

Odds curve for two classification distributions

Chong Sun Hong^{1,a}, Se Hyeon Oh^a, Tae Gyu Oh^a

^aDepartment of Statistics, Sungkyunkwan University

Abstract

The ROC, TOC, and TROC curves, which are visually descriptive methods of exploring the performance of the binary classification model, are implemented with TP, TN, FP, FN which consist of the confusion matrix, as well as their ratios TPR, TNR, FPR, FNR. In this study, we consider two types odds and then propose an odds curve representing these odds. And show the relationship between the odds curve and ROC curve. Based on the odds curve, we propose not only two statistics that measure the discriminant power of the odds curve but also the criteria for validation ratings of the odds curve. According to the shape of the odds curves, two classification distributions can be estimated and a criterion for validation ratings can be determined. The odds curve can be meaningfully used like other visual methods, and two kinds of measures for the discriminant power can be also applied together as an alternative criterion.

Keywords: confusion matrix, odds, sensitivity, specificity, validation rating

1. 서론

분류 모형을 연구하는 의학통계와 신용평가 분야 등에서 환자를 정상(non-disease / negative)상태와 질병(disease / positive)상태로 구분하면, 분류점(threshold, cut-off point)으로부터의 이항 결과(binary result)를 2×2 의 분할표인 혼동행렬(confusion matrix)로 나타낼 수 있다. 임의의 분류점에 대하여 true positive (TP)와 true negative (TN)은 각각 정확하게 분류된 질병집단과 정상집단의 수를 그리고 false positive (FP)와 false negative (FN)은 각각 질병을 정상으로 예측한 집단의 수와 정상을 질병으로 예측한 집단의 수를 의미한다. 표본의 크기 $p = TP + FN$ 인 질병상태를 나타내는 확률변수 X 의 누적분포함수 $F_d(x)$ 와 크기 $q = FP + TN$ 인 정상상태를 나타내는 분포함수 $F_n(x)$ 로 설정한다. 본 연구에서는 모든 x 에 대하여 $F_d(x) \geq F_n(x)$ 을 가정한다 (Metz와 Kronman, 1980; Swets, 1988; Hsieh와 Turnbull, 1996; Provost와 Fawcett, 1997, 2001; Engelmann 등, 2003; Pepe, 2003; Pepe 등, 2003; Fawcett, 2004, 2006; Stein, 2005; Sonogo 등, 2008; Hong 등, 2010; Hong과 Wu, 2014; Cho와 Hong, 2015; Hong 등, 2019a, 2019b).

혼동행렬의 각 칸의 비율인 TP rate (TPR, sensitivity)과 FP rate (FPR, $1 - \text{specificity}$) 그리고 FN rate (FNR)과 TN rate (TNR)은 다음과 같이 정의한다.

$$\begin{aligned} \text{TPR} &= \frac{\text{TP}}{p} = \hat{F}_d(x), & \text{FPR} &= \frac{\text{FP}}{q} = \hat{F}_n(x), \\ \text{TNR} &= \frac{\text{TN}}{q} = 1 - \hat{F}_n(x), & \text{FNR} &= \frac{\text{FN}}{p} = 1 - \hat{F}_d(x). \end{aligned}$$

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

의학통계와 신용평가 분야 등에서 많이 사용하는 receiver operating characteristic (ROC) 곡선은 이진분류 모형(binary classification model)의 성능(performance)을 탐색하는 유용한 시각적인 방법으로, 모든 분류점에서의 TPR과 FPR의 변화를 나타내는 곡선이다 (Altman과 Bland, 1994; Green과 Swets, 1966; Bamber, 1975; Egan와 Egan, 1975; Metz, 1978; Hanley와 McNeil, 1982; Swets, 1988; Centor, 1991; Zweig와 Campbell, 1993; Vuk와 Curk, 2006; Tasche, 2008).

두 표본의 크기인 p 와 q 가 매우 상이하여 표본이 치우쳐 있거나(class skew) 또는 불균형 자료(imbalanced data)인 경우에 ROC 곡선은 분류결과의 특징을 충분히 표현하지 못하며 (Fawcett, 2006; Jeni 등, 2013), 빈도수로 표현되는 혼동행렬의 정보를 설명하지 못하는 단점들을 보완하기 위하여 Pontius와 Si (2014)는 빈도수로 표현되는 total operating characteristic (TOC) 곡선을 제안하였다. TOC 곡선의 좌표는 임의의 분류점에 대하여 $(TP + FP, TP)$ 이며, 빗변의 기울기가 1(각도가 45°)인 평행사변형 안에서의 곡선으로 구현된다. TOC 곡선의 임의의 좌표에서부터 각 평행사변형의 변까지의 길이는 혼동행렬의 빈도수인 TP, FP, TN, FN을 나타내므로 ROC 곡선에서 유추할 수 없는 혼동행렬의 값들을 식별할 수 있다는 장점이 있다.

TOC 곡선은 혼동행렬의 TP, FP, TN, FN의 값들을 추론가능하나 평행사변형으로 구현되어 분류결과에 대한 특징을 파악하기가 어려운 문제점을 개선하여 Hong과 Lee (2018)가 제안한 total ROC (TROC) 곡선은 세로와 가로축을 혼동행렬의 빈도수인 TP와 FP로 구현하여, 세로와 가로의 길이가 각각 p 와 q 인 직사각형 안에서의 곡선으로 표현된다. 각 집단의 표본크기를 파악할 수 있는 TROC 곡선에서의 임의의 점으로부터 각 직사각형의 변까지의 길이는 혼동행렬의 모든 빈도수인 TP, FP, TN과 FN를 의미하므로 혼동행렬의 빈도수들을 쉽게 추정할 수 있다.

ROC 곡선은 비율로 표현하고, TOC와 TROC 곡선은 빈도수로 구현하지만, 본 연구에서는 다음과 같은 두 종류의 비율비인 오즈(odds)를 고려한다.

$$\frac{FPR}{TPR} = \frac{\hat{F}_n(x)}{\hat{F}_d(x)}, \quad \frac{FNR}{TNR} = \frac{1 - \hat{F}_d(x)}{1 - \hat{F}_n(x)}.$$

일반적인 가정 $F_d(x) \geq F_n(x)$ 에서 오즈 각각은 0과 1 사이의 값을 가지므로, 가로와 세로축의 좌표를 두 오즈인 $(F_n(x)/F_d(x), \{1 - F_d(x)\}/\{1 - F_n(x)\})$ 으로 설정하여 단위면적인 정사각형에서의 곡선으로 구현한다. 본 연구에서는 오즈로 표현하는 곡선을 오즈 곡선(odds curve)이라고 제안하고 연구한다.

