Boston-Keio Workshop 2016 (Probability and Statistics)

Model ~~evaluation~~comparison based on sensitivity and specificity

Kenichi Hayashi (Ken)
Keio University, Japan

August 18, 2016

Outline & contents

Outline

- Evaluation/Comparison of binary regression models
- Keywords: AUC, IDI, sensitivity, specificity
- Joint work with Dr.Eguchi (ISM)

Contents

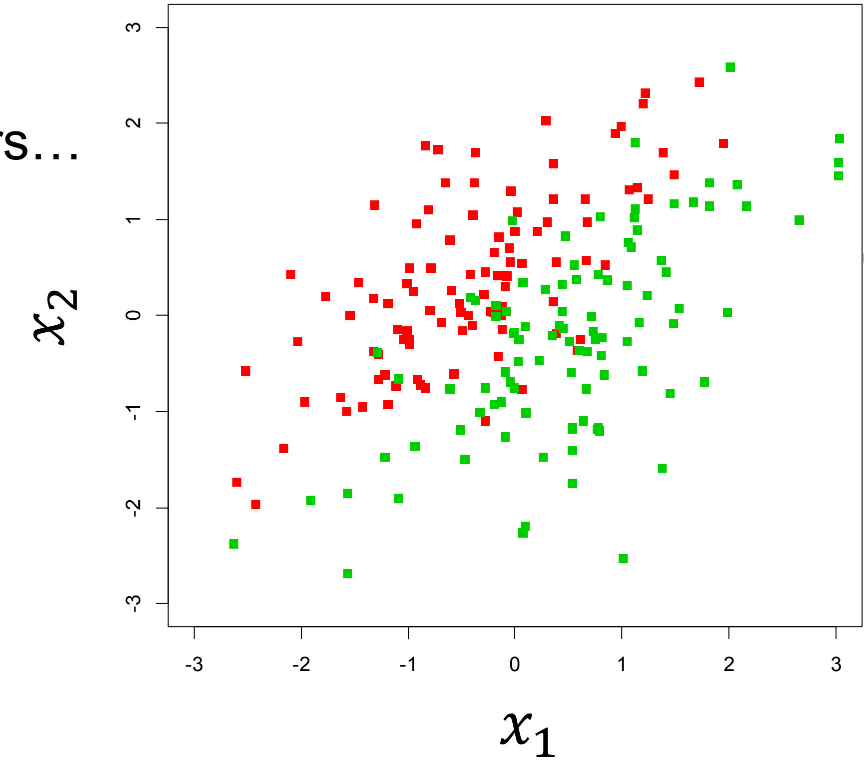
1. Introduction: binary regression models and their evaluation
2. Problems in existing indices of predictability increment
3. A solution: modification of IDI

1. Introduction

binary regression models and
their evaluation

Regression models for binary response

- D: a binary random variable
  - $D = 1$: event, $D = 0$: non-event
- $X = (X_1, \dots, X_d)'$: covariates
  - Patient characteristics, biomarkers…
- $p(x) = \mathrm{P}[D = 1 | X = x]$
- $\pi_d = \mathrm{P}[D = d]$



<u>Examples</u>

- Logistic regression model

$$p(x;\alpha) = \frac{\exp(x'\alpha)}{1 + \exp(x'\alpha)}$$

- Probit regression model

$$p(x;\alpha) = \int_{-\infty}^{x'\alpha} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$$

Objective

- Comparison of new and old models for binary response D
  - $p_{\mathrm{old}}(x) = \mathrm{P}[D = 1 | X_{\mathrm{old}} = x_{\mathrm{old}}]$: conventionally used
  - $p_{\mathrm{new}}(x) = \mathrm{P}[D = 1 | X_{\mathrm{old}} = x_{\mathrm{old}},\ X_{\mathrm{new}} = x_{\mathrm{new}}]$
    - $X = \left(X_{\mathrm{old}}^{\top},\ X_{\mathrm{new}}^{\top}\right)^{\top}$
    - X_{new}: biomarker(s) that can improve predictability
- Examples
  - Breast cancer

    Old: $\mathrm{Age} + \mathrm{BMI} + \cdots + \mathrm{Chemotherapy}$

    New: $\mathrm{Age} + \mathrm{BMI} + \cdots + \mathrm{Chemotherapy} + \mathrm{Estrogen\ reseptor}$
  - Acute coronary syndromes

    Old: $\mathrm{Age} + \mathrm{Sex} + \cdots + \mathrm{e - GFR}$

    New: $\mathrm{Age} + \mathrm{Sex} + \cdots + \mathrm{e - GFR} + \mathrm{hs - TnT}$
- Interest: quantify predictability increment by X_{new}

Model evaluation

How well a binary reg. model $p(x)$ fits to data?

For data $\{(x_i, d_i); i = 1, \dots n\}$ and $p_i = p(x_i)$

- Mean squared error (Brier score): $n^{-1} \sum_{i=1}^{n} (d_i - p_i)^2$

Other residuals

  - Deviance res. ➡ $\sqrt{-2 \log p_i}$ (if $d_i = 1$), $-\sqrt{-2 \log(1 - p_i)}$ (if $d_i = 0$)
  - Pearson res. ➡ $(d_i - p_i)/\sqrt{p_i(1 - p_i)}$
  - Anscombe res.

- Hosmer-Lemeshow test

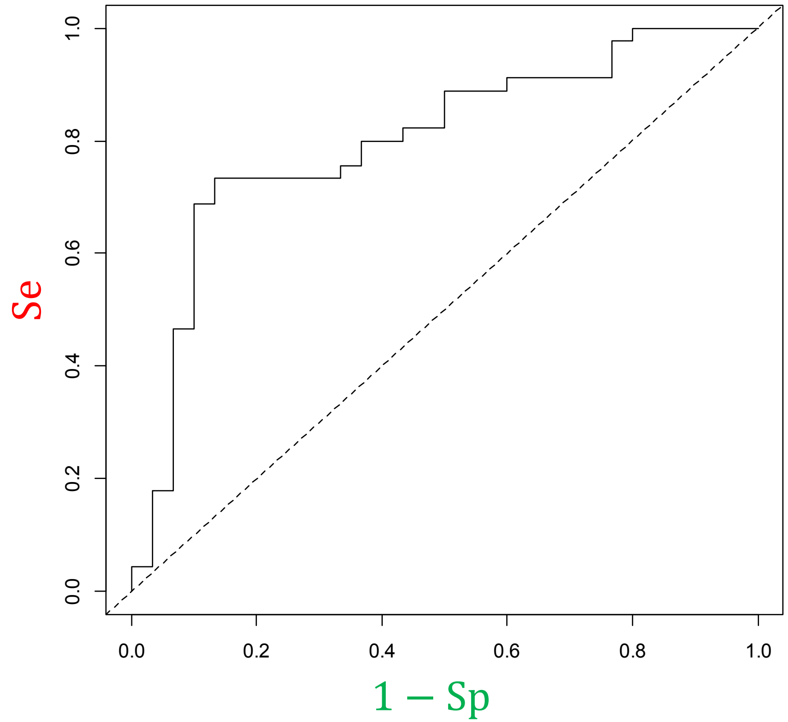
Sensitivity/Specificity: basic measures

- Sensitivity: $\mathrm{Se}(t;p) = \mathrm{P}[p(X) > t | D = 1]$
  - True positive rate (TPR)
- Specificity: $\mathrm{Sp}(t;p) = \mathrm{P}[p(X) \leq t | D = 0]$
  - False positive rate (FPR) = $1 - \mathrm{Sp}$

- ROC curve:

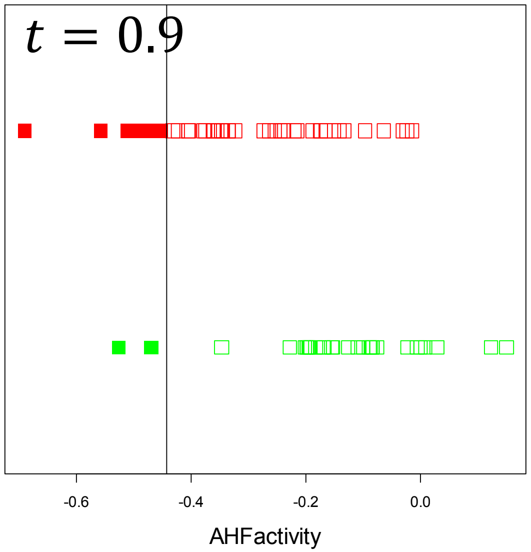
  $\{(1 - \mathrm{Sp}(t;p), \mathrm{Se}(t;p)); t \in [0,1]\}$

