

# COHEN의 합치도의 두 가지 역설을 해결하기 위한 새로운 합치도의 제안

박미희<sup>1)</sup> 박용규<sup>2)</sup>

## 요약

두 평정자가 있는 이차원 분류표에서 우연에 의한 합치 비율을 보정한 Cohen의 합치도( $\kappa$ )는 문제점이 있는 측도로 알려져 있다.  $\kappa$ 는 평정자의 분류 비율(주변확률)에 매우 민감하고, 합치도로서의 조건도 만족하지 못한다. 그러나  $\kappa$ 를 비롯한 기존 합치도들은 주변확률이 0.5에 가까울 때 안정되며 비슷한 값을 갖는다. 본 연구에서는 이차원 분류표의 불균형적 주변분포를 보정함으로써  $\kappa$ 의 역설을 해결하는 새로운 합치도  $H$ 를 제안한다. 또한 예제를 통해 기존의 합치도들과 제안된 합치도를 비교한다.

주요용어: 합치도, 우연 합치 비율, 역설, 균형적 주변분포, 조화평균.

## 1. 서론

동일한 개체들에 대해 두 명의 평정자(rater)가 두 가지 범주로 평가한 경우, 평가가 서로 일치하는 정도는 합치도(measure of agreement)로 나타낸다. 여러 종류의 합치도가 Bennet 등(1954), Scott(1955), Cohen(1960), Holley와 Guilford(1964) 등에 의해 제안되었으나, 그 중 Cohen의 합치도  $\kappa$ 가 주로 사용된다.

$\kappa$ 는 평정자간 평가의 관찰된 합치 비율(observed agreement proportion)에서 우연에 의해 기대되는 합치 비율(chance expected agreement proportion)을 보정한 측도로, 확률 모형에 의존하지 않고 계산이 간편하다는 장점을 가지고 있다. 그러나 우연에 의한 합치에 대한 정의가 불명확하고, 주변분포에 매우 민감한 측도라는 단점이 있다(Feinstein과 Cicchetti, 1990; Gwet, 2001).

Cicchetti와 Feinstein(1990), Byrt 등(1993), Lantz와 Nebenzahl(1996)은  $\kappa$ 를 사용할 때,  $\kappa$ 의 역설이 나타나게 된 원인을 설명해주는 측도도 함께 제시할 것을 제안하였다. 그러나 이때에는 하나의 측도를 사용할 때 갖는 단순성은 없다. 또한 Aickin(1990)은 constant predictive probability(CPP) 모형하에서 우연에 의하기 보다는 원인에 의해 분류한다는 새로운 정의를 이용하여  $\alpha$ 를 제안하였고, Andres와 Marzo(2004)는 다중 선택 검정(multiple choice test)으로부터 우연이 아닌 합치의 비율인  $\Delta$ 를 제안하였다. 이들은  $\kappa$ 처럼 하나의 측도이지만, 계산이 복잡하고 확률 모형에 의존한다는 단점을 가지고 있다.

1) (137-701) 서울시 서초구 반포동 505, 가톨릭대학교 대학원 의학통계학과, 석사과정  
E-mail: sheherazard7@hanmail.net

2) (교신저자)(137-701) 서울시 서초구 반포동 505, 가톨릭대학교 대학원 의학통계학과, 교수  
E-mail: ygpark@catholic.ac.kr

본 연구에서는 두 평정자에 의해 두 명목형 범주로 교차 분류된 이차원 분할표에서의  $\kappa$ 의 문제점과  $\kappa$ 와 표현식이 유사한 기존 합치도들의 문제점을 제시하고, 그 문제점을 해결하는 새로운 합치도를 제안한다. 2절에서는 많은 연구에서 사용되고 있는 Cohen의  $\kappa$ 를 비롯한 기존 합치도와 주변분포가 이들에 미치는 영향을 설명한다. 3절에서는 이러한 문제점을 해결하기 위한 새로운 합치도를 제안한다. 4절에서는 예제를 통해 기존 합치도와 제안된 합치도를 비교하고, 그래프를 통하여 제안된 합치도의 성질을 제시한다. 마지막으로 5절에서는 본 연구의 결과와 의의를 정리한다.

## 2. 여러가지 합치도와 그 문제점

### 2.1. 기호와 조건

평정자 A와 평정자 B가 전체  $N$ 개의 개체를 두 개의 가능한 반응 범주 중 하나로 각각 분류하는 신뢰도 실험을 고려한다. 이 때 두 반응 범주는 'Yes'와 'No'라고 하고 서로 겹치지 않는다고 가정한다. 표 2.1에는 이러한 신뢰도 실험의 결과가  $2 \times 2$  분할표로 제시되어 있다.

표 2.1: 평정자와 반응 범주에 의한 전체  $N$  개체의 분포

평정자 B	평정자 A		Total
	Yes	No	
Yes	$a$	$b$	$g_1$
No	$c$	$d$	$g_2$
Total	$f_1$	$f_2$	$N$

$a$ 는 두 평정자가 모두 범주 Yes에 분류한 개체 수이고,  $d$ 는 범주 No에 분류한 개체 수이다.  $b$ 는 평정자 A와 B가 각각 범주 No와 Yes에 분류한 개체 수이고, 유사하게  $c$ 는 평정자 A는 Yes에, 평정자 B는 No에 분류한 개체 수이다. 평정자 A가 범주 Yes와 No에 분류한 총 개체 수는 각각  $f_1$ 과  $f_2$ 로 나타내기로 한다. 주변합  $g_1$ 과  $g_2$ 도 이와 유사한 방법으로 정의한다.

관찰된 합치 비율  $p_o$ 는 두 평정자가 같은 범주로 분류한 개체들의 비율로, 식 (2.1)과 같다.

$$p_o = \frac{a + d}{N}. \quad (2.1)$$

그러나 두 평정자가 개체들을 우연히 같은 범주로 분류하는 경우가 있으므로, 이를 보정한 합치도를 사용하며 그 식은 (2.2)와 같다.

$$\frac{p_o - p_e}{1 - p_e}. \quad (2.2)$$

여기서  $p_e$ 는 우연에 의해 기대되는 합치 비율을 나타내며, 식 (2.2)는 관찰된 합치 비율과 완벽한 합치 비율( $N/N = 1$ ) 각각에서 우연에 의한 합치 비율을 빼 값의 비로 정의된다.

