

Partial AUC using the sensitivity and specificity lines

Chong Sun Hong^{a,1} · Dong Hwan Jang^a

^aDepartment of Statistics, Sungkyunkwan University

(Received May 20, 2020; Revised June 12, 2020; Accepted June 14, 2020)

Abstract

The receiver operating characteristic (ROC) curve is expressed as both sensitivity and specificity; in addition, some optimal thresholds using the ROC curve are also represented with both sensitivity and specificity. In addition to the sensitivity and specificity, the expected usefulness function is considered as disease prevalence and usefulness. In particular, partial the area under the ROC curve (AUC) on a certain range should be compared when the AUCs of the crossing ROC curves have similar values. In this study, partial AUCs representing high sensitivity and specificity are proposed by using sensitivity and specificity lines, respectively. Assume various distribution functions with ROC curves that are crossing and AUCs that have the same value. We propose a method to improve the discriminant power of the classification models while comparing the partial AUCs obtained using sensitivity and specificity lines.

Keywords: AUC, sensitivity, specificity, threshold, utility

1. Introduction

본 연구는 의학 분야의 관점에서 두 범주로 구분할 수 있는 질병(disease, positive; d)과 정상(normal, negative; n)인 피험자(subject)에 대한 혼합분포를 분류하는 연구를 설명한다. 통계적 의사결정 및 여러 분야에서 폭넓게 사용되며 평가모형의 비교 및 타당성을 검정하는 대표적인 방법 중에서 성과(performance)를 기반으로 분류자(classifier)를 시각화하여 손쉽게 파악할 수 있는 유용한 방법으로 receiver operating characteristic (ROC) 곡선을 고려한다. ROC 곡선은 실제 정상을 질병으로 잘못 예측하는 비율(false positive rate; FPR) ($1 - \text{specificity}$)와 실제 질병을 질병으로 정확히 예측하는 비율(true positive rate; TPR) (sensitivity)을 각각 수평과 수직축 좌표로 표현된다 (Metz, 1978; Zweig와 Campbell, 1993; Greiner 등, 2000; Tasche, 2006). ROC 곡선에 대한 판별력을 측정하는 통계량으로 ROC 곡선의 아래의 면적을 나타내는 area under the ROC curve (AUC)를 많이 사용한다. 그리고 제약된 특정 구간에서 분류모형의 성능을 비교할 때는 부분 AUC (partial AUC; pAUC)를 이용한다 (Centor, 1991; Jiang 등, 1996; Bradley, 1997; Swets 등, 2000; Fawcett, 2003; Krzanowski와 Hand, 2009).

ROC 곡선에서 두 분포 사이의 판별력을 최대화 시켜주는 최적분류점(optimal threshold)을 구하는 다양한 방법으로 ‘closest to point(0, 1)’ (Perkins와 Schisterman, 2006), Youden 지수 (Youden, 1950), sum of sensitivity and specificity (Koepsell과 Connell, 1985), accuracy area (Brasil, 2010) 등이

¹Corresponding author: Department of Statistics, Sungkyunkwan University, 25-2, Sungkyunkwan-Ro, Jongno-Gu, Seoul 03063, Korea. E-mail: cshong@skku.edu

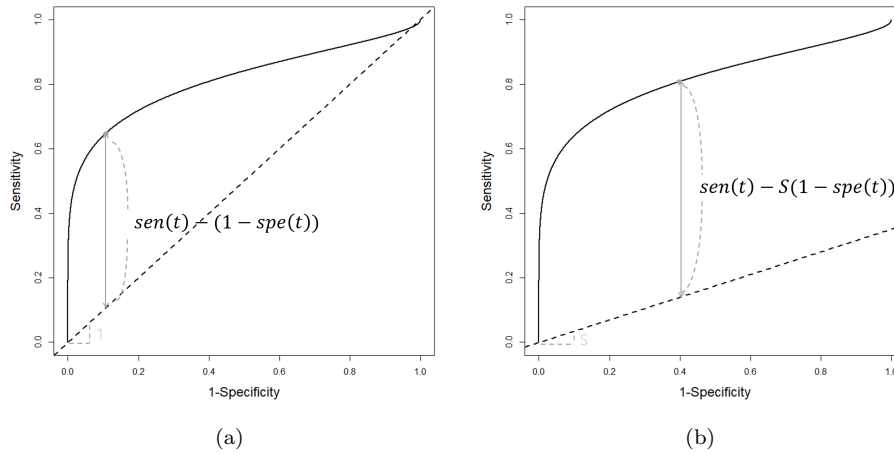


Figure 1.1. ROC curve and expected utility.

있다 (Irwin와 Irwin, 2012; Dodd와 Pepe, 2003). 이러한 방법들은 오직 민감도(sensitivity)와 특이도(specificity)만을 반영할 뿐 질병률(disease prevalence)과 같은 임상적 결과를 고려하지 않는다 (Sox 등, 1988). 대안적인 방법으로 Subtil과 Rabilloud (2015)은 질병률과 분류로부터 발생할 수 있는 효용(utility, benefit)을 고려하는 분류 방법을 제안하였다.

예측모형의 성과를 쉽게 파악할 수 있는 지표로서 2×2 혼동행렬(confusion matrix)로 나타내는데 각 칸은 true positive (TP), false positive (FP), false negative (FN), true negative (TN)으로 구성된다. 혼동행렬의 각 칸에 대응하는 효용을 나타낸 효용행렬(utility matrix)로 표현할 수 있으며 이에 대응하는 U_{TP} 와 U_{TN} 은 정분류 효용이며 U_{FP} 와 U_{FN} 은 오분류 효용이다.