  - receiver operating characteristics
  - Hanley and McNeil (1982)

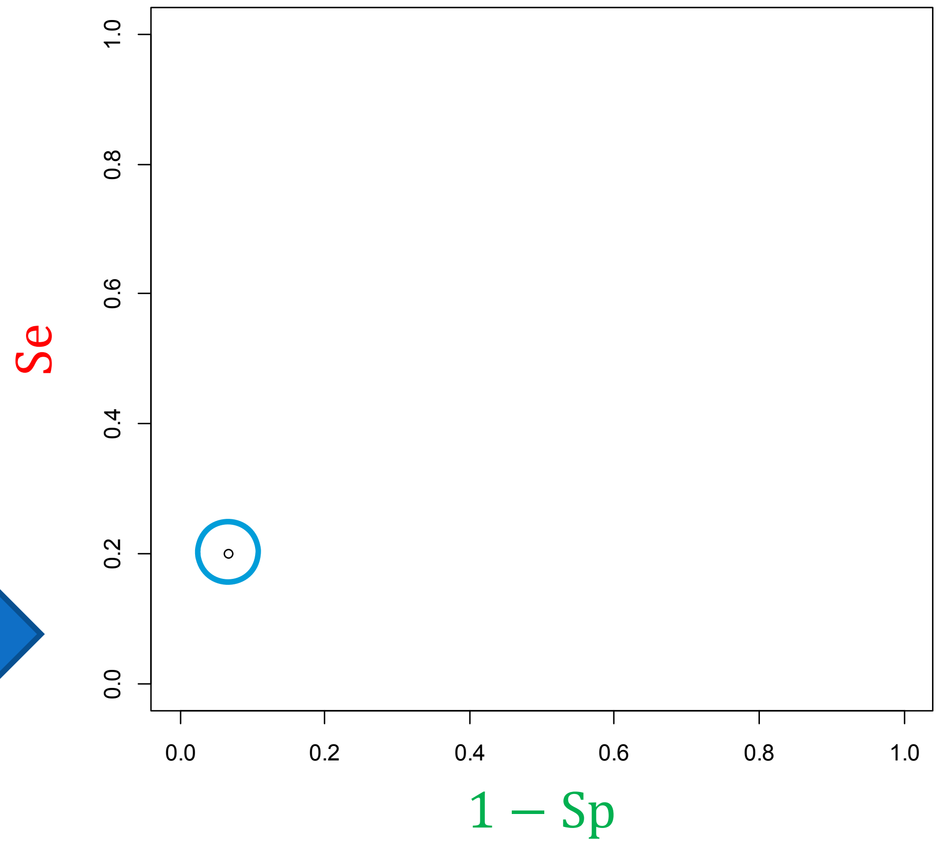


ROC curve 1/5

$$\{(1 - \mathrm{Sp}(t;p), \mathrm{Se}(t;p)); t \in [0,1]\}$$

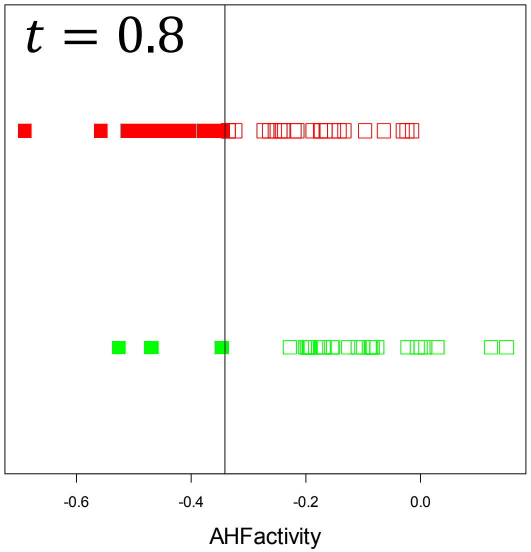


1-Sp=0.07
Se=0.20

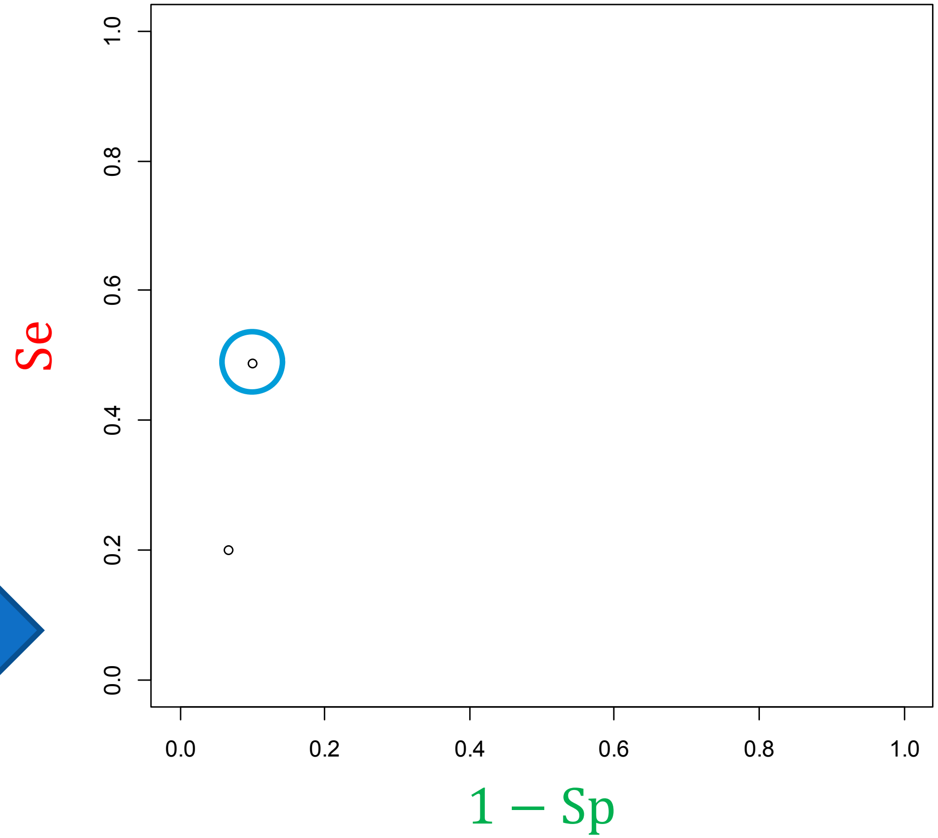


ROC curve 2/5

$$\{(1 - \mathrm{Sp}(t;p), \mathrm{Se}(t;p)); t \in [0,1]\}$$

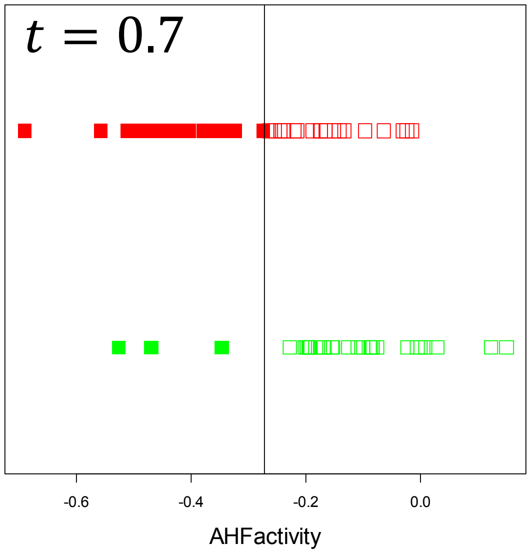


1-Sp=0.10
Se=0.49

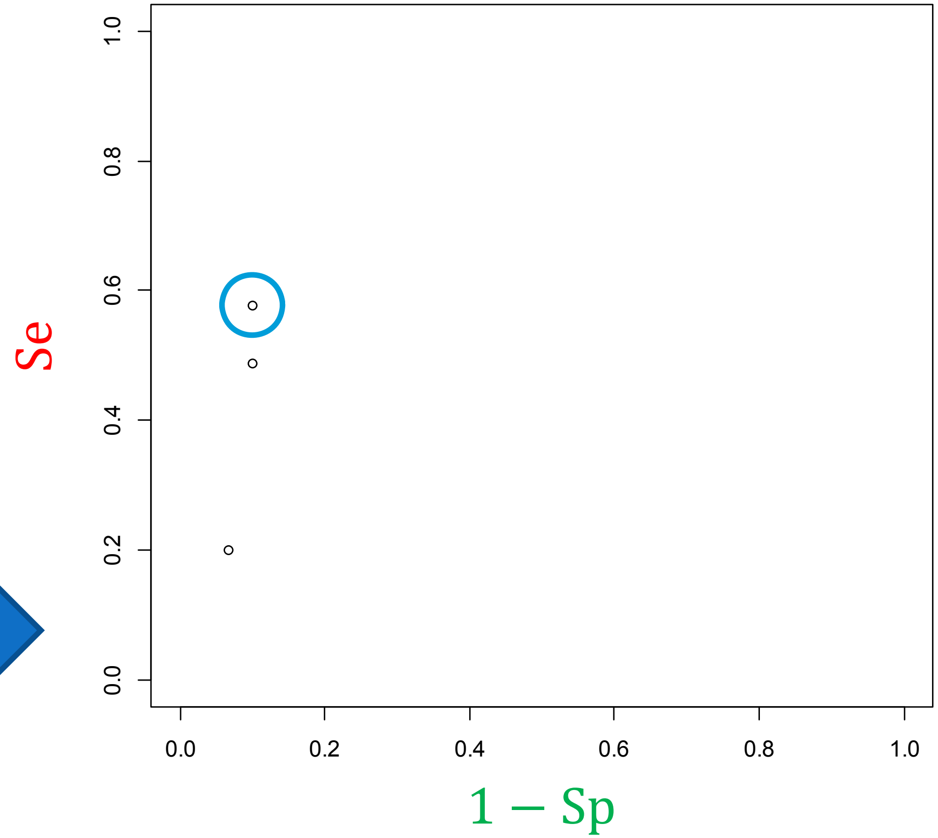


ROC curve 3/5

$$\{(1 - \mathrm{Sp}(t;p), \mathrm{Se}(t;p)); t \in [0,1]\}$$

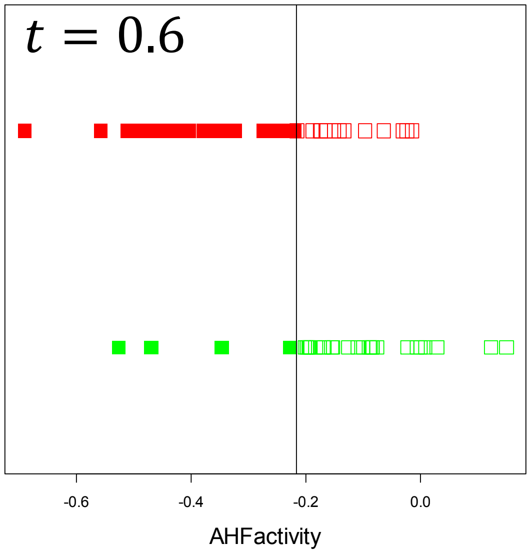


1-Sp=0.10
Se=0.58

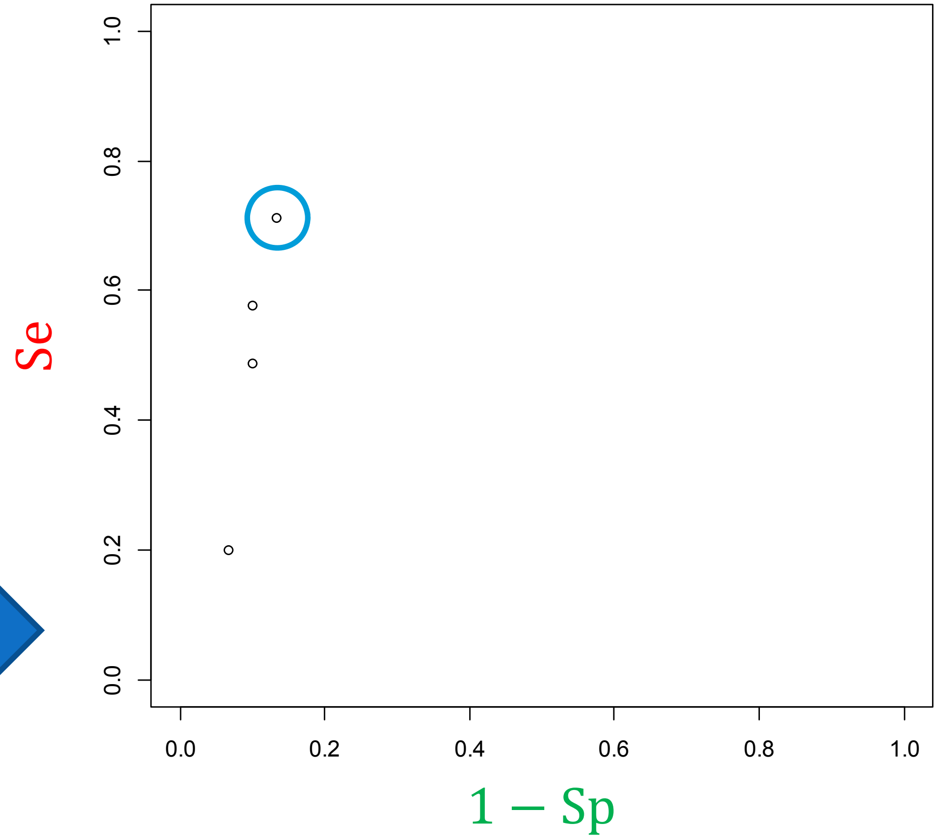


ROC curve 4/5

$$\{(1 - \mathrm{Sp}(t;p), \mathrm{Se}(t;p)); t \in [0,1]\}$$

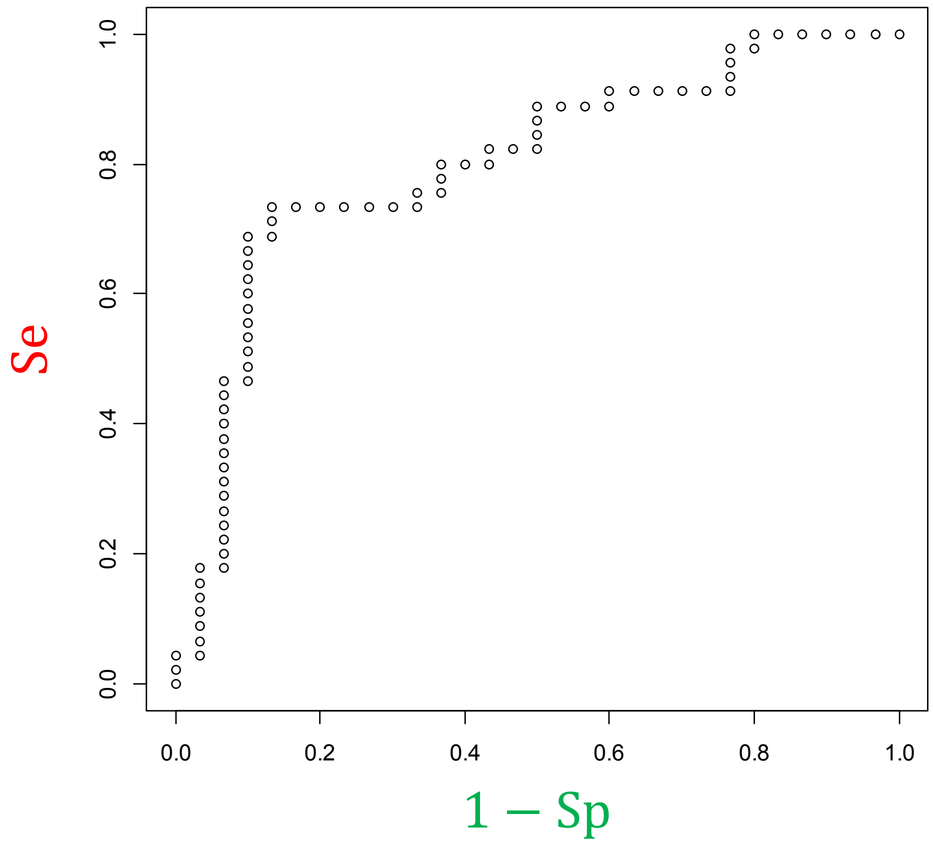


1-Se=0.13
Sp=0.71



ROC curve 5/5

$$\{(1 - \mathrm{Sp}(t;p), \mathrm{Se}(t;p)); t \in [0,1]\}$$



Related measures to sens/spec

ROC curve: $\{(1 - \mathrm{Sp}(t;p), \mathrm{Se}(t;p)); t \in [0,1]\}$

- Hit rate: $\mathrm{HR}(t;p) = \mathrm{P}[\mathbb{I}\{p(X) > t\} = D]$

