

Comparación de pruebas diagnósticas desde la curva ROC

Comparing Diagnostic Tests from ROC Curve

PABLO MARTÍNEZ-CAMBLOR^a

FUNDACIÓN CAUBET-CIMERA ILLES BALEARS, MALLORCA, ESPAÑA

Resumen

Se aborda el problema de comparar el poder de clasificación de métodos diferentes a partir de la curva ROC. Por un lado, se propone un método de comparación basado en la medida del supremo y, por otro, una solución al problema de comparar más de dos pruebas diagnósticas a través del *área bajo la curva* ROC (AUC) a partir de sus propiedades asintóticas. También se comprueba la validez de los estimadores propuestos para muestras pequeñas a partir del método *bootstrap*. Finalmente, se aplican los métodos propuestos en la predicción de diagnósticos sépticos (infecciosos) en pacientes admitidos en la Unidad de Cuidados Intensivos Pediátricos (UCIP) del Hospital Central de Asturias.

Palabras clave: curvas ROC, sensibilidad, especificidad, AUC, bootstrap.

Abstract

We study the problem of comparing the power of classification of different methods from the ROC Curve. On one hand, we propose a method based on the supremum measure and, on the other hand, we study the problem of comparing two or more ROC curves from the asymptotic properties of area under ROC curves (AUC). We study the performance of proposed estimators to small samples problems with Bootstrap method and we apply them to differentiate two classes of patients of the Pediatric Intensive Care Unit (PICU) of the Hospital Central de Asturias.

Key words: ROC curves, Sensitivity, Specificity, AUC, Bootstrap.

^aPrograma de epidemiología e investigación clínica. E-mail: martinez@caubet-cimera.es

1. Introducción

Un problema que se da en muchos ámbitos es el de clasificar individuos con base en una o varias variables predictoras. Cuando el número de grupos es dos, a partir de los modelos de regresión logística, y dada una colección de variables, se pueden obtener pesos para cada variable cuyos valores pueden interpretarse como las probabilidades de pertenencia a uno de los grupos fijado como objetivo. En este tipo de problemas, hay dos fuentes de error: el que se comete cuando a un individuo no séptico se le clasifica erróneamente (la proporción de individuos no positivos clasificados correctamente se conoce como *especificidad*), y por otro, el que se comete cuando a un individuo séptico (infectado por un virus) se le clasifica como no séptico, siendo la *sensibilidad* la proporción de individuos positivos clasificados de forma adecuada. Un buen sistema de clasificación será aquel que maximice la sensibilidad y la especificidad; obviamente, hay que tener en cuenta que los errores cometidos no tienen siempre la misma importancia, siendo labor del experto valorar el efecto de los mismos. Uno de los métodos utilizados habitualmente para determinar la calidad diagnóstica es la conocida curva ROC, que representa sobre un sistema de coordenadas la sensibilidad y la especificidad, y que ha sido ampliamente estudiada por autores como Lloyd (1998, 1999), Zhou & Castelluccio (2003) o Cai & Pepe (2002), entre muchos otros.

En este trabajo se hace un repaso de algunos de los métodos utilizados para la estimación de la curva ROC (sección 2). La sección 3 se dedica al estudio del área bajo la curva ROC (AUC). En la sección 4 se propone un método para la comparación de k -pruebas diagnósticas que se aplica a un caso práctico en la sección 5. Finalmente, en la sección 6 se propone y estudia la medida del supremo para comparar dos pruebas curvas ROC.

2. La curva ROC

Se tiene una medida determinada X (recordar que esta medida puede resultar de realizar una regresión logística sobre varias variables de distintas naturalezas) realizada sobre una población de positivos X_P , y otra de negativos X_N , con función de distribución G y F , respectivamente. Suponiendo que $E(X_N) \leq E(X_P)$, para clasificar a los individuos en uno u otro grupo, se debe fijar un criterio, punto de corte, a partir del cual un individuo será considerado positivo. Por tanto, fijado un punto de corte t , la sensibilidad de la prueba vendrá determinada por $1 - G(t)$, siendo $F(t)$ su especificidad y quedando por tanto determinada la curva ROC por las coordenadas del vector $(1 - F(t), 1 - G(t))$, o, equivalentemente, por la función $1 - G(F^{-1}(1 - t))$, $t \in [0, 1]$. A modo de ejemplo, en la figura 1 se presenta un modelo de dos distribuciones para una hipotética muestra de positivos y otra de negativos y la curva ROC resultante.

Como siempre, el problema surge cuando se desconocen las distribuciones de la variable en las poblaciones de positivos y negativos y deben estimarse a partir de muestras aleatorias. Una de las posibilidades es suponer que las poblaciones siguen algún modelo paramétrico, el gaussiano usualmente, o bien, aplicar algún

