United States
Department of
Agriculture

Forest Service

**Rocky Mountain
Forest and Range
Experiment Station**

Fort Collins,
Colorado 80526

**Research Paper
RM-316**

Variance Approximations for Assessments of Classification Accuracy

Raymond L. Czaplewski

[25] $$\hat{\mathrm{Var}}(\hat{\kappa}_w) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\begin{matrix}(\overline{w}_{i\cdot}+\overline{w}_{\cdot j})(\hat{p}_o-1)\\ +w_{ij}(1-\hat{p}_c)\end{matrix}\right]\sum_{r=1}^{k}\sum_{s=1}^{k}\hat{E}[\varepsilon_{ij}\varepsilon_{rs}]\left[\begin{matrix}(\overline{w}_{r\cdot}+\overline{w}_{\cdot s})(\hat{p}_o-1)\\ +w_{rs}(1-\hat{p}_c)\end{matrix}\right]}{(1-\hat{p}_c)^4}$$

$$\hat{\mathrm{Var}}_o(\hat{\kappa}_w) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\begin{matrix}-(\overline{w}_{i\cdot}+\overline{w}_{\cdot j})(1-\hat{p}_c)\\ +w_{ij}(1-\hat{p}_c)\end{matrix}\right]\sum_{r=1}^{k}\sum_{s=1}^{k}\hat{E}_o[\varepsilon_{ij}\varepsilon_{rs}]\left[\begin{matrix}-(\overline{w}_{r\cdot}+\overline{w}_{\cdot s})(1-\hat{p}_c)\\ +w_{rs}(1-\hat{p}_c)\end{matrix}\right]}{(1-\hat{p}_c)^4}$$

[28] $$\hat{\mathrm{Var}}_o(\hat{\kappa}_w) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k}(w_{ij}-\overline{w}_{i\cdot}-\overline{w}_{\cdot j})\sum_{r=1}^{k}\sum_{s=1}^{k}\hat{E}_o[\varepsilon_{ij}\varepsilon_{rs}](w_{rs}-\overline{w}_{r\cdot}-\overline{w}_{\cdot s})}{(1-\hat{p}_c)^2}$$

$$\hat{\mathrm{Cov}}_o(\hat{\mathbf{p}}_{(i|\cdot i)} - \hat{\mathbf{p}}_{i\cdot}) = [\mathbf{I}-\mathbf{I}][\hat{\mathrm{Cov}}_o(\mathbf{p}_{\cdot i})][\mathbf{I}-\mathbf{I}].$$ [101]

$$\hat{\mathrm{Var}}(\hat{p}_{i|j\cdot}) = \frac{(\hat{p}_{i\cdot}-\hat{p}_{ji})^2}{\hat{p}_{j\cdot}^4}\hat{E}[\varepsilon_{ji}^2] - 2\frac{(\hat{p}_{i\cdot}-\hat{p}_{ji})\hat{p}_{ji}}{\hat{p}_{j\cdot}^4}\sum_{\substack{r=1\\ r\neq i}}^{k}\hat{E}[\varepsilon_{ji}\varepsilon_{jr}] + \frac{\hat{p}_{ji}^2}{\hat{p}_{i\cdot}^4}\sum_{\substack{r=1\\ r\neq i}}^{k}\sum_{\substack{s=1\\ s\neq i}}^{k}\hat{E}[\varepsilon_{jr}\varepsilon_{js}]$$ [87]

$$\mathrm{Cov}(\hat{\kappa}_{\cdot i}) = \mathbf{d}'_{k\cdot i}\left(\hat{\mathrm{Cov}}(vec\hat{\mathbf{P}})\right)\mathbf{d}_{k\cdot i}$$ [80]

[85] $$\varepsilon^2_{P(i|\cdot j)} \approx \left(\frac{\hat{p}_{\cdot j}-\hat{p}_{ij}}{\hat{p}_{\cdot j}^2}\varepsilon_{ij} - \frac{\hat{p}_{ij}}{\hat{p}_{\cdot j}^2}\sum_{\substack{r=1\\ r\neq i}}^{k}\varepsilon_{rj}\right)\left(\frac{\hat{p}_{\cdot j}-\hat{p}_{ij}}{\hat{p}_{\cdot j}^2}\varepsilon_{ij} - \frac{\hat{p}_{ij}}{\hat{p}_{\cdot j}^2}\sum_{\substack{s=1\\ s\neq i}}^{k}\varepsilon_{sj}\right)$$

$$\hat{\mathrm{Var}}(\hat{p}_{i|\cdot j}) = \frac{(\hat{p}_{\cdot j}-\hat{p}_{ij})^2}{\hat{p}_{\cdot j}^4}\hat{E}[\varepsilon_{ij}^2] - 2\frac{(\hat{p}_{\cdot j}-_{ij})\hat{p}_{ij}}{\hat{p}_{\cdot j}^4}\sum_{\substack{r=1\\ r\neq i}}^{k}\hat{E}[\varepsilon_{ij}\varepsilon_{rj}] + \frac{\hat{p}_{ij}^2}{\hat{p}_{\cdot j}^4}\sum_{\substack{r=1\\ r\neq i}}^{k}\sum_{\substack{s=1\\ s\neq i}}^{k}\hat{E}[\varepsilon_{rj}\varepsilon_{sj}]$$

Abstract

Czaplewski, R. L. 1994. Variance approximations for assessments of classification accuracy. Res. Pap. RM-316. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. 29 p.

Variance approximations are derived for the weighted and unweighted kappa statistics, the conditional kappa statistic, and conditional probabilities. These statistics are useful to assess classification accuracy, such as accuracy of remotely sensed classifications in thematic maps when compared to a sample of reference classifications made in the field. Published variance approximations assume multinomial sampling errors, which implies simple random sampling where each sample unit is classified into one and only one mutually exclusive category with each of two classification methods. The variance approximations in this paper are useful for more general cases, such as reference data from multiphase or cluster sampling. As an example, these approximations are used to develop variance estimators for accuracy assessments with a stratified random sample of reference data.

Keywords: Kappa, remote sensing, photo-interpretation, stratified random sampling, cluster sampling, multiphase sampling, multivariate composite estimation, reference data, agreement.

USDA Forest Service
Research Paper RM-316

September 1994

Variance Approximations for Assessments of Classification Accuracy

Raymond L. Czaplewski
USDA Forest Service
Rocky Mountain Forest and Range Experiment Station[1]

[1]Headquarters is in Fort Collins, in cooperation with Colorado State University.

