Chapter 7

Within the sample Comparison of prediction performances of models and sub-models. Application to Alzheimer disease.

Abstract. Our objective is to compare the predictive ability of several nested models. It stems from the following problem in epidemiology: the occurrence of a certain disease is to be predicted to happen within a fixed period of time thanks to the values of a number of items measured on the observed patients. It may happen that one or several items, proved to be relevant for the best fitting model, have a non significant contribution to the prediction of who is at risk of developing the disease. The indices we use to compare the respective predictive ability of two models are the Integrated Discrimination Improvement (IDI) and the BRier's score Improvement (BRI). Estimation of the models and their relative IDI and BRI are conducted on the same sample, and their respective asymptotic properties are proved. We apply the results to Alzheimer disease.

7.1. Introduction

When the objective of modeling a data set is explanatory it is most appropriate to choose the best fitting model using the usual model selection procedures. But if the objective is to predict and not to explain the facts, and some of the factors selected by the "best fitting model" are not available for all subjects, as is sometimes the case for certain genetic markers, one can compare models by their prediction qualities rather than by their goodness of fit to the data. In this paper, we consider the case of two competing, nested, probability predicting models, similar

---

Chapter written by Catherine HUBER-CAROL and Shulamith TOAFF-GROSS and Annick ALPÉROVITCH .

to Whittemore's (2010) example, and consistent with Pencina et al. (2008) setting, [M.J 08]. The nested model contains traditional factors, and the larger model contains in addition some expensive, or generally hard to obtain, often genetic, relevant markers.

7.2. Framework

7.2.1. *General description of the data set and the models to be compared*

The data set is as follows. $X = (Y, \mathbf{Z})$ is a random variable such that the response variable Y is binary with values in $\{0, 1\}$, and $\mathbf{Z}$ is a k-dimensional real variable. Observed are $\mathbf{X} = (X_1, \cdots, X_n)$, n i.i.d. observations of X, and two models for predicting Y on the basis of $\mathbf{Z}$ are to be compared:

$$\begin{array}{lrcl} \text{Model 1} & P(Y=1|\mathbf{Z}=\mathbf{z}) & = & p_1(\mathbf{z}) \\ \text{Model 2} & P(Y=1|\mathbf{Z}=\mathbf{z}) & = & p_2(\mathbf{z}) \end{array}$$

while the true distribution of Y given $Z = z$, which will remain unknown all along, is given by

$$P(Y=1|\mathbf{Z}=\mathbf{z}) \quad = \quad p(\mathbf{z})$$

This setting originates from a problem in epidemiology: Y_i is the indicator of the occurrence of a specfic disease for subject i within a given period of time. The prediction of occurrence of this event is based on the value $\mathbf{z_i}$ of $\mathbf{Z}$ observed on subject i. In the special case of linear logistic models, p_1 and p_2 are denoted g_1 and g_2 in the sequel. While g_1 is including all k components of $\mathbf{Z}$, g_2 is obtained by dropping k" components of $\mathbf{Z}$, keeping thus only $k' = k - k" < k$ of them. Without restriction of the generality, we treat in detail the case when one drops only one factor. The theoretical aim of this work is to derive the asymptotic properties of the estimators of the IDI and the BRI in order to obtain for them confidence intervals within the sample used to estimate the two models.

7.2.2. *Definition of the performance prediction criteria: IDI and BRI*

From Pencina et al (op. cit.), the IDI of model 2 with respect to model 1 is

$$\begin{array}{rcl} IDI_{2/1} & = & E[p_2(\mathbf{Z}) - p_1(\mathbf{Z})|Y=1] \\ & & -E[p_2(\mathbf{Z}) - p_1(\mathbf{Z})|\mathbf{Y}=\mathbf{0}] \end{array}$$

where E denotes the expectation with respect to the distribution of X. We denote

$$\pi \quad = \quad P(Y=1) \quad := \quad E[p(\mathbf{Z})]\ ,$$

the population prevalence of the event under study. We derive a simpler expressions for $IDI_{2/1}$ to be used later:

$$IDI_{2/1} = E\left[(p_2(\mathbf{Z}) - p_1(\mathbf{Z}))\ (\frac{Y-\pi}{\pi\,(1-\pi)})\right] \tag{7.2.1}$$

