ASYMPTOTIC DISTRIBUTION OF AGREEMENT SCORE (KAPPA) WITH FIXED MARGINAL TOTALS

Y. S. Lin, Booz, Allen & Hamilton, Inc.

## I. INTRODUCTION

One particular form of contingency table (ordered I x I table) gives rise to a special problem of statistical interest - measurement of agreement. Suppose two raters independently categorize items or responses among the same set of nominal categories, and we wish to develop a measure of agreement for these raters. This problem can be viewed as one of measuring the reliability between two raters. Goodman and Kruskal (1954) suggested that for the situation when each of the r raters independently assigns N responses (one to each of the N objects) among I categories a measure of agreement, adjusted for chance, among r raters is needed.

Many coefficients of relative agreement measure have been proposed within the last two decades. The more widely used agreement coefficient has been the one called Kappa that was suggested by Cohen (1960) and others. Kappa coefficient for a two rater is defined as:

$$\kappa = (\theta_1 - \theta_2)/(1-\theta_2) \qquad (1)$$

where $\theta_1 = \sum_i P_{ii}$ and $\theta_2 = \sum_i P_{i.}\, P_{.i}$ ,

$P_{ij}$ = true proportion that an object is assigned by rater 1 to category i and in category j by rater 2 .

Let $X_{ij}$ be the number of objects assigned to (i,j) cell in the ordered I x I contingency table and $N = \sum_i \sum_j X_{ij}$. The maximum likelihood estimator for $\kappa$ under the multinomial sampling situation is:

$$\hat{\kappa} = (\hat{\theta}_1 - \hat{\theta}_2)/(1-\hat{\theta}_2) \qquad (2)$$

where $\hat{\theta}_1 = \sum_i X_{ii}/N$ and $\hat{\theta}_2 = \sum_i X_{i.}\, X_{.i}/N^2$.

## II. ASYMPTOTIC DISTRIBUTION OF $\hat{\kappa}$ WITH FIXED MARGINAL TOTALS

The asymptotic variance of $\hat{\kappa}$ as given by some authors (e.g., Cohen {1960, 1968}, Fleiss {1971}, Marx and Light {1973}) is of the form:

$$\mathrm{Var}\,(\hat{\kappa}) = \begin{cases} \dfrac{\theta_1(1-\theta_1)}{N\,(1-\theta_2)^2} & \text{for non-null case} \\[2ex] \dfrac{\theta_2}{N\,(1-\theta_2)} & \text{for null case.} \end{cases} \qquad (3)$$

It was later shown (by Fleiss, Cohen & Everitt {1969} and Bishop, Fienberg & Holland {1975}, etc.) that the expressions in (3) are not correct for sampling situations without fixed marginal totals under both null (raters are independent) and non-null cases. One might think that the asymptotic variances given in (3) are appropriate for the situation with fixed marginal totals. In this paper, we obtain the conditional (on both marginals) asymptotic variances for $\hat{\kappa}$ for both null and non-null cases and compare them to that of (3).

Because of the computational difficulty, we use the simplest case of two raters using only two rating categories (I=2). Let $P_{ij}$ (i,j = 1,2) be the probabilities for (i,j) cell and $X_{ij}$ denote the (i,j) cell counts obtained in the experiment. Assume $X_{1.}$, $X_{2.}$, $X_{.1}$ and $X_{.2}$ are fixed with $X_{1.} + X_{2.} = N$. Using the $\hat{\kappa}$ as defined in (2) to obtain the variance of $\hat{\kappa}$, we need to get the variance of $(X_{11} + X_{22})$ conditional on the marginal totals: $X_{1.}$ and $X_{.1}$. The conditional distribution of $X_{11}$ given the marginal totals is obtained by Harkness and Katz {1964} to be the "extended hypergeometric distribution".

$$f(X_{11} \mid X_{1.}, X_{.1}) = g(X_{1.}, X_{.1}, t) \binom{X_{1.}}{X_{11}} \binom{X_{2.}}{X_{.1} - X_{11}} t^{X_{11}} \qquad (4)$$

where

$$g(X_{1.}, X_{.1}, t) = \left[ \sum_a \binom{X_{1.}}{a} \binom{X_{2.}}{X_{.1} - a} t^a \right]^{-1}$$

and $t = P_{11}P_{22}/P_{12}P_{21}$; with $o < t < \infty$. In the general non-null situation, we can replace $P_{11}$ by $\lambda\, P_{1.}P_{.1}$, $P_{12}$ by $P_{1.}\,(1-\lambda P_{.1})$, $P_{21}$ by $P_{.1}(1-\lambda P_{1.})$ and $P_{22}$ by $1-P_{1.}-P_{.1}+\lambda P_{1.}P_{.1}$, where

$$\mathrm{Max}\left(o, \frac{P_{1.} + P_{.1} - 1}{P_{1.}\, P_{.1}}\right) < \lambda < \mathrm{Min}\left(\frac{1}{P_{1.}}, \frac{1}{P_{.1}}\right) ,$$

$$\text{and } t = \frac{\lambda\, (1 - P_{1.} - P_{.1} + \lambda\, P_{1.}\, P_{.1})}{(1 - \lambda\, P_{1.})\,(1 - \lambda\, P_{.1})} .$$