본 연구의 구성은 다음과 같다. 2절에서는 본 연구에서 제안한 오즈 곡선의 성격과 특징을 설명한다. 다양한 분류 분포의 형태에 따라 구현되는 오즈 곡선들의 특징에 대하여 정리하고, ROC 곡선과의 관계를 살펴본다. 오즈 곡선을 실제의 분류 모형에 적용하고 활용하기 위하여 두 종류의 측정 통계량을 3절에서 제안한다. ROC 곡선에 대한 area under the ROC curve (AUC)와 유사한 오즈 곡선의 아래 면적인 측도와 오즈 곡선이 원점을 지나고 기울기가 1인 직선과의 교차점까지의 정사각형의 넓이 측도를 개발하여, ROC 곡선에서의 판단 기준(validation ratings)에 대응하는 오즈 곡선의 판단 기준을 설정한다. 4절에서는 분포함수에 대한 일반적인 가정이 만족하지 않은 상황에서 오즈 곡선을 설명하고, 오즈 곡선의 형태로부터 두 분류 모형을 추론할 수 있음을 설명한다. 5절에서는 실증 예제를 통해 본 연구에서 제안한 오즈 곡선을 구현하고, 오즈 곡선을 이용한 판별력 측도를 이용하여 판단 기준의 활용성에 대하여 토론한다. 마지막 6절에서는 본 연구의 결론을 유도하고 향후 연구 과제를 제시한다.

2. 오즈 곡선

2.1. 오즈 곡선의 특징

본 연구에서 제안한 오즈 곡선은 가로와 세로축의 좌표를 두 종류의 오즈인 $F_n(x)/F_d(x)$ 와 $\{1 - F_d(x)\}/\{1 - F_n(x)\}$ 로 단위면적인 정사각형 안에서 곡선으로 구현한다.

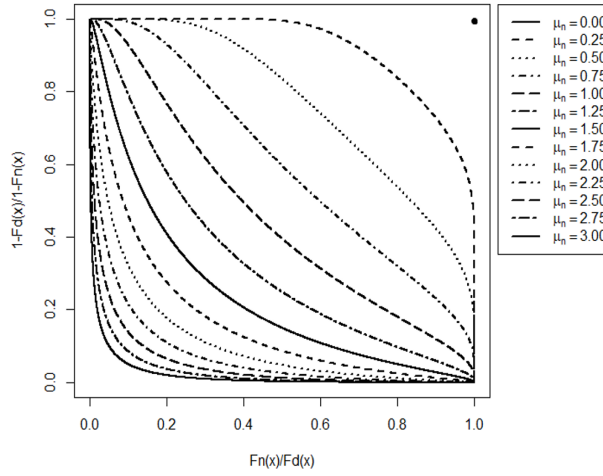


Figure 1: Odds curve when $\sigma_d^2 = \sigma_n^2$.

우선 질병상태를 나타내는 분포함수 $F_d(x)$ 를 표준정규분포 $N(\mu_d = 0, \sigma_d^2 = 1)$ 로 그리고 건강상태를 나타내는 분포함수 $F_n(x)$ 를 평균과 분산이 다양한 $N(\mu_n, \sigma_n^2)$ 로 설정하여, $\sigma_n^2 = 1, 0.5^2, 1.5^2$ 인 세 종류의 경우 각각에 대하여 0보다 큰 값을 갖는 μ_n 의 변화에 대응하는 오즈 곡선을 구현하면서 오즈 곡선의 특징을 살펴 본다.

- 경우 1: 두 분포의 분산이 동일한 경우 ($\sigma_d^2 = \sigma_n^2$)

두 분포의 분산 크기가 1로 동일할 때, Figure 1에 구현된 오즈 곡선은 점 (0, 1)에서 출발하여 점 (1, 0)에 도착하며, 건강상태의 평균 μ_n 이 커져서 두 분포의 평균 차이가 증가할수록 오즈 곡선의 중간 부분이 점 (1, 1)에서 멀어지면서 원점에 접근한다. 가로축 좌표가 0.0에서 1.0으로 변함에 따라 세로축 좌표는 감소하는데 μ_n 이 증가할수록 감소속도가 점점 빨라지면(기울기가 감소) 분류가 잘 되지 않으며, 감소속도가 급격히 빠르다가 천천히 감소하면(기울기가 증가) 분류가 잘 되고 있다고 판단할 수 있다. 따라서 오즈 곡선의 중간 부분이 원점에 가까울수록 분류가 잘 된다고 판단하며, 두 분포의 평균이 일치하여 (μ_n 이 0에 접근) 분류가 무의미(random)한 경우에는 오즈 곡선이 점 (1, 1)에 수렴한다.
- 경우 2: 건강상태의 분산이 작은 경우 ($\sigma_d^2 > \sigma_n^2$)

두 분포의 분산이 동일하지 않으며 건강상태의 분산 σ_n^2 이 σ_d^2 보다 작은 경우에는 Figure 2(a)에서와 같이 오즈 곡선이 점 (0, 1)에서 출발하여 점 (1, 1)에 도착하여, 오즈 곡선의 기울기는 x 값이 증가함에 따라 음수에서 양수로 바뀌는 U자 형태의 곡선으로 나타난다. 건강상태의 평균 μ_n 이 증가하여 평균 차이가 커질수록 오즈 곡선의 중간 부분은 위에서부터 아래의 가로축에 접근한다. 즉, 오즈 곡선의 중간 부분이 가로축에 가까워질수록 분류가 잘 된 것으로 판단한다. 또한 경우 2에서 μ_n 이 0에 접근하면, 오즈 곡선은 단위면적의 정사각형의 윗면에 가까운 곡선으로 표현된다.
- 경우 3: 건강상태의 분산이 큰 경우 ($\sigma_d^2 < \sigma_n^2$)

경우 3인 건강상태의 분산 σ_n^2 이 σ_d^2 보다 클 때에는 Figure 2(b)와 같이 점 (1, 1)에서 출발하여 점 (1, 0)에 도착하는 C자 형태로 구현된다. μ_n 이 증가하여 평균 차이가 커질수록, 오즈 곡선은 오른쪽에서부터 왼쪽

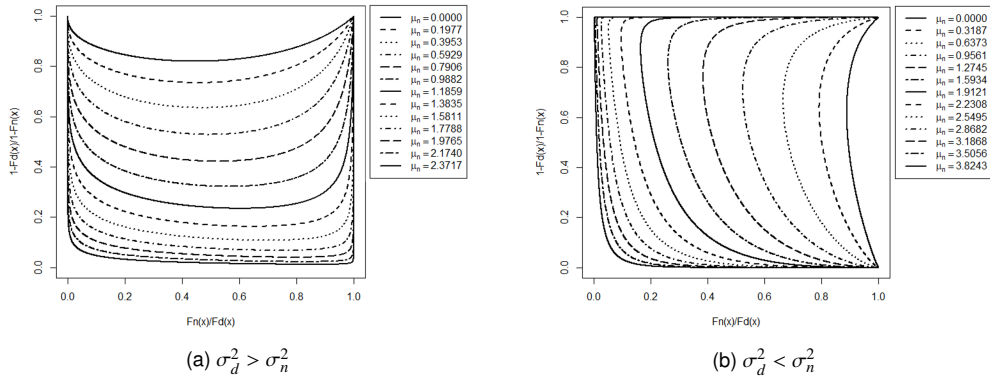


Figure 2: Odds curves when $\sigma_d^2 \neq \sigma_n^2$.