최적분류점은 진단검사(diagnostic test)에 따른 예측모형의 기대효용(expected utility, expected benefit; EU)을 최대로 하는 분류점이며, Sox 등 (1988)이 제안한 효용에 고정효용 U_0 를 추가한 기대효용 함수 $EU(t)$ 는 효용행렬의 각 칸에 발생할 확률과 효용을 곱해서 합한 식 (1.1)과 같이 정의한다.

$$EU(t) = U_0 + \text{sen}(t)\pi U_{TP} + (1 - \text{spe}(t))(1 - \pi)U_{FP} + (1 - \text{sen}(t))\pi U_{FN} + \text{spe}(t)(1 - \pi)U_{TN} \\ \propto \left\{ \text{sen}(t) - \left[\frac{NC}{NB} \times \frac{1 - \pi}{\pi} \right] (1 - \text{spe}(t)) \right\}, \quad (1.1)$$

여기서 $\text{sen}(t)$ 와 $\text{spe}(t)$ 는 분류점 t 에 대한 민감도와 특이도를 각각 나타내며, π 는 질병률이다. 그리고 정상을 질병으로 분류하는 순비용(net cost; NC)은 $U_{TN} - U_{FP}$ 이며, 질병을 질병으로 분류하는 순이익(net benefit; NB)은 $U_{TP} - U_{FN}$ 이므로 NC/NB 는 순비용과 순이익에 대한 오즈(odds)로 해석한다. 식 (1.1)을 최대화하는 최적분류점 t_0 는 다음을 만족한다 (Subtil과 Rabilloud 2015, Equation A4).

$$EU(t_0) = \max_t \{ \text{sen}(t) - S(1 - \text{spe}(t)) \}, \quad (1.2)$$

여기서 $S = NC/NB \times (1 - \pi)/\pi$ 이다. 최적분류점을 구하는 방법 중에서 많이 사용하는 Youden 지수는 ROC 곡선과 대각선(chance line)의 최대 거리차, $J = \max\{\text{sen}(t) - (1 - \text{spe}(t))\}$ 이며 이를 ROC 곡선에 구현하면 Figure 1.1(a)와 같다. 즉, Youden 지수는 식 (1.2)의 기대효용함수에서 기울기 $S = 1$ 인 경우이다. 따라서 기대효용의 관점에서 살펴보면, Figure 1.1(b)와 같이 기울기가 S 인 기대효용 직선과 ROC 곡선과의 최대 거리 차가 발생할 때 기대효용이 최대화되고 이에 대응하는 분류점을 최적분류점으로 정한다 (Metz 1978).

ROC 곡선을 통하여 두 예측모형을 비교할 때, 두 ROC 곡선이 교차하지 않는다면, 큰 AUC와 pAUC 값을 갖는 예측모형이 성능이 우수하고 큰 기대효용을 발생한다고 판단할 수 있다. 그러나 두 ROC 곡선이 교차하는 경우에는 두 예측모형의 비교는 교차하지 않는 특정한 범위에 대하여 살펴보아야 한다 (McClish, 1989; Dodd와 Pepe, 2003). 이런 문제를 해결하기 위하여 Subtil과 Rabilloud (2015)은 ROC 곡선에 민감도와 특이도 직선을 고려하여, 높은 특이도 또는 민감도를 갖는 범위에 대한 ROC 곡선에서의 pAUC를 통하여 예측모형의 성능을 비교 방법을 실증 예제를 통해 제안하였다.

본 연구에서는 ROC 곡선들이 교차하고 유사한 AUC 값을 가진 모형들의 판별력을 비교하기 위해서, 정상 및 질병 상태의 다양한 분포함수들을 고려한다. 평균은 동일하지만 분산이 다른 단봉과 쌍봉의 분포함수들을 다양하게 설정하여, 특이도 직선으로부터 높은 특이도를 참조하는 pAUC와 높은 민감도를 고려하는 민감도 직선을 이용하여 pAUC를 비교하면서 모형의 판별력을 파악하는 방법을 제안한다.

본 논문의 2절에서는 Subtil과 Rabilloud (2015)가 제안한 민감도 직선과 특이도 직선을 정의하고, ROC 곡선과 각각의 직선에 의해 결정되는 pAUC에 대해 설명한다. Samawi 등 (2017)이 제시한 ROC 곡선들이 교차하고 유사한 AUC 값을 가진 모형들을 포함한 다양한 분포함수를 설정하여, 3절에서는 특이도 직선 그리고 4절에서는 민감도 직선을 바탕으로 구한 pAUC를 비교하면서 판별력을 탐색한다. 5절에서는 국내 은행의 신용평가자료를 바탕으로 ROC 곡선을 구현하고, 특이도와 민감도 직선을 설정하여 pAUC를 구하고 분석한다. 마지막 6절에서는 결론을 유도하고 활용 방법에 대하여 토론한다.

2. Specificity line and sensitivity line

예측모형은 분류성능에 따라 특정한 효용을 발생시키고 이 효용이 적절한지에 대해서 극단적인 두 가지 가정을 설정한다.

가정 1: 모든 피험자를 정상으로 간주하는 경우

가정 2: 모든 피험자를 질병으로 간주하는 경우

가정 1에서는 민감도 = 0 그리고 특이도 = 1이므로 기대효용을 $EU_{\text{treat-nobody}}$ 로 표현하고, 가정 2에서는 민감도 = 1 그리고 특이도 = 0이므로 기대효용을 $EU_{\text{treat-all}}$ 로 나타낸다. 식 (1.1)의 기대효용함수를 각 가정에 대하여 정리하면 식 (2.1)과 같다 (Subtil과 Rabilloud, 2015).

$$\begin{aligned} EU_{\text{treat-nobody}} &= 0 \\ EU_{\text{treat-all}} &= 1 - \frac{NC}{NB} \times \frac{1 - \pi}{\pi} = 1 - S \end{aligned} \quad (2.1)$$

식 (1.1)의 기대효용함수를 $EU(t) \equiv EU_{\text{test-based}}(t)$ 으로 표현하며, 이 기대효용이 가정 1과 2의 기대효용보다 높은 값을 가질 때에 유효한 예측 모형이다 (Vickers과 Elkin 2006). 우선, 가정 1에서의 기대효용보다 예측모형의 기대효용이 높으면 (즉, $EU_{\text{test-based}}(t) \geq EU_{\text{treat-nobody}}$), $sen(t) \geq S(1 - spe(t))$ 으로 나타나며, 여기서 직선 $sen(t) = S \cdot (1 - spe(t))$ 은 특이도 직선(specificity line)으로 정의하였다. 또한 가정 2에서의 기대효용보다 예측모형의 기대효용이 높은 경우에는 (즉, $EU_{\text{test-based}}(t) \geq EU_{\text{treat-all}}$), $sen(t) - 1 \geq S(1 - spe(t) - 1)$ 으로 나타나며, 여기서 직선 $sen(t) = -S \cdot spe(t) + 1$ 은 민감도 직선(sensitivity line)으로 정의하였다 (Subtil과 Rabilloud 2015). 즉,