$$= \pi_1 \mathrm{Se}(t;p) + \pi_0 \mathrm{Sp}(t;p)$$

- AUC: area under the ROC curve

$$\mathrm{AUC}(p) = \int_0^1 \mathrm{Se}(t;p)\, d(1 - \mathrm{Sp}(t;p))$$

$$= \mathrm{P}[p(X_1) > p(X_0)]$$

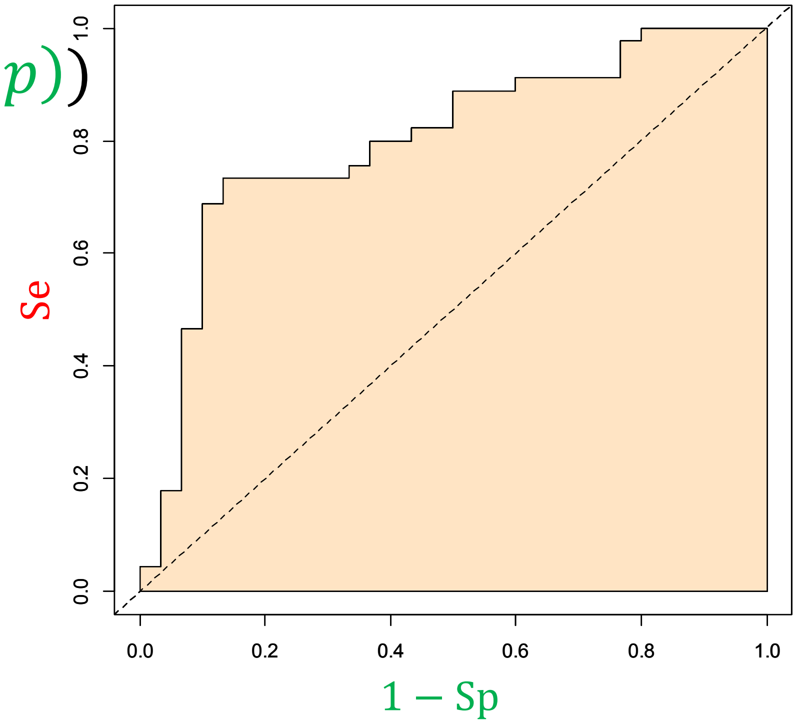
  - $X_d = X|_{D=d}, d = 1,0$

- Integrated sens/spec

$$\mathrm{IS}(p) = \int_0^1 \mathrm{Se}(t;p)dt = \mathrm{E}[p(X_1)]$$

$$= \mathrm{E}[p(X_1)]$$

$$\mathrm{IP}(p) = \mathrm{E}[1 - p(X_0)]$$



Model comparison

$p_{\text{new}}(x)$ vs. $p_{\text{old}}(x)$: is the new model better?

- AUC difference

$$\Delta\text{AUC}(p_{\text{new}}, p_{\text{old}}) = \text{AUC}(p_{\text{new}}) - \text{AUC}(p_{\text{old}})$$

- IDI: integrated discrimination improvement

$$\begin{aligned}\text{IDI}(p_{\text{new}}, p_{\text{old}}) &= \text{E}[p_{\text{new}}(X_1) - p_{\text{old}}(X_1)] \\ &\quad + \text{E}[q_{\text{new}}(X_0) - q_{\text{old}}(X_0)] \\ &= \big(\text{IS}(p_{\text{new}}) - \text{IP}(p_{\text{new}})\big) \\ &\quad - \big(\text{IS}(p_{\text{old}}) - \text{IP}(p_{\text{old}})\big)\end{aligned}$$

  - Pencina et al. (2008, 2011)