Contents

Variance Approximations for Assessments of Classification Accuracy

Raymond L. Czaplewski

INTRODUCTION

Assessments of classification accuracy are important to remote sensing applications, as reviewed by Congalton and Mead (1983), Story and Congalton (1986), Rosenfield and Fitzpatrick-Lins (1986), Campbell (1987, pp. 334–365), Congalton (1991), and Stehman (1992). Monserud and Leemans (1992) consider the related problem of comparing different vegetation maps. Recent literature favors the kappa statistic as a method for assessing classification accuracy or agreement.

The kappa statistic, which is computed from a square contingency table, is a scalar measure of agreement between two classifiers. If one classifier is considered a reference that is without error, then the kappa statistic is a measure of classification accuracy. Kappa equals 1 for perfect agreement, and zero for agreement expected by chance alone. Figure 1 provides interpretations of the magnitude of the kappa statistic that have appeared in the literature. In addition to kappa, Fleiss (1981) suggests that conditional probabilities are useful when assessing the agreement between two different classifiers, and Bishop et al. (1975) suggest statistics that quantify the disagreement between classifiers.

Existing variance approximations for kappa assume multinomial sampling errors for the proportions in the contingency table; this implies simple random sampling

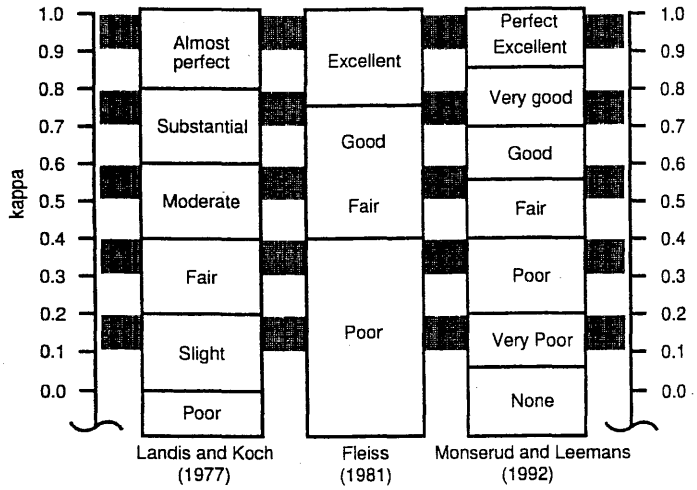


Figure 1. — Interpretations of kappa statistic as proposed in past literature. Landis and Koch (1977) characterize their interpretations as useful benchmarks, although they are arbitrary; they use clinical diagnoses from the epidemiological literature as examples. Fleiss (1981, p. 218) bases his interpretations on Landis and Koch (1977), and suggests that these interpretations are suitable for "most purposes." Monserud and Leemans (1992) use their interpretations for global vegetation maps.

where each sample unit is classified into one and only one mutually exclusive category with each of the two methods (Stehman 1992). This paper considers more general cases, such as reference data from stratified random sampling, multiphase sampling, cluster sampling, and multistage sampling.

KAPPA STATISTIC (κ_w)

The weighted kappa statistic (κ_w) was first proposed by Cohen (1968) to measure the agreement between two different classifiers or classification protocols. Let p_{ij} represent the probability that a member of the population will be assigned into category i by the first classifier and category j by the second. Let k be the number of categories in the classification system, which is the same for both classifiers. κ_w is a scalar statistic that is a nonlinear function of all k^2 elements of the $k \times k$ contingency table, where p_{ij} is the ijth element of the contingency table. Note that the sum of all k^2 elements of the contingency table equals 1:

$$\sum_{i=1}^{k}\sum_{j=1}^{k} p_{ij} = 1. \qquad [1]$$

Define w_{ij} as the value which the user places on any partial agreement whenever a member of the population is assigned to category i by the first classifier and category j by the second classifier (Cohen 1968). Typically, the weights range from $0 \le w_{ij} \le 1$, with $w_{ii} = 1$ (Landis and Koch 1977, p. 163). For example, w_{ij} might equal 0.67 if category i represents the large size class and j is the medium size class; if r represents the small size class, then w_{ir} might equal 0.33; and w_{js} might equal 0.0 if s represents any other classification. The unweighted kappa statistic uses $w_{ii} = 1$ and $w_{ij} = 0$ for $i \neq j$ (Fleiss 1981, p. 225), which means that the agreement must be exact to be valued by the user.

Using the notation of Fleiss et al. (1969), let:

$$p_{i\cdot} = \sum_{j=1}^{k} p_{ij} \qquad [2]$$

$$p_{\cdot j} = \sum_{i=1}^{k} p_{ij} \qquad [3]$$

$$p_0 = \sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij} p_{ij} \qquad [4]$$

$$p_c = \sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij} p_{i\cdot} p_{\cdot j} \quad [5]$$

Using this notation, the weighted kappa statistic (κ_w) as defined by Cohen (1968) is given as:

$$\kappa_w = \frac{p_0 - p_c}{1 - p_c}. \quad [6]$$

Estimated Weighted Kappa ($\hat{\kappa}_w$)

The true proportions p_{ij} are not known in practice, and the true κ_w must be estimated with estimated proportions in the contingency table ($\hat{p}_{ij}$):

$$\hat{\kappa}_w = \frac{\hat{p}_0 - \hat{p}_c}{1 - \hat{p}_c}, \quad [7]$$

where $\hat{p}_o$ and $\hat{p}_c$ are defined as in Eqs. 2, 3, 4, and using $\hat{p}_{ij}$ in place of p_{ij}.