이 식에서 사용된  $p_e$ 와 합치도 값은 다음과 같은 두 가지 조건을 만족해야 한다. 첫 번째 조건은 완벽한 합치일 때와 완벽한 불합치일 때의 합치도 값에 관한 것으로, 표 2.1에서  $a + d = N$ 이 성립하는 경우, 즉  $p_o = 1$ 인 경우에는  $p_e$ 값에 관계없이 합치도 값은 1이 된다. 반면에  $a + d = 0$ 이 성립하는 경우, 즉  $p_o = 0$ 인 경우 완벽한 불합치이므로 합치도 값은 -1이 되는데, 이 조건은  $p_e = 0.5$ 일 때만 성립한다. 두 번째 조건은  $p_e$ 의 범위에 관한 것으로, 두 평정자가 개체를 독립적이고 랜덤하게 평가할 경우, 우연에 의한 합치 비율의 최대 값은 반응 범주의 수의 역수가 된다(Gwet, 2001). 즉, 반응 범주가 두 개인 경우에는 두 평정자의 랜덤한 평가가 전체 평가의 50%를 넘게 합치할 수 없다. 따라서  $p_e$ 는  $[0, 0.5]$ 의 범위를 갖는다.

### 2.2. Cohen의 $\kappa$ 와 그 문제점

식 (2.2)에서  $p_e$ 의 정의에 따라 여러가지 합치도가 정의된다. Cohen의  $\kappa$ 는  $p_e$ 를 식 (2.3)과 같이 대응하는 주변확률의 곱으로 표현한 합치도이다.

$$p_e(\kappa) = \left(\frac{f_1}{N}\right) \left(\frac{g_1}{N}\right) + \left(\frac{f_2}{N}\right) \left(\frac{g_2}{N}\right). \tag{2.3}$$

$p_e(\kappa)$ 의 첫 번째 항은 두 평정자가 범주 Yes에 한 개체를 독립적으로 분류할 우연을 추정하는 것이고, 두 번째 항은 범주 No로 한 개체의 독립적인 분류 확률을 추정한 것이다.

식 (2.2)에서  $p_o$ 값이 동일 하다면,  $p_e(\kappa)$ 가 작을수록  $\kappa$ 는 커진다. 따라서  $p_o$ 가 1에 가까운 값이라 해도,  $p_e(\kappa)$ 가 크다면  $\kappa$ 는 매우 작은 값을 가지며, 이것이 Feinstein과 Cicchetti(1990)가 말한  $\kappa$ 의 첫 번째 역설이다.

표 2.2:  $\kappa$ 의 첫번째 역설의 예

(a) 대칭적 균형 주변분포				(b) 대칭적 불균형 주변분포			
평정자 B	평정자 A		Total	평정자 B	평정자 A		Total
	Yes	No			Yes	No	
Yes	40	9	49	Yes	80	10	90
No	6	45	51	No	5	5	10
Total	46	54	100	Total	85	15	100
$p_o = 0.85, p_e(\kappa) = 0.50, \kappa = 0.70$				$p_o = 0.85, p_e(\kappa) = 0.78, \kappa = 0.32$			

표 2.2의 (a)와 (b)는 모두  $p_o$ 가 0.85로 비교적 높은 관찰된 합치 비율을 보이고 있으나, (b)의  $\kappa$ 값은 (a)의 절반 정도로 낮게 나타난다. 식 (2.3)에서 보듯이  $p_e(\kappa)$ 는 분할표에서의 주변확률의 곱으로 추정되고, 주변분포가 다르기 때문에 나타난  $p_e(\kappa)$ 값의 차이가  $\kappa$ 값에 영향을 미친 것이다. (a)와 같이 주변확률이 각각 0.5에 가까운 경우를 균형적 주변분포라

하고, (b)와 같이 주변확률이 0.5보다 상당히 크거나 작은 경우를 불균형적 주변분포라 한다. 불균형적 주변분포일 때 균형적 주변분포보다  $p_e(\kappa)$ 가 더 큰 값을 가지게 되고,  $\kappa$ 는 작아진다.

평정자 A와 평정자 B의 각 범주에서의 주변분포가 같을 때, 즉  $f_1 = g_1$ 이고  $f_2 = g_2$ 일 때를 대칭이라고 한다. Feinstein과 Cicchetti(1990)가 말한  $\kappa$ 의 두 번째 역설은 불균형적 주변분포에서 주변분포가 비대칭적일 때가 대칭적일 때보다, 불완전한 대칭일 때가 완전한 대칭일 때보다  $p_e(\kappa)$ 의 값이 작아  $\kappa$ 의 값이 커진다는 것이다.

표 2.3:  $\kappa$ 의 두 번째 역설의 예

(a) 대칭적 불균형 주변분포				(b) 비대칭적 불균형 주변분포			
평정자 A				평정자 A			
평정자 B	Yes	No	Total	평정자 B	Yes	No	Total
Yes	45	15	60	Yes	25	35	60
No	25	15	40	No	5	35	40
Total	70	30	100	Total	30	70	100
$p_o = 0.60, p_e(\kappa) = 0.54, \kappa = 0.13$				$p_o = 0.60, p_e(\kappa) = 0.46, \kappa = 0.26$			
(c) 완전한 대칭적 불균형 주변분포				(d) 불완전한 대칭적 불균형 주변분포			
평정자 A				평정자 A			
평정자 B	Yes	No	Total	평정자 B	Yes	No	Total
Yes	85	5	90	Yes	70	10	80
No	5	5	10	No	0	20	20
Total	90	10	100	Total	70	30	100
$p_o = 0.90, p_e(\kappa) = 0.82, \kappa = 0.44$				$p_o = 0.90, p_e(\kappa) = 0.62, \kappa = 0.74$			

표 2.3에서 (a)와 (b)는 모두  $p_o$ 가 0.60이며, (a)와 같은 대칭적 불균형 주변분포는 (b)와 같은 비대칭적 불균형 주변분포보다  $\kappa$ 의 값이 더 작게 나타난다. 또한 (c)와 (d)는 모두  $p_o = 0.90$ 이고 대칭적 불균형 주변분포이지만, (c)는  $f_1 = g_1 = 90$ 이고  $f_2 = g_2 = 10$ 인 완전한 대칭이고, (d)는 주변분포가 각각  $f_1 = 70, g_1 = 80, f_2 = 30, g_2 = 20$ 인 불완전한 대칭이다. 이처럼 동일한 관찰 합치 비율을 갖고 불균형적 주변분포가 완전 대칭인 경우에는, 불완전 대칭일 때보다  $\kappa$ 값이 더 작게 나타났다.

그리고  $\kappa$ 는 앞에서 언급한 우연에 의한 합치 비율을 보정한 합치도의 두 가지 조건을 만족하지 않는다.  $p_e(\kappa)$ 의 범위는  $[0, 1]$ 이 되며, 주변분포에 영향을 받는  $p_e(\kappa)$ 이기 때문에 완벽한 불합치일 때 합치도는 -1이 되지 않는다. 또한 표 2.1에서  $a = N$ 이거나  $d = N$ 인 경우  $p_o = 1$ 이지만  $p_e(\kappa)$ 도 1이 된다.