<u>Other methods</u>

  - Likelihood ratio test, information criteria (AIC, BIC,…)

Estimators

Observation: $\{(x_1, d_1), \ldots, (x_n, d_n)\}$

- AUC difference

$$\widehat{\Delta\text{AUC}}(p_{\text{new}}, p_{\text{old}}) = \frac{1}{n_1 n_0} \sum_{i,j}^{n} \mathbb{I}\{p_{\text{new}}(x_i) > p_{\text{new}}(x_j)\} d_i (1 - d_j) - \frac{1}{n_1 n_0} \sum_{i,j}^{n} \mathbb{I}\{p_{\text{old}}(x_i) > p_{\text{old}}(x_j)\} d_i (1 - d_j)$$

  - $n_1 = \sum_{i=1}^{n} d_i, \; n_0 = n - n_1$
  - Test statistic: DeLong et al. (1988)

- IDI

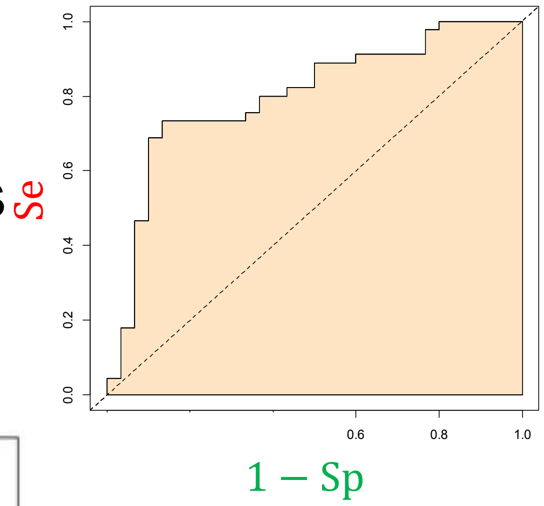
$$\widehat{\text{IDI}}(p_{\text{new}}, p_{\text{old}}) = \frac{1}{n_1} \sum_{i=1}^{n} (p_{\text{new}}(x_i) - p_{\text{old}}(x_i)) d_i + \frac{1}{n_0} \sum_{i=1}^{n} (q_{\text{new}}(x_i) - q_{\text{old}}(x_i))(1 - d_i)$$

  - Test statistic: standardized mean

2. Problems in existing indices of predictability increment

Problems in AUC differences

- Insensitive to detect a difference
  - Cook (2007)
- Distribution of a test statistics degenerates
  - Demler et al. (2012)