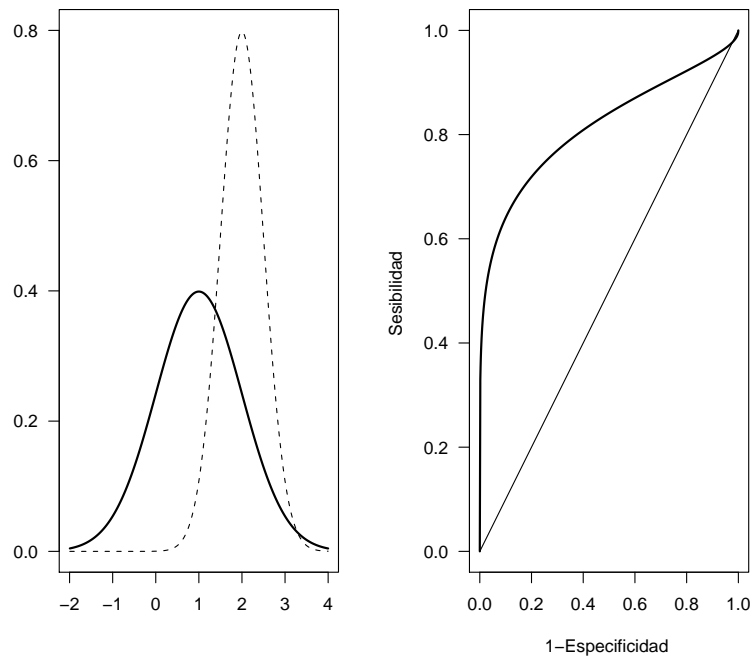


FIGURA 1: En la gráfica de la izquierda se ven las densidades de las poblaciones de origen. En la gráfica de la derecha se muestra la curva ROC resultante.

método no paramétrico, siendo los más frecuentes sustituir las funciones de distribución desconocidas por sus Funciones de Distribución Empíricas (FDE) o por las Funciones de Distribución Empírica Suavizadas (FDES).

2.1. El modelo binormal

El modelo binormal es introducido inicialmente por Dorfman & Alf (1969) y tratado más recientemente por Hsieh & Turnbull (1996). En él se supone que existe una función monótona H de modo que transforme a las dos poblaciones de origen en normales, esto es, tanto $H(X_N)$ como $H(X_P)$ siguen una distribución normal con medias μ_1 y μ_2 , y varianzas σ_1^2 y σ_2^2 , respectivamente. Así las cosas, si se denota por Φ la función de distribución normal estándar, la curva ROC tiene la expresión

$$\begin{aligned}
 ROC(t) &= 1 - G(F^{-1}(1 - t)) \\
 &= 1 - GHH^{-1}F^{-1}(1 - t) = 1 - \Phi\left(\frac{\mu_1 + \sigma_1\Phi^{-1}(1 - t) - \mu_2}{\sigma_2}\right) \quad (1)
 \end{aligned}$$

Su estimador natural sustituye las medias y varianzas involucradas en (1) por sus estimadores máximo verosímiles, resultando

$$ROC_p(t) = 1 - \Phi\left(\frac{\hat{\mu}_1 + \hat{\sigma}_1\Phi^{-1}(1 - t) - \hat{\mu}_2}{\hat{\sigma}_2}\right) \quad (2)$$

Algunas de las propiedades del estimador “semiparamétrico” de la curva ROC son tratadas en Hsieh & Turnbull (1996). Cai & Pepe (2002) también estudian las posibilidades de los métodos paramétricos y semiparamétricos, realizando estudios de simulación y construyendo bandas de confianza en ambos casos.

2.2. Estimación empírica de la curva ROC

Si no se hace ninguna suposición sobre la distribución de las variables, el método más frecuente de estimación consiste en sustituir las funciones de distribución desconocidas por sus correspondientes Funciones de Distribución Empíricas. De la expresión (1), se tiene que, dadas muestras de positivos X_P y de negativos X_N de tamaños m y n y distribuciones G y F , respectivamente, la estimación empírica para la curva ROC viene dada por

$$\widehat{ROC}(t) = 1 - \widehat{G}_m(\widehat{F}_n^-(1-t)) \quad (3)$$

donde \widehat{G}_m es la función de distribución empírica asociada a la muestra X_P y $\widehat{F}_n^- = \inf\{y : F_n^-(y)\}$, siendo \widehat{F}_n la función de distribución empírica asociada a la muestra X_N . Cuando las distribuciones de origen F y G son absolutamente continuas con densidades f y g , respectivamente, se tiene que para cualquier subintervalo (a, b) de $(0, 1)$ $f(G^{-1}(t))/g(G^{-1}(t))$ está acotada y $n/m \rightarrow \lambda > 0$. Hsieh & Turnbull (1996) prueban la consistencia fuerte de la curva ODC¹ (*Ordinal Dominance Curve*) y que existe un espacio probabilístico en el que se pueden definir dos puentes brownianos independientes $B_1^{(n)}$ y $B_2^{(n)}$ de modo que

$$\begin{aligned} \sqrt{n}(\widehat{F}_n(\widehat{G}_m^-(t)) - F(G^{-1}(t))) = \\ \sqrt{\lambda}B_1^{(n)}(F(G^{-1}(t))) + \frac{f(G^{-1}(t))}{g(G^{-1}(t))}B_2^{(n)}(t) + o(n^{-1/2}(\log n)^2) \quad c.s. \end{aligned} \quad (4)$$

uniformemente en $[a, b]$. Este resultado es extendible directamente a la curva ROC.