The true κ_w equals the estimated $\hat{\kappa}_w$ plus an unknown random error ε_k:

$$\kappa_w = \hat{\kappa}_w + \varepsilon_k. \quad [8]$$

If $\hat{\kappa}_w$ is an unbiased estimate of κ_w, then $E[\varepsilon_k] = 0$ and $E[\hat{\kappa}_w] = \kappa_w$. By definition, $E[\varepsilon_k^2] = E[(\kappa_w - \hat{\kappa}_w)^2]$, and the variance of $\hat{\kappa}_w$ is:

$$\mathrm{Var}(\hat{\kappa}_w) = E[\varepsilon_k^2] - E^2[\varepsilon_k] = E\left[\varepsilon_k^2\right]. \quad [9]$$

Taylor Series Approximation for Var ($\hat{\kappa}_w$)

κ_w is a nonlinear, multivariate function of the k^2 elements (p_{ij}) in the contingency table (Eqs. 2, 3, 4, 5, and 6). The multivariate Taylor series approximation is used to produce an estimated variance $\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa}_w)$. Let $\varepsilon_{ij} = (p_{ij} - \hat{p}_{ij})$, and $(\partial k_w / \partial p_{ij})_{|p_{ij}=\hat{p}_{ij}}$ be the partial derivative of $\hat{\kappa}_w$ with respect to p_{ij} evaluated at $p_{ij} = \hat{p}_{ij}$. The multivariate Taylor series expansion (Deutch 1965, pp. 70–72) of κ_w is:

$$\begin{aligned}\kappa_w &= \hat{\kappa}_w \\ &+ \varepsilon_{11}\left(\frac{\partial k_w}{\partial p_{11}}\right)_{|p_{ij}=\hat{p}_{ij}} + \mathrm{L} + \varepsilon_{1k}\left(\frac{\partial k_w}{\partial p_{1k}}\right)_{|p_{ij}=\hat{p}_{ij}} + \\ &+ \varepsilon_{21}\left(\frac{\partial k_w}{\partial p_{21}}\right)_{|p_{ij}=\hat{p}_{ij}} + \mathrm{L} + \varepsilon_{2k}\left(\frac{\partial k_w}{\partial p_{2k}}\right)_{|p_{ij}=\hat{p}_{ij}} + \mathrm{L} \\ &+ \varepsilon_{k1}\left(\frac{\partial k_w}{\partial p_{k1}}\right)_{|p_{ij}=\hat{p}_{ij}} + \mathrm{L} + \varepsilon_{kk}\left(\frac{\partial k_w}{\partial p_{kk}}\right)_{|p_{ij}=\hat{p}_{ij}} + R \end{aligned} \quad [10]$$

where R is the remainder. In addition, assume that $\hat{p}_{ij}$ is nearly equal to p_{ij} (i.e., $p_{ij} \approx \hat{p}_{ij}$); hence, $\varepsilon_{ij} \approx 0$ because $\varepsilon_{ij} = (p_{ij} - \hat{p}_{ij})$, the higher-order products of ε_{ij} in the Taylor series expansion are assumed to be much smaller than ε_{ij}, and the R in Eq. 10 is assumed to be negligible. Eq. 10 is linear with respect to all $\varepsilon_{ij} = (p_{ij} - \hat{p}_{ij})$.

The Taylor series expansion in Eq. 10 provides the following linear approximation after ignoring the remainder R:

$$\varepsilon_k = (\kappa_w - \hat{\kappa}_w) \approx \sum_{i=1}^{k}\sum_{j=1}^{k} \varepsilon_{ij}\left(\frac{\partial k_w}{\partial p_{ij}}\right)_{|p_{ij}=\hat{p}_{ij}}. \quad [11]$$

The squared random error approximately equals ε_k^2 from Eq. 11:

$$\varepsilon_k^2 \approx \left[\sum_{i=j}^{k}\sum_{j=i}^{k} \varepsilon_{ij}\left(\frac{\partial k_w}{\partial p_{ij}}\right)_{|p_{ij}=\hat{p}_{ij}}\right]\left[\sum_{r=1}^{k}\sum_{s=1}^{k} \varepsilon_{rs}\left(\frac{\partial k_w}{\partial p_{rs}}\right)_{|p_{rs}=\hat{p}_{rs}}\right]$$

$$\varepsilon_k^2 \approx \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{r=1}^{k}\sum_{s=1}^{k} \varepsilon_{ij}\varepsilon_{rs}\left(\frac{\partial k_w}{\partial p_{rs}}\right)_{|p_{rs}=\hat{p}_{rs}}\left(\frac{\partial k_w}{\partial p_{ij}}\right)_{|p_{ij}=\hat{p}_{ij}}. \quad [12]$$

From Eqs. 9 and 12, Var ($\hat{\kappa}_w$) is approximately:

$$\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa}_w) \approx \sum_{i=1}^{k}\sum_{j=1}^{k}\left(\frac{\partial k_w}{\partial p_{ij}}\right)_{|p_{ij}=\hat{p}_{ij}} \hat{\mathrm{E}}[\varepsilon_{ij}\varepsilon_{rs}] \sum_{r=1}^{k}\sum_{s=1}^{k}\left(\frac{\partial k_w}{\partial p_{rs}}\right)_{|p_{rs}=\hat{p}_{rs}}. \quad [13]$$

This corresponds to the approximation using the delta method (e.g., Mood et al. 1963, p. 181; Rao 1965, pp. 321–322). The partial derivatives needed for $\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa}_w)$ in Eq. 13 are derived in the following section.

Partial Derivatives for Var ($\hat{\kappa}_w$) Approximation

The partial derivative of κ_w in Eq. 13 is derived by rewriting κ_w as a function of p_{ij}. First, p_o in Eq. 6 is expanded to isolate the p_{ij} term using the definition of p_o in Eq. 4:

$$p_0 = w_{ij} p_{ij} + \sum_{\substack{r=1 \\ \{rs\} \neq \{ij\}}}^{k}\sum_{s=1}^{k} w_{rs} p_{rs}. \quad [14]$$

The partial derivative of p_o with respect to p_{ij} is simply:

$$\frac{\partial p_0}{\partial p_{ij}} = w_{ij}. \quad [15]$$

As the next step in deriving the partial derivative of κ_w in Eq. 13, p_c in Eq. 6 is expanded to isolate the p_{ij} term using the definition of p_c in Eq. 5:

$$p_c = \sum_{r=1}^{k}\sum_{s=1}^{k} w_{rs} p_{r\cdot} p_{\cdot s}'$$

$$p_c = \sum_{r=1}^{k}\left[w_{rj} p_{r\cdot} p_{\cdot j} + \sum_{\substack{s=1 \\ s\neq j}}^{k} w_{rs} p_{r\cdot} p\cdot s \right],$$

$$p_c = \sum_{\substack{r=1 \\ r\neq i}}^{k}\left[w_{rj} p_{r\cdot}\left(\sum_{u=1}^{k} p_{uj}\right) + \sum_{\substack{s=1 \\ s\neq j}}^{k} w_{rs} p_{r\cdot} p_{\cdot s} \right] + w_{ij} p_{i\cdot} p_{\cdot j} + \sum_{\substack{s=1 \\ s\neq j}}^{k} w_{is} p_{i\cdot} p_{s},$$