### 2.3. 다른 합치도와 그 문제점

#### 2.3.1. Holley와 Guilford의 $G$

Bennet 등(1954)은 우연에 의해 기대되는 합치 비율의 가장 좋은 추정치가 0.5라 보고  $p_e$ 가 0.5인  $S$ 를 제안하였다. 이후 Holley와 Guilford(1964), Janson과 Vegelius(1979), Brennan과 Prediger(1981) 등도  $p_e$ 를 0.5로 둔 여러 측도들을 제안하였다. Bennet 등(1954)에 의해 우연에 의한 합치 비율의 추정은, 반응 범주 수의 역수를 이용한다는 아이디어가 처음 제안되었지만,  $\kappa$ 와 같은 신뢰도의 평가로서의 방법은 Holley와 Guilford의  $G$ 에서 처음 제안되었으므로, 이 측도를 앞으로는  $G$ 라고 한다.

$$p_e(G) = 0.5. \quad (2.4)$$

식 (2.4)를 식 (2.2)에 대입하면  $G = 2p_o - 1$ 이 되고, 두 평정자가 같은 범주로 분류한 개체 수가 약 50%라면  $G$ 는 0 또는 0에 근사한 값을, 같은 범주로 분류한 개체 수가 50%보다 작다면 항상 음의 값을 갖게 된다. 또한 완벽한 불합치일 때에는  $-1$ , 완벽한 합치일 때에는 1의 합치도 값을 갖지만,  $p_o$ 의 크기에 상관없이 항상  $p_e(G)$ 는 0.5로 고정되어 있는 합치도 값이므로 앞에서 언급한 조건을 만족한다고는 볼 수 없다.

#### 2.3.2. Scott의 $\pi$

Scott(1955)이 제안한  $\pi$ 의  $p_e(\pi)$ 는 두 평정자가 주어진 범주에 같은 확률로 개체를 분류할 것이라는 가정 하에서 추정되며,  $p_e(\pi)$ 의 각 항은 각 평정자의 분류 비율의 곱으로 얻어진다.

$$p_e(\pi) = P_1^2 + P_2^2. \quad (2.5)$$

여기서  $P_1$ 과  $P_2$ 는 각각  $P_1 = [(f_1/N + g_1/N)/2]$ ,  $P_2 = [(f_2/N + g_2/N)/2]$ 이다.

그러나  $p_e(\pi)$ 의 최소값이 0.5이고 최대값이 1이 되는 특성을 가지므로 앞의 조건에 어긋나는 측도이다. 그리고  $\kappa$ 에서와 같이 두 평정자가 모든 개체를 하나의 같은 범주로 분류하는 경우  $p_e(\pi)$ 값도 1이 된다. 따라서 완벽한 합치일 때 합치도가 1이 되지 못한다.

#### 2.3.3. Gwet의 $AC_1$

Gwet(2001)이 제안한  $AC_1$ 은 평정자내 변동(intraobserver variation)을 나타내는 불확실성 계수(coefficient of uncertainty)를 이용하여  $p_e(AC_1)$ 을 구한 것으로, 그 식은 (2.6)과 같다.

$$p_e(AC_1) = 2P_1(1 - P_1). \quad (2.6)$$

여기서  $P_1$ 과  $P_2$ 는 식 (2.5)에서와 동일하다.

$p_e(AC_1)$ 은 범위가  $[0, 0.5]$ 이고, (Yes, Yes) 또는 (No, No)의 칸에 모든 평가가 집중되어 있을 때  $p_e(AC_1) = 0$ 이 되어, 앞에서 제시한 조건을 만족하는 합치도이다.

그림 2.1은 주변분포가  $\kappa$ ,  $\pi$ ,  $AC_1$ 의 기대 합치 비율에 미치는 영향을 나타낸 것이다.

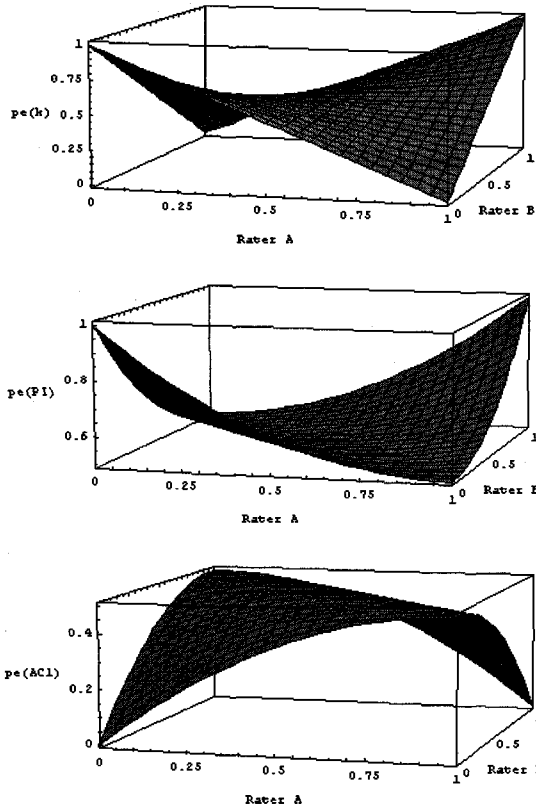


그림 2.1: 범주 Yes의 주변확률과  $p_e$ 의 변화

### 3. 제안 합치도

2절에서 살펴본 바와 같이 대부분의 합치도는 주변분포가 균형적일 때 가장 안정되고 좋은 추정량을 갖는다. 그러나 각 범주의 개체 수를 동일하게 구성하기 어려울 뿐 아니라, 동일하게 구성한다 해도 평정자들이 눈가림(blind) 상태에서 주관적인 평가를 하므로, 분류된 대상자의 수가 같은 경우가 드물다. 이 절에서는 개체의 불균형한 모집과 평정자의 평가 비율의 같지 않음으로 인한  $\kappa$ 의 역설을 해결할 수 있는 새로운 합치도를 제안한다.

#### 3.1. 제안 합치도

비율의 평균을 구하는 문제에서는 비율의 어떤 요인이 불변인지에 따라 사용하는 평균이 달라진다. 만약  $A/B$ 라는 비율이 불변하는 값이면 산술평균을 사용하고,  $B$ 에 의하여 영향을 받게 되어 변하는 값이면 조화평균을 사용한다(Ferger, 1931).

표 2.1에서 각 평정자가 각 범주로 분류한 개체 수를 전체 개체 수  $N$ 으로 나눈 값은, 이 실험에서 얻어낼 수 있는 각 평정자의 평가 비율이다. 그리고 주변확률의 분포를 균형적 주변분포로 보정하기 위해 이 비율의 평균을 사용한다.

균형적 주변분포는 표 2.2의 (a)와 같이 평정자의 각 범주에 같은 수의 개체가 분류되는 것이다. 이러한 균형적 주변분포로 보정하기 위해서는 평정자의 평가 비율의 비대칭성과 불균형성을 함께 보정해야 한다.