2.3. Estimación suavizada de la curva ROC

Si se supone que las variables aleatorias en estudio son continuas, y se desea que la estimación de la distribución conserve estas propiedades, se reemplaza la función de distribución real por la Función de Distribución Empírica Suavizada (FDES) en lugar de su FDE. La FDES propuesta por Nadaraya (1964) queda definida por

$$\tilde{F}_n(t) = \frac{1}{nh_n} \sum_{i=1}^n \int_{-\infty}^t K\left(\frac{x_i - s}{h_n}\right) ds \quad (5)$$

donde K es una función núcleo y h_n una secuencia de números positivos usualmente denominada “ancho de banda”. Un amplio estudio de este tipo de estimaciones puede verse en Silverman (1986) o Wand & Jones (1995).

¹Curva resultante de invertir los ejes de la curva ROC, definida por Bamber (1975) y con las mismas propiedades y forma que la curva ROC pero de expresión más sencilla.

Así las cosas, la estimación suavizada de la curva ROC tiene la expresión

$$\widetilde{ROC}(t) = 1 - \widetilde{G}_m(\widetilde{F}_n^{-1}(1-t)) \quad (6)$$

Lloyd & Yong (1999) estudian las propiedades de este estimador y calculan su esperanza y varianza bajo ciertas condiciones. Hall & Hyndman (2003) plantean criterios para optimizar la elección del parámetro h_n y Hall et al. (2004) muestran métodos para construir bandas de confianza para el estimador suavizado de la curva ROC.

3. El área bajo la curva ROC

Una vez fijado un criterio de pertenencia, la curva ROC da una medida de la calidad de clasificación que se obtiene con la variable aleatoria en estudio. Así pues, considerando todos los puntos conjuntamente, se obtiene una medida de la “bondad global” de las predicciones realizadas. El área bajo la curva ROC (AUC) se define como:

$$AUC = \int_0^1 (1 - G(F^{-1}(1-t))) dt \quad (7)$$

Nótese que solo haciendo un cambio de variable se obtiene

$$\begin{aligned} AUC &= 1 - \int_0^1 G(F^{-1}(t)) dt \\ &= 1 - \int G(s) dF(s) = 1 - \iint_{-\infty}^y dG(x) dF(y) = \mathcal{P}\{\mathcal{X}_{\mathcal{P}} > \mathcal{X}_{\mathcal{N}}\} = \theta \end{aligned}$$

En función de la estimación usada para la curva ROC, se obtiene una estimación diferente para el AUC, cuando la estimación es la dada en (3), el estimador resultante para el área bajo la curva es proporcional al estadístico de Mann-Whitney,

$$\widehat{AUC} = \int \widehat{F}_n(s) d\widehat{G}_m(s) = \frac{1}{nm} \sum_{i=1}^m \sum_{j=1}^n I_{\{X_i < Y_j\}} \quad (8)$$

Hsieh & Turnbull (1996) demuestran que, bajo las condiciones pedidas para (5), se da la siguiente igualdad:

$$\sqrt{n}(\widehat{AUC} - \theta) = \mathcal{N}(0, \sigma) + o_{\mathcal{P}}(1) \quad (9)$$

donde $\sigma^2 = \|FG^{-1}\|^* + \|GF^{-1}\|^*$ y $\|h\|^* = \int_0^1 h^2 dt - (\int_0^1 h dt)^2$.

Por otro lado, si se utiliza la función de distribución suavizada como estimador de la función de distribución real, el estimador para el AUC es

$$\widetilde{AUC} = \int \widetilde{F}_n(s) d\widetilde{G}_m(s) \quad (10)$$

En condiciones bastante generales, se comprueba que \widetilde{AUC} y \widehat{AUC} tienen la misma distribución asintótica. Basta tener en cuenta el siguiente resultado:

Proposición 1. *Dada una muestra aleatoria simple de tamaño n procedente de una variable aleatoria, X , cuya función de densidad f , está acotada, se anula fuera de un intervalo compacto y tiene segunda derivada continua y acotada. Dada \tilde{F}_n la FDES definida en (5). Si la función núcleo K es de variación acotada y se anula fuera de un intervalo compacto y h_n cumple*

$$\frac{\log h_n^{-1}}{nh_n^3} \xrightarrow{n} 0$$

entonces,

$$\sup_{t \in \mathbb{R}} \left| \tilde{F}_n(X, t) - F_n(X, t) \right| = o(h_n) \quad c.s. \quad (11)$$

Demostración. Se supondrá que $f(x) = 0 \forall x \notin [a, b]$ y que $K(x) = 0 \forall x \notin [-s, s]$ luego si $J = (a - sh_n, b + sh_n)$ se da la igualdad

$$\sup_{t \in \mathbb{R}} \left| \tilde{F}_n(X, t) - F_n(X, t) \right| = \sup_{t \in J} \left| \tilde{F}_n(X, t) - F_n(X, t) \right|$$

por otro lado, siempre es cierta la desigualdad

$$\begin{aligned} \sup_{t \in J} \left| \tilde{F}_n(X, t) - F_n(X, t) \right| &\leq \sup_{t \in J} \left| \tilde{F}_n(X, t) - E(\tilde{F}_n(X, t)) \right| \\ &+ \sup_{t \in J} \left| E(\tilde{F}_n(X, t)) - E(F_n(X, t)) \right| + \sup_{t \in J} \left| E(F_n(X, t)) - F_n(X, t) \right| \end{aligned} \quad (12)$$