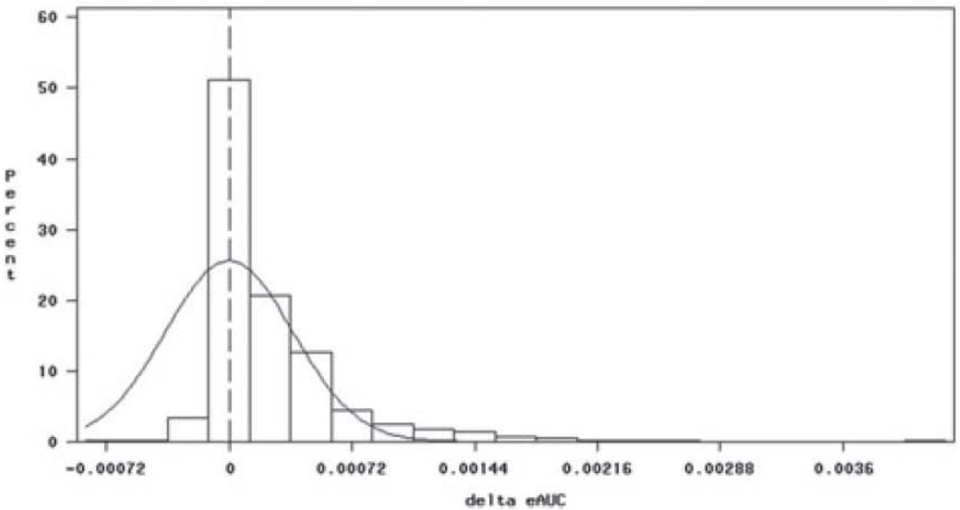


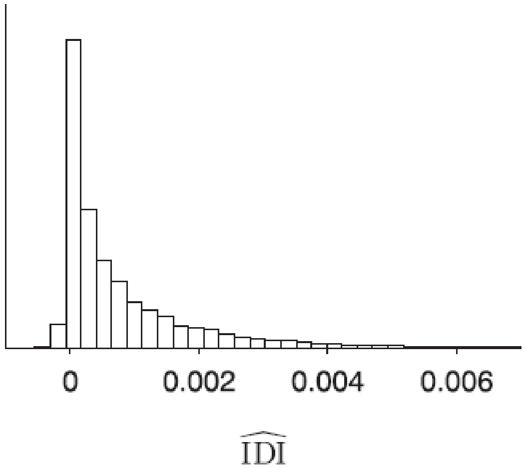
Figure 2. Histogram of change in eAUC under null hypothesis for multivariate normal data and sample size of 8365 with superimposed plot of corresponding distribution function used by DeLong test.

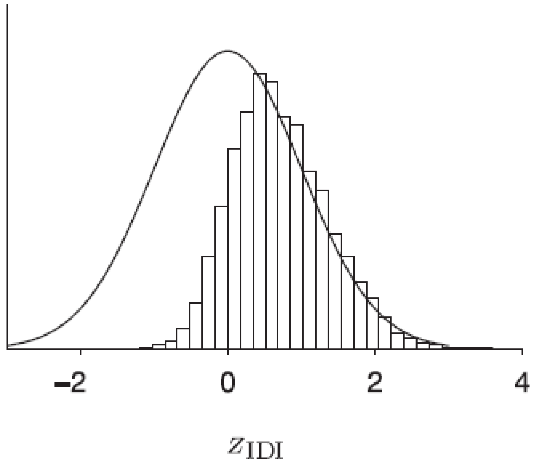
Problems in IDI

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

$$\mathrm{IDI}(p_{\mathrm{new}}, p_{\mathrm{old}}) = \mathrm{E}[p_{\mathrm{new}}(X_1) - p_{\mathrm{old}}(X_1)] + \mathrm{E}[q_{\mathrm{new}}(X_0) - q_{\mathrm{old}}(X_0)]$$

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

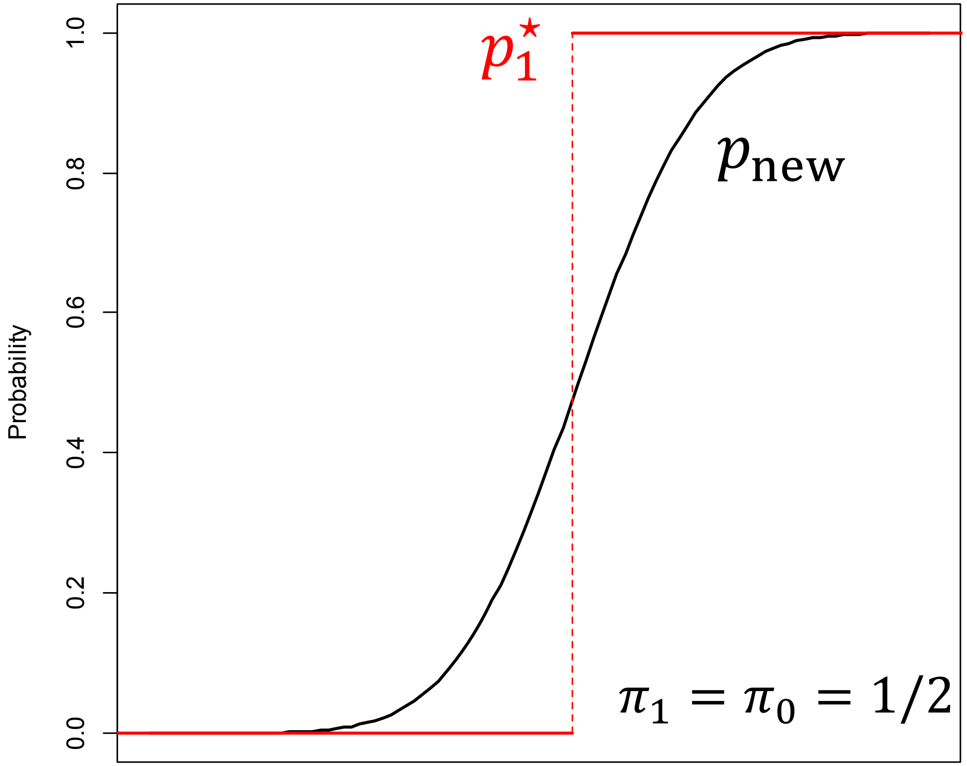




True model does not maximize IDI

- $\mathrm{IDI}(p_{\mathrm{new}}, p_{\mathrm{old}}) \leq \mathrm{IDI}(p_1^\star, p_{\mathrm{old}})$ even when p_{new} is true

$$p_1^\star(x) = \begin{cases} 1 & \text{if } p_{\mathrm{new}}(x) > \lambda q_{\mathrm{new}}(x) \\ 1/2 & \text{if } p_{\mathrm{new}}(x) = \lambda q_{\mathrm{new}}(x) \\ 0 & \text{if } p_{\mathrm{new}}(x) < \lambda q_{\mathrm{new}}(x) \end{cases} \quad \text{and} \quad \lambda = \frac{\pi_1}{\pi_0} = \frac{\mathrm{P}[D=1]}{\mathrm{P}[D=0]}$$



Proof

- Assumption: p_{new} is true

$$
\begin{aligned}
&\text{IDI}(p_1^\star, p_{\text{old}}) - \text{IDI}(p_{\text{new}}, p_{\text{old}}) \\
&= \text{E}[p_1^\star(X_1) - p_{\text{new}}(X_1)] - \text{E}[p_1^\star(X_0) - p_{\text{new}}(X_0)] \\
&= \int \left(p_1^\star(X_1) - p_{\text{new}}(X_1)\right)\left(f_1(x) - f_0(x)\right)dx \\
&= \int \left(p_1^\star(X_1) - p_{\text{new}}(X_1)\right)\left(\frac{p_{\text{new}}(x)}{\pi_1} - \frac{q_{\text{new}}(x)}{\pi_0}\right)f(x)dx \\
&= \frac{1}{\pi_1\pi_0}\int_{\frac{\pi_0 p_{\text{new}}(x)}{\pi_1 q_{\text{new}}(x)}>1} \left(1 - p_{\text{new}}(x)\right)\left(\pi_0 p_{\text{new}}(x) - \pi_1 q_{\text{new}}(x)\right)f(x)\,dx \\
&\quad + \frac{1}{\pi_1\pi_0}\int_{\frac{\pi_0 p_{\text{new}}(x)}{\pi_1 q_{\text{new}}(x)}<1} \left(-p_{\text{new}}(x)\right)\left(\pi_0 p_{\text{new}}(x) - \pi_1 q_{\text{new}}(x)\right)f(x)dx \\
&\quad \geq 0
\end{aligned}
$$

Properties for predictability increment 1/2

For an index $\mathrm{Idx}(p_{\mathrm{new}}, p_{\mathrm{old}})$ (the larger, the better)

- $p^\star$: maximizer of the index (w.r.t. 1st argument)

$$\mathrm{Idx}(p^\star, p_{\mathrm{old}}) \geq \mathrm{Idx}(p, p_{\mathrm{old}}) \text{ for any } p \text{ and } p_{\mathrm{old}}$$

1. Bayes risk consistency (population ver.)

$$\exists t \in [0,1],\ \mathrm{HR}(t; p^\star) \geq \mathrm{HR}(t; p) \text{ for any } p$$

- $p^\star$ attains the maximum hit rate
- Hit rate: $\mathrm{HR}(t;p) = \mathrm{P}[\mathbb{I}\{p(X) > t\} = D]$
- Prop. $p^\star(x)$ is proportional to $\Lambda(x) = \mathrm{P}[D = 1|x]/\mathrm{P}[D = 0|x]$
- $\Delta\mathrm{AUC}$ and IDI have BRC