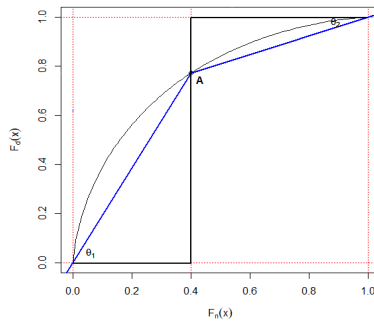


Figure 3: ROC curve and odds.

세로축으로 접근한다. 오즈 곡선의 기울기는 x 값이 증가함에 따라 기울기가 경우 2와 반대로 양수에서 음수로 감소한다. 즉, 경우 3에서는 오즈 곡선의 중간 부분이 세로축에 가까워질수록 분류가 잘 된 것으로 판단한다. 또한 μ_n 이 0에 접근하면, 단위 정사각형 오른쪽 면에 가까운 곡선으로 표현된다.

ROC에서는 곡선이 정사각형의 점 $(0, 1)$ 에, TOC에서는 곡선이 평행사변형의 점 (p, p) 에 그리고 TROC 곡선에서는 직각사각형의 점 $(0, p)$ 에 접근할수록 분류가 잘된다고 판단하지만, 두 분류 분포의 분산이 동일한 경우에는 오즈 곡선의 중간 부분이 정사각형의 원점에 가까울수록 즉, 오즈 곡선의 기울기는 음수에서 0.0으로 수렴하면 분류가 잘 된 것으로 판단할 수 있다. 그리고 두 분포의 분산이 일치하지 않는 경우에는 U자 또는 C자 형태의 오즈 곡선이 가로축과 세로축에 접근하면 분류가 잘 된 것으로 파악한다.

반면에 두 종류의 분류가 무의미한 경우에는 ROC 곡선에서는 원점과 점 $(1, 1)$ 을 지나는 대각선의 기울기가 1인 직선으로, TOC 곡선에서는 평행사변형의 대각선인 직선으로 그리고 TROC 곡선에서는 직각사각형의 대각선인 직선으로 표현되지만, 오즈 곡선에서는 두 분포의 분산이 동일한 경우에 점 $(1, 1)$ 에 수렴하거나(경우 1), 두 분포의 분산이 일치하지 않는 경우에는 윗면 또는 오른쪽면에 접근하는 곡선(경우 2와 3)으로 구현된다.

2.2. 오즈 곡선과 ROC 곡선과의 관계

ROC 곡선의 임의의 점 A에서 원점과 점 $(1, 1)$ 를 연결하는 두 직선을 고려한다. Figure 3과 같이 원점과 점 A를 연결하는 선을 빗변으로 하는 직각삼각형에서 빗변과 아래의 수평축과의 각도를 θ_1 이라고 하자. 그리고 점 $(1, 1)$ 과 점 A를 연결하는 선을 빗변으로 하는 직각삼각형에서 빗변과 위의 수평축과의 각도를 θ_2 라고 하면,

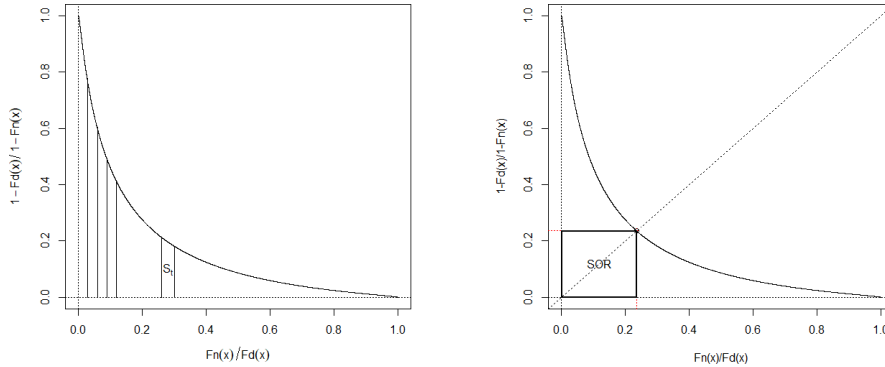


Figure 4: Two measures: AUO and SOR

본 연구에서 제안한 두 종류의 오즈는 두 각도 θ_1 과 θ_2 의 함수로 다음과 같이 정의할 수 있다.

$$\tan \theta_1 = \frac{F_d(x)}{F_n(x)}, \quad \tan \theta_2 = \frac{1 - F_d(x)}{1 - F_n(x)}.$$

따라서 오즈 곡선의 좌표는 ROC 곡선에서 설정한 두 각도의 함수 ($\tan(\pi/2 - \theta_1), \tan \theta_2$)로 해석할 수 있다. 또한, ROC 함수 $ROC(u) = F_d(F_n^{-1}(u))$ 를 이용하여 ROC 함수와 오즈와의 관계를 다음과 같이 표현할 수 있다.

$$\frac{F_d(x)}{F_n(x)} = \frac{ROC(u)}{u}, \quad \frac{1 - F_d(x)}{1 - F_n(x)} = \frac{1 - ROC(u)}{1 - u}.$$

3. 오즈 곡선에 대한 판단 기준

3.1. 판단 기준 척도들

ROC 곡선에서 이진분류모형 또는 분류자의 판별력을 측정하는 대표적인 척도로 ROC 곡선 아래 면적을 나타내는 AUC가 있다 (Hanley와 McNeil, 1982; Swets, 1988; Centor, 1991; Bradley, 1997; Engelmann 등, 2003; Vuk와 Curk, 2006). 일반적인 가정 $F_d(x) \geq F_n(x)$ 에서 AUC는 0.5부터 1까지의 값을 가지며 0.5인 경우 판별력이 전혀 없고, 0.7에서 0.8 사이의 값이면 모형이 채택할만한 수준이며, 0.8에서 1 사이의 값이면 매우 판별력이 좋은 모형이라고 할 수 있다 (Joseph, 2005; Hong 등, 2013).

본 연구에서 제안한 오즈 곡선은 Figure 1과 Figure 2에서와 같은 곡선으로 구현되므로 특히 Figure 1과 Figure 2(a)와 같은 오즈 곡선에서는 ROC 곡선에서의 AUC와 유사하게 오즈 곡선의 아래 면적을 나타내는 area under the odds curve (AUO) 척도를 다음과 같이 제안한다.