**비대칭성의 보정:** 평정자 A와 평정자 B의 Yes로의 평가 비율  $f_1/N$ 와  $g_1/N$ 은 두 평정자가 같은 전체 개체 수에서 구한 비율로, 실험에서 불변하는 값이므로 두 평정자의 평균 평가 비율은 산술평균으로 구할 수 있다. 이 때의 산술평균  $P_1$ 과 반응 범주 No로의 산술평균  $P_2$ 는 식 (3.1)과 (3.2)로 표현된다.

$$\left(\frac{f_1}{N} + \frac{g_1}{N}\right)/2 = P_1, \tag{3.1}$$

$$\left(\frac{f_2}{N} + \frac{g_2}{N}\right)/2 = P_2. \tag{3.2}$$

**불균형성의 보정:** 균형성을 유지하는 것은 각 범주에 같은 수의 개체가 분류되었다는 조건하에서 구해져야 하므로 두 비율  $P_1$ 과  $P_2$ 의 평균은, Ferger가 밝힌 바와 같이, 식 (3.3)과 같은 조화평균을 사용해야 한다.

$$\frac{2}{\frac{1}{P_1} + \frac{1}{P_2}} = 2P_1P_2. \tag{3.3}$$

따라서 주변분포의 비대칭성과 불균형성을 보정한 주변확률 식 (3.3)의 곱의 합으로 표현한, 제안 합치도의  $p_e(H)$ 는 식 (3.4)와 같다.

$$p_e(H) = 2P_1P_2 \times 2P_1P_2 + 2P_1P_2 \times 2P_1P_2 = 2(2P_1P_2)^2. \tag{3.4}$$

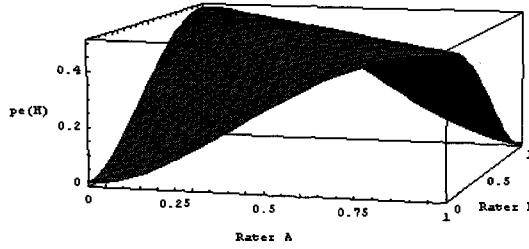
식 (2.2)의  $p_e$ 에  $p_e(H)$ 를 대입하면 식 (3.5)와 같은 제안 합치도  $H$ 를 얻는다.

$$H = \frac{p_o - p_e(H)}{1 - p_e(H)}. \tag{3.5}$$

새로운 합치도는 주변분포의 영향에 민감한  $\kappa$ 와는 달리, 어떠한 주변분포를 가지더라도 완벽한 합치일 때에는 1, 완벽한 불합치일 때에는 -1의 합치도 값을 갖는다. 그리고 그림 3.1에서 보는 바와 같이  $p_e(H)$ 의 범위도  $[0, 0.5]$ 로서 0.5를 넘지 않는다.

### 3.2. 제안 합치도 $H$ 의 최대우도추정량과 분산

$i$ 번째 행과  $j$ 번째 열의 개체 수가  $x_{ij}$ 이고 전체 개체 수  $N$ 이 고정된 다항 표집 모형에서, 식 (3.5)의  $p_o$ 와  $p_e(H)$ 는 각각  $p_o = p_{11} + p_{22}$ ,  $p_e(H) = \{(p_{1+} + p_{+1})(p_{2+} + p_{+2})\}^2/2$  이고, 여기서  $p_{ij}$ 는 결합확률,  $p_{i+}$ 와  $p_{+j}$ 는 주변확률이다. 식 (3.5)에서  $p_{ij}$ 의 최대우도추정량  $\hat{p}_{ij} = x_{ij}/N$ 을 대입하여 얻은 제안 합치도의 최대우도추정량은 식 (3.6)과 같다.

그림 3.1: 범주 Yes의 주변확률과  $p_e(H)$ 의 변화

$$\hat{H} = \frac{2N^3(x_{11} + x_{22}) - \{(x_{1+} + x_{+1})(x_{2+} + x_{+2})\}^2}{2N^4 - \{(x_{1+} + x_{+1})(x_{2+} + x_{+2})\}^2}. \quad (3.6)$$

델타 방법에 의해 구한 제안 합치도의 최대우도추정량의 점근적 대표본 분산은 식 (3.7)과 같으며, 그 유도 과정은 부록에 있다(Bishop 등, 1995).

$$\sigma_{\infty}^2 = \frac{1}{(1-\theta_2)^2 N} \left[ \theta_1(1-\theta_1) + \frac{4\sqrt{2\theta_2}(1-\theta_1)^2\theta_5}{1-\theta_2} + \frac{8(1-\theta_1)^2\theta_2}{(1-\theta_2)^2} \{-\theta_3^2 + (1-\theta_1)(\theta_1 - 2\theta_3) + \theta_4\} \right]. \quad (3.7)$$

여기서  $\theta_1$ - $\theta_5$ 는 식 (3.8)-(3.12)와 같다.

$$\theta_1 = p_{11} + p_{22}, \quad (3.8)$$

$$\theta_2 = \frac{1}{2} \{(p_{1+} + p_{+1})(p_{2+} + p_{+2})\}^2, \quad (3.9)$$

$$\theta_3 = p_{11}(p_{2+} + p_{+2}) + p_{22}(p_{1+} + p_{+1}), \quad (3.10)$$

$$\theta_4 = p_{11}(p_{2+} + p_{+2})^2 + p_{22}(p_{1+} + p_{+1})^2, \quad (3.11)$$

$$\theta_5 = (p_{11} - p_{22})^2. \quad (3.12)$$

### 3.3. 다른 합치도의 $p_e$ 비교

균형적 주변분포로 보정하기 위해 각 합치도에서 사용한  $F_1$ 과  $F_2$ 의 함수와  $p_e$ 를 식 (2.3)의  $p_e(\kappa)$  형태로 분해한 결과는 표 3.1과 같다. 이때 합치도들의  $p_e$ 를 분해한 원소들은, 두 평정자가 같은 범주로 분류한 주변확률의 비대칭성 보정 확률인,  $F_1$ 과  $F_2$ 의 산술, 기하 평균 및 제곱평균으로 표현된다.

같은  $F_1$ 과  $F_2$ 의 함수로서의 산술, 기하, 조화평균 및 제곱평균의 대소관계는 “조화평균  $\leq$  기하평균  $\leq$  산술평균  $\leq$  제곱평균”을 만족하므로, 각 합치도들의  $p_e$ 는 식 (3.13)과 같은 대소관계를 만족한다.