$$p_c = \sum_{\substack{r=1 \\ r\neq i}}^{k}\left[w_{rj} p_{r\cdot}\left(p_{ij} + \sum_{\substack{u=1 \\ u\neq i}}^{k} p_{uj}\right) + \sum_{\substack{s=1 \\ s\neq j}}^{k} w_{rs} p_{r\cdot} p_{s} \right] + w_{ij}\left(p_{ij} + \sum_{\substack{v=1 \\ v\neq j}}^{k} p_{iv}\right)\left(p_{ij} + \sum_{\substack{u=1 \\ u\neq i}}^{k} p_{uj}\right) + \sum_{\substack{s=1 \\ s\neq j}}^{k} w_{is}\left(p_{ij} + \sum_{\substack{w=1 \\ w\neq j}}^{k} p_{iw}\right) p_{\cdot s},$$

$$p_c = \left(\sum_{\substack{r=1 \\ r\neq i}}^{k} w_{rj} p_{r\cdot}\right) p_{ij} + \sum_{\substack{r=1 \\ r\neq i}}^{k}\sum_{\substack{u=1 \\ u\neq i}}^{k} w_{rj} p_{r\cdot} p_{uj} + \sum_{\substack{r=1 \\ r\neq i}}^{k}\sum_{\substack{s=1 \\ s\neq j}}^{k} w_{rs} p_{r\cdot} p_{is} + w_{ij} p_{ij}^2 + w_{ij}\left(\sum_{\substack{v=1 \\ v\neq j}}^{k} p_{iv} + \sum_{\substack{u=1 \\ u\neq i}}^{k} p_{uj}\right) p_{ij} + w_{ij}\sum_{\substack{v=1 \\ v\neq j}}^{k} p_{iv}\sum_{\substack{u=1 \\ u\neq i}}^{k} p_{uj} + \left(\sum_{\substack{s=1 \\ s\neq j}}^{k} w_{is} p_{\cdot s}\right) p_{ij} + \sum_{\substack{s=1 \\ s\neq j}}^{k} w_{is} p_{\cdot s}\sum_{\substack{w=1 \\ w\neq j}}^{k} p_{iw},$$

$$p_c = (w_{ij}) p_{ij}^2 + \left[\sum_{\substack{r=1 \\ r\neq i}}^{k} w_{rj} p_{r\cdot} + w_{ij}\sum_{\substack{v=1 \\ v\neq j}}^{k} p_{iv} + w_{ij}\sum_{\substack{u=1 \\ u\neq i}}^{k} p_{uj} + \sum_{\substack{s=1 \\ s\neq j}}^{k} w_{is} p_{\cdot s} \right] p_{ij} + \left[\begin{array}{l} \sum_{\substack{r=1 \\ r\neq i}}^{k}\sum_{\substack{u=1 \\ u\neq i}}^{k} w_{rj} p_{r\cdot} p_{uj} + \sum_{\substack{r=1 \\ r\neq i}}^{k}\sum_{\substack{s=1 \\ s\neq j}}^{k} w_{rs} p_{r\cdot} p_{\cdot s} + \\ \sum_{\substack{v=1 \\ v\neq j}}^{k}\sum_{\substack{u=1 \\ u\neq i}}^{k} w_{ij} p_{iv} p_{uj} + \sum_{\substack{s=1 \\ s\neq j}}^{k}\sum_{\substack{w=1 \\ w\neq j}}^{k} w_{is} p_{\cdot s} p_{iw} \end{array} \right],$$

$$p_c = (w_{ij}) p_{ij}^2 + (b_{ij}) p_{ij} + c_{ij}, \qquad [16]$$

where

$$b_{ij} = \sum_{\substack{r=1 \\ r\neq i}}^{k} w_{rj} p_{r\cdot} + w_{ij}\sum_{\substack{v=1 \\ v\neq j}}^{k} p_{iv} + w_{ij}\sum_{\substack{u=1 \\ u\neq i}}^{k} p_{uj} + \sum_{\substack{s=1 \\ s\neq j}}^{k} w_{is} p_{\cdot s}, \qquad [17]$$

$$c_{ij} = \left[\begin{array}{l} \sum_{\substack{r=1 \\ r\neq i}}^{k}\sum_{\substack{u=1 \\ u\neq i}}^{k} w_{rj} p_{r\cdot} p_{uj} + \sum_{\substack{r=1 \\ r\neq i}}^{k}\sum_{\substack{s=1 \\ s\neq j}}^{k} w_{rs} p_{r\cdot} p_{\cdot s} + \\ \sum_{\substack{v=1 \\ v\neq j}}^{k}\sum_{\substack{u=1 \\ u\neq i}}^{k} w_{ij} p_{iv} p_{uj} + \sum_{\substack{s=1 \\ s\neq j}}^{k}\sum_{\substack{w=1 \\ w\neq j}}^{k} w_{is} p_{\cdot s} p_{iw} \end{array} \right] \qquad [18]$$

Finally, the partial derivative of p_c with respect to p_{ij} is simply:

$$\frac{\partial p_c}{\partial p_{ij}} = (2w_{ij}) p_{ij} + b_{ij}. \qquad [19]$$

The partial derivative of κ_w (Eq. 6) with respect to p_{ij} is determined with Eqs. 15 and 19:

$$\frac{\partial k_w}{\partial p_{ij}} = \frac{(1-p_c)\left[\frac{\partial p_o}{\partial p_{ij}} - \frac{\partial p_c}{\partial p_{ij}}\right] - (p_o - p_c)\left[-\frac{\partial p_c}{\partial p_{ij}}\right]}{(1-p_o)^2}$$

$$\frac{\partial k_w}{\partial p_{ij}} = \frac{(1-p_c)(w_{ij} - 2w_{ij} p_{ij} - b_{ij}) + (p_o - p_c)(2w_{ij} p_{ij} + b_{ij})}{(1-p_c)^2}. \qquad [20]$$

The b_{ij} term in Eqs. 17 and 20 can be simplified:

$$b_{ij} = \sum_{\substack{r=1 \\ r\neq i}}^{k} w_{rj} p_{r\cdot} + w_{ij}(p_{i\cdot} - p_{ij}) + w_{ij}(p_{\cdot j} - p_{ij}) + \sum_{\substack{s=1 \\ s\neq j}}^{k} w_{is} p_{\cdot s},$$

$$b_{ij} = \sum_{r=1}^{k} w_{rj} p_{r\cdot} - 2w_{ij} p_{ij} + \sum_{s=1}^{k} w_{is} p_{\cdot s}.$$