Definition 1. (오즈 곡선의 아래 면적(AUO)) 오즈 곡선에서의 판별력을 측정하는 척도로써 오즈 곡선의 아래 면적을 다음과 같이 계산하고 이를 AUO로 정의한다.

$$AUO = \sum_{all\ t} S_i = \sum_{all\ t} \left(\frac{F_n(t+1)}{F_d(t+1)} - \frac{F_n(t)}{F_d(t)} \right) \left(\frac{1 - F_d(t)}{1 - F_n(t)} \right).$$

AUO는 Figure 1의 왼쪽 그림에서의 직사각형인 S_i 의 합으로 구할 수 있다.

Figure 1과 Figure 2(a)에서의 오즈 곡선은 AUO를 계산하기 쉽고 활용하기에 매우 적합하지만, Figure 2(b)에서와 같은 오즈 곡선에서는 AUO에 대하여 의미를 부여하기 곤란하다. 그러므로 Figure 1과 Figure 2(a) 뿐만 아니라 Figure 2(b)에서와 같은 다양한 오즈 곡선에서도 판별력을 측정하는 척도로써 오즈비 제곱(squared odds ratio; SOR)을 다음과 같이 제안한다.

Table 1: Validation ratings for the odds curve when case 1

Rating	Meaning	SMD	μ_n	AUC	AUO	SOR
1	random	0	0	0.5		
2	doubtful	0.1768	0.2500	0.5702	0.9227	0.6708
3	poor	0.3536	0.5000	0.6382	0.7701	0.4493
4	marginal	0.5303	0.7500	0.7021	0.6063	0.2998
5	satisfactory	0.7071	1.0000	0.7602	0.4589	0.1991
6	good	0.8839	1.2500	0.8116	0.3370	0.1313
7	very good	1.0607	1.5000	0.8556	0.2414	0.0859
8	strong	1.2374	1.7500	0.8920	0.1689	0.0556
9	very strong	1.4142	2.0000	0.9214	0.1157	0.0356
10	excellent	1.5910	2.2500	0.9442	0.0776	0.0224
11	excellent	1.7678	2.5000	0.9614	0.0510	0.0140
12	excellent	1.9445	2.7500	0.9741	0.0328	0.0085
13	superior	2.1213	3.0000	0.9830	0.0206	0.0051

Definition 2. (오즈비 제공(SOR)) 오즈 곡선에서의 판별력을 측정하는 대안적인 척도로서 오즈 곡선이 원점을 지나고 기울기가 1인 직선과의 교차점까지의 정사각형의 넓이를 다음과 같이 계산하고, 오즈비 제공 척도로 정의한다.

$$SOR = \left(\frac{F_n(t)}{F_d(t)} \right)^2,$$

여기서 t 는 $F_n(t)/F_d(t) = (1 - F_d(t))/(1 - F_n(t))$ 를 만족하며 Figure 4의 오른쪽 그림에서와 같이 두 종류의 오즈가 동일할 때를 의미한다.

Definition 1과 2에서 제안한 판별력을 측정하는 두 종류의 척도는 0.0과 1.0 사이의 값을 가지므로 이를 이용하여 오즈 곡선의 판단 기준(validation ratings)을 설정한다. 오즈 곡선은 ROC 곡선과 밀접한 관계가 있으므로 ROC 곡선의 AUC의 판단 기준을 바탕으로 오즈 곡선의 판단 기준을 설정한다.

3.2. 판단 기준

Joseph (2005)이 제안한 ROC 곡선에서의 판단 기준은 두 분류 분포를 정규분포로 설정하고 두 분포의 표준화된 평균 차이(standardized mean difference; $SMD = (\mu_n - \mu_d) / \sqrt{(\sigma_d^2 + \sigma_n^2)}$)를 기반으로 하고 있으므로 본 연구에서도 유사하게 설정하여 질병상태를 나타내는 분포함수 $F_d(x)$ 를 평균과 분산이 $\mu_d = 0, \sigma_d^2 = 1$ 인 표준정규분포로 그리고 정상상태를 나타내는 분포함수 $F_n(x)$ 를 평균과 분산이 각각 μ_n, σ_n^2 인 정규분포로 설정하고, 두 분포의 분산이 동일한 경우와 건강상태의 분산이 서로 다른 두 종류의 경우에 대하여 오즈 곡선의 판단 기준을 제시한다.

- 경우 1: 두 분포의 분산이 동일한 경우 ($\sigma_d^2 = \sigma_n^2$)

두 분포의 분산이 동일한 경우 1에서는 μ_n 를 0부터 3까지 0.25씩 증가시킨 $SMD = \mu_n / \sqrt{(1^2 + 1^2)}$ 를 기반으로 하는 AUC의 13단계의 ROC 곡선의 판단 기준을 바탕으로 오즈 곡선의 판단 기준을 설정한다. 경우 1에서는 판별력을 측정하는 척도로 AUO와 SOR 모두 사용할 수 있으며 AUC와 더불어 AUO와 SOR 척도의 판단 기준을 Table 1에 정리하였다.

Table 1를 통하여 분산이 동일한 경우에 표준화된 평균 차이가 증가할수록, AUO와 SOR값은 작아지는데 AUO는 1.0에 가까운 값부터 0.0206까지 감소하며, SOR은 AUO보다 작은 값을 가지며 감소하는 것을 파악할 수 있다. 특히, 분산이 동일할 때 SOR은 최대값을 갖는다. AUC값이 0.8 을 초과하는 6번째 단계인

Table 2: Validation ratings for the odds curve when case 2

Rating	Meaning	SMD	μ_n	AUC	AUO	SOR
1	random	0	0	0.5	0.8630	
2	doubtful	0.1768	0.1977	0.5702	0.7920	0.6564
3	poor	0.3536	0.3953	0.6382	0.7045	0.4301
4	marginal	0.5303	0.5929	0.7021	0.6042	0.2807
5	satisfactory	0.7071	0.7906	0.7602	0.4973	0.1821
6	good	0.8839	0.9882	0.8116	0.3920	0.1172
7	very good	1.0607	1.1859	0.8556	0.2957	0.0746
8	strong	1.2374	1.3835	0.8920	0.2140	0.0470
9	very strong	1.4142	1.5811	0.9214	0.1491	0.0292
10	excellent	1.5910	1.7788	0.9442	0.1005	0.0178
11	excellent	1.7678	1.9765	0.9614	0.0675	0.0107
12	excellent	1.9445	2.1740	0.9741	0.0419	0.0063
13	superior	2.1213	2.3717	0.9830	0.0261	0.0036

'good'부터 AUO값은 0.3370의 작은 값을 갖는다. 따라서 경우 1에서 AUO값이 0.40 이하이면, 판별력이 좋다고 판단할 수 있다. 또한 SOR값이 0.13 이하이면 6번째 단계의 'good' 상황이며, 0.10 미만의 값을 가지면, 7단계인 'very good' 이상으로 판별력이 좋은 상황이라고 파악할 수 있다.