Gu and Pepe, [GU 09], define the PEV (proportion of explained variance) of a model :

$$PEV = \frac{var(P(Y=1|Z))}{\pi\,(1-\pi)} \tag{7.2.2}$$

from which it is clear that:

$$IDI_{2/1} = PEV_2 - PEV_1 \tag{7.2.3}$$

For a single model p the Brier score is defined as

$$BR(p) = E[(Y - p(\mathbf{Z}))^2] \tag{7.2.4}$$

As the bigger the Brier's score the worst is the model, we define the BRier's score improvement for model 2 with respect to model 1 as

$$\begin{aligned} BRI_{2/1} = & \quad BR(p_1) - BR(p_2) \\ = & \quad E[(p_1(\mathbf{Z}) - p_2(\mathbf{Z}))\times \\ & \quad (p_1(\mathbf{Z}) + p_2(\mathbf{Z}) \ - \ 2\,Y)] \end{aligned} \tag{7.2.5}$$

A negative $IDI_{2/1}$ as well as a negative $BRI_{2/1}$ means that the predictive properties of model 2 are not as good as those of model 1. The ranges of IDI and BRI are respectively $-2\,,\ +2$ and $-1\,,\ +1$.

7.3. Estimation of IDI and BRI

Now we assume that we have a sample of size n of X, and the two prediction models $p_j(\theta_j, \mathbf{z})$, defined for $j = 1, 2$ through the respective parameters $\theta_{\mathbf{1}}, \theta_{\mathbf{2}}$, as:

$$\begin{array}{rcl} P(Y = 1|\text{ model 1}) & = & p_1(\theta_{\mathbf{1}}, \mathbf{z}) \\ P(Y = 1|\text{ model 2}) & = & p_2(\theta_{\mathbf{2}}, \mathbf{z}) \end{array}$$

7.3.1. *General estimating equations for IDI and BRI*

Using (7.2.1) and (7.2.5), and denoting for simplicity

$$\widehat{p_{ji}} := p_j(\widehat{\theta_j}, \mathbf{z_i}) \quad \text{for } \mathbf{j = 1, 2, i = 1, \cdots, n}. \tag{7.3.6}$$

where $\widehat{\theta_j}, j = 1, 2$ are maximum likelihood estimates for the parameters of models 1 and 2, natural estimates of $IDI_{2/1}$ and $BRI_{2/1}$ are respectively

$$\widehat{IDI_{2/1}} \ = \ \frac{1}{n}\sum_{i=1}^{n}(\widehat{p_{2i}} - \widehat{p_{1i}})\frac{y_i - \overline{y}}{\overline{y}(1-\overline{y})} \tag{7.3.7}$$

$$\widehat{BRI_{2/1}} \ = \ \frac{2}{n}\sum_{i=1}^{n}[(\widehat{p_{1i}} - \widehat{p_{2i}})(\frac{(\widehat{p_{1i}} + \widehat{p_{2i}})}{2} \ - \ Y)] \tag{7.3.8}$$

Under usual regularity conditions on models 1 and 2, their parameters estimates are asymptotically consistent and normal.

7.3.2. *Estimation of IDI and BRI in the logistic case*

Let $u =< \theta, \mathbf{z} >$ be the scalar product of two $k+1$ dimensional real vectors $\theta = (\theta_0, \cdots, \theta_k)$, $\mathbf{z} = (1, z_1, \cdots, z_k)$) and g the function

$$g(u) = \frac{e^u}{1+e^u} \tag{7.3.9}$$

Models 1 and 2 are logistic : $p_1 \equiv g_1$ and $p_2 \equiv g_2$:

$$\begin{array}{rcl} g_1(\mathbf{z}) & = & g(< \theta_{\mathbf{1}}, \mathbf{z} >) \\ g_2(\mathbf{z}) & = & g(< \theta_{\mathbf{2}}, \mathbf{z} >) \end{array}$$

where some components of $\theta_{\mathbf{1}}$ and some (possibly different) components of $\theta_{\mathbf{2}}$ are predefined. If we refer to the motivating example of the introduction :

$$\begin{array}{rcl} \theta_{\mathbf{1}} & = & (\theta_0, \theta_1, \cdots, \theta_{k-1}, \theta_k) \\ \theta_{\mathbf{2}} & = & (\theta'_0, \theta'_1, \cdots, \theta'_{k-1}, 0) \end{array}$$