Maximum hit rate

- Prop. $p^{\star}(x)$ is proportional to $\Lambda(x) = \frac{\mathrm{P}[D=1|x]}{\mathrm{P}[D=0|x]} = \frac{\pi_1 f_1(x)}{\pi_0 f_0(x)}$

  ➡ When $p(x) = \xi(\Lambda(x))$, $t = \xi^{-1}(1)$ is optimal

  - ξ: monotone increasing function

$$
\begin{aligned}
&\mathrm{HR}(\xi^{-1}(1+\varepsilon), p^{*}) \\
&= \mathrm{P}[\mathbb{I}\{p(X) > 1+\varepsilon\} = D] \\
&= \mathrm{P}[\mathbb{I}\{p(X) > 1+\varepsilon\} = D, D = 1] + \mathrm{P}[\mathbb{I}\{p(X) > 1+\varepsilon\} = D, D = 0] \\
&= \int_{\{x;\Lambda(x)>1+\varepsilon\}} \pi_1 f_1(x)\,dx + \int_{\{x;\Lambda(x)\le 1\}} \pi_0 f_0(x)\,dx + \int_{\{x;1<\Lambda(x)\le 1+\varepsilon\}} \pi_0 f_0(x) dx \\
&\le \int_{\{x;\Lambda(x)>1+\varepsilon\}} \pi_1 f_1(x)\,dx + \int_{\{x;\Lambda(x)\le 1\}} \pi_0 f_0(x)\,dx + \int_{\{x;1<\Lambda(x)\le 1+\varepsilon\}} \pi_1 f_1(x) dx \\
&= \int_{\{x;\Lambda(x)>1\}} \pi_1 f_1(x)\,dx + \int_{\{x;\Lambda(x)\le 1\}} \pi_0 f_0(x)\,dx = \mathrm{HR}(\xi^{-1}(1), p^{*})
\end{aligned}
$$

Properties for predictability increment 2/2

For an index $\mathrm{Idx}(p_{\mathrm{new}}, p_{\mathrm{old}})$ (the larger, the better)

- $p^\star$: maximizer of the index (w.r.t. 1st argument)

$$\mathrm{Idx}(p^\star, p_{\mathrm{old}}) \geq \mathrm{Idx}(p, p_{\mathrm{old}}) \text{ for any } p \text{ and } p_{\mathrm{old}}$$

2. Fisher consistency
   - $p^\star(x)$ is the true model and the inequality holds iff $p \equiv p^\star$
   - Remark: $\Delta\mathrm{AUC}$ and IDI *does not* have FC
3. Interpretability

3. A solution

modification of IDI

Modification of IDI

$$F_\beta(p_{\mathrm{new}}, p_{\mathrm{old}}) = \frac{1}{\beta}\mathrm{E}\left[p_{\mathrm{new}}(X_1)^\beta - p_{\mathrm{old}}(X_1)^\beta\right] + \frac{1}{\beta}\mathrm{E}\left[q_{\mathrm{new}}(X_0)^\beta - q_{\mathrm{old}}(X_0)^\beta\right] \quad (\beta \in (0,1])$$

- $p_\beta^\star$: maximizer of $F_\beta(p, p_{\mathrm{old}})$ w.r.t. p
  - When p_{new} is true,

$$p_\beta^\star(x) = \frac{(p_{\mathrm{new}}(x)/\pi_1)^{1/(1-\beta)}}{(p_{\mathrm{new}}(x)/\pi_1)^{1/(1-\beta)} + (q_{\mathrm{new}}(x)/\pi_0)^{1/(1-\beta)}}$$

- F_β has Bayes risk consistency
  - Maximizer is proportional to $\Lambda(x) = p_{\mathrm{new}}(x)/q_{\mathrm{new}}(x)$
- But still $F_\beta(p_{\mathrm{new}}, p_{\mathrm{old}}) \le F_\beta(p_\beta^\star, p_{\mathrm{old}})$

Further modification: our proposal

$$\mathrm{IDI}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}}) = \frac{1}{\beta}\mathrm{E}\left[\tilde{p}_{\mathrm{new}}(X_1)^\beta - \tilde{p}_{\mathrm{old}}(X_1)^\beta\right] + \frac{1}{\beta}\mathrm{E}\left[\tilde{q}_{\mathrm{new}}(X_0)^\beta - \tilde{q}_{\mathrm{old}}(X_0)^\beta\right] \quad (\beta \in (0,1])$$

where $\tilde{p}_\square(x) = \dfrac{(p_\square(x)/\pi_1)^{1/(1-\beta)}}{(p_\square(x)/\pi_1)^{1/(1-\beta)} + (q_\square(x)/\pi_0)^{1/(1-\beta)}}$

and $\tilde{q}_\square(x) = 1 - \tilde{p}_\square(x)$

- IDI_β has Fisher consistency: $\mathrm{IDI}_\beta(p, p_{\mathrm{old}}) \le \mathrm{IDI}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}})$
  - Equality holds iff $p = p_{\mathrm{new}}$
- Relation to power divergence with $\gamma = \beta/(1-\beta)$

$$C_\beta(f,g) \propto -\frac{\int f(x)^{\beta/(1-\beta)} g(x)dx}{\left\{\int f(x)^{1/(1-\beta)}dx\right\}^\beta}$$: power cross entropy for f and g

(Eguchi et al., 2011)

IDI_β: empirical version

- Observation: $\{(x_1, d_1), \ldots, (x_n, d_n)\}$

$$\widehat{\mathrm{IDI}}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}}) = \frac{1}{n_1\beta}\sum_{i=1}^{n}\left(\tilde{p}_{\mathrm{new}}(x_i)^\beta - \tilde{p}_{\mathrm{old}}(x_i)^\beta\right)d_i + \frac{1}{n_0\beta}\sum_{i=1}^{n}\left(\tilde{q}_{\mathrm{new}}(x_i)^\beta - \tilde{q}_{\mathrm{old}}(x_i)^\beta\right)(1 - d_i)$$

  - $n_1 = \sum_{i=1}^n d_i$, $n_0 = n - n_1$
  - $\hat{\pi}_d = n_d/n$ $(d = 0{,}1)$ is plugged-in to π_d in $\tilde{p}_{\mathrm{new}}$ and $\tilde{p}_{\mathrm{old}}$
  - Estimation of models p_{new} and p_{old}: maximum likelihood

Numerical experiments

- Aim: comparison of IDI_β with (A) $\Delta\mathrm{AUC}$ and (B) IDI
- Settings
  - $p_{\mathrm{old}}(x) = \mathrm{expit}(\alpha_0 + \alpha_1 x_1 + \cdots + \alpha_d x_d)$
  - $p_{\mathrm{new}}(x) = \mathrm{expit}(\alpha_0 + \alpha_1 x_1 + \cdots + \alpha_d x_d + \alpha_{\mathrm{new}} x_{\mathrm{new}})$

Comparisons

(A) $\mathrm{IDI}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}})$ VS $\Delta\mathrm{AUC}(p_{\mathrm{new}}, p_{\mathrm{old}})$

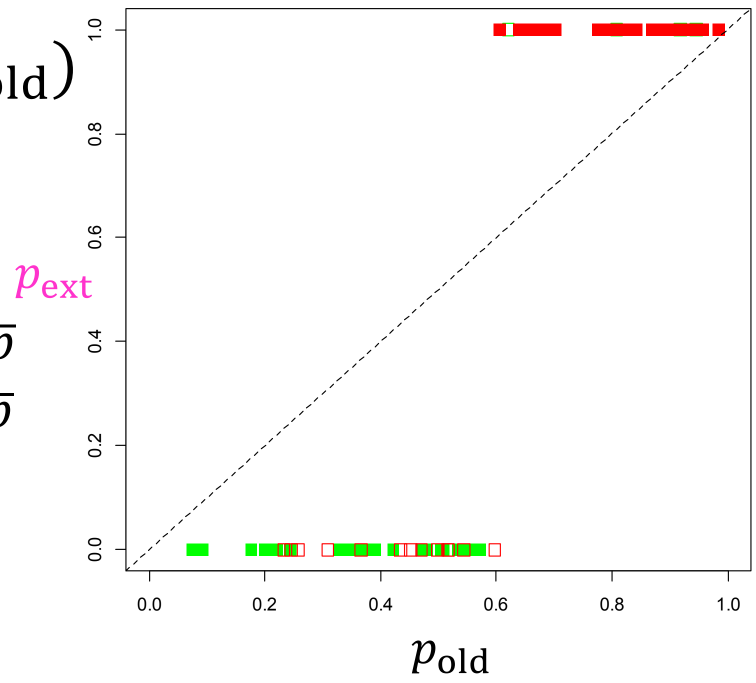
- Detection power

(B) $\mathrm{IDI}_\beta(p_{\mathrm{ext}}, p_{\mathrm{old}})$ VS $\mathrm{IDI}(p_{\mathrm{ext}}, p_{\mathrm{old}})$