Using the notation of Fleiss et al. (1969):

$$b_{ij} = \overline{w}_{.j} - 2w_{ij}p_{ij} + \overline{w}_{i.}, \quad [21]$$

where

$$\overline{w}_{i.} = \sum_{j=1}^{k} w_{ij}p_{.j} = \sum_{j=1}^{k} w_{ij} \sum_{r=1}^{k} p_{rj}, \quad [22]$$

$$\overline{w}_{.i} = \sum_{i=1}^{k} w_{ij}p_{i.} = \sum_{i=1}^{k} w_{ij} \sum_{s=1}^{k} p_{is}. \quad [23]$$

Replacing b_{ij} from Eq. 21 into the partial derivative of $\hat{\kappa}_w$ from Eq. 20:

$$\frac{\partial k_w}{\partial p_{ij}} = \frac{(1-p_c)(w_{ij} - \overline{w}_{.j} - \overline{w}_{i.}) + (p_o - p_c)(\overline{w}_{.j} + \overline{w}_{i.})}{(1-p_c)^2}$$

$$\frac{\partial k_w}{\partial p_{ij}} = \frac{(1-p_c)(w_{ij} + (p_o - 1)(\overline{w}_{.j} + \overline{w}_{i.})}{(1-p_c)^2} \quad [24]$$

Equation 24 contains p_{ij} terms that are imbedded within the $\overline{w}_{i.}$ and $\overline{w}_{.j}$ terms (Eqs. 22 and 24). Any higher-order partial derivatives should use Eq. 20 rather than Eq. 24.

First-Order Approximation of Var ($\hat{\kappa}_w$)

The first-order variance approximation for $\hat{\kappa}_w$ is determined by combining Eqs. 13 and 24:

$$\hat{\mathrm{Var}}(\hat{\kappa}_w) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\begin{matrix}(\overline{w}_{i.} + \overline{w}_{.j})(\hat{p}_o - 1) \\ +w_{ij}(1-\hat{p}_c)\end{matrix}\right]\sum_{r=1}^{k}\sum_{s=1}^{k}\hat{E}[\varepsilon_{ij}\varepsilon_{rs}]\left[\begin{matrix}(\overline{w}_{r.} + \overline{w}_{.s})(\hat{p}_o - 1) \\ +w_{rs}(1-\hat{p}_c)\end{matrix}\right]}{(1-\hat{p}_c)^4}. \quad [25]$$

The multinomial distribution is typically used for $\hat{E}[\varepsilon_{ij}\varepsilon_{rs}]$ in Eq. 25 (see also Eq. 104). However, other types of covariance matrices are possible, such as the covariance matrix for a stratified random sample (see Eqs. 124 and 125), the sample covariance matrix for a simple random sample of cluster plots (see Eq. 105), or the estimation error covariance matrix for multivariate composite estimates with multiphase or multistage samples of reference data (see Czaplewski 1992).

$\hat{\mathrm{Var}}_o(\hat{\kappa}_w)$ Assuming Chance Agreement

In many accuracy assessments, the null hypothesis is that the agreement between two different protocols is no greater than that expected by chance, which is stated more formally as the hypothesis that the row and column classifiers are independent. Under this hypothesis, the probability of a unit being classified as type i with the first protocol is independent of the classification with the second protocol, and the following true population parameters are expected (Fleiss et al. 1969):

$$p_{ij} = p_{i.}p_{.j}. \quad [26]$$

Substituting Eq. 26 into Eq. 4 and using the definition of p_c in Eq. 5, the hypothesized true value of p_o under this null hypothesis is:

$$p_o = \sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij}p_{i.}p_{.j} = p_c. \quad [27]$$

Substituting Eqs. 26 and 27 into Eq. 25, the approximate variance of $\hat{\kappa}_w$ expected under the null hypothesis is:

$$\hat{\mathrm{Var}}_o(\hat{\kappa}_w) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\begin{matrix}-(\overline{w}_{i.} + \overline{w}_{.j})(1-\hat{p}_c) \\ +w_{ij}(1-\hat{p}_c)\end{matrix}\right]\sum_{r=1}^{k}\sum_{s=1}^{k}\hat{E}_o[\varepsilon_{ij}\varepsilon_{rs}]\left[\begin{matrix}-(\overline{w}_{r.} + \overline{w}_{.s})(1-\hat{p}_c) \\ +w_{rs}(1-\hat{p}_c)\end{matrix}\right]}{(1-\hat{p}_c)^4}$$

$$\hat{\mathrm{Var}}_o(\hat{\kappa}_w) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k}(w_{ij} - \overline{w}_{i.} - \overline{w}_{.j})\sum_{r=1}^{k}\sum_{s=1}^{k}\hat{E}_o[\varepsilon_{ij}\varepsilon_{rs}](w_{rs} - \overline{w}_{r.} - \overline{w}_{.s})}{(1-\hat{p}_c)^2}. \quad [28]$$

The covariances $\hat{E}_o[\varepsilon_{ij}\varepsilon_{rs}]$ in Eq. 28 need to be estimated under the conditions of the null hypothesis, namely that $p_{ij} = p_{i.}p_{.j}$ (see Eqs. 113, 114, and 117).

Unweighted Kappa ($\hat{K}$)