표 3.1: 보정확률과  $p_e$ 의 분해

보정방법	보정확률	$p_e$ 의 분해
산술평균	$\frac{1}{2}$	$p_e(G) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2}$
기하평균	$\sqrt{P_1 P_2}$	$p_e(AC_1) = \sqrt{P_1 P_2} \times \sqrt{P_1 P_2} + \sqrt{P_1 P_2} \times \sqrt{P_1 P_2}$
제곱평균	$\sqrt{\frac{P_1^2 + P_2^2}{2}}$	$p_e(\pi) = \sqrt{\frac{P_1^2 + P_2^2}{2}} \times \sqrt{\frac{P_1^2 + P_2^2}{2}} + \sqrt{\frac{P_1^2 + P_2^2}{2}} \times \sqrt{\frac{P_1^2 + P_2^2}{2}}$

$$p_e(H) \leq p_e(AC_1) \leq p_e(G) \leq p_e(\pi). \tag{3.13}$$

같은  $P_1$ 과  $P_2$ 하에서  $H$ 의  $p_e$ 가 가장 작고,  $AC_1$ ,  $G$  그리고  $\pi$ 의 순서로  $p_e$ 가 크게 추정된다. 그리고  $P_1 = P_2$ 일 때, 즉 두 평정자의 각 범주로의 분류확률의 합이 1이 될 때,  $P_1$ 과  $P_2$ 의 평균들은 등호관계가 성립하며, 식 (3.13)의 모든  $p_e$ 도 0.5가 된다.

따라서 같은  $p_o$ 를 가질 때, 각 합치도의 대소관계는 식 (3.14)와 같다.

$$\pi \leq G \leq AC_1 \leq H. \tag{3.14}$$

## 4. 예제와 제안 합치도의 성질

### 4.1. Feinstein 과 Cicchetti(1990)의 예제

$\kappa$ 의 역설에 관한 Feinstein 과 Cicchetti의 연구에서 사용된 2.2절의 표 2.2와 표 2.3의 자료에서 구한 각 합치도들의 값과  $p_e$ 가 표 4.1에 제시되어 있다.

표 2.2의 (a)에서와 같이 주변확률이 0.5로 균형적일 때에는, 모든 합치도들의 우연에 의한 합치 비율은 0.5가 되고, 합치도 값은 0.7로 같아진다. 그러나 (b)처럼 주변분포가 불균형적일 때에는  $\pi$  또는  $\kappa$ 에 비해  $AC_1$ 과  $H$ 의  $p_e$ 가 더 작게 추정되고, 합치도 값도 관찰 합치 비율에 더 가깝게 된다. 표 2.2의 (a)와 (b)는 동일한  $p_o$ 이지만  $\pi$ 와  $\kappa$ 값은 약 두 배의 차이를 보이고 있다. 따라서  $AC_1$ 과  $H$ 는 불균형 주변분포로 인해 발생하는  $\kappa$ 의 첫 번째 역설을 해결하는 합치도라 할 수 있다.

대칭적 불균형 주변분포에서의 합치도가 비대칭적 불균형 주변분포에서보다 클 것으로 기대하지만,  $\pi$ 와  $\kappa$ 는 반대의 결과를 보이고,  $AC_1$ 과  $H$ 는 대칭적 불균형 주변분포에서 더 높다는 사실을 표 2.3의 (a)와 (b)에 대한 결과를 통해 알 수 있다. (c)와 (d)의 예제에서  $p_o$ 는 0.90으로 매우 높음에도 불구하고  $\pi$ 와  $\kappa$ 의 값은 작게 나타났고, 완전한 대칭적 불균형 주변분포에서의 값이 불완전한 대칭적 불균형 주변분포에서의 값보다 훨씬 작은 값이 나왔다. 그러나  $AC_1$ 과  $H$ 는  $p_o$ 와 거의 같은 값을 가지며, 완전한 대칭일 때가 불완전한 대칭일 때보다 더 큰 값을 갖는다.

표 4.1:  $\kappa$ 의 역설 예제에서의 합치도 값

표	$p_o$	$p_e(G)$	$p_e(\pi)$	$p_e(\kappa)$	$p_e(AC_1)$	$p_e(H)$	$G$	$\pi$	$\kappa$	$AC_1$	$H$
2.2 (a)	0.85	0.5	0.50	0.50	0.50	0.50	0.70	0.70	0.70	0.70	0.70
(b)			0.78	0.78	0.22	0.10		0.31	0.32	0.81	0.83
2.3 (a)	0.60		0.55	0.54	0.46	0.41	0.20	0.12	0.13	0.27	0.32
(b)			0.51	0.46	0.50	0.49		0.19	0.26	0.21	0.22
(c)	0.90		0.82	0.82	0.18	0.06	0.80	0.44	0.44	0.88	0.89
(d)			0.63	0.62	0.38	0.28		0.73	0.74	0.84	0.86

$G$ 는 항상 같은 우연에 의한 합치 비율을 사용하기 때문에 주변분포를 반영하는 합치도 값이라 할 수 없다.

#### 4.2. $p_e$ 의 그래프적 비교

표 2.1에서 주변확률을 모두  $f_1$ 과  $f_2$ 에 관한 식으로 나타낼 때,  $g_1$ 과  $g_2$ 는 각각  $g_1 = f_2 + (a - d)$ ,  $g_2 = f_1 - (a - d)$ 이다. 이들을 각 합치도의  $p_e$ 에 대입하면 식 (4.1)-(4.4)와 같다.

$$p_e(\kappa) = \frac{2f_1f_2 + (f_1 - f_2)(a - d)}{N^2}, \quad (4.1)$$

$$p_e(\pi) = \frac{1}{2} \left[ 1 + \left( \frac{a - d}{N} \right)^2 \right], \quad (4.2)$$

$$p_e(AC_1) = \frac{1}{2} \left[ 1 - \left( \frac{a - d}{N} \right)^2 \right], \quad (4.3)$$

$$p_e(H) = \frac{1}{2} \left[ 1 - \left( \frac{a - d}{N} \right)^2 \right]^2. \quad (4.4)$$

$\kappa$ 는 균형적 주변분포인 경우  $f_1 - f_2 = 0$ 이므로  $(a - d)$ 의 방향에 영향을 받지 않지만, 불균형적 주변분포를 가지는 경우, 즉  $f_1 - f_2 \neq 0$ 인 경우에는  $(f_1 - f_2)$ 과  $(a - d)$ 의 방향에 영향을 받는다. 같은 방향일 경우에는  $(f_1 - f_2)(a - d)$ 가 양수가 되어  $p_e(\kappa)$ 가 커지므로  $\kappa$ 가 작아지지만, 반대 방향일 경우에는  $(f_1 - f_2)(a - d)$ 가 음수가 되어  $p_e(\kappa)$ 는 작아져  $\kappa$ 는 커진다. 따라서  $\kappa$ 의 두 번째 역설에서도 지적했던 바와 같이 대칭성의 여부가  $p_e(\kappa)$ 의 크기를 좌우한다.