Dropping the index j for simplicity of notation, we get the log-likelihood L_n:

$$\begin{aligned} L_n(\theta) &= \frac{1}{n}\sum_{i=1}^{n} y_i \log(g(< \theta, \mathbf{z_i} >)) \\ &\quad +(1-y_i)\log(1-g(< \theta, \mathbf{z_i} >)) \\ &= \frac{1}{n}\sum_{i=1}^{n} [y_i < \theta, \mathbf{z_i} > - \log(1+e^{<\theta, \mathbf{z_i}>})] \end{aligned} \tag{7.3.10}$$

Then we estimate the parameters θ_1 and θ_2 of the two logistic models. Those two estimators, obtained through classical maximum likelihood equations, are proved to be consistent and asymptotically normal.

7.3.2.1. *Asymptotics of $\widehat{IDI_{2/1}}$ for logistic predictors*

For simplicity, and since there is no ambiguity as we always consider how model 2 behaves with respect to model 1, we now drop the index $2/1$.
Define $\widetilde{IDI}$:

$$\widetilde{IDI} = \frac{1}{n}\sum_{i=1}^{n}(g_{2i} - g_{1i})\frac{y_i - \overline{y}}{\overline{y}(1-\overline{y})} \tag{7.3.11}$$

which is the estimate of IDI when the two models are perfectly known, or more realistically, estimated "out of the sample", and $\widehat{g_{ji}}$:

$$\widehat{g_{ji}} = g(< \widehat{\theta_j}, z_i >) \quad , \quad j = 1, 2 \tag{7.3.12}$$

the estimates of the two models. From equation (7.3.7), by considering $(\widehat{IDI} - \widetilde{IDI}) + \widetilde{IDI}$, we get

$$\begin{aligned}
\widehat{IDI} &= \frac{1}{n}\sum_{i=1}^{n}\left[(\hat{g}_{2i} - \hat{g}_{1i}) - (g_{2i} - g_{1i})\right]\frac{y_i - \overline{y}}{\overline{y}(1-\overline{y})} \\
&+ \frac{1}{n}\sum_{i=1}^{n}[g_{2i} - g_{1i}]\,\frac{y_i - \overline{y}}{\overline{y}(1-\overline{y})} \\
&:= T_{1n} + \widetilde{IDI}
\end{aligned}$$

THEOREM.– *[Consistency of $\widehat{IDI}$]* $\widehat{IDI} \xrightarrow[n\to\infty]{a.s.} IDI.$

Theorem 7.3.2.1 results from the almost sure convergence of T_{1n} to 0, due to the CLT for the model estimators $\widehat{g_{ji}}$ and the boundedness of the derivative of g, $(g(u)(1-g(u))u')$, and the fact that

$$\widetilde{IDI} \xrightarrow[n\to\infty]{a.s.} IDI$$

THEOREM.– *[CLT of $\widehat{IDI}$]*

$$\sqrt{n}(\widehat{IDI} - IDI) \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}(0, \sigma^2). \tag{7.3.13}$$

where $\sigma^2 = (\frac{1}{(1-\pi)\,\pi})^2\, var(V)$ with V defined as

$$\begin{aligned}
V &= (g(\theta_2, \mathbf{Z_2}) - g(\theta_1, \mathbf{Z_1}) - E_\Delta)(Y - \pi) \\
&+ (Y - g(\theta_2, \mathbf{Z_2}))\; {}^t(\mathbf{Z_2})(I^{-1}(\theta_2))E_2 \\
&- (Y - g(\theta_1, \mathbf{Z_1}))^t(\mathbf{Z_1})(I^{-1}(\theta_1))E_1 \\
&+ IDI\,(2\pi - 1)\,(Y - \pi)
\end{aligned}$$

where

$$\begin{aligned}
E_\Delta &= E(g(\theta_2, \mathbf{Z_2}) - g(\theta_1, \mathbf{Z_1})) \\
E_j &= E[g_j(1-g_j)(Y-\pi)\mathbf{Z_j}] \;\;,\;\; j = 1, 2 \\
\widehat{\sigma^2} &= (\frac{1}{\overline{Y}(1-\overline{Y})})^2\, \widehat{var(V)}
\end{aligned}$$

where $\widehat{\sigma^2}$ is a consistent estimator of σ^2.