The unweighted kappa ($\hat{\kappa}$) treats any lack of agreement between classifications as having no value or weight. $\hat{\kappa}$ is used in remote sensing more often than the weighted kappa ($\hat{\kappa}_w$). κ is a special case of $\hat{\kappa}_w$ (Fleiss et al. 1969), in which $w_{ii} = 1$ and $w_{ij} = 0$ for $i \neq j$. In this case, $\hat{\kappa}_w$ is defined as in Eq. 6 (Fleiss et al. 1969) with the following intermediate terms in Eqs. 4, 5, 22, and 23 equal to:

$$\hat{p}_o = \sum_{i=1}^{k} \hat{p}_{ii} \quad [29]$$

$$\hat{p}_c = \sum_{i=1}^{k} \hat{p}_{i.}\hat{p}_{.i} \quad [30]$$

$$\overline{w}_{i.} = \hat{p}_{.i} \quad [31]$$

$$\overline{w}_{.j} = \hat{p}_{j.}. \quad [32]$$

Replacing Eqs. 29, 30, 31, and 32 into $\hat{\mathrm{Var}}(\hat{\kappa}_w)$ in Eq. 25, where $w_{ij} = 0$ if $i \neq j$ and $w_{ii} = 1$, the variance of the unweighted kappa is:

$$\hat{\mathrm{Var}}(\hat{\kappa}) = \frac{\left[\begin{matrix}\sum_{i=1}^{k}\sum_{j=1}^{k}(\hat{p}_{.i} + p_{j.})(\hat{p}_o - 1) \\ +\sum_{i=1}^{k}\sum_{j=1}^{k}w_{ij}(1-\hat{p}_c)\end{matrix}\right]\left[\begin{matrix}\sum_{r=1}^{k}\sum_{s=1}^{k}\hat{E}[\varepsilon_{ij}\varepsilon_{rs}](\hat{p}_{.r} + \hat{p}_{s.})(\hat{p}_o - 1) \\ +\sum_{r=1}^{k}\sum_{s=1}^{k}\hat{E}[\varepsilon_{ij}\varepsilon_{rs}]w_{rs}(1-\hat{p}_c)\end{matrix}\right]}{(1-\hat{p}_c)^4}$$

$$\hat{\mathrm{Var}}(\hat{\kappa}) = \frac{\left[(\hat{p}_o - 1)\sum_{i=1}^{k}\sum_{\substack{j=1 \\ j=i}}^{k}(\hat{p}_{\cdot i} + p_{j\cdot}) + (1-\hat{p}_c)\sum_{i=1}^{k}1\right]\left[(\hat{p}_o - 1)\sum_{r=1}^{k}\sum_{\substack{s=1 \\ s=r}}^{k}\hat{E}[\varepsilon_{ij}\varepsilon_{rs}](\hat{p}_{\cdot r} + p_{s\cdot}) + (1-\hat{p}_c)\sum_{r=1}^{k}\hat{E}[\varepsilon_{ij}\varepsilon_{rr}]\right]}{(1-\hat{p}_c)^4}$$

$$\hat{\mathrm{Var}}(\hat{\kappa}) = \frac{\left[\begin{array}{l}(\hat{p}_o - 1)^2\sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{r=1}^{k}\sum_{s=1}^{k}\hat{E}[\varepsilon_{ij}\varepsilon_{rs}](\hat{p}_{\cdot i} + \hat{p}_{j\cdot})(\hat{p}_{\cdot r} + \hat{p}_{s\cdot}) \\ +(\hat{p}_o - 1)(1-\hat{p}_c)\sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{r=1}^{k}\hat{E}[\varepsilon_{ij}\varepsilon_{rr}](\hat{p}_{\cdot i} + \hat{p}_{j\cdot}) \\ +(1-\hat{p}_c)(\hat{p}_o - 1)\sum_{i=1}^{k}\sum_{r=1}^{k}\sum_{s=1}^{k}\hat{E}[\varepsilon_{ii}\varepsilon_{rs}](\hat{p}_{\cdot r} + \hat{p}_{s\cdot}) \\ +(1-\hat{p}_c)^2\sum_{i=1}^{k}\sum_{r=1}^{k}\hat{E}[\varepsilon_{ii}\varepsilon_{rr}]\end{array}\right]}{(1-\hat{p}_c)^4}$$

$$\hat{\mathrm{Var}}(\hat{\kappa}) = \frac{\left[\begin{array}{l}(1-\hat{p}_o)^2\sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{r=1}^{k}\sum_{s=1}^{k}\hat{E}[\varepsilon_{ij}\varepsilon_{rs}](\hat{p}_{\cdot i} + p_{j\cdot})(\hat{p}_{\cdot r} + \hat{p}_{s\cdot}) \\ -2(1-\hat{p}_o)(1-\hat{p}_c)\sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{r=1}^{k}\hat{E}[\varepsilon_{ij}\varepsilon_{rr}](\hat{p}_{\cdot i} + \hat{p}_{j\cdot}) \\ +(1-\hat{p}_c)^2\sum_{i=1}^{k}\sum_{j=1}^{k}\hat{E}[\varepsilon_{ii}\varepsilon_{jj}]\end{array}\right]}{(1-\hat{p}_c)^4}. \quad [33]$$

Likewise, the variance of the unweighted kappa statistic under the null hypothesis of chance agreement is a simplification of Eq. 28 or 33. Under this null hypothesis, $p_{ij} = p_{i\cdot}p_{\cdot j}$ and $\hat{p}_o = \hat{p}_c$ (see Eq. 27):

$$\hat{\mathrm{Var}}_o(\hat{\kappa}) = \frac{\sum_{i=1}^{k}(1-\hat{p}_{\cdot i} - \hat{p}_{i\cdot})\sum_{r=1}^{k}\hat{E}_o[\varepsilon_{ii}\varepsilon_{rr}](1-\hat{p}_{\cdot r} - \hat{p}_{r\cdot})}{(1-\hat{p}_c)^2}. \quad [34]$$

The covariances $\hat{E}_o[\varepsilon_{ii}\varepsilon_{rr}]$ in Eq. 34 need to be estimated under the conditions of the null hypothesis, namely that $p_{ij} = p_{i\cdot}p_{\cdot j}$ (see Eqs. 113, 114, and 117).

Note that the variance estimators in Eqs. 33 and 34 are approximations since they ignore higher-order terms in the Taylor series expansion (see Eqs. 10, 12, and 13). In the special case of simple random sampling, Stehman (1992) found that this approximation was satisfactory except for sample sizes of 60 or fewer reference plots; these results are based on Monte Carlo simulations with four hypothetical populations.

Matrix Formulation of $\hat{\kappa}$ Variance Approximations

The formulae above can be expressed in matrix algebra, which facilitates numerical implementation with matrix algebra software.