다른 합치도들은 주변확률의 평균을 사용하기 때문에  $\kappa$ 처럼  $(f_1 - f_2)$ 의 크기와 방향에 영향을 받지 않는다. 그리고  $(a - d)$ 의 크기에 영향을 받을 뿐 방향에는 영향을 받지 않는다.  $\pi$ 는  $(a - d)$ 의 크기가 클수록  $p_e(\pi)$ 는 커지고,  $AC_1$ 과  $H$ 는  $(a - d)$ 의 크기가 클수록  $p_e$ 가 작

아진다. 그리고  $\pi$ 와  $AC_1$ 은 0.5를 기준으로  $p_e$ 가 서로 대칭인 그래프를 갖는다.  $p_e(\pi)$ 는 0.5 이상의 값을 가지며,  $p_e(AC_1)$ 는 0.5 이하의 값을 갖는다.  $H$ 는  $AC_1$ 과 비교할 때  $p_e$ 의 그래프가 비슷하지만 더 작은 값을 갖는다.  $AC_1$ 은 블록한 모양을 갖지만,  $H$ 는  $P_1$ 이 0이나 1에 가까운 경우  $AC_1$ 에 비해 더 낮은 꼬리부분을 형성한다. 이들 합치도들의  $P_1$ 에 대한  $p_e$ 의 변화 그래프는 그림 4.1에 나타나 있다.

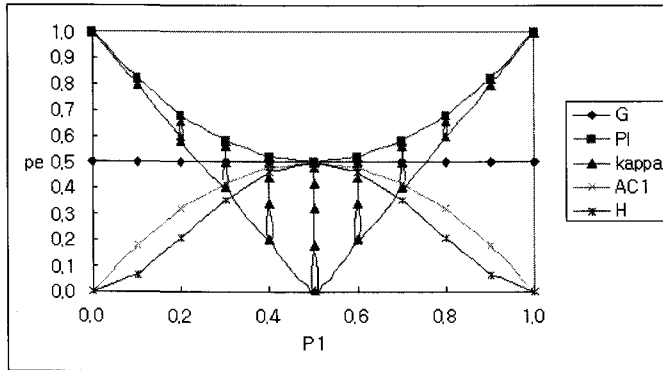


그림 4.1:  $P_1$ 에 대한  $p_e$ 의 변화

### 4.3. 합치도의 그래프적 비교

그림 4.2는 주변확률의 차이에 대한 각 합치도의 민감성을 나타낸다.  $p_o$ 를 0.1부터 0.9까지 0.1씩 증가시킨 각 경우에 대해, 칸 (Yes, Yes)의 비율을  $p_o$ 가 0.6부터 0.9까지는 0.1씩,  $p_o$ 가 0.1부터 0.5까지는 0.05씩 증가시켰을 때, 계산된  $P_1$ 과 각 합치도의 관계를 나타내었다. 주어진  $p_o$ 에서  $G$ ,  $\pi$ ,  $AC_1$ ,  $H$ 는  $P_1$ 에 대하여 일대일 대응 관계가 성립하지만,  $\kappa$ 는 하나의  $P_1$ 을 갖기 위해서는 여러가지 경우의 주변분포 조합이 가능하기 때문에 일대일 대응이 아니다. 즉,  $a$ 와  $d$ 가 고정되어 있을 때,  $b$ 와  $c$ 의 분포에 따라 주변분포는 변하므로, 식 (4.1)에서 보았듯이 주변분포의 영향을 받는  $p_e(\kappa)$ 와 합치도  $\kappa$ 는 같은  $P_1$ 에서 변동이 나타난다.

그림 4.2를 보면  $p_o$ 가 큰 값일수록  $P_1$ 이 매우 작거나 클 때, 주변확률의 균형성 여부에 관계없이  $\kappa$ 나  $\pi$ 는 작은 값을 갖는 경향이 있다. 그리고 주변분포가 0.5에 가까울 때 모든 측도는 서로 근사한 수준의 합치도 값을 갖는다.  $p_o$ 가 높을 때 합치도 값이 크게 되기를 기대하는데,  $AC_1$ 과  $H$ 만이 관찰 합치 비율에 근접한 합치도 값을 제시하고 있고, 이때  $H$ 가  $AC_1$ 보다는  $p_o$ 에 더 근접한 값을 갖는다.  $G$ 는 모든  $P_1$  구간에서 같은 값을 갖는다. 그러나 분할표에서 평정자들의 평가가 한 쪽 칸에 몰리게 되어  $P_1$ 이 매우 높거나 낮은 경우는, 우연에 의한 합치 비율이 감소한다고 볼 수 있어 높은 합치도 값을 갖는다. 이러한 이유로 볼 때 적절한 값을 제시하고 있는 것은  $AC_1$ 과  $H$ 이다.

$p_o$ 가 낮아질수록  $\kappa$ 의 변동은 심하다. 같은  $p_o$ 일 때 같은  $P_1$ 에서 균형적 주변분포와 불균형적 주변분포 간의  $\kappa$ 의 차이는  $p_o$ 가 낮아질수록 커진다. 그러나  $AC_1$ 과  $H$ 는 같은  $p_o$ 하

