Non-monotone transformation of biomarkers

Non-monotone transformation of biomarkers

Gianfranco Adimari[1], Duc Khanh To[1] and Monica Chiogna[2]

Abstract

We comment here on a recent paper in this journal, on a non-monotone transformation of biomarkers aimed at improving diagnostic accuracy. We highlight that, in a binary classification problem, the proposed transformation finds its motivation in the Neyman-Pearson lemma, so that the underlying approach is very general and it is applicable to many parametric families, other than the normal one.

Keywords

Diagnostic medicine, classification accuracy, ROC analysis, Neyman-Pearson lemma

1 Introduction

In a recent paper in this journal, Yang et al.[1] propose a non-monotone transformation for continuous diagnostic tests (or biomarkers), aimed at improving their accuracy.. The Authors focus on a quadratic transformation, whose coefficients are population means and variances. In the two-class case, the transformation aims to increase the value of the area under the ROC curve (AUC) of the test and, as stated by the same Authors, the proposed method is more powerful when test results are assumed to be normally distributed.

In this note, we start our discussion by observing that, to discriminate between diseased and healthy subjects, the best test is the one whose ROC curve dominates all others. Clearly, such a test also has the highest AUC value. From this perspective, the best possible transformation for a test stems from the Neyman-Pearson lemma and, ultimately, is provided by the likelihood ratio (see also McIntosh and Pepe[2], Gasparini and Sacchetto[3]). Therefore, our first objective is to show that the transformation proposed by Yang et al.[1], under the assumption of normality, is just based on the likelihood ratio. Then, we highlight the generality of the method, illustrating its application to some parametric families, other than the normal one.

[1]Department of Statistical Sciences, University of Padova, Italy
[2]Department of Statistical Sciences “Paolo Fortunati”, University of Bologna, Italy

Corresponding author:
Gianfranco Adimari, Department of Statistical Sciences, University of Padua, Via C. Battisti, 241 - 35121 Padua, Italy.
Email: gianfranco.adimari@unipd.it

Prepared using sagej.cls *[Version: 2017/01/17 v1.20]*

2 Optimal binary classification rule

Let X_1 and X_2 denote the (continuous) test results for a pair of randomly selected diseased and healthy subjects, respectively. Let X denote the test result from a new subject to classify. By the Neyman-Pearson lemma, the best classification rule based on X is given by the likelihood ratio $R(X) = f_1(X)/f_2(X)$, where $f_1(\cdot)$ and $f_2(\cdot)$ denote the density functions of X_1 and X_2, respectively. Following the optimal rule, a value x of the test suggests positivity if $R(x)$ is greater than a threshold c, suitably chosen.

Assume that $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$, with $\sigma_1^2 \neq \sigma_2^2$. Because the ROC curve is invariant with respect to monotone transformations (in particular, increasing transformations), the optimal rule based on

$$R(X) = \frac{\sigma_1^{-2}\, e^{\sigma_1^{-2}(X-\mu_1)^2/2}}{\sigma_2^{-2}\, e^{\sigma_2^{-2}(X-\mu_2)^2/2}} = \frac{\sigma_2^2}{\sigma_1^2}\, e^{-[\sigma_1^{-2}(X-\mu_1)^2 - \sigma_2^{-2}(X-\mu_2)^2]/2}$$

is equivalent to a rule based on $e^{-[\sigma_1^{-2}(X-\mu_1)^2 - \sigma_2^{-2}(X-\mu_2)^2]/2}$ and, then, to a rule based on

$$g(X) = (\sigma_1^2 - \sigma_2^2)X^2 + 2(\sigma_2^2\mu_1 - \sigma_1^2\mu_2)X.$$

Transformation $g(X)$ is the same as the one proposed by Yang et al.[1] for the binary classification problem, when one avoids the complication of dividing by $(\sigma_1^2 - \sigma_2^2)$. Clearly, $g(X)$ depends on the parameters $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ that, in practice, must be estimated, leading to the estimated version $\hat{g}(X) = (\hat{\sigma}_1^2 - \hat{\sigma}_2^2)X^2 + 2(\hat{\sigma}_2^2\hat{\mu}_1 - \hat{\sigma}_1^2\hat{\mu}_2)X$.

The above discussed approach easily generalizes to many parametric families, i.e., to different choices for the distributions of X_1 and X_2. In the following, we give some possibile examples.