aplicando la desigualdad de Smirnov, y de la relación entre h_n y n impuesta en el enunciado de la proposición, se tiene que

$$\sup_{t \in J} \left| E(F_n(X, t)) - F_n(X, t) \right| \leq o(h_n^2) \quad c.s. \quad (13)$$

dado que la segunda derivada de f es continua y acotada, se verifica que

$$\sup_{t \in J} \left| E(\tilde{F}_n(X, t)) - E(F_n(X, t)) \right| \leq o(h_n) \quad (14)$$

del teorema de Fubini y el del valor medio del cálculo integral se tiene que existe un $\xi_t \in (a - sh, t)$ tal que

$$\begin{aligned} \sup_{t \in J} \left| \tilde{F}_n(X, t) - E(\tilde{F}_n(X, t)) \right| &= \\ \sup_{t \in J} \left| (t - a + sh_n)(f_n(X, \xi_t) - E(f_n(X, \xi_t))) \right| &\leq \\ L \sup_{t \in J} \left| f_n(X, t) - E(f_n(X, t)) \right| & \end{aligned}$$

donde $L = \sup_n(b - a + 2sh_n) < \infty$.

Dado que la derivada F esta acotada, se tiene que

$$\sup_{\substack{x, y \in J \\ x \neq y}} \frac{|F(x) - F(y)|}{|x - y|} < \infty \quad (15)$$

la función es de variación acotada y se anula fuera de un intervalo acotado, satisfaciendo así (15), luego se podrá aplicar la ley del logaritmo iterado dada por Stute (1982) para obtener

$$\limsup_n \sup \sqrt{\frac{nh_n}{2 \log \frac{1}{h_n}}} \sup_{t \in J} |f_n(X, t) - E(f_n(X, t))| = C \quad c.s. \quad (16)$$

para cierta constante real C .

De lo anterior se obtiene que existe una constante C_1 tal que

$$\lim_n \frac{\sup_{t \in J} |\tilde{F}_n(X, t) - E(\tilde{F}_n(X, t))|}{h_n} \leq \lim_n \frac{C_1 \sqrt{\frac{2 \log \frac{1}{h_n}}{nh_n}}}{h_n} \quad c.s.$$

Además,

$$\lim_n \frac{\sqrt{\frac{2 \log h_n^{-1}}{nh_n}}}{h_n} = 0 \iff \lim_n \frac{\log h_n^{-1}}{nh_n^3} = 0 \quad (17)$$

quedando demostrado el resultado. \square

4. Comparación de k -curvas ROC

En esta sección se aborda el problema de comparar simultáneamente dos o más curvas ROC. Para ello se construye un estadístico basado en los AUC y se estudia su distribución asintótica. Posteriormente, se realizan algunas simulaciones que muestran su comportamiento.

Dados m individuos positivos y n negativos, se toman k -medidas sobre cada uno de ellos cuyas distribuciones k -dimensionales vienen dadas por \mathbf{F} y \mathbf{G} , respectivamente. Si F_1, \dots, F_k y G_1, \dots, G_k son las funciones de probabilidad marginales, se tiene que, si cada curva ROC es $F_i(G_i^{-1}(t))$ y se denota por θ_i a su AUC, el objetivo será contrastar

$$\begin{cases} H_0 : \theta_1 = \dots = \theta_k \\ H_1 : \text{no } H_0 \end{cases} \quad (18)$$

para cada $i \in 1, \dots, k$ se define $\hat{a}_i = \int \hat{F}_{m,i}(t) d\hat{G}_{n,i}(t)$, $a_i = \int F_i(t) dG_i(t)$, y

$$\hat{\mathbf{a}}_{m,n}(k) = (\hat{a}_1, \dots, \hat{a}_k) \quad (19)$$

$$\vec{\mathbf{a}}(k) = (a_1, \dots, a_k) \quad (20)$$

En el siguiente resultado se demuestra la normalidad asintótica de la expresión $\sqrt{n}(\hat{\mathbf{a}}_{m,n}(k) - \vec{\mathbf{a}}(k))$.

Teorema 1. *Dada una variable k -dimensional, \mathbf{X} , una muestra de m positivos, $\mathbf{X}_P = \{\mathbf{X}_{P_1}, \dots, \mathbf{X}_{P_m}\}$, donde para cada $i \in 1, \dots, m$ se tiene $\mathbf{X}_{P_i} = \{x_{P_i,1}, \dots, x_{P_i,k}\}$, y una muestra de n negativos $\mathbf{X}_N = \{\mathbf{X}_{N_1}, \dots, \mathbf{X}_{N_n}\}$, nuevamente, para cada $i \in 1, \dots, n$ se tiene $\mathbf{X}_{N_i} = \{x_{N_i,1}, \dots, x_{N_i,k}\}$. Sean \mathbf{F} y \mathbf{G}*

las funciones de distribución en los positivos y en los negativos, respectivamente, entonces, según las hipótesis de la proposición 1,

$$\sqrt{n}(\hat{\mathbf{a}}_{m,n}(k) - \vec{\mathbf{a}}(k)) \xrightarrow{\mathcal{L}} \mathcal{N}_k(\mathbf{0}, \mathbf{\Sigma}) \quad (21)$$

donde $\mathbf{\Sigma}$ es la matriz de varianzas y covarianzas definida por

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k,1} & \sigma_{k,2} & \cdots & \sigma_{k,k} \end{bmatrix} \quad (22)$$

donde para cada $i, j \in 1, \dots, k$,

$$\begin{aligned} \sigma_{i,j} = & \sqrt{\lambda} \left(\int_0^1 F_i(G_i^{-1}(t)) F_j(G_j^{-1}(t)) dt - \int_0^1 F_i(G_i^{-1}(t)) dt \int_0^1 F_j(G_j^{-1}(t)) dt \right) \\ & + \left(\int_0^1 G_i(F_i^{-1}(t)) G_j(F_j^{-1}(t)) dt - \int_0^1 G_i(F_i^{-1}(t)) dt \int_0^1 G_j(F_j^{-1}(t)) dt \right) \end{aligned} \quad (23)$$