$$\text{특이도 직선: } sen(t) = S \cdot (1 - spe(t)),$$

$$\text{민감도 직선: } sen(t) = -S \cdot spe(t) + 1,$$

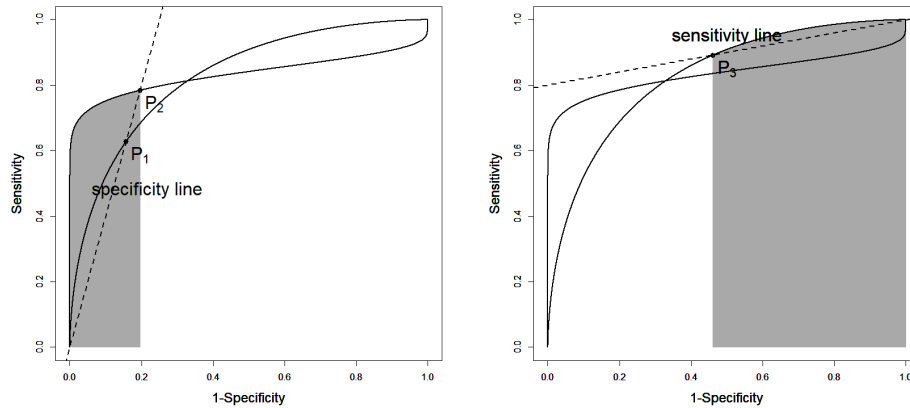


Figure 2.1. ROC curve with specificity and sensitivity lines.

여기서 특이도 직선은 ROC 곡선의 원점 $(0, 0)$ 을 통과하고 기울기 S 인 직선이며, 민감도 직선은 ROC 곡선의 점 $(1, 1)$ 을 통과하며 기울기 S 인 직선이다.

Figure 2.1(a)는 기울기가 1보다 큰 특이도 직선이 교차하는 두 ROC 곡선을 통과한다. 특이도 직선이 통과하는 점 P_2 에서의 ROC 곡선의 수평축인 $1 -$ 특이도의 값 $(= 0.196)$ 이 특이도 직선이 통과하는 다른 점 P_1 에 대응하는 $1 -$ 특이도의 값 $(= 0.157)$ 보다 높다. 따라서 점 P_2 에 대응하는 $1 -$ 특이도의 범위 $(0, 0.196)$ 에서 pAUC(회색 부분)를 구한다. 즉, 점 P_2 에 대응하는 특이도 값 $(= 0.804)$ 보다 높은 특이도를 가지는 범위 $(0.804, 1.0)$ 에 대하여 두 모형의 pAUC를 구하고 비교 분석한다. 따라서 높은 특이도 범위에서의 pAUC 중에서 큰 값을 갖는 모형이 판별력이 우수하다고 판단한다.

Figure 2.1(b)는 기울기가 1보다 작은 민감도 직선이 두 ROC 곡선 중 하나의 곡선을 통과한다. 민감도 직선과 ROC 곡선이 만나는 점 P_3 에 대응하는 민감도 $(= 0.891)$ 보다 높은 민감도에 대응하는 $1 -$ 특이도의 범위 $(0.4608, 1.0)$ 에 대하여 pAUC를 각각 구하고 비교한다. 그러므로 높은 민감도를 가지는 범위에 대응하는 두 pAUC 중에서 큰 값을 갖는 모형이 우수한 판별력을 갖는다고 결론 내릴 수 있다.

3. Specificity line and pAUC

3절에서는 교차하는 ROC 곡선들을 구현하고 특이도 직선이 통과하는 경우에 대하여 특이도 직선과 만나는 ROC 곡선들 중에서 특이도의 특성을 제일 잘 반영하는 ROC 곡선에 대응하는 특이도 범위에 대한 pAUC를 비교하고 분석한다.

높은 특이도를 갖는 경우는 특이도 직선의 기울기 S 가 1보다 큰 경우이며, 질병률 π 가 0.5 이하인 일반적인 경우에는 $NC > NB$ 로 순비용이 순이익보다 큰 경우로 해석된다. 그리고 $NC > NB$ 는 $U_{TN} + U_{FN} > U_{TP} + U_{FP}$ 이므로 만약 $U_{FN} = U_{FP} = 0$ 이면, $U_{TN} > U_{TP}$ 이 되어 특이도에 대응되는 효용 U_{TN} 이 민감도에 대응하는 효용 U_{TP} 보다 큰 경우를 의미한다. 그러므로 질병률 π 가 0.5 이하인 일반적인 경우에서 $S > 1$ 인 경우를 Figure 3.1(a)와 (b)에서 설명한다. 그러나 질병률 π 가 0.5 이상인 경우에도 NC/NB 가 1보다 매우 크다면, 기울기 S 가 1보다 클 수 있으므로 이 경우에 대해서는 Figure 3.1(c)에서 살펴본다.

본 논문에서는 교차하는 ROC 곡선을 구현하기 위하여 Samawi 등 (2017)이 제시한 분포를 확장하여, 두 질병 및 정상 상태의 분포가 모두 단봉형이고 평균이 다른 상황을 ‘경우 1(Case 1)’로, 두 분포가 모두 단봉형이며 평균은 동일하나 분산이 서로 다른 상황을 ‘경우 2(Case 2)’로, 그리고 정상 상태의 분포

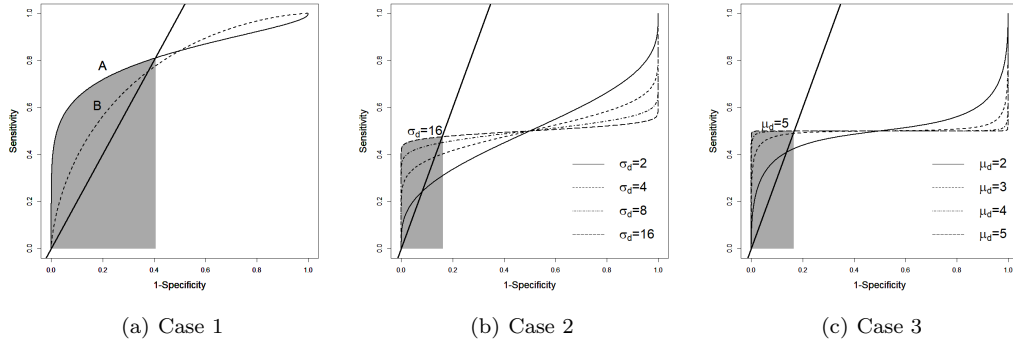


Figure 3.1. ROC curves and specificity lines.