7.3.2.2. *Asymptotics of $\widehat{BRI}_{2/1}$ for logistic predictors*

Again, we drop the index 2/1, and consider the estimated BRI (7.3.8).

$$\begin{aligned}\widehat{BRI} &= \frac{2}{n}\sum_{i=1}^{n}[(g(<\hat{\theta}_1, Z_i>) - g(<\hat{\theta}_2, Z_i>))\times \\ &\times \left(\frac{(g(<\hat{\theta}_1, Z_i>) + g(<\hat{\theta}_2, Z_i>))}{2} - Y_i\right)]\end{aligned}$$

Using the same method as for IDI, we get

$$\sqrt{n}(\widehat{BRI} - BRI) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(W_i - BRI) + o_P(1)$$

where the random variables W_i are defined as

$$\begin{aligned}W_i &= 2[(g(<\widehat{\theta_1}, \mathbf{Z}_i>) - g(<\widehat{\theta_2}, \mathbf{Z}_i>))\times \\ &\times \left(\frac{(g(<\widehat{\theta_2}, \mathbf{Z}_i>) + g(<\widehat{\theta_1}, \mathbf{Z}_i>))}{2} - Y_i\right)] \\ &- (Y_i - g(<\widehat{\theta_2}, \mathbf{Z}_{2i}>))^t Z_{2i} I^{-1}(\widehat{\theta_2})(2E_2 - E_4) \\ &+ (Y_i - g(<\widehat{\theta_1}, \mathbf{Z}_{1i}>))^t Z_{1i} I^{-1}(\widehat{\theta_1})(2E_1 - E_3)\end{aligned}$$

where the expectations $E_j, j = 1, \cdots, 4$ are defned as

$$\begin{aligned}E_1 &= E[g(<\theta_1, \mathbf{Z}_1>)(1 - g(<\theta_1, \mathbf{Z}_1>))\mathbf{Z}_1\times \\ &\times \left(\frac{g(<\theta_2, \mathbf{Z}_2>) + g(<\theta_1, \mathbf{Z}_1>)}{2} - Y\right)] \\ E_2 &= E[g(<\theta_2, \mathbf{Z}_2>)(1 - g(<\theta_2, \mathbf{Z}_2>))\mathbf{Z}_2\times \\ &\times \left(\frac{(g(<\theta_1, \mathbf{Z}_1>) + g(<\theta_2, \mathbf{Z}_2>))}{2} - Y\right))] \\ E_3 &= E[g(<\theta_1, \mathbf{Z}_1>)(1 - g(<\theta_1, \mathbf{Z}>))\mathbf{Z}_1(g(<\theta_2, \mathbf{Z}_1>) \\ &\quad -g(<\theta_1, \mathbf{Z}_1>)) \\ E_4 &= E[g(<\theta_2, \mathbf{Z}>)(1 - g(<\theta_2, \mathbf{Z}_2>))\mathbf{Z}_2(g(<\theta_2, \mathbf{Z}_2>) \\ &\quad -g(<\theta_1, \mathbf{Z}_2>))\end{aligned}$$

Actually, we have,

THEOREM.– *[Consistency of the estimated BRI]*

$$\widehat{BRI} \xrightarrow[n\to\infty]{a.s.} BRI \tag{7.3.14}$$

THEOREM.– *[CLT for $\widehat{BRI}$]*

$$\sqrt{n}(\widehat{BRI} - BRI) \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}(0, \sigma_B^2). \tag{7.3.15}$$

A consistent estimate of $\sigma_B^2 = var(W)$ is given by

$$\widehat{\sigma_B^2} = \frac{1}{n-1}\sum_{i=1}^{n}(\widehat{W_i} - \hat{\overline{W}})^2.$$

The estimated W's are obtained by replacing all parameters by their maximum likelihood estimates and expectations by sample averages.

