

(一) 研究背景與前言

臨床上經常藉由 Receiver Operating Characteristic Curve (ROC curve) 來藉由生物標誌 (biomarkers) 作為去區分出健康與患病之個體。若個體之生物標誌值大於某個特定的閾值 (specified threshold value)，則會將該個體給評估為有病(或沒病)。與閾值有直接相關的即是敏感度 (sensitivity) 與特異度 (specificity)，ROC curve 正是劃出所有可能閾值下以敏感度與 1-特異度去畫出來的遞增曲線。在診斷此 ROC curve 時廣用的指標之一則是 area under the ROC curve (AUC)。

然而針對個體是否患病的診斷 (diagnosis) 與篩檢 (screening) 中，經常不會只有一個生物標誌，而是有多個生物標誌。正如同 Amy (2018) 等人提出在診斷是否患有肺腺癌的患者中，BRCAness gene and driver mutational gene 都是對於肺腺癌患者是很好的生物標誌因子。因此對於如何結合多個生物標誌來對個體之患病與否的診斷與篩檢就很重要，因為能夠結合多個生物標誌應相較於單一生物標記理應能夠提供更精確的篩檢與診斷。若生物標誌對於有病與否本身服從多維常態 (multivariate normal)，則利用 linear discriminant analysis (LDA) 則可最大化 AUC，其他的方法諸如 regression tree, artificial neural network, nearest neighbor 也都被發展得很完善。然而並沒有任何一個方法被證實其具有用於醫學篩檢與診斷等目的時生物標誌有最佳組合 (optimal combination)，因此找尋一個可以使得在面對醫學篩檢與診斷時，以多個生物標誌去預測個體是否得疾病可以有最佳 (optimal) 決策的準則具有必要性。

(二) likelihood ratio rule is uniformly most sensitive (UMS) screening test

首先定義 $D \in \{0,1\}$ 代表疾病狀態的二元指標，其中 1 代表有病 0 代表沒病；生物標誌則定義為 $Y = (Y_1, \dots, Y_k)$ 代表有 k 個生物標誌，True positive rate (sensitivity, TPR) 與 False positive rate (1-specificity, FPR) 分別定義為

$$\begin{aligned} TPR &= P(\text{Screenpositive}|\text{disease}) \\ FPR &= P(\text{screenpositive}|\text{notdisease}) \end{aligned}$$

其中一個去建構篩檢或診斷依據的方法即是藉由 likelihood ratio rule 來進行，其代表

$$LR(Y) > c(f_0)$$

其中

$$LR(Y) = \frac{P(Y|D = 1)}{P(Y|D = 0)}$$
$$f_0 = P(LR(Y) > c(f_0)|D = 0)$$
$$LR\left(\frac{P(Y|D = 1)}{P(Y|D = 0)}\right) > c(P(LR(Y) > c(f_0)|D = 0))$$

若我們設定虛無假設與對立假設分別為 $D = 0, D = 1$ ，在此情境中即可發現

$$\alpha = P(\text{Reject } D = 0 | \text{Donotreject } D = 0) = P(\text{screenpositive} | \text{notdisease}) = \text{FPR}$$

$$\beta = P(\text{Donotreject } D = 0 | \text{Reject } D = 0)$$

$$1 - \beta = P(\text{reject } D = 0 | \text{Reject } D = 0) = P(\text{Screenpositive} | \text{disease}) = \text{TPR}$$

這樣的設定促使我們得以藉由 Neyman-Pearson Lemma (Neyman & Pearson, 1933) 來證實出此 likelihood ratio decision 在給定 FPR (type I error) 下，likelihood ratio test 是均勻最佳檢定 (uniformly most powerful test)，即可以達到最大的 TPR (power)。這代表了以此為基礎的生物標誌來進行二元事件之篩檢或診斷工具，必然會是最優的，其他的方法諸如 LDA, neural network, regression tree 表現再好，它們也只是漸進優到 binary regression 罷了。這個非常強力的結果就是 McIntosh and Pepe (2002) 所發現的，而它們稱此結果為 uniformly most sensitive (UMS) screening test based on the multiple markers.