Demostración. Se define inicialmente,

$$\mathbb{P}\hat{\mathbf{a}}_{m,n}(k) = (\mathbb{P}\hat{a}_1, \dots, \mathbb{P}\hat{a}_k) \quad (24)$$

donde para cada $l \in 1, \dots, k$,

$$\mathbb{P}\hat{a}_l = \frac{1}{m} \sum_{i=1}^m (F_l(x_{N_{i,l}}) - E[F_l(x_{N_{i,l}})]) - \frac{1}{n} \sum_{j=1}^n (G_l(x_{P_{j,l}}) - E[G_l(x_{P_{j,l}})]) \quad (25)$$

sin más que aplicar el teorema del límite central k -dimensional se tiene que si $\lim_n n/m = \lambda$,

$$\sqrt{n} \mathbb{P}\hat{\mathbf{a}}_{m,n}(k) \xrightarrow{\mathcal{L}} \mathcal{N}_k(\mathbf{0}, \mathbf{\Sigma}) \quad (26)$$

con un poco de cálculo y desde la independencia de los sumandos involucrados en (25) se obtiene que $\mathbf{\Sigma}$ es la matriz de varianzas y covarianzas definida en (22).

Por otro lado, para cada $i \in 1, \dots, k$ y directamente desde la proposición 1 se tiene que la matriz de varianzas y covarianzas del estadístico $\sqrt{n}(\hat{\mathbf{a}}_{m,n}(k) - \vec{\mathbf{a}}(k) - \mathbb{P}\hat{\mathbf{a}}_{m,n}(k))$ tiende a ser la matriz nula y, por tanto,

$$\sqrt{n}(\hat{\mathbf{a}}_{m,n}(k) - \vec{\mathbf{a}}(k) - \mathbb{P}\hat{\mathbf{a}}_{m,n}(k)) \xrightarrow{\mathcal{P}} 0 \quad (27)$$

Aplicando el lema de Slutsky se tiene que

$$\sqrt{n}(\hat{\mathbf{a}}_{m,n}(k) - \vec{\mathbf{a}}(k)) \xrightarrow{\mathcal{L}} \mathcal{N}_k(\mathbf{0}, \mathbf{\Sigma}) \quad (28)$$

concluyendo así la demostración del resultado. \square

Nótese que si \mathbf{I}_k es la matriz identidad k -dimensional y se define la matriz

$$\mathbf{U} = \mathbf{I}_k - \frac{1}{k} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \quad (29)$$

\mathbf{U} es una matriz simétrica de rango $k - 1$ y, por tanto, bajo las condiciones del teorema 1 y, directamente de este,

$$\sqrt{n}(\hat{\mathbf{a}}_{m,n}(k) - \bar{\mathbf{a}}(k))\mathbf{U} \xrightarrow{\mathcal{L}} \mathcal{N}_{k-1}(\mathbf{0}, \mathbf{U}\Sigma\mathbf{U}) \quad (30)$$

Según la hipótesis nula dada en (18), se tiene que $\bar{\mathbf{a}}(k)\mathbf{U} = \mathbf{0}$; por tanto, asumiendo esta hipótesis y si $\mathbf{U}\Sigma\mathbf{U}^-$ denota la matriz pseudoinversa de $\mathbf{U}\Sigma\mathbf{U}$, se sigue que

$$\sqrt{n}\hat{\mathbf{a}}_{m,n}(k)\mathbf{U}\Sigma\mathbf{U}^- \sqrt{n}\hat{\mathbf{a}}_{m,n}(k)^t \xrightarrow{\mathcal{L}} \chi_{k-1}^2 \quad (31)$$

deduciéndose así que

$$R.C. = \{ \mathbf{X}_P \in \mathbb{R}^{m \times k}, \mathbf{X}_N \in \mathbb{R}^{n \times k} \text{ tal que } \hat{\mathbf{a}}_{m,n}(k)\mathbf{U}\Sigma\mathbf{U}^- \hat{\mathbf{a}}_{m,n}(k)^t > \chi_{k-1,\alpha}^2 \} \quad (32)$$

es una región crítica asintótica con significación α para contrastes del tipo (18).

Para ilustrar los resultados que la aplicación de este test puede tener, se dan a continuación los resultados de algunas simulaciones realizadas.