7.4. Simulation studies

We do two different simulations: one to check the behaviour of our estimates of IDI and BRI, and the the other parallel to Gu and Pepe's.

7.4.1. *First simulation*

Our basic model to simulate the data is a logistic one with two independent covariates: $Z_1 \in \{-1, 0, 1\}$ with respective probabilities $(.2, .4, .4)$, and $Z_2 \in \{0, 1\}$ with respective probabilities $(.2, .8)$, with $\theta = (0, 2, 1)$ so that

$$\mathscr{L}(Y_i|Z_i) = \text{Bernoulli}(\frac{\exp(2Z_{1i} + Z_{2i})}{1 + \exp(2Z_{1i} + Z_{2i})}).$$

This is the true model used to simulate the data.
We generate 1000 samples of 1000 triplets (Y_i, Z_{1i}, Z_{2i}). On each sample thus obtained, three logistic models are considered : model 1 including both covariates, $\theta_{\mathbf{1}} = (\theta_{10}, \theta_{11}, \theta_{12})$, model 2 including only Z_1, $\theta_{\mathbf{2}} = (\theta_{20}, \theta_{21}, 0)$ and model 2' including only Z_2, $\theta_{\mathbf{2}'} = (\theta_{2'0}, 0, \theta_{2'2})$. We know the true values of $\theta_{\mathbf{1}}$, $\theta_{\mathbf{2}}$ and $\theta'_{\mathbf{2}}$: the first one is equal to θ of the true model:

$$\theta_{\mathbf{1}} = (0\ ,\ 2\ ,\ 1).$$

and the two other ones are obtained by minimizing the Kullback distance between the true law and the logistic based on covariate Z_1 alone and Z_2 alone respectively:

$$\begin{aligned} \theta_{\mathbf{2}} &= (0.096\ ,\ 1.969\ ,\ 0) \\ \theta_{\mathbf{2}'} &= (0.307\ ,\ 0\ ,\ 0.674) \end{aligned}$$

The true values of $IDI_{2/1}$ and $IDI_{2'/1}$ can thus be computed :

$$\begin{aligned} IDI_{2/1} &= -0.01037. \\ IDI_{2'/1} &= -0.3282. \end{aligned}$$

Table 7.1. *Comparison of the estimated and simulated standard errors of $\widehat{IDI_{2/1}}$ and $\widehat{IDI_{2'/1}}$.*

	Asymptotic	Empirical
standard error of $IDI_{2/1}$	0.0057029	0.0052650
standard error of $IDI_{2'/1}$	0.025556	0.026306

Table 7.2. *95% Confidence intervals for $IDI_{2/1}$ and $IDI_{2'/1}$.*

	95% confidence	interval	True value
$IDI_{2/1}$	-0.02140714	-0.00076843	- 0.01037
$IDI_{2'/1}$	-0.38141	-0.27829	- 0.3282

Table 7.3. *Brier's scores and BRI for the three models.*

	Brier's score	BRI with respect to model 1
model 1	0.1603	0
model 2	0.16281	- 0.002511
model 2'	0.23962	- 0.079316

Table 7.4. *95% Confidence intervals for $BRI_{2/1}$ and $BRI_{2'/1}$.*

	95% confidence	interval	True value
$BRI_{2/1}$	-0.00535781	-0.00067996	-0.002511
$BRI_{2'/1}$	-0.090107	-0.069466	-0.079316

For each sample we estimated the three models and computed $\widehat{IDI_{2/1}}$ and $\widehat{IDI_{2'/1}}$, as well as our estimates of their standard errors. In table 1 below, we compare the mean estimated asymptotic standard errors of the IDI's, and the empirical standard errors of the estimated IDI's. The true values of the Brier's score of the three models lead to the true values of the BRI of models 2 and 2' with respect to model 1:

It can be seen from the QQ plots below, that the estimated IDI for model 2' with respect to model 1 is close to the normal while the estimated IDI for model 2 with respect to model 1, which is very small, is more erratic. The same occurs for the QQ plots of the BRI's.