- 경우 2: 건강상태의 분산이 작은 경우 ($\sigma_d^2 > \sigma_n^2 = 0.5^2$)

건강상태의 분산이 작은 경우 2에서도 SMD를 기반으로 13단계 판단 기준을 제시한다. 여기서 $SMD = \mu_n / \sqrt{(1^2 + 0.5^2)}$ 이다. 경우 2에서도 경우 1과 마찬가지로 판별력을 측정하는 측도 AUO와 SOR 모두 사용할 수 있으며 Table 2에 판단 기준을 정리하였다. Table 2의 SMD와 AUC는 Table 1과 동일하게 설정하고 이에 대응하는 AUO와 SOR 측도를 계산하였다.

경우 2에서도 표준화된 평균 차이가 증가할수록 AUO와 SOR은 작아짐을 Table 2를 통하여 파악할 수 있다. 13단계 판단 기준의 낮은 단계에서 경우 2의 AUO는 경우 1의 AUO보다 작은 값을 가지나, 5단계부터는 경우 2의 AUO값이 경우 1의 AUO보다 큰 값을 가지며 감소한다. AUC값이 0.8을 초과하는 6번째 판단 기준인 'good'부터는 AUO값은 0.40 이하의 작은 값을 갖는다. 따라서 경우 1과 2에서 AUO값이 0.40 이하이면 판별력이 좋다고 판단할 수 있다. 또한 경우 2의 SOR값은 경우 1보다 조금씩 작은 값을 가지며 감소한다. 그러므로 경우 1과 2를 통해서 SOR이 0.12과 0.13 사이의 값을 가지면 6번째 단계의 'good'이며, 0.10 미만의 값을 가지면, 7단계인 'very good' 이상으로 판별력이 좋다고 파악할 수 있다.

- 경우 3: 건강상태의 분산이 큰 경우 ($\sigma_d^2 < \sigma_n^2 = 1.5^2$)

건강상태의 분산이 큰 경우 3에서도 경우 1과 2에서 구한 $SMD = \mu_n / \sqrt{(1^2 + 1.5^2)}$ 를 기반으로 판단 기준을 제시한다. 경우 3에서는 판별력을 측정하는 측도 AUO는 정의할 수 없고 SOR 측도만 사용할 수 있으므로 Table 3에 SOR의 판단 기준을 정리하였다.

Table 3을 통하여 표준화된 평균 차이 SMD가 증가할수록 SOR값은 작아짐을 알 수 있는데 경우 3의 SOR값은 경우 1보다 작지만, 경우 2보다는 큰 값을 가져서 경우 1과 2 사이의 값을 갖는 것을 파악할 수 있다. AUC값이 0.8을 초과하는 6번째 판단 기준인 'good'부터는 경우 1부터 3까지의 상황에서의 SOR값이 0.12와 0.13 정도의 작은 값을 갖는다. 그러므로 경우 1부터 3까지에서 일반적인 SOR값이 0.10 이하의 값을 가지면 판별력이 좋다고 판단할 수 있다.

Table 3: Validation ratings for the odds curve when case 3

Rating	Meaning	SMD	μ_n	AUC	SOR
1	random	0	0	0.5	
2	doubtful	0.1768	0.3187	0.5702	0.6655
3	poor	0.3536	0.6373	0.6382	0.4421
4	marginal	0.5303	0.9561	0.7021	0.2927
5	satisfactory	0.7071	1.2745	0.7602	0.1927
6	good	0.8839	1.5934	0.8116	0.1259
7	very good	1.0607	1.9121	0.8556	0.0816
8	strong	1.2374	2.2308	0.8920	0.0523
9	very strong	1.4142	2.5495	0.9214	0.0331
10	excellent	1.5910	2.8682	0.9442	0.0206
11	excellent	1.7678	3.1868	0.9614	0.0127
12	excellent	1.9445	3.5056	0.9741	0.0076
13	superior	2.1213	3.8243	0.9830	0.0045

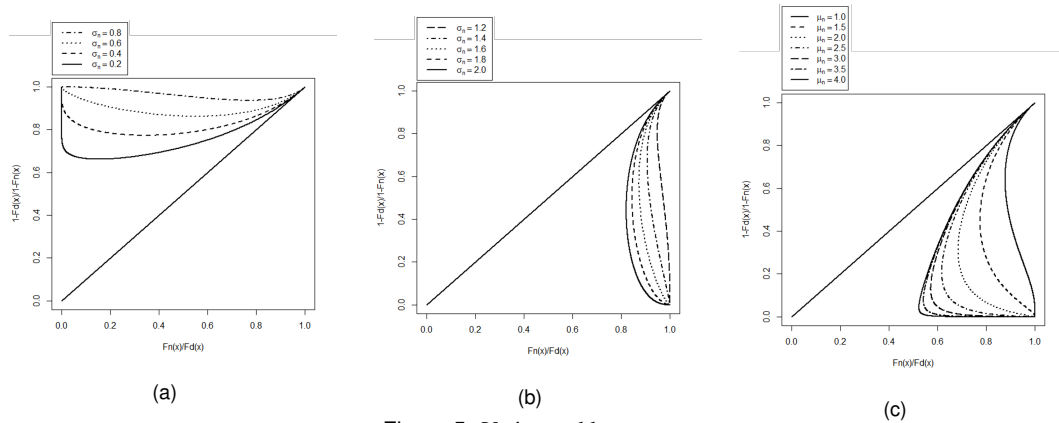


Figure 5: Various odds curves.

4. 가정이 만족하지 않은 경우의 오즈 곡선들

본 연구에서 설정한 일반적인 가정 $F_d(x) \geq F_n(x)$ 이 만족하지 않으며 $AUC = 0.5$ 로 판별이 안 되는 대표적인 세 가지 종류의 경우를 고려하고, 각각의 경우에 대응하는 오즈 곡선을 구현하고 특징을 살펴본다. 경우 4부터 6까지 모두 질병상태의 분포는 경우 1부터 3까지와 동일하게 표준정규분포로 설정하고, 정상상태의 분포의 평균은 0으로 동일하지만 분산 σ_n^2 이 다른 정규분포 또는 정규혼합분포로 설정한다. 따라서 이 경우들에서는 확률변수 X 의 값이 양수 또는 음수인 경우에만 가정 $F_d(x) \geq F_n(x)$ 이 만족한다.