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

$$\bar{p} = \frac{1}{n}\sum_{i=1}^{n} p_{\mathrm{old}}(x_i), \varepsilon \sim U(0, 0.5)$$

- False detection



Results

- $n = 100$, $d = 4$, X_d: normal dist.
  - $\mathrm{E}[X_1] = (0, \dots, 0)^\top$, $\mathrm{E}[X_0] = (0.65, -0.74, 0.48, 0.62, 0.42)^\top$
  - mimicks Framingham Heart Study (Demler et al., 2012)

(A) $\mathrm{IDI}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}})$ VS $\Delta\mathrm{AUC}(p_{\mathrm{new}}, p_{\mathrm{old}})$

	β (in IDI_β)						IDI	$2\Delta\mathrm{AUC}$
	0.0	0.2	0.4	0.6	0.8	1.0		
Value (%)	0.51	0.50	0.49	0.47	0.44	0.38	29.71	0.43
Positive rate (%)	99.30	95.64	89.11	80.17	69.31	52.65	98.69	83.03

Results

- $n = 100$, $d = 4$, X_d: normal dist.
  - $\mathrm{E}[X_1] = (0, \dots, 0)^\top$, $\mathrm{E}[X_0] = (0.65, -0.74, 0.48, 0.62, 0.42)^\top$
  - mimicks Framingham Heart Study (Demler et al., 2012)

(A) $\mathrm{IDI}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}})$ VS $\Delta\mathrm{AUC}(p_{\mathrm{new}}, p_{\mathrm{old}})$

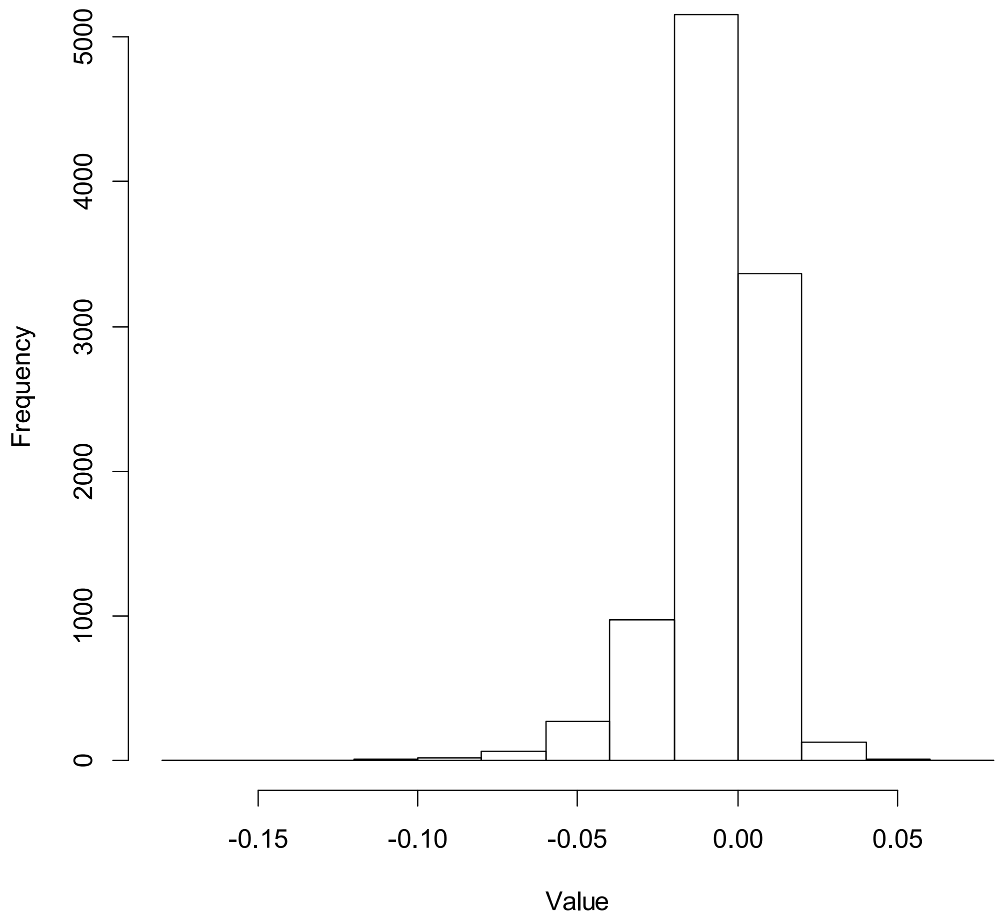
	β (in IDI_β)						IDI	2ΔAUC
	0.0	0.2	0.4	0.6	0.8	1.0		
Value (%)	0.51	0.50	0.49	0.47	0.44	0.38	29.71	0.43
Positive rate (%)	99.30	95.64	89.11	80.17	69.31	52.65	98.69	83.03

(B) $\mathrm{IDI}_\beta(p_{\mathrm{ext}}, p_{\mathrm{old}})$ VS $\mathrm{IDI}(p_{\mathrm{ext}}, p_{\mathrm{old}})$

	β (in IDI_β)						IDI	2ΔAUC
	0.0	0.2	0.4	0.6	0.8	1.0		
Value (%)	—	-105.48	-43.40	-24.45	-16.34	-12.13	31.88	-8.02
Positive rate (%)	—	0.00	0.00	0.01	1.17	8.29	99.94	1.08

Null distribution

- $n = 200, \beta = 0.3$



Summary

- IDI_β: for predictability improvement of p_{new} from p_{old}
  - Power transformation of the original IDI
  - Powerful than $\Delta\mathrm{AUC}$
  - Safer than the original IDI: avoid positive detection

- Further issues
  - Optimization w.r.t. β
  - Statistical hypothesis: non-normal (Kerr et al., 2011)
  - Extension to multi-category response

Referenes

- Cook (2007) *Circulation*
- Demler et al. (2012) *Statistics in Medicine*
- Eguchi et al. (2011) *Entropy*
- Hilden and Gerds (2014) *Statistics in Medicine*
- Pencina et al. (2008; 2011) *Statistics in Medicine*

Thank you for your attention

Backup slides

NRI; net reclassification improvement

- Also proposed in Pencina et al. (2008,2011)

$$\mathrm{NRI}(p_{\mathrm{new}}, p_{\mathrm{old}}) = \mathrm{E}[\mathrm{sign}(p_{\mathrm{new}}(X_1) - p_{\mathrm{old}}(X_1))] + \mathrm{E}[\mathrm{sign}(q_{\mathrm{new}}(X_0) - q_{\mathrm{old}}(X_0))]$$

  - $q_{\square}(x) = 1 - p_{\square}(x)$
  - $X_d \sim X$ under $D = d$ $(d = 1,0)$
- NRI does not have Fisher consistency
- Modification? ➡ No