7.4.2. *Second simulation along Gu and Pepe's*

Gu and Pepe's example is as follows: they assume that there is only one covariate, so that $Z = (Z_0, Z_1)$, Z_0 being fixed and equal to 1, that the laws of Z_1, conditionally on $Y = 0$ and

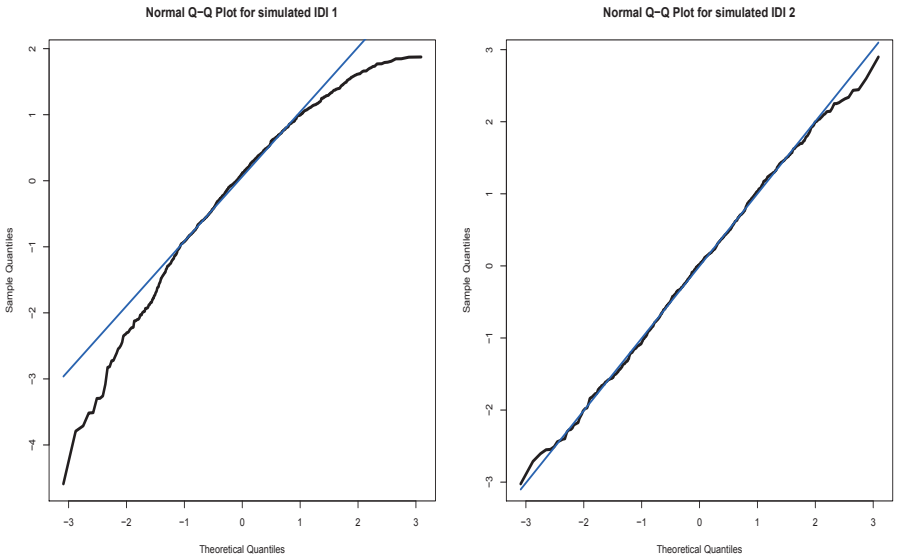

Figure 7.1. *qqplot for* $IDI_{2/1}$ *and* $IDI_{2'/1}$

$Y = 1$, are normal and that $P(Y = 1) = 0.2$:

$$\begin{array}{rcl} P(Y=1) & = & 0.2 \\ \mathscr{L}(Z_1|Y=1) & = & \mathscr{N}(1,1) \\ \mathscr{L}(Z_1|Y=0) & = & \mathscr{N}(0,1) \end{array}$$

Computing the law of Y conditional on Z_1, we get the following logistic model

$$\begin{array}{rcccc} P(Y=1) & = & p_{i1} & = & 0.2 \\ \theta & = & (\theta_0,\theta_1) & = & (log(p_{i1}/(1-p_{i1}))-.5 \ , \ 1) \end{array}$$

