{\square}(y|\boldsymbol{x}) = \dfrac{\left(p_{\square}(y|\boldsymbol{x})/\pi_y\right)^{1/(1-\beta)}}{(p_{\square}(1|\boldsymbol{x})/\pi_1)^{1/(1-\beta)} + (p_{\square}(0|\boldsymbol{x})/\pi_0)^{1/(1-\beta)}}$

and $\beta \in (0,1)$

- IDI_β has Fisher consistency
  - Equality holds iff $p_{\mathrm{new}} \equiv p^\star$
- Relationship to power divergence: with $\gamma = \beta/(1-\beta)$

$C_\beta(f,g) \propto -\dfrac{\int f(\boldsymbol{x})^{\beta/(1-\beta)} g(\boldsymbol{x}) d\boldsymbol{x}}{\left\{\int f(\boldsymbol{x})^{1/(1-\beta)} d\boldsymbol{x}\right\}^\beta}$: power cross entropy between f & g

(Eguchi et al., 2011)

Possible interpretation 3