- *Exponential family*: Assume that X_1 and X_2 have distributions that belong to some exponential family. Then, the density functions $f_1(\cdot)$ and $f_2(\cdot)$ can be written as $f_j(x) = h(x)\exp\{\sum_{k=1}^{p}\eta_k(\theta_j)T_k(x) - A(\theta_j)\}$, for $j = 1, 2$, where θ_1 and θ_2 are parameter vectors, and $h(\cdot)$, $\eta_k(\cdot)$, $T_k(\cdot)$ and $A(\cdot)$ are suitable functions. It is easy to verify that, in this case, the optimal rule based on the likelihood ratio is equivalent to the rule based on $g(X) = \sum_{k=1}^{p}[\eta_k(\theta_1) - \eta_k(\theta_2)]T_k(X)$. In particular, for instance, if X_1 and X_2 have gamma distribution with shape parameters α_1 and α_2 ($\alpha_1 \neq \alpha_2$) and scale parameters β_1 and β_2 ($\beta_1 \neq \beta_2$), respectively, then $g(X) = (\alpha_1 - \alpha_2)\log(X) + X(\beta_1 - \beta_2)/\beta_1\beta_2$, so that the optimal transformation is no longer a quadratic trasformation.

- *Weibull distributions*: Let X_1 and X_2 be Weibull random variables, with parameters α_1, β_1 and α_2, β_2, respectively, with $\alpha_1 \neq \alpha_2$. Then, $f_j(x) = (\alpha_j/\beta_j)(x/\beta_j)^{\alpha_j - 1}\exp\{-(x/\beta_j)^{\alpha_j}\}$, $j = 1, 2$, and the likelihood ratio $R(X)$ is equivalent to $g(X) = (\alpha_1 - \alpha_2)\log(X) - (X/\beta_1)^{\alpha_1} + (X/\beta_2)^{\alpha_2}$. We remark that the Weibull's family of distributions is not an exponential family.

- *Distributions not belonging to the same family*: The Neyman-Pearson lemma holds provided that X_1 and X_2 have mutually absolutely continuous measures[3]. In practice, it is sufficient that X_1 and X_2 have the same support, i.e., the same set of values where the corresponding density functions, $f_1(\cdot)$ and $f_2(\cdot)$, are strictly positive. Hence, for instance, if X_1 has Weibull distribution with parameters α and β, and X_2 has log-normal distribution with parameters μ and σ^2, then the optimal classification rule based on the likelihood ratio is equivalent to the rule based on $g(X) = (\alpha - (\mu/\sigma^2))\log(X) - (X/\beta)^{\alpha} + 0.5\log^2(X)/\sigma^2$.

For the above considered cases, Table 1 provides AUC values for the test X before and after the optimal transformation $g(\cdot)$ is applied, in some fixed settings. Moreover, for a visual inspection of the effect of

the transformation on the ROC curve, Figure 1 shows the graphs of the ROC curve in the last two of these settings, before and after the transformation. Without loss of generality, to compute the AUC of the unprocessed X-based rule, we assumed that test X was positively associated with the disease status, i.e., that high test values indicate suspected disease.

Table 1. Original AUC (AUC_0) and AUC of transformed diagnostic tests (AUC_g) in some fixed settings.

	Distribution of X_2	Distribution of X_1	AUC_0	AUC_g
Setting 1	$\mathcal{N}(1, 3^2)$	$\mathcal{N}(1.2, 1.5^2)$	0.524	0.706
Setting 2	$\mathcal{N}(1, 1)$	$\mathcal{N}(3, 3^2)$	0.736	0.837
Setting 3	$\mathrm{Gamma}(2, 0.9)$	$\mathrm{Gamma}(1, 2.5)$	0.541	0.644
Setting 4	$\mathrm{Weibull}(1.5, 1.8)$	$\mathrm{Weibull}(1.1, 2)$	0.512	0.601
Setting 5	$\mathrm{Weibull}(0.8, 1)$	$\mathrm{Weibull}(3, 1.8)$	0.741	0.842
Setting 6	$\mathrm{LogNormal}(0.3, 0.6^2)$	$\mathrm{Weibull}(1.1, 2)$	0.505	0.677