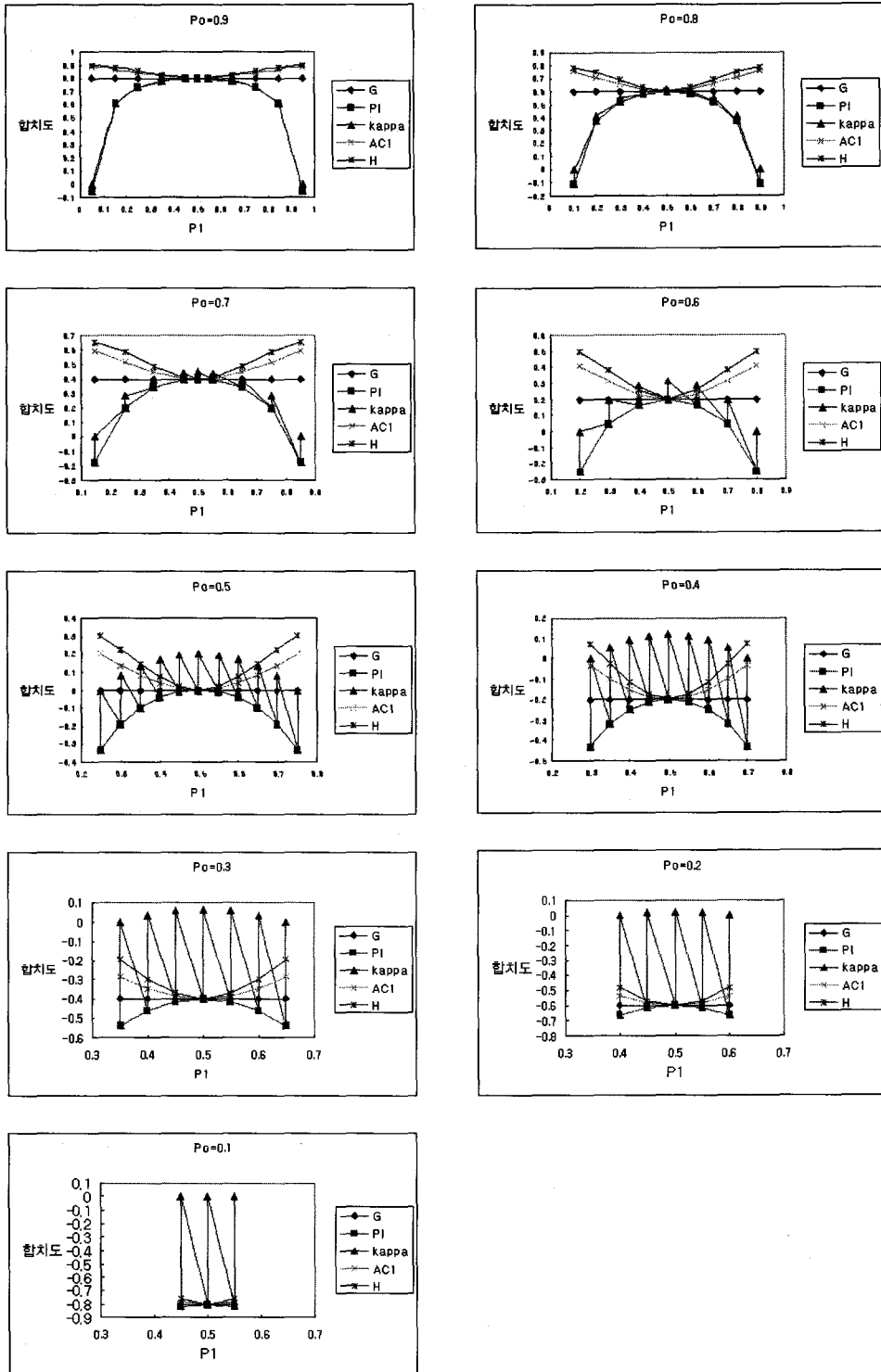


그림 4.2:  $P_1$ 에 대한 합치도의 변화

에서 평정자의 불합치한 평가가 어느 칸에 있는지 동일한  $P_1$ 에서는 일관된 합치도 값을 갖는다.

### 5. 결론

Cohen의  $\kappa$ 는 두 평정자에 의한 이차원 분할표에서 평정자의 평가의 합치 정도를 측정하는 척도로 널리 사용되고 있지만, 주변분포에 민감하다는 단점을 가지고 있다. 본 연구에서는 이러한  $\kappa$ 의 문제점을 해결하고, 합치도로서의 조건을 만족하는 새로운 합치도  $H$ 를 제안하였다. 주변분포의 균형성과 대칭성으로 인한 문제점을 해결하기 위해 비율의 산술평균과 조화평균의 개념을 이용하여 새로운 우연에 의한 합치 비율  $p_e(H)$ 를 도출하였다.  $p_e(H)$ 는  $[0, 0.5]$ 의 범위를 가지며,  $p_e(H)$ 를 적용한 합치도  $H$ 는 항상 완벽한 합치일 때에는 1, 완벽한 불합치일 때는 -1의 값이 된다는 것을 확인하였다. 또한  $\kappa$ 와 표현식이 유사한 기존 합치도들을 함께 비교하였고,  $H$ 는 이들 합치도의 문제점도 해결하는 유용한 척도라고 할 수 있다. 추후 연구 과제로 이차원 분할표에서 나타난  $\kappa$  역설을 해결하기 위해 제안된 조화평균을 이용한 합치도  $H$ 를  $K \times K$  분할표로 확장해 볼 수 있을 것이다.

### 부록

$2 \times 2$  분할표에서 칸 확률의 벡터는  $p = [p_{11}, p_{12}, p_{21}, p_{22}]$ 이며, 식 (3.6)의 점근적 대표본 분산은 델타 방법에 의해 구할 수 있다.

$H(p)$ ,  $H(\theta_1)$ ,  $H(\theta_2)$ 가 식 (3.5)를 각각  $p$ ,  $\theta_1 (= p_o)$ ,  $\theta_2 (= p_e(H))$ 에 관한 함수로 표현한 것이라 할 때, 연쇄법칙(Chain rule)에 의해 다음 식이 성립한다.

$$\begin{aligned} \frac{d}{dp} H(p) &= \frac{dH(\theta_1)}{d\theta_1} \frac{d\theta_1}{dp} + \frac{dH(\theta_2)}{d\theta_2} \frac{d\theta_2}{dp} \\ &= \frac{1}{1-\theta_1} \frac{d\theta_1}{dp} + \left\{ \frac{-(1-\theta_1)}{(1-\theta_2)^2} \right\} \frac{d\theta_2}{dp}. \end{aligned} \tag{A.1}$$

위 식 (A.1)을 델타 공식에 대입하면 아래와 같이 식 (A.2)를 얻는다.

$$\begin{aligned} \text{Var}(H(\hat{p})) &= \left\{ \frac{d}{dp} H(p) \right\}^2 \text{Cov}(\hat{p}) \\ &= \left[ \frac{1}{1-\theta_1} \frac{d\theta_1}{dp} + \left\{ \frac{-(1-\theta_1)}{(1-\theta_2)^2} \right\} \frac{d\theta_2}{dp} \right]^2 \text{Cov}(\hat{p}) \\ &= \frac{1}{1-\theta_1} \frac{d\theta_1}{dp} \text{Cov}(\hat{p}) \frac{d\theta_1}{dp} - \frac{2(1-\theta_1)}{(1-\theta_2)^3} \frac{d\theta_1}{dp} \text{Cov}(\hat{p}) \frac{d\theta_2}{dp} \\ &\quad + \frac{(1-\theta_1)^2}{(1-\theta_2)^4} \frac{d\theta_2}{dp} \text{Cov}(\hat{p}) \frac{d\theta_2}{dp}. \end{aligned} \tag{A.2}$$

여기서

$$\frac{d\theta_1}{dp} = [1, 0, 0, 1],$$

$$\frac{d\theta_2}{dp} = 2(p_{1+} + p_{+1})(p_{2+} + p_{+2})[(p_{2+} + p_{+2}), 1, 1, (p_{1+} + p_{+1})],$$

$$\text{Cov}(\hat{p}) = \begin{bmatrix} \frac{p_{11}(1-p_{11})}{N} & -\frac{p_{11}p_{12}}{N} & -\frac{p_{11}p_{21}}{N} & -\frac{p_{11}p_{22}}{N} \\ -\frac{p_{12}p_{11}}{N} & \frac{p_{12}(1-p_{12})}{N} & -\frac{p_{12}p_{21}}{N} & -\frac{p_{12}p_{22}}{N} \\ -\frac{p_{21}p_{11}}{N} & -\frac{p_{21}p_{12}}{N} & \frac{p_{21}(1-p_{21})}{N} & -\frac{p_{21}p_{22}}{N} \\ -\frac{p_{22}p_{11}}{N} & -\frac{p_{22}p_{12}}{N} & -\frac{p_{22}p_{21}}{N} & \frac{p_{22}(1-p_{22})}{N} \end{bmatrix}$$

이다.