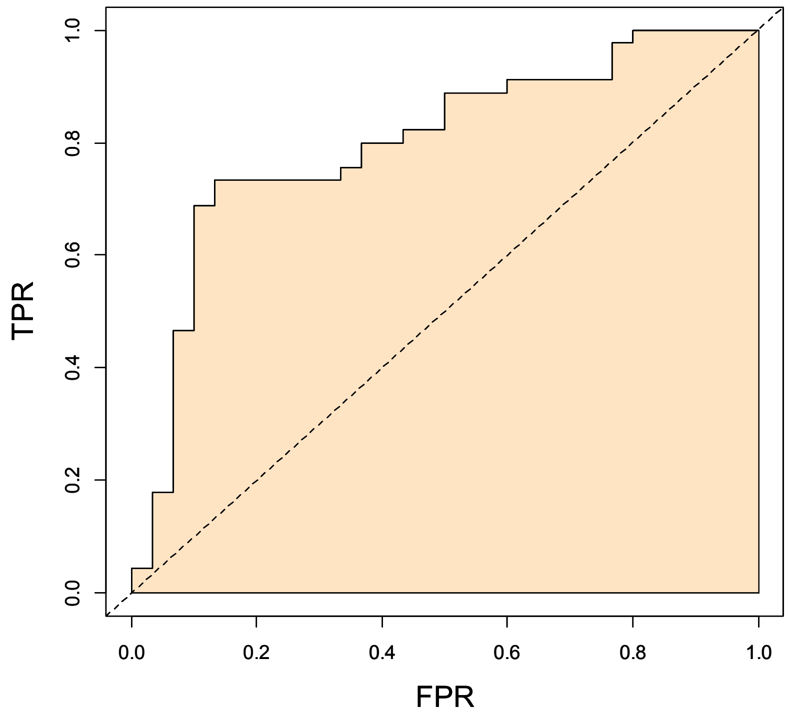
AUC difference

- An index of improvement based on TPR & FPR:

$$\Delta\mathrm{AUC}(p_{\mathrm{new}}, p_{\mathrm{old}}) = \mathrm{AUC}(p_{\mathrm{new}}) - \mathrm{AUC}(p_{\mathrm{old}})$$

where $\mathrm{AUC}(p) = \int_0^1 \mathrm{TPR}(t;p)d\mathrm{FPR}(t;p)$

  - AUC: area under the ROC curve
- Problems
  - Insensitive to the difference
    - Cook (2007)
  - Test unsuitable for nested models
    - $p_{\mathrm{old}} \subset p_{\mathrm{new}}$
    - Demler et al. (2012)



IDI: integrated discrimination improvement

- An index of improvement based on TPR & FPR:

$$\mathrm{IDI}(p_{\mathrm{new}}, p_{\mathrm{old}}) = \mathrm{E}[p_{\mathrm{new}}(X_1) - p_{\mathrm{old}}(X_1)]$$
$$+\mathrm{E}[q_{\mathrm{new}}(X_0) - q_{\mathrm{old}}(X_0)]$$

where $X_d = X|_{D=d}$, $q_\clubsuit(x) = 1 - p_\clubsuit(x)$

- Based on integrated T/FPR: $\int_0^1 \mathrm{TPR}(t; p_{\mathrm{new}})dt = \mathrm{E}[p_{\mathrm{new}}(X_1)]$
- Problem
  - "…*use is not always safe*" (Hilden and Gerds, 2014)
  - Model can be improved without adding measured information
  - Even when p_{new} is true, $\mathrm{IDI}(p_{\mathrm{new}}, p_{\mathrm{old}}) \leq \mathrm{IDI}(p_1^\star, p_{\mathrm{old}})$

where $p_1^\star(x) = \begin{cases} 1 & \text{if } p_{\mathrm{new}}(x) > \lambda q_{\mathrm{new}}(x) \\ 1/2 & \text{if } p_{\mathrm{new}}(x) = \lambda q_{\mathrm{new}}(x) \\ 0 & \text{if } p_{\mathrm{new}}(x) < \lambda q_{\mathrm{new}}(x) \end{cases}$ and $\lambda = \dfrac{\pi_1}{\pi_0} = \dfrac{\mathrm{P}[D=1]}{\mathrm{P}[D=0]}$

Results (2)

- $n = 100$, $d = 2$, X_d: t-dist. (d.f.=4)

(1) $\mathrm{IDI}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}})$ VS $\Delta\mathrm{AUC}(p_{\mathrm{new}}, p_{\mathrm{old}})$

	β (in IDI_β)						IDI	$2\Delta\mathrm{AUC}$
	0.0	0.2	0.4	0.6	0.8	1.0		
Value (%)	-0.16	0.29	0.68	1.02	1.21	0.76	1.47	0.03
Positive rate (%)	55.23	57.97	60.84	62.81	62.23	50.33	77.28	55.12

(2) $\mathrm{IDI}_\beta(p_{\mathrm{ext}}, p_{\mathrm{old}})$ VS $\mathrm{IDI}(p_{\mathrm{ext}}, p_{\mathrm{old}})$

	β (in IDI_β)						IDI	$2\Delta\mathrm{AUC}$
	0.0	0.2	0.4	0.6	0.8	1.0		
Value (%)	—	-151.82	-67.69	-39.84	-26.91	-22.84	0.08	-6.05
Positive rate (%)	—	0.07	0.36	1.29	2.99	3.26	50.99	14.69

Power divergence with $\gamma = \beta/(1-\beta) > 0$

$$C^{\gamma}(f,g) = -\frac{\int f(x)^{\gamma} g(x)dx}{\left\{\int f(x)^{1+\gamma} dx\right\}^{\gamma/(1+\gamma)}} = -\frac{\int f(x)^{\beta/(1-\beta)} g(x)dx}{\left\{\int f(x)^{1/(1-\beta)} dx\right\}^{\beta}}$$

Further modification: our proposal

$$\mathrm{IDI}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}}) = \frac{1}{\beta}\mathrm{E}\left[\tilde{p}_{\mathrm{new}}(X_1)^\beta - \tilde{p}_{\mathrm{old}}(X_1)^\beta\right] + \frac{1}{\beta}\mathrm{E}\left[\tilde{q}_{\mathrm{new}}(X_0)^\beta - \tilde{q}_{\mathrm{old}}(X_0)^\beta\right] \quad (\beta \in (0,1])$$

- $\tilde{p}_\square(x) = \frac{(p_\square(x)/\pi_1)^{1/(1-\beta)}}{(p_\square(x)/\pi_1)^{1/(1-\beta)} + (q_\square(x)/\pi_0)^{1/(1-\beta)}}$, $\tilde{q}_\square(x) = 1 - \tilde{p}_\square(x)$
- Remark: simple transformation is not enough

$$F_\beta(p_{\mathrm{new}}, p_{\mathrm{old}}) = \frac{1}{\beta}\mathrm{E}\left[p_{\mathrm{new}}(X_1)^\beta - p_{\mathrm{old}}(X_1)^\beta\right] + \frac{1}{\beta}\mathrm{E}\left[q_{\mathrm{new}}(X_0)^\beta - q_{\mathrm{old}}(X_0)^\beta\right]$$

- $p_\beta^\star$: maximizer of $F_\beta(p, p_{\mathrm{old}})$ w.r.t. p
  - When p_{new} is true,

Properties of IDI_β

- IDI_β has Fisher consistency: $\mathrm{IDI}_\beta(p, p_{\mathrm{old}}) \le \mathrm{IDI}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}})$
  - Equality holds iff $p = p_{\mathrm{new}}$
- Relation to ~~β-divergence~~power divergence with $\gamma = \beta/(1-\beta)$

$$C_\beta(f,g) = -\frac{\int f(x)^{\beta/(1-\beta)} g(x)dx}{\left\{\int f(x)^{1/(1-\beta)}dx\right\}^\beta}$$: cross entropy for f and g

(Eguchi et al., 2011)

- Empirical version: for $\{(x_1, d_1), \ldots, (x_n, d_n)\}$

$$\widehat{\mathrm{IDI}}_\beta(p_{\mathrm{new}}, p_{\mathrm{old}}) = \frac{1}{n_1\beta}\sum_{i=1}^{n}\left(\tilde{p}_{\mathrm{new}}(x_i)^\beta - \tilde{p}_{\mathrm{old}}(x_i)^\beta\right)d_i + \frac{1}{n_0\beta}\sum_{i=1}^{n}\left(\tilde{q}_{\mathrm{new}}(x_i)^\beta - \tilde{q}_{\mathrm{old}}(x_i)^\beta\right)(1-d_i)$$

  - $n_1 = \sum_{i=1}^n d_i$, $n_0 = n - n_1$
  - $\hat{\pi}_d$ $(d = 0,1)$ is plugged-in to π_d in $\tilde{p}_{\mathrm{new}}$ and $\tilde{p}_{\mathrm{old}}$
  - Estimation of models p_{new} and p_{old}: maximum likelihood