- 경우 4: 정상상태 분포의 분산이 작은 경우 ($\sigma_d^2 > \sigma_n^2$).
- 경우 5: 정상상태 분포의 분산이 큰 경우 ($\sigma_d^2 < \sigma_n^2$).
- 경우 6: 정상상태 분포가 정규혼합분포인 경우 ($0.5 N(\mu_n, 1) + 0.5 N(-\mu_n, 1)$)

경우 4부터 6까지에서는 두 종류의 오즈가 1.0보다 큰 값을 갖는 영역이 발생하므로 $F_d(x) \geq F_n(x)$ 가정을 만족하는 확률변수의 부분영역에서만 오즈 곡선을 Figure 5에 구현하였다. Figure 5(a)는 경우 4의 σ_n 을 0.8부터 0.2까지 0.2씩 감소시키고, Figure 5(b)는 경우 5의 σ_n 을 1.2부터 2.0까지 0.2씩 증가시키면서 오즈 곡선을 구현하였고, Figure 5(c)는 경우 6에서 μ_n 을 1.0부터 4.0까지 0.5씩 증가시킨 오즈 곡선이다.

Table 4: RS rating data of K bank

RS grade	1	2	3	4	5	6	7	8	9	10	
Default	745	1424	1837	2514	38060	5893	7676	8985	9455	5676	
Non-default	0	0	0	55	56	175	189	279	518	421	
RS grade	11	12	13	14	15	16	17	18	19	20	Total
Default	3759	2854	1546	1078	5	16	12	0	0	0	57335
Non-default	875	761	631	1102	555	965	347	0	0	0	6929

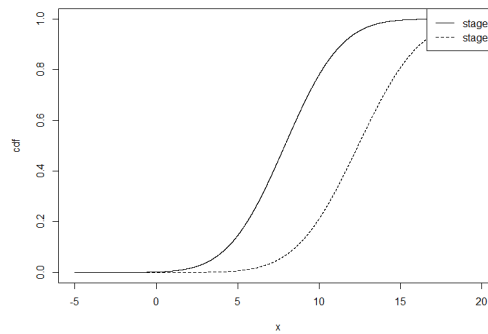


Figure 6: Two cdf's.

경우 4에서는 가정 $F_d(x) \geq F_n(x)$ 이 만족하는 확률변수 X 의 음수 영역에서는 $0.5 \leq 1 - F_d(x) \leq 1 - F_n(x) \leq 1$ 이므로 $(1 - F_d(x))/(1 - F_n(x)) \geq 0.5$ 를 만족하여, 오즈 곡선들은 Figure 5(a)에서와 같이 대각선 오른쪽의 위에 위치한다. 가정이 만족하는 확률변수 X 의 영역이 양수인 경우 4와 5에서는 $0.5 \leq F_d(x) \leq F_n(x) \leq 1$ 이므로 $F_n(x)/F_d(x) \geq 0.5$ 를 만족하여, 오즈 곡선들은 Figure 5(b)와 (c)에서 살펴보듯이 대각선 아래의 오른쪽에 위치한다. 또한, 경우 5의 정규혼합분포에서 평균 μ_n 이 커지면 혼합분포의 분산이 경우 4보다 크고, 이에 대응하는 Figure 5(c)의 오즈 곡선이 Figure 5(b)보다 원점을 향하면서 최소의 $F_n(x)/F_d(x)$ 값이 0.5에 접근하고 있음을 파악할 수 있다.

Figure 5에서 구현된 경우 4, 5, 6의 모든 오즈 곡선에서 AUO와 SOR 측도를 사용할 수 없으므로 3절에서 제시한 판단 기준을 적용할 수 없다. 그럼에도 불구하고 경우 4에서 σ_n 이 감소할수록, 경우 5에서 σ_n 이 증가할수록 그리고 경우 6에서 μ_n 이 증가할수록, 두 분류 분포의 판별이 가능해지는 것을 파악할 수 있으며 이에 대응하는 오즈 곡선의 중간 부분이 원점에 가까워지는 것을 식별할 수 있다. 그러므로 오즈 곡선의 특성을 설명한 2절과 3절에서 유도한 판단 기준 측도들의 성격과 일치하므로, 오즈 곡선의 중간 부분이 원점에 가까워지면 판별이 잘 된 것으로 파악할 수 있다. 그러므로 오즈 곡선의 위치와 형태를 통해서도 두 분류 분포의 성격과 특징을 판단할 수 있다.

5. 실증 예제

한국 K은행의 신용평가모형을 위한 2018년 6월에 수집한 RS 등급(risk score rating) 자료를 이용한다 (Hong과 Jang, 2020; Hong과 Na, 2020; Hong, 2021). 표본 크기는 65,455로 세 종류의 집단으로 구분하며 높을수록 파산확률이 높다고 간주하는 신용을 20개의 RS 등급으로 분류했다. 본 연구에서는 표본크기가 작은 세 번째 집단을 제외하고 64,264개의 표본을 첫 번째와 두 번째 집단을 부도(default)와 정상(non-default)상태로 간주하고 Table 4에 정리하였다.

Hong과 Jang (2020) 그리고 Hong과 Na (2020)는 Table 4의 자료에 적합한 분포를 각각 $N(7.89, 2.74^2)$

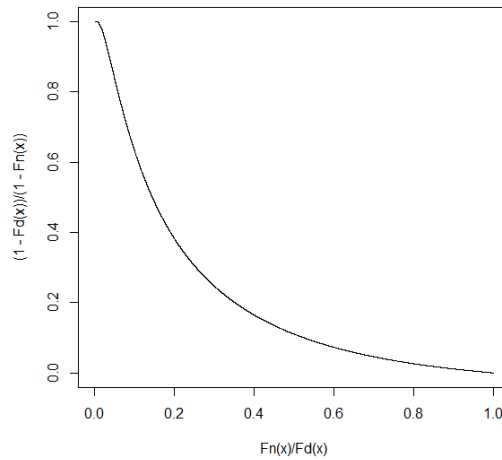


Figure 7: Odds curve.

과 $N(12.4, 2.99^2)$ 로 추정하였으므로 본 연구에서도 동일한 분포에 적합하다고 설정한다. 그리고 부도상태의 분포를 $N(7.89, 2.74^2)$ 를 표준정규분포로 표준화시키면, 정상상태의 분포의 평균은 부도상태의 평균만큼 줄어들고, 정상상태의 분산은 부도상태의 분산으로 나누어져 정상상태의 분포를 $N(4.59, 1.09^2)$ 로 간주할 수 있다. 두 분포의 누적분포함수를 Figure 6에 구현하였고 이에 대응하는 오즈 곡선은 Figure 7에 구현하였다.