는 단봉형이고 질병 상태의 분포는 쌍봉형인 상황을 ‘경우 3(Case 3)’로 구분하여 탐색한다. 각 경우에 대해서 다음과 같은 분포를 설정한다. 경우1에서 비교하는 두 모형을 ‘A’와 ‘B’로 설정한다.

$$\begin{aligned}
 \text{Case 1 : } & A : F_n(t) = \Phi(t; 1, 1), F_d(t) = \Phi(t; 3, 2^2), \\
 & B : F_n(t) = \Phi(t; 1, 2^2), F_d(t) = \Phi(t; 3, 2^2), \\
 \text{Case 2 : } & F_n(t) = \Phi(t; 0, 1), F_d(t) = \Phi(t; 0, \sigma_d^2); \sigma_d = \{2, 4, 8, 16\}, \\
 \text{Case 3 : } & F_n(t) = \Phi(t; 0, 1), F_d(t) = 0.5\Phi(t; -\mu_d, 1) + 0.5\Phi(t; \mu_d, 1); \mu_d = \{2, 3, 4, 5\}.
 \end{aligned}$$

세 경우의 분포함수에 대하여, 수평과 수직 좌표가 $(1 - F_n(t), 1 - F_d(t))$ 인 ROC 곡선들과 기울기 S 가 모두 1보다 큰 특이도 직선을 추가하여 Figure 3.1(a), (b), (c)에 구현하였다. Figure 3.1(a)는 특이도 직선이 경우1에 대응하는 두 ROC 곡선들을 통과하는 경우이다. 질병률은 0.5보다 작은 $\pi = 0.4$ 이고 $NC/NB = 1.33$ 으로 순비용이 순이익보다 크게 설정하여 기울기 $S = 2$ 인 특이도 직선을 고려한다. 두 ROC 곡선과 특이도 직선이 만나는 점을 살펴보면, 모형 A의 ROC 곡선에서 만나는 점이 모형 B의 ROC 곡선보다 오른쪽에 위치하며 ROC 곡선의 수평축인 $1 -$ 특이도의 값 $(= 1 - 0.5952)$ 이 모형 B의 ROC 곡선과 만나는 $1 -$ 특이도의 값보다 높다. 즉, 모형 A의 ROC 곡선과 특이도 직선이 만나는 점에 대응하는 특이도보다 높은 특이도를 가지는 범위는 모형 A가 모형 B보다 넓으며 이는 특이도의 특성을 더 많이 반영하는 것을 의미한다. 따라서 모형 A와 특이도 직선이 교차하는 점에 대응하는 특이도 값 $(= 0.5952)$ 보다 높은 특이도를 가지는 범위 $(0.5952, 1.0)$ 에 대하여 pAUC를 살펴보면, 모형 A의 pAUC(회색 부분)가 모형 B보다 큰 것을 파악할 수 있다. 그러므로 특이도 직선을 바탕으로 구한 pAUC를 이용하여 특이도가 높은 범위에서는 모형 A의 판별력이 모형 B보다 좋다고 판단할 수 있다.

Figure 3.1(b)는 특이도 직선이 평균은 동일하지만 분산이 다른 경우2에 대응하는 네 ROC 곡선들을 통과하는 경우이다. Figure 3.1(a)와 유사하게 질병률은 0.5보다 작은 $\pi = 0.4$ 이고 $NC/NB = 2.0$ 으로 설정하여 기울기 $S = 3$ 인 특이도 직선을 고려한다. 네 ROC 곡선들과 특이도 직선이 만나는 점들에 대응하는 $1 -$ 특이도 중에서, 가장 큰 $1 -$ 특이도의 값 $(= 1 - 0.8387)$ 을 나타내는 ROC 곡선은 표준 편차가 가장 큰 분포에 대응하는 ROC 곡선이다. 따라서 높은 특이도를 가지는 범위가 가장 넓은 ROC 곡선은 질병 상태의 $\sigma_d = 16$ 일 때로 특이도의 성질을 가장 많이 포함한다. 질병 상태의 $\sigma_d = 16$ 일 때 ROC 곡선과 특이도 직선이 교차하는 점에 대응하는 특이도 값 $(= 0.8387)$ 보다 높은 특이도를 가지는 범위 $(0.8387, 1.0)$ 에 대한 pAUC(회색 부분)를 비교하면, 질병 상태의 σ_d 가 증가할수록 pAUC가 커지므로 $\sigma_d = 16$ 일 때 pAUC가 가장 크다. 그러므로 특이도 직선을 바탕으로 구한 pAUC를 이용하여 평균은 동일하지만 분산이 다른 분포함수들의 ROC 곡선들 중에서 분산이 가장 큰 모형의 판별력이 특이

Table 3.1. pAUCs for three cases

Model	Case 1	Case 2		Case 3	
	pAUC	σ_d	pAUC	μ_d	pAUC
A	0.2806	2	0.0370	2	0.0561
		4	0.0569	3	0.0753
B	0.2119	8	0.0685	4	0.0814
		16	0.0745	5	0.0825

도가 높은 범위에서는 가장 우수하다고 결론내릴 수 있다.

Figure 3.1(c)에서는 특이도 직선이 평균이 다른 경우3에 대응하는 네 ROC 곡선들을 통과하는 경우이다. 질병률이 0.5보다 큰 $\pi = 0.6$ 이지만 NC/NB = 4.5으로 순비용이 순이익보다 매우 크게 설정하여 기울기 $S = 3$ 인 특이도 직선을 고려한다. 네 ROC 곡선들과 특이도 직선이 만나는 점들에 대응하는 $1 -$ 특이도 중에서, 가장 높은 값 ($= 1 - 0.8348$)을 나타내는 ROC 곡선은 평균 차이가 가장 큰 분포에 대응하는 ROC 곡선이다. 그러므로 특이도가 높은 범위를 갖는 ROC 곡선은 질병 상태의 $\mu_d = 5$ 일 때 가장 넓으며 특이도의 특성을 가장 많이 반영한다. 질병 상태의 $\mu_d = 5$ 일 때의 ROC 곡선과 특이도 직선이 교차하는 점에 대응하는 특이도 값 ($= 0.8348$)보다 높은 특이도를 가지는 범위 (0.8348, 1.0)에 대한 pAUC를 살펴보면, 질병 상태의 μ_d 가 증가할수록 pAUC가 커지므로 $\mu_d = 5$ 일 때 pAUC가 가장 크다. 따라서 특이도 직선을 바탕으로 구한 pAUC를 이용하여 정상 상태의 분포는 단봉형이고 질병 상태의 분포는 쌍봉형인 분포함수들의 ROC 곡선들 중에서 쌍봉형인 분포함수들의 평균 차이가 가장 큰 모형이 특이도가 높은 범위에서는 판별력이 가장 좋다고 파악할 수 있다.