(三) Risk Score and Binomial regression

儘管有如此強力的結果，但這代表需要先得知 $P(Y|D)$ 機率架構為何才可以有強力的結果，然而這通常是複雜的機率架構並包含了未知參數，甚至可能整個機率架構我們也毫無先驗資訊，因此就可能產生 mis-specification 與估計參數上的問題。McIntosh and Pepe (2002) 在範例中也提出了卵巢癌的兩個生物標誌之聯合分配在控制組與有病組分別為多維常態分配與 mixture 多維常態分配。McIntosh and Pepe (2002) 發現了

$$P(D = 1|Y) = \frac{P(Y|D = 1)P(D = 1)}{[P(Y|D = 1)P(D = 1) + P(Y|D = 0)P(D = 0)]} = \frac{LR(Y)q}{LR(Y)q + 1}$$

其中 $q = \frac{P(D=1)}{P(D=0)}$ is the odds of diseases in the population。既然有這件事，那我們只要去針對 risk score $P(Y) = P(D = 1|Y)$ 即可而不需要針對 likelihood ratio

function，而且這 risk score 其實就是 binomial regression tools。上述的發現表示 risk score is a monotonic increasing function of $LR(Y)$ 。因此先前的 likelihood ratio rule 就可以表示成

$$p(Y) = c^*(f_0)$$

where f_0 is chosen to yield $FPR=f_0$, i.e.,

$$c^*(f_0) = \frac{c(f_0)q}{c(f_0)q + 1}$$

如此一來針對 risk score 去用多個生物標誌來作為個體篩檢與診斷之依據也可以具有 UMS，即也可以得到最大的 AUC，如此一來便可以避免需要先得知 $P(Y|D)$ 機率架構之問題。並且隨後 Pepe (2003) 也證實了以此為基礎的準則實際上可以將 overall misclassification rate 與 expected cost of false-negative and false positive errors 都給最小化。

在估計上基於此機率模型的架構，看似必須要用 cross-sectional cohort sampling 去估計 optimal screening rule，但也可以用 binary regression 去進行，也就是 logistic regression form with intercept term

$$\text{logit}(p(Y)) = \beta_0 + h(\beta, Y)$$

which is also optimal because it is a monotone transformation of the risk score，且實際上 intercept 有沒有都不會影響，也就是 rules based on $h(\beta, Y)$ are also optimal. 這就代表了 case-control design can be used to derive optimal marker combinations with logistic regression. 那麼 threshold for the risk score 該怎麼選？由於這是由 FPR 所定義的，由於

$$f_0 = P(P(Y) > c^*(f_0) | D = 0)$$

因此 $c^*(f_0)$ 實際上只是 $1 - f_0$ quantile of the risk score $p(Y)$. in the nondiseased population. 因此 Empirical or parametric estimates of the $1 - f_0$ quantile of $P(Y)$ can be calculated to approximate $c^*(f_0)$

(四) Conclusion and Future Work

總結而言，McIntosh and Pepe (2002) 證明出了給定生物標誌下二元變數的 risk

score 基於 Neyman-Pearson Lemma 實際上正是分類二元變數的一個最佳函數，此分類法可以使得給定 FPR 下 TPR 最大，因而使得 AUC 最大。此 risk score 可以用 logistic regression 來 approximation，而後者是已經很廣泛使用的統計技巧，在各現有軟體中大多都可進行計算，進而開拓了一個以多個共變數來預測二元分類器上一個最佳的理論基礎。

然而仍有多個限制之處，諸如 logistic regression 有假設生物標誌與疾病之間關係為 logit scale，若此關係不滿足時就會有 model mis-specification 的問題，以及用 logistic regression 來進行 approximation 也會牽涉到估計的是否 flexible 等問題，以及在二元分類上對於罕見事件 (rare event) 的估計可能也需要一些額外的校正方法。這些問題在後續的學者實際上也正在探討中。

(五) Reference

Jan A, Lin V, Lian P. (2018) Landscape on the Association between Lung Cancer Driver Mutations and Tumor Mutational Burden. Paper presented at the Implementation Research and Science for Health for all, Kyoto university.

Jin H, Lu Y. A procedure for determining whether a simple combination of diagnostic tests may be noninferior to the theoretical optimum combination. Med Decis Making. 2008 Nov-Dec;28(6):909-16.

McIntosh MW, Pepe MS. Combining several screening tests: optimality of the risk score. Biometrics. 2002;58:657-664

Neyman J, Pearson ES. On the problem of the most efficient tests of statistical hypothesis. Philos Trans R Soc Lond A. 1933; 231:289-337.

McIntosh MW, Pepe MS. Combining several screening tests: optimality of the risk score. Biometrics. 2002;58:657-664

Yu B. Approximating the risk score for disease diagnosis using MARS. J Appl Stat. 2009;36(7):769-778.