En el problema *H1* las k poblaciones de negativos proceden de distribuciones normales estandarizadas y las k poblaciones de positivos de poblaciones cuya función de densidad viene determinada por $\varphi_{1,1}$ (siendo $\varphi_{\mu,\sigma}$ la densidad de una normal con media μ y varianza σ^2); por tanto, la hipótesis nula es cierta. En el problema *H2*, las k poblaciones de negativos vuelven a proceder de una distribución gaussiana estandarizada pero, en este caso, una de las poblaciones de positivos procede de una distribución con densidad $\varphi_{2,1}$ mientras que las $k - 1$ restantes proceden de poblaciones con densidades $\varphi_{1,1}$. En la tabla 1 se muestran las proporciones de rechazos observadas en 2000 repeticiones para distintos valores de k y siendo todos los tamaños muestrales iguales a n .

En las simulaciones realizadas se observa que el estadístico sobreestima el tamaño del test y que el cómputo de la aproximación se hace lento cuando aumenta el número de curvas, si bien para la hipótesis alternativa propuesta, se obtiene una potencia elevada.

5. Aplicación práctica

Se presenta a continuación una de las múltiples aplicaciones prácticas que tienen los métodos descritos en este trabajo; en concreto, el problema que inicialmente lo motivó, fruto de la colaboración entre el Departamento de Estadística de la Universidad de Oviedo y la Unidad de Cuidados Intensivos Pediátricos del Hospital Central de Asturias (Rey et al. 2007).

TABLA 1: Potencias estimadas.

| | $n = 25$ | | | | $n = 50$ | | | |
|---------|-----------------|-------|-----------------|-------|-----------------|-------|-----------------|-------|
| | $\alpha = 0.05$ | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.01$ | |
| | $H1$ | $H2$ | $H1$ | $H2$ | $H1$ | $H2$ | $H1$ | $H2$ |
| $k = 3$ | 0.073 | 0.653 | 0.024 | 0.415 | 0.053 | 0.914 | 0.017 | 0.785 |
| $k = 4$ | 0.075 | 0.653 | 0.022 | 0.478 | 0.066 | 0.937 | 0.016 | 0.928 |
| $k = 5$ | 0.089 | 0.686 | 0.031 | 0.501 | 0.078 | 0.983 | 0.027 | 0.939 |

| | $n = 75$ | | | | $n = 100$ | | | |
|---------|-----------------|-------|-----------------|-------|-----------------|-------|-----------------|-------|
| | $\alpha = 0.05$ | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.01$ | |
| | $H1$ | $H2$ | $H1$ | $H2$ | $H1$ | $H2$ | $H1$ | $H2$ |
| $k = 3$ | 0.056 | 0.983 | 0.011 | 0.928 | 0.055 | 0.999 | 0.017 | 0.989 |
| $k = 4$ | 0.060 | 0.989 | 0.015 | 0.956 | 0.055 | 0.998 | 0.010 | 0.993 |
| $k = 5$ | 0.057 | 0.990 | 0.016 | 0.964 | 0.065 | 0.999 | 0.014 | 0.993 |

El objetivo es determinar qué variable, entre tres propuestas, discrimina mejor a los pacientes sépticos de los no sépticos. Dos de las variables son usualmente utilizadas para este propósito; la Procalcitonina (PCT) y la Proteína C-Reactiva (PCR), la tercera es el nivel de Leucocitos (LEU). Se trata de determinar si estas variables tienen la misma capacidad de determinar si un paciente tiene o no sepsis.

La muestra de la que se dispone consta de 115 pacientes *sépticos* y 215 *no infecciosos*. Algunas de las estadísticas descriptivas para la muestra son:

TABLA 2: Estadísticas descriptivas.

| | | Promedio | Desviación típica | Mín. | $P(50)$ | Máx. | n |
|-----------|-----|----------|-------------------|-------|---------|---------|-----|
| No Infec. | PCT | 1.497 | 4.116 | 0.010 | 0.300 | 39.010 | 215 |
| | PCR | 6.319 | 10.361 | 0.000 | 3.000 | 109.000 | 215 |
| | LEU | 12649 | 5811 | 1300 | 11900 | 29900 | 215 |
| Infec. | PCT | 22.767 | 40.393 | 0.110 | 10.575 | 347.100 | 115 |
| | PCR | 15.640 | 15.893 | 0.100 | 8.250 | 115.00 | 115 |
| | LEU | 14390 | 9497 | 200 | 13200 | 54900 | 115 |

Las áreas bajo las curvas ROC construidas a partir de las variables PCT, PCR y LEU son de 0.913, 0.755 y 0.528, respectivamente. La matriz de varianzas y covarianzas definida en (22) para este caso viene dada por

$$M = \begin{bmatrix} 0.00026 & 0.00011 & -0.00008 \\ 0.00011 & 0.00073 & 0.00011 \\ -0.00008 & 0.00011 & 0.00129 \end{bmatrix} \quad (33)$$

Utilizando los resultados obtenidos en la sección 4, se obtiene que el valor del estadístico es 140.768 para una significación asintótica menor de 0.0001; por tanto, la hipótesis nula será rechazada. Las diferencias encontradas entre los métodos son significativas; el método basado en la PCT es mejor que el resto ya que su AUC es notablemente mayor que el de los otros dos biomarcadores estudiados.