Figure 3.1(a), (b), (c)를 통하여 각 경우에서 ROC 곡선들과 특이도 직선이 만나는 점들 중에서 가장 큰 $1 -$ 특이도 값을 선택하고, ROC 곡선들이 왼쪽 부분에서 교차하므로 0부터 가장 큰 $1 -$ 특이도 값까지 즉, 특이도의 측면에서는 특이도 값이 큰 범위에서의 pAUC를 구하고 비교하였다. 다양한 분포함수들의 ROC 곡선들 중에서 특이도 직선을 바탕으로 구한 pAUC가 가장 큰 모형에서 판별력이 가장 우수하다고 판단할 수 있다.

Figure 3.1(a)에서 두 모형 A와 B의 pAUC와 Figure 3.1(b)에서 $\sigma_d = \{2, 4, 8, 16\}$ 에 따른 pAUC 그리고 Figure 3.1의 (c)에서 $\mu_d = \{2, 3, 4, 5\}$ 에 따른 pAUC를 Table 3.1에 정리하였다. Table 3.1와 Figure 3.1를 통해 각 경우의 pAUC를 살펴보면, Figure 3.1(a)에서는 모형 A의 pAUC가 0.2806으로 모형 B의 pAUC = 0.2119보다 크다. 따라서 교차하지 않는 범위에서 pAUC가 높으면 기대효용이 높으므로 모형 A가 선호된다. Figure 3.1(b)에서 질병 상태의 표준편차 σ_d 가 증가할수록 pAUC가 증가하는 것을 알 수 있고 Figure 3.1(c)에서 질병 상태의 평균 μ_d 가 증가할수록 pAUC가 증가하는 것을 확인할 수 있다. 이는 σ_d 와 μ_d 가 증가할수록 정상과 질병 상태의 분포가 겹치는 면적이 감소하므로 판별력이 증가하기 때문에 특이도 직선을 바탕으로 구한 pAUC가 가장 큰 모형의 판별력이 최고라고 결론내릴 수 있다.

4. Sensitivity line and pAUC

3절에서 고려한 분포함수들과 동일하게 설정하여 교차하는 ROC 곡선들과 민감도 직선이 통과하는 경우에 대하여 민감도 직선과 만나는 ROC 곡선들 중에서 민감도의 특성을 제일 잘 반영하는 ROC 곡선에서 $1 -$ 특이도의 범위에서의 민감도에 대응하는 pAUC를 토론한다.

높은 민감도를 갖는 경우는 민감도 직선의 기울기 S 가 1보다 작은 경우이며, 질병률 π 가 0.5 이상인 경우에는 $NC < NB$ 으로 순이익이 순비용보다 큰 경우로 해석한다. $NC < NB$ 이면 $U_{TN} + U_{FN} <$

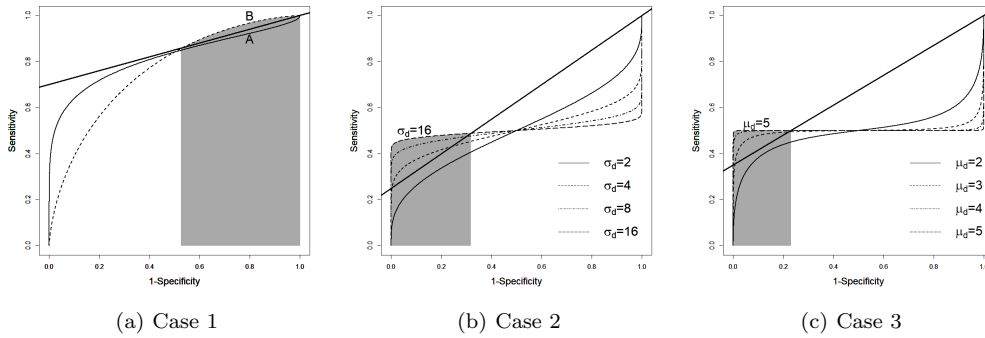


Figure 4.1. ROC curves and sensitivity line.

$U_{TP} + U_{FP}$ 이므로 만약 $U_{FN} = U_{FP} = 0$ 이면, $U_{TN} < U_{TP}$ 이므로 민감도에 대응되는 효용 U_{TP} 이 특이도에 대응하는 효용 U_{TN} 보다 큰 경우를 의미한다. 따라서 질병률 π 가 0.5 이상인 경우에서 기울기 $S < 1$ 인 경우를 Figure 4.1(a)와 (c)에서 살펴본다. 그러나 질병률 π 가 0.5 이하인 경우에도 NC/NB가 1보다 충분히 작다면, 기울기 S 가 1보다 작을 수 있으므로 이 경우에 대해서는 Figure 4.1(b)를 통해 설명한다.

3절과 동일한 ROC 곡선들을 Figure 4.1(a), (b), (c)에 구현하고, 기울기 S 가 모두 1보다 작은 민감도 직선을 추가하여, ROC 곡선들과 민감도 직선이 교차하는 경우를 살펴본다. Figure 4.1(a)는 경우 1에 대응하는 두 ROC 곡선들을 민감도 직선이 통과하는 경우이다. 질병률은 0.5보다 큰 $\pi = 0.6$ 이고 $NC/NB = 0.45$ 로 순이익이 순비용보다 크게 설정하여 기울기 $S = 0.3$ 인 민감도 직선을 고려한다. 민감도 직선이 모형 A의 ROC 곡선과는 만나지 않고 모형 B의 ROC 곡선과는 민감도 값이 0.8104인 점 ($1 - 특이도 = 0.5277$)에서 만난다. 따라서 모형 B의 ROC 곡선과 민감도 직선이 만나는 점보다 높은 민감도를 갖는 범위에 대응하는 $1 - 특이도$ 의 범위 ($0.5277, 1.0$)인 ROC 곡선의 오른쪽을 참조하고 이에 대응하는 pAUC(회색 부분)가 오른쪽 부분에 나타난다. 민감도가 높은 범위에서 모형 B의 ROC 곡선이 모형 A의 ROC 곡선보다 위에 있기 때문에 높은 민감도를 가지는 범위가 넓고 민감도의 특성을 더 포함하는 것으로 해석된다. 따라서 모형 B의 pAUC가 모형 A보다 큰 것을 알 수 있다. 그러므로 민감도 직선을 바탕으로 구한 pAUC를 이용하여 민감도가 높은 범위에서는 모형 B의 판별력이 모형 A보다 좋다고 판단할 수 있다. 참고로 특이도가 높은 범위에서 모형 A의 판별력이 모형 B보다 좋다고 판단하였다.