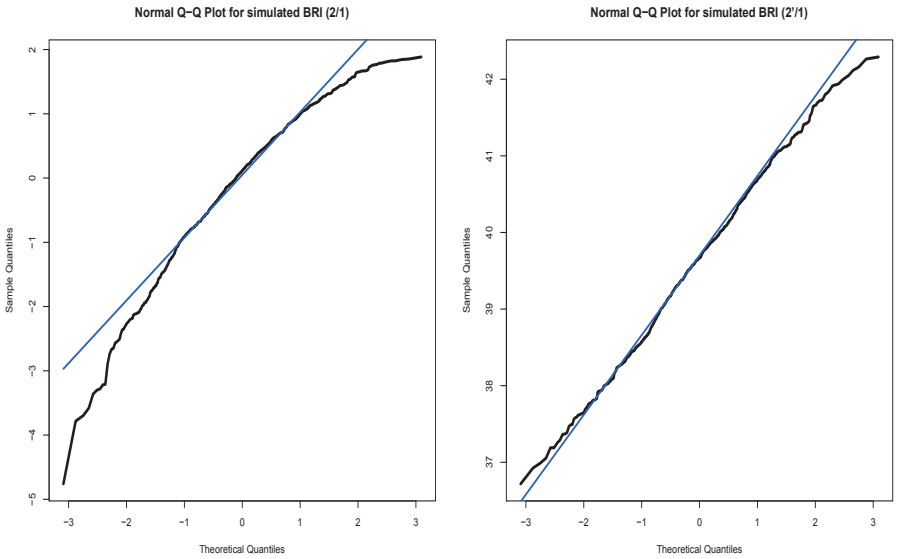

Figure 7.2. *qqplot for* $BRI_{2/1}$ *and* $BRI_{2'/1}$

$$P \quad (Y = 1|Z_1 = z_1) = \frac{\exp(z_1 + (log(p_{i1}/(1-p_{i1})) - .5))}{1 + \exp(z_1 + (log(p_{i1}/(1-p_{i1})) - .5)}$$

$$P \quad (Y = 0|Z_1 = z_1) = \frac{1}{1 + \exp(z_1 + (log(p_{i1}/(1-p_{i1})) - .5)}$$

Three different expressions are given by Gu and Pepe for the PEV:

$$\begin{aligned} PEV &= \frac{var(P(Y=1|Z_1))}{p_{i1}(1-p_{i1})} \\ &= \frac{var(Y) - E(var(Y|Z_1))}{var(Y)} \\ &= corr(Y, P(Y=1|Z_1)) \end{aligned}$$

It seems that the third definition is not correct as the result for the three definitions given by 1000 simulations of a sample of size 1000, are respectively 0.15433, 0.15503 and 0.39384

7.5. The Three City Study of Alzheimer disease.

The Three-City (3C) study is a cohort study conducted in three cities in France (Bordeaux, Dijon, and Montpellier), aiming to estimate the risk of dementia and cognitive impairment attributable to vascular factors.1 A sample of non-institutionalized subjects aged 65 years and older was selected randomly from the electoral rolls of each city (The 3C Study Group. Vascular factors and risk of dementia. Design of the Three-City Study and baseline characteristics of the study population. Neuroepidemiology. 2003; 22:316-325.) Each follow-up examination included cognitive testing and dementia diagnosis.

We have obtained official permission to use the data for the purpose of studying the contribution of the genetic marker APOE4 to the prediction of developing dementia in the next four years beyond the traditional variables: age at recruitment into the study, gender, education, indicator for depression, indicator of previous cardiologic complications, indicator of consumption of psychotropic medication, and indicator of incapacitation of any sort. We were given a sub-sample of n = 4486 men and women among which 162 developed Alzheimer within four years.

After removing records with missing data we ended up with n=4214 records. Based on these data we fit the best model. It included the genetic marker APOE4:

1) Age at recruitment in 3 classes: $65 \leq Age < 71$,$71 \leq Age < 78$,≥ 78,
2) Education level,
3) Past history of vascular disease,
4) Use of psychotropic drugs,
5) Incapacities in activities of daily living,
6) The genetic marker APOE4.

All the covariates are highly significant.

For the model without the genetic marker:

Table 7.5. *Logistic model 1 including the genetic marker APOE4.*

	Estimate	Std. Error	Pr(>|z|)
(Intercept)	-2.944	0.176	< 2e-16
age.fac.31	-2.089	0.330	2.3e-10
age.fac.32	-0.984	0.191	2.5e-07
nivetudes.spec	-0.430	0.180	0.0167
card	0.616	0.233	0.0081
depress	0.786	0.201	9.5e-05
incap	1.180	0.206	1.1e-08
APOE4	0.634	0.195	0.0012
AIC	=	1094	

Table 7.6. *model 2 : Logistic model without the genetic marker APOE4.*

	Estimate	Standard Error	p-value
(Intercept)	-2.797	0.168	< 2e-16
age.fac.31	-2.060	0.330	4.0e-10
age.fac.32	-0.963	0.190	4.2e-07
nivetudes.spec	-0.434	0.179	0.0155
card	0.677	0.231	0.0034
depress	0.805	0.201	6.2e-05
incap	1.124	0.206	4.6e-08
AIC	=	1102	

Clearly, model 1 that includes APOE4 is significantly a better fit to the data than model 2: it has a smaller AIC and the coefficient of APOE4 is highly significant. We turn now to check whether including APOE4 improves also significantly the prediction ability of model 2, by estimating the indices $IDI_{2/1}$ and $BRI_{2/1}$. We estimate the IDI and the BRI of model 2 with respect to model 1 directly from the data, and then ran a bootstrap with 1000 repetitions to obtain a competing estimate of the standard errors of the indices, as well as 95% confidence intervals for the true IDI and BRI for the Dementia data.