Under the null hypothesis of independence where $P_{ij} = P_{i.}P_{.j}$ (i,j - 1,2), then t = 1 (i.e., $\lambda$=1) and the expression (4) reduces to the ordinary hypergeometric distribution

$$f\,(X_{11}\; X_{1.}, X_{.1}) = \frac{\binom{X_{1.}}{X_{11}} \binom{N - X_{.1}}{X_{.1} - X_{11}}}{\binom{N}{X_{.1}}} \qquad (5)$$

The conditional distribution given in (5) is generally used to perform the exact test of independence for a 2x2 contingency table with small samples. For the large sample case (as $N \to \infty$) with $X_{1.}/N \to P_{1.}$ and $X_{.1}/N \to P_{.1}$, Harkness and Katz {1964} obtain the asymptotic mean and variance of $X_{11}$ as

$$E(X_{11} \mid X_{1.}, X_{.1}) = X_{1.}Q = \lambda^* \frac{X_{1.}X_{.1}}{N},$$

$$\mathrm{Var}\,(X_{11} \mid X_{1.}, X_{.1}) = \left(\sum_{i,\,j=1}^{2} \frac{1}{P^*_{ij}}\right)^{-1}, \qquad (6)$$

where

$$Q = \frac{-d + \left[d^2 + 4X_{1.}X_{.1}t\,(1-t)\right]^{\frac{1}{2}}}{2\,(1-t)\,X_{1.}}$$

$$d = N - (X_{1.} + X_{.1})\,(1-t)$$

$$\lambda^* = \frac{NQ}{X_{.1}},$$

and $\{P_{ij}^*\}$ are $\{P_{ij}\}$ expressed in terms of $\lambda^*$, $X_{.1}/N$ in place of $\lambda$, $P_{i.}$ and $P_{.j}$. In the null case, $t = 1$, (i.e., $\lambda^* = 1$) then

$$E_0\,(X_{11} \mid X_{1.}, X_{.1}) = \frac{X_{1.}X_{.1}}{N}$$

and

$$\mathrm{Var}_0\,(X_{11} \mid X_{1.}, X_{.1}) = \frac{X_{1.}X_{.1}X_{2.}X_{.2}}{N^2\,(N-1)}. \qquad (7)$$

Another way of obtaining the conditional variance of $X_{11}$ given the marginal totals $X_{1.}$ and $X_{.1}$ under the null case is to use a lemma due to Hinkley (1974).

Lemma (Hinkley)

If $S = S(X)$ is complete minimal sufficient for such that $E_X\,(a\,(X);\,\theta) = b\,(\theta)$
and $E_S\,(c\,(S);\,\theta) = b\,(\theta)$,
then $E_{X|S}\,(a\,(x) \mid S) = c\,(s)$. (8)

Using the lemma above, we need to show that

$$E_{X|S}\,(X_{11} - E_{11})^2 = \frac{X_{1.}X_{.1}X_{2.}X_{.2}}{N^2\,(N-1)},$$

where

$$E_{11} = E_0\,(X_{11} \mid X_{1.}, X_{.1}) = \frac{X_{1.}X_{.1}}{N}.$$

$$\text{Let } Q = \left(X_{11} - \frac{X_{1.}X_{.1}}{N}\right)^2.$$

$$\text{and } E\,(Q \mid S) = E\,(X_{11}^2 \mid S) - \left(\frac{X_{1.}X_{.1}}{N}\right)^2,$$

then $E(X_{11}^2) = NP_{1.}P_{.1} + N\,(N-1)\,P_{1.}^2P_{.1}^2 = b\,(\theta)$. Now we need to construct $c\,(s)$ such that (8) is satisfied. This is accomplished by first showing that

$$E\,(X_{1.}^2X_{.1}) = N^2\,(N-1)\,P_{11}P_{1.} + N^2\,P_{11},$$

$$E\,(X_{1.}X_{.1}^2) = N^2\,(N-1)\,P_{11}P_{.1} + N^2P_{11},$$

and

$$E(X_{1.}^2X_{.1}^2) = N^2\,(N-1)^2P_{11}^2 + N^2\,(N-1)P_{11}(P_{1.}+P_{.1}) + N^2P_{11}.$$

Hence

$$c(s) = (X_{1.}^2X_{.1}^2 - X_{1.}X_{.1}^2 - X_{1.}^2X_{.1} + NX_{1.}X_{.1})/(N(N-1)).$$

thus,

$$E\,(c\,(s)\,;\,\theta) = b\,(\theta).$$

Substituting $c\,(s)$ into $E\,(Q \mid S)$ above, we get:

$$E\,(Q \mid S) = \frac{X_{1.}^2X_{.1}^2}{N}\left(\frac{1}{N-1} - \frac{1}{N}\right) + \frac{NX_{1.}X_{.1} - X_{1.}X_{.1} - X_{1.}X_{.1}^2}{N\,(N-1)}$$

$$= \frac{X_{1.}X_{.1}X_{2.}X_{.2}}{N^2\,(N-1)},$$

which is the conditional variance of $X_{11}$ under null case as given in (7). Since

$$\mathrm{Var}\,(X_{11}+X_{22} \mid X_{1.}, X_{.1})$$
$$= \mathrm{Var}\left(X_{11} + (N - X_{1.} - X_{.1} + X_{11}) \mid X_{1.}, X_{.1}\right)$$
$$= \mathrm{Var}\,(2X_{11} \mid X_{1.}, X_{.1}) = 4\,\mathrm{Var}\,(X_{11} \mid X_{1.}, X_{.1})$$

we get the asymptotic variance of $\hat{\kappa}$, given $X_{1.}$ and $X_{.1}$, as

$$\mathrm{Var}\left(\hat{K} \mid X_{1.}, X_{.1}\right) = \frac{4\mathrm{Var}\left(X_{11} \mid X_{1.}, X_{.1}\right)}{N^2\left(1-\sum_i \frac{X_{i.}X_{.i}}{N^2}\right)} \qquad (9)$$

Where var $(X_{11} | X_{1.}, X_{.1})$ are given in either (7) or (6) depending upon whether or not the hypothesis of independence is assumed. Thus, for the null case, the asymptotic variance of $\hat{K}$ given the marginal totals is

$$\mathrm{Var}_0\left(\hat{K} \mid X_{1.}, X_{.1}\right) = \frac{4\, X_{1.}X_{.1}X_{2.}X_{.2}}{N^4(N-1)\left(1-\sum_i \frac{X_{i.}X_{.i}}{N^2}\right)} \qquad (10)$$

$$= \frac{4\, P_{1.}P_{.1}P_{2.}P_{.2}}{(N-1)\left(1-\sum_i \frac{X_{i.}X_{.i}}{N^2}\right)} .$$

Comparing (10) to the second expression in (3), we see that the asymptotic variance for null case as given by some authors referred to earlier is incorrect even for the conditional situation.

Applying (6) to (9), we obtain the non-null asymptotic variance of $\hat{K}$ with fixed marginals to be

$$\mathrm{Var}\left(\hat{K}\ X_{1.}, X_{.1}\right) = \frac{4}{N^2\left(\sum_{i,j} \frac{1}{P_{ij}^*}\right)\left(1-\sum_i \frac{X_{i.}X_{.i}}{N^2}\right)}, \qquad (11)$$

where $P_{ij}^*$ are defined as before.

This conditional asymptotic variance of $\hat{K}$ for the non-null case, as given in (10), is also different from the first expression of (3). Thus, we concluded that the asymptotic variances as given in (3) are not correct either for the unconditional or for the conditional cases.

## REFERENCES


Goodman, L.A. and Kruskal, W.H. (1954), "Measures of Association for Cross Classifications", JASA, 49, 732-64.

Cohen, J. (1960), "A Coefficient of Agreement for Nominal Scales", Education and Psychological Measurement, 20, No. 1, 37-46.

Cohen, J. (1968), Weighted Kappa: Nominal Scale Agreement With Provision For Scaled Disagreement or Partial Credit", Psychological Bulletin, 70, No. 4,213-20.

Fleiss, J.L. (1971), "Measuring Nominal Scale Agreement Among Many Raters",Psychological Bulletin, 76, No. 5, 378-82.

Light, R.J. (1971), "Measures of Response Agreement for Qualitative Data: Some Generalizations and Alternations", Psychological Bulletin, 76, No.5, 365-77.

Marx, T.J. and Light, R.J. (1973), "A Many Observer Agreement Measure for Qualitative Response Data", Mimeograph, Laboratory of Human Development, Harvard Graduate School of Education, Cambridge, Massachusetts.

Fleiss, J.L. Cohen, J. and Everitt, B.S. (1969) "Large Sample Standard Errors for Kappa and Weighted Kappa", Psychological Bulletin, 72, No.5, 323-27.

Bishop, Y.M.M. Fienberg, S.E. and Holland, P.W. (1975), Discrete Multivariate Analysis - Theory and Practices, MIT Press, Cambridge Massachusetts.

Harkness, W.L. and Katz, L. (1964), "Comparison of the Power Functions For The Test of Independence In 2x2 Contingency Tables", Annal of Math. Stat., 1115-27.

Hinkley, D. (1974), "On Expectation Conditional On A Sufficient Statistic", University of Minnesota School of Statistics, Technical Report No. 237.