Let **P** represent the $k \times k$ matrix in which the ijth element of **P** is the scalar p_{ij}. In remote sensing jargon, **P** is the "error matrix" or "confusion matrix." Note that k is the number of categories in the classification system. Let $\mathbf{p}_{i\cdot}$ be the $k \times 1$ vector in which the ith element is the scalar $p_{i\cdot}$ (Eq. 2), and $\mathbf{p}_{\cdot j}$ be the $k \times 1$ vector in which the ith element is $p_{\cdot i}$ (Eq. 3). From Eqs. 2 and 3:

$$\mathbf{p}_{i\cdot} = \mathbf{P}\mathbf{1}, \quad [35]$$

$$\mathbf{p}_{\cdot j} = \mathbf{P}'\mathbf{1}, \quad [36]$$

where **1** is the $k \times 1$ vector in which each element equals 1, and $\mathbf{P}'$ is the transpose of **P**. The expected matrix of joint classification probabilities, analogous to **P**, under the hypothesis of chance agreement between the two classifiers is the $k \times k$ matrix $\mathbf{P}_c$, where each element is the product of its corresponding marginal:

$$\mathbf{p}_c = \mathbf{p}_{i\cdot}\mathbf{p}'_{\cdot j}. \quad [37]$$

Let **W** represent the $k \times k$ matrix in which the ijth element is w_{ij} (i.e., the weight or "partial credit" for the agreement when an object is classified as category i by one classifier and category j by the other classifier). From Eqs. 4 and 5,

$$p_o = \mathbf{1}'(\mathbf{W} \otimes \mathbf{P})\mathbf{1}, \quad [38]$$

$$p_c = \mathbf{1}'(\mathbf{W} \otimes \mathbf{P}_c)\mathbf{1}, \quad [39]$$

where $\otimes$ represents element-by-element multiplication (i.e., the ijth element of $\mathbf{A} \otimes \mathbf{B}$ is $a_{ij}b_{ij}$, and matrices **A** and **B** have the same dimensions). The weighted kappa statistic (κ_w) equals Eq. 6 with p_o and p_c defined in Eqs. 38 and 39.

The approximate variance of κ_w can be described in matrix algebra by rewriting the $k \times k$ contingency table as a $k^2 \times 1$ vector, as suggested by Christensen (1991). First, rearrange the $k \times k$ matrix **P** into the following $k^2 \times 1$ vector denoted *vec***P**; if $\mathbf{p}_j$ is the $k \times 1$ column vector in which the ith element equals p_{ij}, then $\mathbf{p} = [\mathbf{p}_1 | \mathbf{p}_2 | \cdots | \mathbf{p}_k]$, and $vec\mathbf{P} = [\mathbf{p}'_1 | \mathbf{p}'_2 | \cdots | \mathbf{p}'_k]'$. Let **Cov**(*vec***P**) denote the $k^2 \times k^2$ covariance matrix for the estimate $\hat{vec\mathbf{P}}$ of *vec***P**, such that the uvth element of **Cov**(*vec***P**) equals $E[(vec\mathbf{P}_u - \hat{vec\mathbf{P}}_u)(vec\mathbf{P}_v - \hat{vec\mathbf{P}}_v)]$ and $vec\mathbf{P}_u$ represents the uth element of *vec***P**. Define the $k \times 1$ intermediate vectors:

$$\mathbf{w}_{i\cdot} = \mathbf{W}\,\mathbf{p}_{\cdot j}, \quad [40]$$

$$\mathbf{w}_{\cdot j} = \mathbf{W}'\,\mathbf{p}_{j\cdot}, \quad [41]$$

and from Eq. 25 , the $k^2 \times 1$ vector:

$$\mathbf{d}_k = (1-p_c)vec\mathbf{W} - (1-p_o)\left\{\begin{bmatrix}\mathbf{w}_{1\cdot} \\ \mathbf{w}_{2\cdot} \\ \vdots \\ \mathbf{w}_{k\cdot}\end{bmatrix} + vec\begin{bmatrix}\mathbf{w}'_{\cdot 1} \\ \mathbf{w}'_{\cdot 2} \\ \vdots \\ \mathbf{w}'_{k}\end{bmatrix}\right\} \quad [42]$$

where *vec***W** is the $k^2 \times 1$ vector version of the weighting matrix **W**, which is analogous to *vec***P** above. Examples of $\mathbf{w}_{i\cdot}$, $\mathbf{w}_{\cdot j}$, and $\mathbf{d}_k$ are given in tables 1and 2.

Table 1. — Example data[1] from Fleiss et al. (1969, p. 324) for weighted kappa ($\hat{\kappa}_w$), including vectors used in matrix algebra formulation.

		Classifier A				
Classifier B	**Statistic**	**1**	**2**	**3**	$\hat{p}_{i\cdot}$	$w_{i\cdot}$
i=1	w_{1j}	1	0	0.4444	0.60	0.6944
	$\hat{p}_{1j}$	0.53	0.05	0.02		
	$\hat{p}_{i\cdot}\hat{p}_{\cdot j}$	0.39	0.15	0.06		
	$w_{i\cdot}+w_{\cdot j}$	1.3389	1.0611	1.2611		
i=2	w_{2j}	0	1	0.6667	0.30	0.3167
	$\hat{p}_{2j}$	0.11	0.14	0.05		
	$\hat{p}_{2\cdot}\hat{p}_{\cdot j}$	0.195	0.075	0.03		
	$w_{2\cdot}+w_{\cdot j}$	0.9611	0.6833	0.8833		
i=3	w_{3j}	0.4444	0.6667	1	0.10	0.5555
	$\hat{p}_{3j}$	0.01	0.06	0.03		
	$\hat{p}_{3j}\hat{p}_{\cdot j}$	0.065	0.025	0.01		
	$w_{3\cdot}+w_{\cdot j}$	1.1200	0.9222	1.1222		
	$\hat{p}_{\cdot j}$	0.65	0.25	0.10	1.00	
	$w_{\cdot j}$	0.6444	0.3667	0.5667		

Weighted κ from Fleiss et al. (1969, p. 324)

$\hat{\kappa}_w$ = **0.5071** $\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa}_w)$ = **0.003248** $\hat{\mathrm{V}}\mathrm{ar}_o(\hat{\kappa}_w)$ = **0.004269**
$\hat{p}_o$ = **0.8478** $\hat{p}_c$ = **0.6756**

$\hat{\mathbf{P}}$ subscripts							
i	j		$vec\hat{\mathbf{P}}$	$vec\,\mathbf{W}$	$(w_1 \vert w_2 \vert \mathrm{L} \vert w_k)'$	$vec(w_1' \vert w_2 \vert \mathrm{L} \vert w_k)'$	$\mathbf{d}_k$
1	1	1	0.53	1	0.6944	0.6444	0.1472
1	2	2	0.11	0	0.3167	0.6444	–0.2050
1	3	3	0.01	0.4444	0.5555	0.6444	–0.0637
2	1	4	0.05	0	0.6944	0.3667	–0.2264
2	2	5	0.14	1	0.3167	0.3667	0.2870
2	3	6	0.06	0.6667	0.5555	0.3667	0.0918
3	1	7	0.02	0.4444	0.6944	0.5667	–0.0767
3	2	8	0.05	0.6667	0.3167	0.5667	0.1001
3	3	9	0.03	1	0.5555	0.5667	0.1934