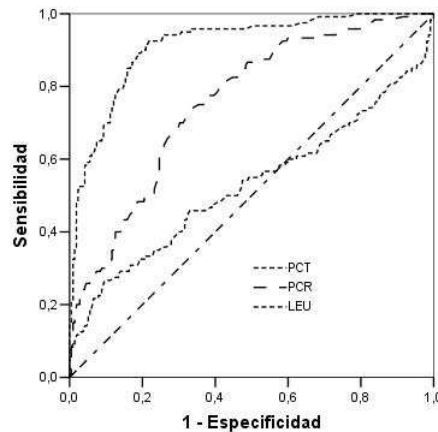


FIGURA 2: Curvas ROC para los biomarcadores considerados.

6. Comparación de dos curvas ROC

La curva ROC es una herramienta útil para determinar la calidad diagnóstica de un grupo de variables. En ocasiones, se plantea el problema de comparar dos métodos diagnósticos. En la sección anterior se vio cómo el área bajo la curva ROC puede utilizarse para este propósito, considerándose la diferencia entre estas áreas para determinar la calidad diagnóstica. Con este método no se comparan directamente las curvas ROC sino que se hace a través de sus respectivos AUC, existiendo funciones muy diferentes con idénticos AUC. Una posible solución sería el uso de la distancia del supremo (Kolmogorov-Smirnov) entre las curvas. Como se puede ver en las siguientes simulaciones, este método es menos potente en el caso de que las curvas sean iguales, pero diferencia casos “patológicos” en los que las curvas sean diferentes pero sus AUC iguales.

A continuación, se muestra la proporción de rechazos obtenidos en 2000 simulaciones por la distancia del supremo, SUP y por la diferencias entre las áreas bajo la curva ROC, AUC, cuando se estima la función de distribución real mediante la función de distribución empírica aproximando la distribución de los estadísticos mediante el método *bootstrap*. En los tres problemas que se proponen, la función de densidad de la población de negativos viene determinada por $\varphi_{0,1}(t)$. En $H1$ las dos distribuciones de positivos, G_1 y G_2 , proceden de variables con función de densidad $\varphi_{1,1}(t)$ siendo ambas curvas iguales. En el problema $H2$, G_1 procede de una variable con densidad $\varphi_{1,1}(t)$ y siendo $\varphi_{2,1}$ la densidad de G_2 . Ambas tienen diferentes AUC. Por último, en $H3$ se han buscado curvas que, a pesar de tener el mismo AUC, son distintas, viniendo determinadas por distribuciones con densidades, $0.6\varphi_{0,5,2}(t) + 0.4\varphi_{3,426,1}(t)$ y $\varphi_{1,1}$. En todos los casos se ha hecho que $n_1 = n_2 = n$ realizando las simulaciones para distintos valores de n .

En los resultados de las simulaciones, mostrados en la tabla 3 se aprecia cómo el test basado en el área bajo la curva ROC no detecta curvas distintas del tipo

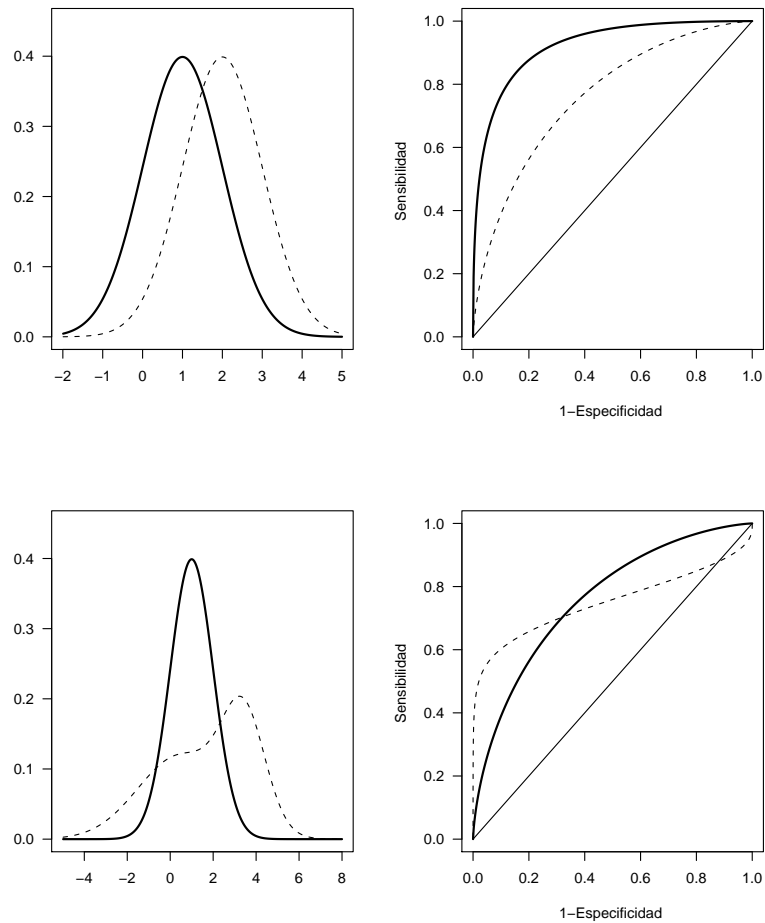


FIGURA 3: En las gráficas de la izquierda se ven las densidades de las dos poblaciones de negativos en $H2$ (arriba) y las dos poblaciones de negativos en $H3$ (abajo). A la derecha se muestran las respectivas curvas ROC resultantes.

dado en la hipótesis $H3$ (caso patológico en el que dos curvas distintas tienen el mismo AUC), siendo más potente cuando los AUC son realmente distintos (hipótesis $H2$). Cabe destacar también que los tamaños obtenidos cuando se utilizan la aproximación *bootstrap* (caso en el que la hipótesis nula es cierta, considerado en la hipótesis $H1$) para la distancia del supremo son bastantes menores de lo esperado.