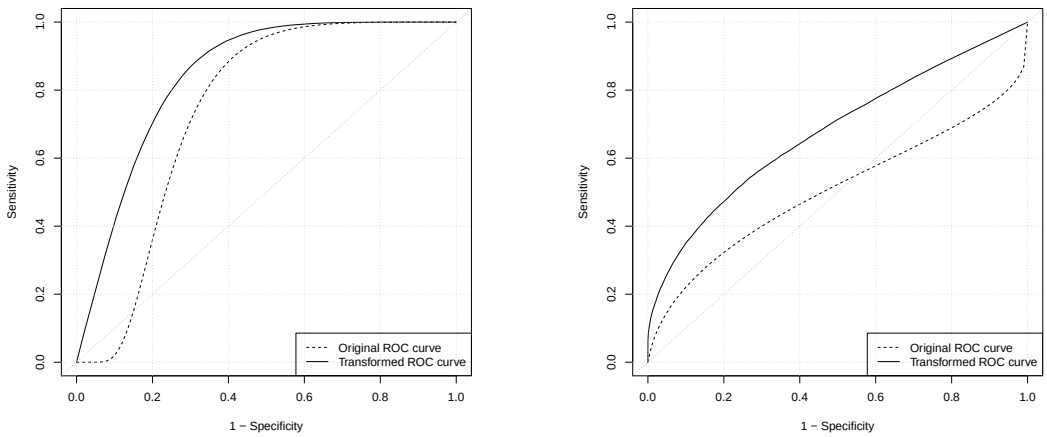

Figure 1. Original ROC curve and ROC curve after transformation, in Setting 5 (two Weibull distributions, left panel) and Setting 6 (lognormal distribution and Weibull distribution, right panel) of Table 1.

Both, the AUC values in the table and the graphs in the figure, clearly show the improvement that can be induced by the transformation $g(\cdot)$ of X. The considered settings were chosen so as to make this effect evident; nevertheless, it is important to highlight that effectiveness of the transformation depends on the true (and typically unknown) distributions of the test in the populations of healthy and diseased subjects, and, *a priori*, it might not be so remarkable. Indeed, even in the above considered parametric models, one could have several cases where the effect of the transformation is essentially negligible. For example, when $X_1 \sim N(2.5, 1.5^2)$ and $X_2 \sim N(1, 1)$, the AUC for the original test X is equal to 0.797 and the AUC for $g(X)$ is 0.799. Finally, the graphs in Figure 1 also highlight an important feature of the likelihood ratio-based classification rule, that is, the associated ROC curve is always concave[3].

As already mentioned, in practice the transformation $g(\cdot)$ must be estimated, i.e., the involved unknown coefficients need to be estimated. A natural estimator (for instance $(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2)$ in the normal case) is

the maximum likelihood estimator, derived from some pair of simple random samples, $x_{11}, x_{12}, \ldots, x_{1n_1}$ and $x_{21}, x_{22}, \ldots, x_{2n_2}$, of size n_1 and n_2, from X_1 and X_2, respectively. To evaluate the impact of the inference on the effectiveness of the transformation, one should explore the behaviour of $\hat{g}(X)$ with respect to that of $g(X)$. We carried out this exploration with a simulation study aimed at examining expected discrepancies, measured in terms of AUC, between the classification rule based on $\hat{g}(X)$ and that based on the true $g(X)$. For each setting in Table 1, we generated test values for simulated samples of diseased and healthy subjects, with sample sizes $n_1 = n_2 = n$, and $n \in \{10, 20, 50, 150\}$. For each pair of generated samples, the maximum likelihood estimates of the coefficients involved in $g(X)$ were obtained. Then, the AUC of $\hat{g}(X)$ was evaluated by computing the empirical AUC over other two samples, of large size ($m_1 = m_2 = m = 1000$), of simulated test values for diseased and healthy individuals, as

$$\mathrm{AUC}_{\hat{g}} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{m} I\left(\hat{g}(x_{1i}) > \hat{g}(x_{2j})\right)}{m^2}.$$

Here $I(\cdot)$ denotes the indicator function. Each simulation experiment was based on 10000 replications. Simulation results are shown in Table 2, which provides Monte Carlo averages of the $\mathrm{AUC}_{\hat{g}}$ values. Such averages can be considered as reliable evaluations of the unknown true AUC values for the classification rules based on $\hat{g}(X)$ in the considered settings. Therefore, by a comparison with the AUC values for rules based on $g(X)$ provided in Table 1, results in Table 2 seem to indicate that the effect of the inference on $g(X)$ (a decrease in the effectiveness of the transformation) can be significant for small sample sizes ($n \leq 20$).

Table 2. Monte Carlo means of the $\mathrm{AUC}_{\hat{g}}$ values over 10,000 replications, for the settings considered in Table 1, in samples of sizes $n_1 = n_2 = n$.