Figure 4.1(b)는 민감도 직선이 분산이 다른 경우 2에 대응하는 네 ROC 곡선들을 통과하는 경우이다. 질병률은 0.5보다 작은 $\pi = 0.4$ 이지만 $NC/NB = 0.5$ 으로 순이익이 순비용보다 충분히 크게 설정하여 기울기 $S = 0.75$ 인 민감도 직선을 고려한다. 네 ROC 곡선들과 민감도 직선이 만나는 점들에 대응하는 민감도 중에서 가장 높은 민감도를 나타내는 ROC 곡선은 표준편차가 가장 큰 분포에 대응하는 ROC 곡선이며 이때 민감도 값은 0.4881이다 ($1 - 특이도 = 0.3173$). 경우 2에서는 민감도 직선이 ROC 곡선의 왼쪽 부분에서 교차하므로 ROC 곡선의 왼쪽을 참조하고 이에 대응하는 pAUC는 왼쪽 부분(회색 부분)을 탐색한다. 이 경우에는 질병 상태의 표준편차가 $\sigma_d = 16$ 일 때의 ROC 곡선과 민감도 직선이 교차하는 점에 대응하는 $1 - 특이도$ 까지의 범위 ($0, 0.3173$)에 대하여 네 모형의 pAUC를 비교한다. Figure 3.1(b)와 마찬가지로 질병 상태의 표준편차 σ_d 가 증가할수록 민감도가 증가하므로 질병 상태의 $\sigma_d = 16$ 일 때의 ROC 곡선에서 민감도가 가장 높고 이에 따라 pAUC가 가장 크다. 그러므로 민감도 직선을 바탕으로 구한 pAUC를 이용하여 평균은 동일하지만 분산이 다른 분포함수들의 ROC 곡선들 중에서 분산이 가장 큰 모형의 판별력이 민감도 측면에서 가장 우수하다고 결론내릴 수 있다.

Table 4.1. pAUCs for three cases

Model	Case 1	Case 2		Case 3	
	pAUC	σ_d	pAUC	μ_d	pAUC
A	0.4317	2	0.0933	2	0.0847
		4	0.1239	3	0.1074
B	0.4476	8	0.1409	4	0.1141
		16	0.1497	5	0.1152

Figure 4.1(c)에서는 민감도 직선이 경우 3에 대응하는 네 ROC 곡선들을 통과하는 경우이다. Figure 4.1(a)와 유사하게 질병률은 0.5보다 큰 $\pi = 0.6$ 이고 $NC/NB = 0.975$ 으로 설정하여 기울기 $S = 0.65$ 인 민감도 직선을 고려한다. 네 ROC 곡선들과 민감도 직선이 만나는 점들에 대응하는 민감도 중에서 가장 높은 민감도를 나타내는 ROC 곡선은 쌍봉형인 분포함수들의 평균 차이가 가장 큰 분포에 대응하는 ROC 곡선이며 이때 민감도는 0.4999이다 ($1 - \text{특이도} = 0.2305$). Figure 4.1(b)와 마찬가지로 민감도 직선과 ROC 곡선이 만나는 점들이 ROC 곡선의 왼쪽부분에서 교차하므로 pAUC는 ROC 곡선의 왼쪽부분(회색 부분)을 계산한다. 이때 질병 상태의 $\mu_d = 5$ 일 때의 ROC 곡선과 민감도 직선이 교차하는 점에 대응하는 $1 - \text{특이도}$ 까지의 범위 ($0, 0.2305$)에 대하여 네 모형의 pAUC를 구하여 비교한다. Figure 3.1(c)와 마찬가지로 질병 상태의 μ_d 가 증가할수록 민감도가 증가하므로 질병상태의 $\mu_d = 5$ 일 때의 ROC 곡선에서 민감도가 가장 높으면서 pAUC가 가장 크다. 그러므로 민감도 직선을 바탕으로 구한 pAUC를 이용하여 정상 상태의 분포는 단봉형이고 질병 상태의 분포는 쌍봉형인 분포함수들의 ROC 곡선들 중에서 쌍봉형인 분포함수들의 평균 차이가 가장 큰 모형이 민감도 측면에서 판별력이 좋다고 파악할 수 있다.

Figure 4.1의 각 경우에서 ROC 곡선들과 민감도 직선이 만나는 점들을 살펴보면, Figure 4.1(a)에서는 ROC 곡선이 민감도 직선과 오른쪽 부분에서 교차하는 점에 대응하는 민감도 보다 높은 민감도 범위에서 pAUC를 구하였다. 그리고 Figure 4.1(b)와 (c)에서는 ROC 곡선들이 민감도 직선과 왼쪽 부분에서 교차하는 경우로, 0부터 가장 큰 민감도까지의 범위에서 pAUC를 파악하였다. 따라서 민감도 직선의 기울기 S 가 1보다 작은 경우는 pAUC가 오른쪽 또는 왼쪽 부분에 나타날 수 있음을 알 수 있다. 각 경우에 대한 pAUC를 Table 4.1에 정리하였다. Table 4.1과 Figure 4.1를 통해 각 경우의 pAUC를 살펴보면, Figure 4.1(a)에서 모형 B의 pAUC가 0.4476으로 모형 A의 pAUC = 0.4317보다 크다. 따라서 3절에서 특이도 직선을 고려하여 모형 A를 선호한 것과 다르게 민감도 직선을 고려하는 경우 모형 B가 선호된다. Figure 4.1(b)의 질병 상태의 표준편차 σ_d 와 Figure 4.1(c)의 질병 상태의 평균 μ_d 가 각각 증가할수록 pAUC는 증가한다. 이는 σ_d 와 μ_d 가 증가할수록 정상과 질병 상태의 분포가 겹치는 면적이 감소하므로 판별력이 증가하기 때문에 3절과 동일하게 민감도 직선을 바탕으로 구한 pAUC가 가장 큰 모형의 판별력이 우수하다고 결론내릴 수 있다.

5. Illustrative example

국내 I은행으로부터 신용평가모형을 위한 2018년 6월의 자료를 활용한다. 수집한 65,455개의 표본을 대기업(large enterprises), 중소기업(small and medium corporates), 개인 자료로 구분하고, 신용을 나타내는 확률변수를 20개의 등급으로 분류하였다. 본 논문에서는 이 자료 중에서 대기업, 중소기업에 대한 자료만을 발췌하여 Table 5.1에 정리한 64, 264개의 표본만을 활용하여 분석한다. 대기업과 중소기업의 신용평가 자료를 각각 $N(7.89, 2.74^2)$ 과 $N(12.4, 2.99^2)$ 인 정규분포로 적합시킨 후 표본분포함수와 정규누적분포함수를 중복시킨 Figure 5.1를 살펴보면 매우 적합하다고 판단할 수 있다.