표준화된 정상상태의 분산 $\sigma_n = 1.09$ 은 1.0과 유사한 값으로 간주하여 두 집단의 분산이 같을 경우에 대한 Table 1의 판단 기준을 적용할 수 있다. 두 분포에 대한 SMD값은 1.1120이고 이에 대응하는 AUO값은 0.2120 그리고 SOR값은 0.0756으로 산정되었다. Table 1에 따르면 AUO는 7단계에 해당하는 구간 (0.1689, 0.2414)에 속하며, SOR 역시 7단계에 해당하는 구간 (0.0556, 0.0859)에 속하여 'very good'으로 동일하게 판단할 수 있다. 여기서 AUC값은 0.8669로 판단 기준 7단계에 해당하며 위의 결과와 동일한 해석을 보여준다.

6. 결론

이진분류모형의 성능을 탐색하는 시각적인 방법으로 ROC 곡선과 TOC 곡선 그리고 TROC 곡선이 존재한다. 이 곡선들은 혼동행렬을 구성하는 TP, TN, FP, FN 그리고 이들의 비율인 TPR, TNR, FPR, FNR으로 구현된다. 본 연구에서는 두 종류의 비율비인 오즈 FPR/TPR 와 FNR/TNR 을 고려한다. 일반적인 가정에서 오즈 각각은 0과 1 사이의 값을 가지므로, 가로와 세로축의 좌표를 두 오즈인 (FPR/TPR , FNR/TNR)으로 설정하여 단위면적인 정사각형에서 구현하는 오즈 곡선을 제안하였다.

두 분포의 분산 크기가 동일할 때, 오즈 곡선은 점 (0, 1)에서 출발하여 점 (1, 0)에 도착하며, 두 분포의 평균 차이가 증가할수록 오즈 곡선의 중간 부분이 점 (1, 1)에서 멀어지면서 원점에 접근한다. 즉, 가로축 좌표가 0.0에서 1.0으로 변함에 따라 감소속도가 점점 빨라지면 분류가 잘 되지 않으며, 감소속도가 급격히 빠르다가 천천히 감소하면 분류가 잘 되고 있다고 판단할 수 있다.

두 분포의 분산이 동일하지 않으며 건강상태의 분산이 질병상태의 분산보다 작은 경우에는 오즈 곡선이 점 (0, 1)에서 출발하여 점 (1, 1)에 도착하는 U자 형태를 보인다. 평균 차이가 커질수록 오즈 곡선의 중간 부분은 위에서부터 아래의 가로축에 접근한다. 즉, 오즈 곡선의 중간 부분이 가로축에 가까워질수록 분류가 잘 된 것으로 판단한다.

건강상태의 분산이 질병상태의 분산보다 클 때에는 점 (1, 1)에서 출발하여 점 (1, 0)에 도착하는 C자 형태

로 구현된다. 평균 차이가 커질수록, 오즈 곡선은 오른쪽에서부터 왼쪽 세로축으로 접근한다. 즉, 오즈 곡선의 중간 부분이 세로축에 가까워질수록 분류가 잘 된 것으로 판단한다.

반면에 두 종류의 분류가 무의미한 경우에는 ROC 곡선에서는 원점과 점 (1, 1)을 지나는 대각선의 기울기가 1인 직선으로, TOC 곡선에서는 평행사변형의 대각선인 직선으로 그리고 TROC 곡선에서는 직각사각형의 대각선인 직선으로 표현되지만, 오즈 곡선에서는 두 분포의 분산이 동일한 경우에 점 (1, 1)에 수렴하거나, 두 분포의 분산이 일치하지 않는 경우에는 윗면 또는 오른쪽면에 접근하는 곡선으로 구현된다.

또한 본 연구에서 제안한 두 종류의 오즈는 ROC 곡선의 임의의 점에 대한 각도들의 함수로 표현할 수 있으므로 ROC 곡선과의 관계를 설명하였다.

오즈 곡선에서 판별력을 측정하는 측도로 ROC 곡선의 AUC와 유사하게 오즈 곡선의 아래 면적을 나타내는 AUO와 두 종류의 오즈가 동일할 때의 점까지의 정사각형의 넓이인 오즈비 제곱 SOR을 제안하였다. 표준화된 평균 차이를 기반으로 하는 AUC의 판단 기준을 바탕으로 AUO와 SOR 측도를 이용하여 오즈 곡선의 판단 기준을 설정하였다. 두 분류 분포의 평균 차이가 증가할수록 AUO는 1.0에 가까운 값에서 출발하여 0.0206까지 감소하며 SOR은 AUO보다 작은 값을 가지며 감소한다. 특히, 분산이 동일할 때 SOR은 최대값을 갖는다. AUC값이 0.8을 초과하는 13단계 중에서 6번째 단계인 'good'과 동등하게 AUO값이 0.40 이하이면, 판별력이 좋다고 판단할 수 있다. 또한 SOR값이 0.12과 0.13 사이의 값을 가지면 6번째 단계의 'good'이며, 0.10 미만의 값을 가지면, 7단계인 'very good' 이상으로 판별력이 좋다고 파악할 수 있다. 그러므로 SOR 측도만 사용할 수 있는 경우에도 0.10 이하의 값을 가지면 판별력이 충분히 좋다고 판단할 수 있다.

본 연구에서 설정한 일반적인 가정이 만족하지 않으며 분류가 무의미하여 판별이 안 되는 경우의 오즈 곡선에서는 AUO와 SOR 측도를 사용할 수 없으므로 판단 기준을 적용할 수는 없다. 그러나 오즈 곡선의 중간 부분이 원점에 가까워지면 판별이 잘 된 것으로 파악할 수 있으므로 오즈 곡선의 위치와 형태를 통해서도 두 분류 분포의 특성을 판단할 수 있으며 이를 신용평가모형을 위한 실증 예제를 통해서 살펴보았다.

다양한 오즈 곡선들의 형태를 바탕으로 두 분류 분포를 예측하고 추정할 수 있으며, 판별력을 측정하는 두 종류의 측도 AUO와 SOR을 이용한 13단계의 판단 기준으로 판별력을 측정하고 판단할 수 있다. 그러므로 본 연구에서 제안한 시각적 방법인 오즈 곡선은 ROC, TOC 그리고 TROC 곡선 등과 같이 두 분포의 분류 성능을 탐색하는 방법으로 유용하게 사용할 수 있으며, 오즈 곡선의 판별력을 측정하는 두 종류의 측도 AUO와 SOR은 ROC 곡선에 대한 AUC의 대안적인 측도로서 분류 성능을 판단하는 방법으로 같이 사용할 수 있다.

본 연구의 3절에서 두 분포의 분산이 다른 두 가지 경우만을 연구하였는데, 향후 연구에서는 분산이 동일하지 않은 다양한 크기의 분산에 대하여 탐색하고, 이에 대응하는 판단 기준에 대한 추가적인 연구가 필요하다. 그리고 ROC 곡선에서 여러 종류의 정확도 측도(accuracy measures)들을 이용하여 최적 분류점을 발견하는 방법들과 유사하게, 오즈 곡선에서 최적 분류점을 발견하기 위한 방법을 탐색하고 이 방법에 대응하는 정확도 측도들에 대한 연구를 향후 과제로 남겨둔다.