	$n = 10$	$n = 20$	$n = 50$	$n = 150$
Setting 1	0.671	0.691	0.701	0.704
Setting 2	0.824	0.832	0.836	0.837
Setting 3	0.596	0.618	0.635	0.642
Setting 4	0.550	0.567	0.586	0.596
Setting 5	0.827	0.836	0.840	0.842
Setting 6	0.631	0.654	0.670	0.676

However, in many situations, the practitioner typically only has a pair of samples of test results from diseased and healthy populations (let's say again $x_{11}, x_{12}, \ldots, x_{1n_1}$ and $x_{21}, x_{22}, \ldots, x_{2n_2}$), which he can use to estimate the coefficients of the optimal transformation $g(\cdot)$ for X. Then, since a sharp effectiveness of the transformation is not ensured *a priori*, and the effectiveness itself may be significantly reduced by inference, it is important to recover an adequate estimate of the AUC of the classification rule based on $\hat{g}(X)$. To do this, by using the same data, the practitioner may resort to cross-validation methods. In particular (see, for instance, Huang et al.[4]), the leave-one-pair-out estimate of the AUC of $\hat{g}(X)$ is

$$\widehat{\mathrm{AUC}}_{\hat{g}} = \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I\left(\hat{g}^{(-ij)}(x_{1i}) > \hat{g}^{(-ij)}(x_{2j})\right)}{n_1 n_2},$$

where $\hat{g}^{(-ij)}(\cdot)$ denotes the transformation estimated from the data without the pair (x_{1i}, x_{2j}). Table 3 show the results of a simulation study conducted to asses the behavior of the estimator $\widehat{\mathrm{AUC}}_{\hat{g}}$. It provides Monte Carlo means and standard deviations, computed by simulated samples of diseased and healthy subjects, with sample sizes $n_1 = n_2 = n$ and $n \in \{10, 20, 50\}$, for each setting in Table 1. Each simulation experiment is based on 10000 replications. Taking as reference values those stored in Table 2, results in Table 3 show that the bias of the estimator $\widehat{\mathrm{AUC}}_{\hat{g}}$ is negligible in all settings. Moreover, as expected, the accuracy of the estimator improves as the sample size grows.

Table 3. Monte Carlo means (MCM) and Monte Carlo standard deviations (SD) of the estimates $\widehat{\mathrm{AUC}}_{\hat{g}}$ over 10,000 replications, for the settings considered in Table 1, at different sample sizes $n = n_1 = n_2$.

	$n = 10$		$n = 20$		$n = 50$	
	MCM	SD	MCM	SD	MCM	SD
Setting 1	0.665	0.149	0.691	0.092	0.701	0.054
Setting 2	0.821	0.099	0.831	0.067	0.836	0.041
Setting 3	0.592	0.170	0.615	0.112	0.635	0.069
Setting 4	0.549	0.176	0.567	0.124	0.584	0.070
Setting 5	0.826	0.101	0.835	0.067	0.840	0.041
Setting 6	0.629	0.154	0.654	0.100	0.669	0.057

3 Conclusion

The non-monotone transformation proposed by Yang et al.[1] for a binary classification problem leads to a classification rule that, ultimately, is equivalent to the optimal rule based on the likelihood ratio. We emphasized the generality of this approach, identifying the optimal transformation $g(X)$ under some parametric models, other than the normal one. The optimal rule is characterized by a ROC curve that dominates all curves associated with possible other rules based on functions of the original diagnostic test X, and that is always concave. Moreover, the classification rule based on $g(X)$ is always the same, i.e., it classifies a subject with test value x as diseased if $g(x)$ is greater than a suitably chosen threshold c, irrespective of whether high values of the test are associated with the disease or not.

However, effectiveness of the transformation $g(\cdot)$ is not always significant, and the fact that typically $g(\cdot)$ must be estimated, reduces its effectiveness at least when inference is based on small samples. For these reasons, we suggest that practitioners assess the actual effectiveness of the (estimated) transformation by resorting to $\widehat{\mathrm{AUC}}_{\hat{g}}$ to estimate the AUC of the corresponding rule.

References

1. Yang J, Kuan PF and Li J. Non-monotone trasformation of biomarkers to improve diagnostic and screening accuracy in a dna methylation study with trichotomous phenotypes. *Stat Meth Med Res* ; DOI: 10.1177/0962280219882047.
2. McIntosh MW and Pepe MS. Combining severel screening test: optimality of the risk score. *Biometrics* 2002; 58: 657–664.

3. Gasparini M and Sacchetto L. Proper likelihood ratio based roc curves for general binary classification problems. *arXiv preprint arXiv:180900694* 2018; .
4. Huang X, Qin G and Fang Y. Optimal combinations of diagnostic tests based on auc. *Biometrics* 2011; 67(2): 568–576.