Table 5.1. Credit evaluation data for large enterprises and small and medium corporates

Grade	Large enterprises	Small and medium corporates
1	745	0
2	1424	0
3	1837	0
4	2514	55
5	3860	56
6	5893	175
7	7676	189
8	8985	279
9	9455	518
10	5676	421
11	3759	875
12	2854	761
13	1546	631
14	1078	1102
15	5	555
16	16	965
17	12	347
18	0	0
19	0	0
20	0	0
Total	57335	6929

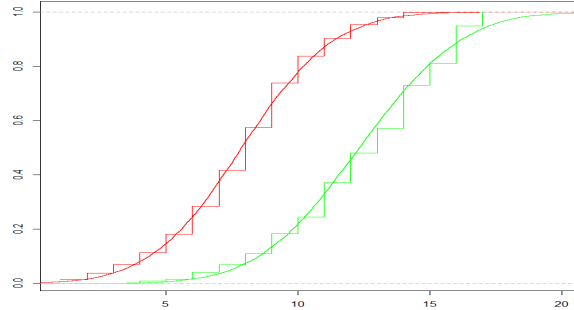


Figure 5.1. Sample and normal cumulative distribution functions.

Figure 5.2의 왼쪽 그림에서 부도율 π 는 $6929/(6929 + 57335) = 0.1078$ 이며, 네 종류의 효용을 $U_{TN} = 95$, $U_{FP} = -10$, $U_{TP} = 60$, $U_{FN} = -40$ 으로 설정하면, $NC = 105$, $NB = 100$ 이며 $NC/NB = 1.05$ 이다. 따라서 특이도 직선의 기울기 $S = 8.6902$ 로 기울기가 1보다 매우 큰 특이도 직선이 ROC 곡선을 통과하는 경우이다. 특이도에 대응되는 효용이 민감도에 대응되는 효용보다 크므로 특이도가 높은 범위를 고려하는 것이 선호되어 ROC 곡선과 특이도 직선이 만나는 점에 대응하는 특이도의 값 (= 0.9380)보다 높은 특이도를 가지는 범위 (0.9380, 1)에서 pAUC(회색 부분)는 0.0243이다. 만약 Figure 5.2의 ROC 곡선의 AUC (= 0.8669)와 유사한 값을 가진 다른 ROC 모형이 존재한다면, 이 모형의 특이도 범위 (0.9380, 1)에서 pAUC를 비교할 필요가 있다. Figure 5.2의 오른쪽 그림에서는 $U_{TN} = 20$,

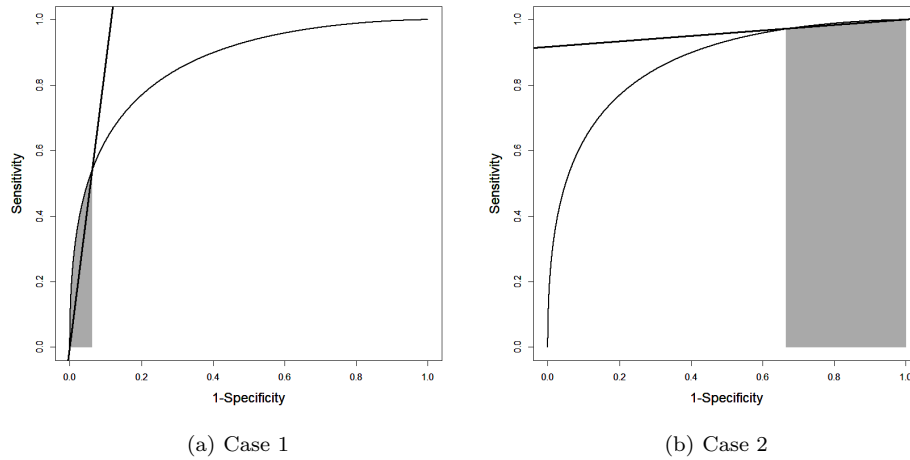


Figure 5.2. Specificity and sensitivity lines on ROC curve.

$U_{FP} = 19$, $U_{TP} = 80$, $U_{FN} = -20$ 으로 설정하면, $NC = 1$, $NB = 100$ 이고 $NC/NB = 0.01$ 이 된다. 따라서 민감도 직선의 기울기 $S = 0.0827$ 으로 기울기가 1보다 매우 작은 민감도 직선이 ROC 곡선을 통과하는 경우이다. 민감도에 대응되는 효용이 특이도에 대응되는 효용보다 크므로 민감도가 높은 범위를 고려하는 것이 선호된다. ROC 곡선과 민감도 직선이 만나는 점에서의 민감도 ($= 0.9715$)보다 높은 민감도를 가지는 범위에 대응하는 $1 -$ 특이도의 범위 $(0.6667, 1)$ 에서 pAUC(회색 부분)는 0.3298이다. 만약 유사한 AUC 값을 가진 다른 ROC 모형이 존재한다면, 이 모형의 높은 민감도에 대응하는 $1 -$ 특이도 범위 $(0.6667, 1)$ 에서 pAUC를 비교하면서 분석한다.

6. Conclusion

ROC 곡선은 예측모형의 성능을 손쉽게 파악할 수 있는 시각적 도구이며 ROC 곡선 아래의 면적인 AUC를 통해 모형의 성능을 평가한다. 교차하는 두 ROC 곡선들과 이들의 AUC를 이용하여 비교할 때는 교차하지 않는 제한된 범위에 대한 pAUC의 비교가 필요하다. Subtil과 Rabilloud (2015)은 효용과 질병률을 고려한 특이도 직선과 민감도 직선을 정의하고, 두 직선을 고려하여 높은 특이도 또는 민감도를 갖는 범위에 대해서 pAUC를 통하여 모형을 비교하는 방법을 제안하였다.

본 연구에서는 질병 및 정상 상태의 분포함수를 다양한 정규분포 및 정규혼합분포들에 대하여 AUC 값이 유사하면서 교차하는 ROC 곡선들을 설정하고, 효용과 질병률을 다양하게 고려한 특이도 직선과 민감도 직선을 바탕으로 특이도와 민감도에 제한된 범위에서의 pAUC를 탐색하였다.

기울기가 1보다 큰 특이도 직선이 ROC 곡선의 왼쪽을 통과하며 만나는 점에 대응하는 특이도가 높은 범위인 ROC 곡선의 왼쪽에서의 pAUC를 구한다. ROC 곡선과 특이도 직선이 만나는 점에 대응하는 특이도보다 높은 특이도를 가지는 범위가 가장 넓다면, 모형이 특이도의 특성을 더 많이 반영하는 것을 파악하였다. 따라서 높은 특이도를 갖는 범위에서 pAUC가 제일 큰 모형이 판별력이 가장 우수하다는 것을 발견하였다.