식 (A.2)의 각 항에서

$$\frac{d\theta_1}{dp} \text{Cov}(\hat{p}) \frac{d\theta_1}{dp} = \frac{(p_{11} + p_{22})\{1 - (p_{11} + p_{22})\}}{N} = \frac{\theta_1(1 - \theta_1)}{N}, \quad (\text{A.3})$$

$$\begin{aligned} \frac{d\theta_1}{dp} \text{Cov}(\hat{p}) \frac{d\theta_2}{dp} &= -\frac{2}{N}(p_{1+} + p_{+1})(p_{2+} + p_{+2})(p_{11} - p_{22})^2(p_{12} + p_{21}) \\ &= -\frac{2(1 - \theta_1)\sqrt{2\theta_2}\theta_5}{N}, \end{aligned} \quad (\text{A.4})$$

$$\begin{aligned} \frac{d\theta_2}{dp} \text{Cov}(\hat{p}) \frac{d\theta_2}{dp} &= \frac{2^2}{N}(p_{1+} + p_{+1})^2(p_{2+} + p_{+2})^2 \times \\ &\quad \left[ (-1)\{p_{11}(p_{2+} + p_{+2}) + p_{22}(p_{1+} + p_{+1})\}^2 \right. \\ &\quad \left. + (p_{12} + p_{21})\{(p_{11} + p_{22}) - 2\{p_{11}(p_{2+} + p_{+2}) + p_{22}(p_{1+} + p_{+1})\}\} \right. \\ &\quad \left. + p_{11}(p_{2+} + p_{+2})^2 + p_{22}(p_{1+} + p_{+1})^2 \right] \\ &= \frac{8\theta_2\{-\theta_3^2 + (1 - \theta_1)(\theta_1 - 2\theta_3) + \theta_4\}}{N} \end{aligned} \quad (\text{A.5})$$

이고, 여기서

$$\begin{aligned} \theta_1 &= p_{11} + p_{22}, \\ \theta_2 &= \frac{1}{2}\{(p_{1+} + p_{+1})(p_{2+} + p_{+2})\}^2, \\ \theta_3 &= p_{11}(p_{2+} + p_{+2}) + p_{22}(p_{1+} + p_{+1}), \\ \theta_4 &= p_{11}(p_{2+} + p_{+2})^2 + p_{22}(p_{1+} + p_{+1})^2, \\ \theta_5 &= (p_{11} - p_{22})^2 \end{aligned}$$

이다.

식 (A.2)에 식 (A.3)-(A.5)를 대입하면, 아래와 같은 점근적 대표본 분산을 얻을 수 있

다.

$$\sigma_{\infty}^2 = \frac{1}{(1-\theta_2)^2 N} \left[ \theta_1(1-\theta_1) + \frac{4\sqrt{2\theta_2}(1-\theta_1)^2\theta_5}{1-\theta_2} + \frac{8(1-\theta_1)^2\theta_2}{(1-\theta_2)^2} \{ -\theta_3^2 + (1-\theta_1)(\theta_1 - 2\theta_3) + \theta_4 \} \right].$$

### 참고문헌

- Aickin, M. (1990). Maximum likelihood estimation of agreement in the constant predictive probability model, and its relation to Cohen's kappa, *Biometrics*, **46**, 293-302.
- Andres, A. M. and Marzo, P. F. (2004). Delta: A new measure of agreement between two raters, *The British Journal of Mathematical and Statistical Psychology*, **57**, 1-19.
- Bennet, E. M., Alpert, R. and Goldstein, A. C. (1954). Communications through limited response questioning, *Public Opinion Quarterly*, **18**, 303-308.
- Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1995). *Discrete Multivariate Analysis*, MIT Press, Cambridge, Mass.
- Brennan, R. L. and Prediger, D. (1981). Coefficient kappa: Some uses, misuses, and alternatives, *Educational and Psychological Measurement*, **41**, 687-699.
- Byrt, T., Bishop, J. and Carlin, J. B. (1993). Bias, prevalence and kappa, *Journal of Clinical Epidemiology*, **46**, 423-429.
- Cicchetti, D. V. and Feinstein, A. R. (1990). High agreement but low kappa: 2. Resolving the paradoxes, *Journal of Clinical Epidemiology*, **43**, 551-558.
- Cohen, J. (1960). A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, **20**, 37-46.
- Feinstein, A. R. and Cicchetti, D. V. (1990). High agreement but low kappa: 1. The problems of two paradoxes, *Journal of Clinical Epidemiology*, **43**, 543-549.
- Ferger, W. F. (1931). The nature and use of the harmonic mean, *Journal of the American Statistical Association*, **26**, 36-40.
- Gwet, K. (2001). *Handbook of inter-rater reliability*, STATAXIS Publishing company, Gaithersburg.
- Holley, J. W. and Guilford, J. P. (1964). A note on the G index of agreement, *Educational and Psychological Measurement*, **24**, 749-753.
- Janson, S. and Vegelius, J. (1979). On generalizations of the G index and the PHI coefficient to nominal scales, *Multivariate Behavioral Research*, **14**, 255-269.
- Lantz, C. A. and Nebenzahl, E. (1996). Behavior and interpretation of the  $\kappa$  statistic: Resolution of the two paradoxes, *Journal of Clinical Epidemiology*, **49**, 431-434.
- Scott, W. A. (1955). Reliability of content analysis: The case of nominal scale coding, *Public Opinion Quarterly*, **19**, 321-325.

[ 2006년 8월 접수, 2006년 10월 채택 ]

## A New Measure of Agreement to Resolve the Two Paradoxes of Cohen's Kappa

Mi-Hee Park<sup>1)</sup> Yong-Gyu Park<sup>2)</sup>

### ABSTRACT

In a  $2 \times 2$  table showing binary agreement between two raters, it is known that Cohen's  $\kappa$ , a chance-corrected measure of agreement, has two paradoxes.  $\kappa$  is substantially sensitive to raters' classification probabilities (marginal probabilities) and does not satisfy conditions as a chance-corrected measure of agreement. However,  $\kappa$  and other established measures have a reasonable and similar value when each marginal distribution is close to 0.5. The objectives of this paper are to present a new measure of agreement,  $H$ , which resolves paradoxes of  $\kappa$  by adjusting unbalanced marginal distributions and to compare the proposed measure with established measures through some examples.

*Keywords:* Measure of agreement, chance expected agreement, paradox, balanced marginal distribution, harmonic mean.

---

1) Graduate Student, Department of Biostatistics, The Catholic University of Korea, Seoul 137-701, Korea

E-mail: sheherazard7@hanmail.net

2) (Corresponding author) Professor, Department of Biostatistics, The Catholic University of Korea, Seoul 137-701, Korea

E-mail: ygpark@catholic.ac.kr