7. Conclusiones

En este trabajo se revisan algunos de los métodos comúnmente utilizados en la estimación y comparación de las curvas ROC. Se presenta un método para la comparación de dos o más curvas construidas desde varios biomarcadores para un mismo diagnóstico estudiándose la distribución asintótica del estadístico resultan-

TABLA 3: Potencias estimadas.

| | $n = 25$ | | | | $n = 50$ | | | |
|------|-----------------|-------|-----------------|-------|-----------------|-------|-----------------|-------|
| | $\alpha = 0.05$ | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.01$ | |
| | AUC | SUP | AUC | SUP | AUC | SUP | AUC | SUP |
| $H1$ | 0.049 | 0.020 | 0.010 | 0.000 | 0.056 | 0.015 | 0.008 | 0.002 |
| $H2$ | 0.534 | 0.385 | 0.236 | 0.169 | 0.839 | 0.608 | 0.580 | 0.344 |
| $H3$ | 0.057 | 0.283 | 0.014 | 0.101 | 0.049 | 0.610 | 0.009 | 0.548 |

| | $n = 75$ | | | | $n = 100$ | | | |
|------|-----------------|-------|-----------------|-------|-----------------|-------|-----------------|-------|
| | $\alpha = 0.05$ | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.01$ | |
| | AUC | SUP | AUC | SUP | AUC | SUP | AUC | SUP |
| $H1$ | 0.049 | 0.019 | 0.010 | 0.001 | 0.054 | 0.024 | 0.013 | 0.003 |
| $H2$ | 0.962 | 0.778 | 0.828 | 0.533 | 1.000 | 0.922 | 0.998 | 0.735 |
| $H3$ | 0.060 | 0.746 | 0.014 | 0.525 | 0.061 | 0.886 | 0.017 | 0.704 |

te. Se muestra el rendimiento que mediante esta técnica se obtiene en un pequeño estudio de simulación y, posteriormente, se aplica a un problema real. Finalmente, se estudia la utilidad y el rendimiento de la medida de Kolmogorov-Smirnov para comparar curvas ROC.

Recibido: diciembre de 2006

Aceptado: julio de 2007

Referencias

- Bamber, D. (1975), ‘The Area Above the Ordinal Dominance Graph and the Area Below the Receiver Characteristic Graph’, *Journal of Mathematical Psychology* **12**, 387–415.
- Cai, T. & Pepe, M. S. (2002), ‘Semiparametric Receiver Operating Characteristic Analysis to Evaluate Biomarkers for Disease’, *Journal of the American Statistical Association* **97**(460), 1099–1107.
- Dorfman, D. D. & Alf, E. (1969), ‘Maximum Likelihood Estimation of Parameters of Signal Detection Theory and Determination of Confidence Interval–Rating Method Data’, *Journal of Mathematical Psychology* **6**, 487–496.
- Hall, P. G. & Hyndman, R. J. (2003), ‘An Improved Method for Bandwidth Selection When Estimating ROC Curves’, *Statistics and Probability Letters* **64**(2), 181–189.
- Hall, P. G., Hyndman, R. J. & Fan, Y. (2004), ‘Nonparametric Confidence Interval for Receiver Operating Characteristic Curve’, *Biometrika* **91**(3), 743–750.
- Hsieh, F. & Turnbull, B. W. (1996), ‘Nonparametric and Semiparametric Estimation of the Receiver Operating Characteristic Curve’, *Annals of Statistics* **24**(1), 25–40.

- Lloyd, C. J. (1998), 'Using Smoothed Receiver Operating Characteristic Curves to Summarize and Compare Diagnostic System', *Journal of the American Statistical Association* **93**(444), 1356–1365.
- Lloyd, C. & Yong, Z. (1999), 'Kernel Estimators of the ROC Curve are Better than Empirical', *Statistics Probability Letters* **44**(3), 221–228.
- Nadaraya, E. A. (1964), 'Some New Estimates for Distributions Functions', *Theory Probability Application* **9**, 497–500.
- Rey, C., Los Arcos, M., Concha, A., Medina, A., Prieto, S., Martínez-Camblor, P. & Prieto, B. (2007), 'Procalcitonin and C-reactive Protein as Markers of Systemic Inflammatory Response Syndrome Severity in Critically ill Children', *Intensive Care Medicine* **33**(3), 477–484.
- Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, Londres.
- Stute, W. (1982), 'A Law of the Logarithm for Kernel Density Estimators', *Annals of Probability* **10**(2), 414–422.
- Wand, M. P. & Jones, M. C. (1995), *Kernel Smoothing*, Chapman and Hall.
- Zhou, X. & Castelluccio, P. (2003), 'Nonparametric Analysis for the ROC Areas of two Diagnostic Test in the Presence of Nonignorable Verification Bias', *Journal of Statistics Planning and Inference* **115**, 193–213.