기울기가 1보다 작은 민감도 직선이 ROC 곡선의 오른쪽 또는 왼쪽에서 통과하는 경우가 발생한다. ROC 곡선의 오른쪽을 참조할 때, 민감도 직선과 ROC 곡선이 만나는 점에 대응하는 민감도보다 높은 민감도를 가지는 범위에서 pAUC를 탐색하였고 높은 민감도를 가지는 범위가 넓은 모형이 민감도의 성

질을 더 포함하므로 pAUC가 더 크게 나타나는 것을 탐색하였다. 그리고 정상 및 질병 상태의 분포가 단봉형으로 평균이 동일한 경우와 정상 상태의 분포는 단봉형이며 질병 상태의 분포는 쌍봉형인 경우에는 ROC 곡선의 왼쪽을 참조하는데 민감도 직선과 ROC 곡선이 만나는 점에 대응하는 민감도가 가장 큰 경우에 대한 모형의 pAUC가 가장 큰 것을 알 수 있었다. 따라서 민감도 직선을 바탕으로 구한 pAUC가 가장 큰 모형이 우수한 판별력을 가지는 것을 확인하였다.

본 연구는 예측모형에 대한 평가 및 비교를 하는 경우에 질병률과 기대효용에 대한 정보를 활용하여 분류 성능을 높이는 경우에 활용할 수 있으며, ROC 곡선이 교차하는 경우에 민감도 직선과 특이도 직선을 고려하고 민감도와 특이도 측면에서 pAUC를 구하여 우수한 판별력을 가진 모형을 선택할 수 있으며 본 연구에서 제안한 방법을 사용하면 판별 성능이 향상되는 것을 기대한다.

References

- Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms, *Pattern Recognitions*, **30**, 1145–1159.
- Brasil, P. (2010). Diagnostic test accuracy evaluation for medical professionals, Package DiagnosisMed in R.
- Centor, R. M. (1991). Signal detectability: the use of ROC analysis, *Med Decision Making*, **11**, 102–106.
- Dodd, L. E. and Pepe, M. S. (2003). Partial AUC estimation and regression, *Med Decision Making*, **59**, 614–623.
- Fawcett, T. (2003). ROC graphs: notes and practical considerations for data mining researchers, *Laboratories, Palo Alto*, **HPL-2003-4**.
- Greiner, M., Pfeiffer, D., and Smith, R. D. (2000). Principles and practical application of the receiver operating characteristic analysis for diagnostic tests, *Veterinary Medicine*, **45**, 23–41.
- Irwin, J. R. and Irwin, C. T. (2012). Appraising credit ratings: does the CAP fit better than the ROC?, *Monetary Fund Working paper*, **WP. 12/122**.
- Jiang, Y., Metz, C., and Nishikawa, R. (1996). A receiver operating characteristic partial area index for highly sensitive diagnostic tests, *Radiology*, **201**, 745–750.
- Koepsell, T. D. and Connell, F. A. (1985). Measures of gain in certainty from a diagnostic test, *American Journal of Epidemiology*, **121**, 744–753.
- Krzanowski, W. J. and Hand, D. J. (2009). *ROC Curves for Continuous Data*, Chapman & Hall/CRC, Boca Raton.
- McClish, D. K. (1989). Analyzing a portion of the ROC curve, *Med Decision Making*, **9**, 190–195.
- Metz, C. E. (1978). Basic principles of ROC analysis, *Seminars in Nuclear Medicine*, **8**, 283–298.
- Perkins, N. J. and Schisterman, E. F. (2006). The inconsistency of “optimal” cutpoints obtained using two criteria based on the receiver operating characteristic curve, *American Journal of Epidemiology*, **163**, 670–675.
- Samawi, H. M., Yin, J., Rochani, H., and Panchal, V. (2017). Notes on the overlap measure as an alternative to the Youden index: How are they related?, *Statistics in Medicine*, **36**, 4230–4240.
- Sox, H. C., Blatt, M. A., Higgins, M. C., and Marton, K. I. (1988). *Medical decision making*, Butterworths, Stoneham.
- Subtil, F. and Rabilloud, M. (2015). An enhancement of ROC curves made them clinically relevant for diagnostic-test comparison and optimal-threshold determination, *Journal of Clinical Epidemiology*, **68**, 752–759.
- Swets, J. A., Dawes, R. M., and Monahan, J. (2000). Better decisions through science, *Scientific American*, 82–87.
- Tasche, D. (2006). Validation of internal rating systems and PD estimates, *The Analytics of Risk Model Validation*, 169–196.
- Vickers, A. J. E. and Elkin, E. B. (2006). Decision curve analysis: a novel method for evaluating prediction models, *Med Decis Making*, **26**, 565–574.

Youden, W. J. (1950). Index for rating diagnostic test, *Cancer*, **3**, 32–35.

Zweig, M. and Campbell, G. (1993). Receiver-operating characteristics (ROC) plots: A fundamental evaluation tool in clinical medicine, *Clinical Chemistry*, **39**, 561–577.

민감도와 특이도 직선을 이용한 부분 AUC

홍중선^{a,1} · 장동환^a

^a성균관대학교 통계학과

(2020년 5월 20일 접수, 2020년 6월 12일 수정, 2020년 6월 14일 채택)

요약

Receiver operating characteristic (ROC) 곡선은 민감도와 특이도로 표현되며, ROC 곡선을 이용하는 최적분류 점도 민감도와 특이도만을 반영하지만, 본 연구에서는 질병률과 효용을 추가하여 고려하는 기대효용함수를 연구한다. 특히 교차하는 ROC 곡선들의 area under the ROC curve (AUC) 값들이 유사한 경우에 특정한 부분의 부분 AUC를 비교해야 한다. 본 연구에서는 정의된 민감도 직선과 특이도 직선을 바탕으로 각각 높은 민감도와 특이도를 나타내는 부분 AUC를 제안한다. ROC 곡선들이 교차하고 동일한 AUC 값을 갖는 다양한 분포함수를 설정하여, 민감도 직선과 특이도 직선을 이용하여 구한 부분 AUC를 비교하면서 모형의 판별력을 향상시키는 방법을 제안한다.

주요용어: 민감도, 분류점, 특이도, 효용, AUC

¹교신저자: (03063) 서울특별시 종로구 성균관로 25-2, 성균관대학교 통계학과. E-mail: cshong@skku.edu