We placed the sample estimates and the bootstrap estimates in the same display.

These results certainly suggest that the dementia model that includes the genetic marker APOE4 is not significantly superior to the model without APOE4 in predicting the development of dementia. This result is obtained both in regard to model discrimination as measured by the IDI, and model calibration as measured by Brier's score difference for the two models.

Table 7.7. *Sample and bootstrap estimates for IDI.*

Sample Mean	Bootstrap Mean	Sample Std Err	Bootstrap Std Err
-.00298	-.00374	0.00303	0.00328

Table 7.8. *Sample and bootstrap estimates for BRI.*

Sample Mean	Bootstrap Mean	Sample Std Err	Bootstrap Std Err
-6.09e-05	-9.02e-05	0.000115	0.000129

Table 7.9. *Asymptotic and Percentile Bootstrap* 95% *Confidence Intervals (CI) of IDI for the 3 cities study*

	95% Confidence Interval for IDI	
Asymptotic	-0.008911	+0.002958
Bootstrap	-0.012012	+0.000389

Table 7.10. *Asymptotic and Percentile Bootstrap* 95% *Confidence Intervals (CI) of BRI for the 3 cities study*

	95% Confidence Interval for BRI	
Asymptotic	-2.87e-04	1.65e-04
Bootstrap	-4.08e-04	9.03e-05

The bootstrap distributions for both the IDI and the Brier score difference remain somewhat of a mystery. Note that in both cases the bootstrap standard error is smaller than the asymptotic standard error as computed from the sample. More significantly, the QQ normal plots below show that in both cases the bootstrap distribution, although based on 1000 repetitions, is markedly non-normal. That explains the fact that the bootstrap confidence intervals in both cases are wider than their corresponding asymptotic confidence intervals based on the indices and their standard errors as estimated from the sample.

7.6. Conclusion.

We present results that enable researchers to do inference on two important indices for measuring the relative effectiveness of two models in predicting the probabilities of future events. Most importantly, we allowed for model estimation prior to index computation on the same data by providing new standard errors for both the IDI and the BRI when the indices are computed on the same data that provided model parameter estimates. There are additional indices

for comparison of models predictiveness accuracy and discrimination. We mention in particular the difference in the area under the ROC curve of two models and Pencina et al's (2008) Net Reclassification Improvement NRI. We are currently working on the asymptotic theory for these indices when parameters are estimated from the same data as the indices.

We have tested our asymptotic results and standard error formulae in simulation studies. We then applied them to Alzheimer data from the French Three Cities study and did not find any evidence to support the effectiveness of the genetic marker APOE4 in predicting occurrence of Alzheimer beyond that achieved by standard non-genetic predictor variables such as age, education, and additional health variables.

One unusual finding is the markedly non-normal bootstrap distribution for these data, based on 1000 bootstrap samples. Although in simulations we found no unusual behavior of the bootstrap distribution, for the dementia data, the bootstrap distribution for the IDI, with in-sample estimation of the models, is visibly right skewed, whereas the bootstrap distribution for the Brier difference, again with in-sample model estimation, is visibly left skewed. We offer no explanation for these phenomena at this time.

7.7. Bibliography

[GU 09] GU W., PEPE M., "Measures to summarize and compare the predictive capacity of markers.", *The International Journal of Biostatistics*, vol. 5, num. 1, 2009.

[M.J 08] M. J. PENCINA R. B. D. S., JR. R. B. D., "Evaluating the added predictive ability of a new marker. From area under the ROC curve to reclassification and beyond.", *Statistics in Medicine*, vol. 27, p. 157–172, 2008.