References

- Altman, D. G. and Bland, J. M. (1994). Diagnostic tests. 1: Sensitivity and specificity, *BMJ: British Medical Journal*, **308**, 1552.
- Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph, *Journal of Mathematical Psychology*, **12**, 387–415.
- Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms, *Pattern Recognition*, **30**, 1145–1159.
- Centor, R. M. (1991). Signal detectability: the use of ROC curves and their analyses, *Medical Decision Making*, **11**, 102–106.

- Cho, M. H. and Hong, C. S. (2015). Two optimal threshold criteria for ROC analysis, *Journal of the Korean Data and Information Science Society*, **26**, 255–260.
- Egan, J. P. and Egan, J. P. (1975). *Signal Detection Theory and ROC-Analysis*, Academic Press.
- Engelmann, B., Hayden, E., and Tasche, D. (2003). Testing rating accuracy, *Risk*, **16**, 82–86.
- Fawcett, T. (2004). ROC graphs: Notes and practical considerations for researchers, *Machine Learning*, **31**, 1–38.
- Fawcett, T. (2006). An introduction to ROC analysis, *Pattern Recognition Letters*, **27**, 861–874.
- Green, D. M. and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (Vol. 1), Wiley, New York.
- Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, **143**, 29–36.
- Hong, C. S. (2021). Confusion plot for the confusion matrix, *The Korean Data & Information Science Society*, **32**, 427–437.
- Hong, C. S. and Jang, D. H. (2020). Partial AUC using the sensitivity and specificity lines, *The Korean Journal of Applied Statistics*, **33**, 541–553.
- Hong, C. S., Joo, J. S., and Choi, J. S. (2010). Optimal thresholds from mixture distributions, *The Korean Journal of Applied Statistics*, **23**, 13–28.
- Hong, C. S., Jung, E. S., and Jung, D. G. (2013). Standard criterion of VUS for ROC surface, *The Korean Journal of Applied Statistics*, **26**, 977–985.
- Hong, C. S., Jung, M. S., and Shin, H. S. (2019a). Partial VUS and optimal thresholds, *The Korean Data & Information Science Society*, **30**, 895–905.
- Hong, C. S., Jung, M. S., and Shin, H. S. (2019b). Three-way partial VUS, *The Korean Data & Information Science Society*, **30**, 445–454.
- Hong, C. S. and Lee, S. J. (2018). TROC curve and accuracy measures, *The Korean Data & Information Science Society*, **29**, 861–872.
- Hong, C. S. and Na, H. R. (2020). Maximum absolute determinant of the confusion matrix, *The Korean Data & Information Science Society*, **31**, 937–950.
- Hong, C. S. and Wu, Z. Q. (2014). Alternative accuracy for multiple ROC analysis, *Journal of the Korean Data & Information Science Society*, **25**, 1521–1530.
- Hsieh, F. and Turnbull, B. W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve, *Annals of Statistics*, **24**, 25–40.
- Jeni, L. A., Cohn, J. F., and De La Torre, F. (2013). Facing imbalanced data—recommendations for the use of performance metrics. In 2013 Humaine association conference on affective computing and intelligent interaction (pp. 245–251). IEEE.
- Joseph, M. P. (2005). A PD validation framework for Basel II internal ratings-based systems, *Credit Scoring and Credit Control IV*.
- Metz, C. E. (1978). *Basic principles of ROC analysis*. In *Seminars in nuclear medicine* (Vol. 8, No. 4, pp. 283–298). WB Saunders.
- Metz, C. E. and Kronman, H. B. (1980). Statistical significance tests for binormal ROC curves, *Journal of Mathematical Psychology*, **22**, 218–243.
- Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford University Press, Oxford.
- Pepe, M. S., Longton, G., Anderson, G. L., and Schummer, M. (2003). Selecting differentially expressed genes from microarray experiments, *Biometrics*, **59**, 133–142.

- Pontius Jr, R. G. and Si, K. (2014). The total operating characteristic to measure diagnostic ability for multiple thresholds, *International Journal of Geographical Information Science*, **28**, 570–583.
- Provost, F. and Fawcett, T. (1997). Analysis and visualization of classifier performance with nonuniform class and cost distributions. In *Proceedings of AAAI-97 Workshop on AI Approaches to Fraud Detection & Risk Management*, 57–63.
- Provost, F. and Fawcett, T. (2001). Robust classification for imprecise environments, *Machine Learning*, **42**, 203–231.
- Sonego, P., Kocsor, A., and Pongor, S. (2008). ROC analysis: applications to the classification of biological sequences and 3D structures, *Briefings in Bioinformatics*, **9**, 198–209.
- Stein, R. M. (2005). The relationship between default prediction and lending profits: Integrating ROC analysis and loan pricing, *Journal of Banking & Finance*, **29**, 1213–1236.
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems, *Science*, **240**, 1285–1293.
- Tasche, D. (2008). Validation of internal rating systems and PD estimates. In *The Analytics of Risk Model Validation*, Academic Press, 169–196.
- Vuk, M. and Curk, T. (2006). ROC curve, lift chart and calibration plot, *Metodoloski Zvezki*, **3**, 89.
- Zweig, M. H. and Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, *Clinical Chemistry*, **39**, 561–577.

Received February 25, 2021; Revised March 24, 2021; Accepted March 25, 2021

두 분류 분포를 위한 오즈 곡선

홍종선^{1,a}, 오세현^a, 오태규^a

^a성균관대학교 통계학과

요 약

이진분류모형의 성능을 탐색하는 시각적인 대표적인 방법인 ROC 곡선과 TOC 곡선 그리고 TROC 곡선은 혼동행렬을 구성하는 TP, TN, FP, FN 그리고 이들의 비율인 TPR, TNR, FPR, FNR으로 구현된다. 본 연구에서는 두 종류의 비율비인 오즈를 고려하여 단위면적인 정사각형에서의 구현하는 오즈 곡선을 제안하고, ROC 곡선과의 관계를 보인다. 오즈 곡선에서 판별력을 측정하는 두 종류의 측도를 제안하고, 오즈 곡선들의 형태를 바탕으로 두 종류의 측도를 이용하여 두 분류 분포의 판단 기준을 설정한다. 본 연구에서 제안한 오즈 곡선은 다른 시각적인 방법 등과 같이 유용하게 사용할 수 있으며, 오즈 곡선의 판별력을 측정하는 두 종류의 측도들은 분류 성능을 판단하는 대안적인 방법으로 같이 이용할 수 있다.

주요용어: 민감도, 비율비, 판단 기준, 특이도, 혼동행렬

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