Unweighted k from Fleiss et al. (1969, p. 326)

$\hat{\kappa}_w$ = **0.4286** $\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa})$ = **0.002885** $\hat{\mathrm{V}}\mathrm{ar}_o(\hat{\kappa})$ = **0.003082**
$\hat{p}_o$ = **0.7000** $\hat{p}_c$ = **0.4750**

$\hat{\mathbf{P}}$ subscripts							
i	j		$vec\hat{\mathbf{P}}$	$vec\,\mathbf{W}$	$(w_1 \vert w_2 \vert \mathrm{L} \vert w_k)'$	$vec(w_1' \vert w_2 \vert \mathrm{L} \vert w_k)'$	$\mathbf{d}_k$
1	1	1	0.53	1	0.65	0.60	0.150
1	2	2	0.11	0	0.25	0.60	–0.255
1	3	3	0.01	0	0.10	0.60	–0.210
2	1	4	0.05	0	0.65	0.30	–0.285
2	2	5	0.14	1	0.25	0.30	0.360
2	3	6	0.06	0	0.10	0.30	–0.120
3	1	7	0.02	0	0.65	0.10	–0.225
3	2	8	0.05	0	0.25	0.10	–0.105
3	3	9	0.03	1	0.10	0.10	0.465

[1] *The covariance matrix for the estimated joint probabilities ($\hat{p}_{ij}$) is estimated assuming the multinomial distribution (see Eqs. 46, 47, 128, 130, 131, and 132.*

Table 2. — Example data[1] from Bishop et al. (1975, p. 397) for unweighted kappa ($\hat{\kappa}$), including vectors used in matrix formulation.

Example contingency table
Bishop et al. (1975, p. 397)

	$\hat{p}_{ij} = \hat{\mathbf{P}}$				
i	j=1	j=2	j=3	$\hat{p}_{i.}$	
1	0.2361	0.0556	0.1111	0.4028	$n = 72$
2	0.0694	0.1667	0	0.2361	$\hat{\kappa} = 0.3623$
3	0.1389	0.0417	0.1806	0.3611	$\hat{\mathrm{Var}}(\hat{\kappa}) = 0.007003$
$\hat{p}_{j.}$	0.4444	0.2639	0.2917		$\hat{\mathrm{Var}}(\hat{\kappa}) = 0.008235$[4]

Vectors used in matrix computations for $\hat{\mathrm{Var}}(\hat{\kappa})$ and $\hat{\mathrm{Var}}_o(\hat{\kappa})$

ij	$vec\hat{\mathbf{P}}$	$vec\hat{\mathbf{P}}_c$	$vec\mathbf{W}$	$(w_{i.} \mid w_{2.} \mid \cdots \mid w_{k.})'$	$vec(w'_{i.} \mid w'_{2.} \mid \cdots \mid w'_{k.})'$
11	0.2361	0.1790	1	0.4444	0.4028
21	0.0694	0.1049	0	0.2639	0.4028
31	0.1389	0.1605	0	0.2917	0.4028
12	0.0556	0.1063	0	0.4444	0.2361
22	0.1667	0.0623	1	0.2639	0.2361
32	0.0417	0.0953	0	0.2917	0.2361
13	0.1111	0.1175	0	0.4444	0.3611
23	0.0000	0.0689	0	0.2639	0.3611
33	0.1806	0.1053	1	0.2917	0.3611

Covariance matrix for $vec\hat{\mathbf{P}}$ assuming multinomial distribution. See $\hat{\mathbf{Cov}}(vec\hat{\mathbf{P}})$ in Eq. 128.

ij	i=1 j=1	i=2 j=1	i=3 j=1	i=1 j=2	i=2 j=2	i=3 j=2	i=1 j=3	i=2 j=3	i=3 j=3
11	0.0025	−0.0002	−0.0005	−0.0002	−0.0005	−0.0001	−0.0004	0	−0.0006
21	−0.0002	0.0009	−0.0001	−0.0001	−0.0002	−0.0000	−0.0001	0	−0.0002
31	−0.0005	−0.0001	0.0017	−0.0001	−0.0003	−0.0001	−0.0002	0	−0.0003
12	−0.0002	−0.0001	−0.0001	0.0007	−0.0001	−0.0000	−0.0001	0	−0.0001
22	−0.0005	−0.0002	−0.0003	−0.0001	0.0019	−0.0001	−0.0003	0	−0.0004
32	−0.0001	−0.0000	−0.0001	−0.0000	−0.0001	0.0006	−0.0001	0	−0.0001
13	−0.0004	−0.0001	−0.0002	−0.0001	−0.0003	−0.0001	0.0014	0	−0.0003
23	0	0	0	0	0	0	0	0	0
33	−0.0006	−0.0002	−0.0003	−0.0001	−0.0004	−0.0001	−0.0003	0	0.0021

Covariance matrix for $vec\hat{\mathbf{P}}$ under null hypothesis of independence between the row and column classifiers and the multinomial distribution. See $\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}})$ in Eqs. 130, 131 and 132.

ij	i=1 j=1	i=2 j=1	i=3 j=1	i=1 j=2	i=2 j=2	i=3 j=2	i=1 j=3	i=2 j=3	i=3 j=3
11	0.0020	−0.0003	−0.0004	−0.0003	−0.0002	−0.0002	−0.0003	−0.0002	−0.0003
21	−0.0003	0.0013	−0.0002	−0.0002	−0.0001	−0.0001	−0.0002	−0.0001	−0.0002
31	−0.0004	−0.0002	0.0019	−0.0002	−0.0001	−0.0002	−0.0003	−0.0002	−0.0002
12	−0.0003	−0.0002	−0.0002	0.0013	−0.0001	−0.0001	−0.0002	−0.0001	−0.0002
22	−0.0002	−0.0001	−0.0001	−0.0001	0.0008	−0.0001	−0.0001	−0.0001	−0.0001
32	−0.0002	−0.0001	−0.0002	−0.0001	−0.0001	0.0012	−0.0002	−0.0001	−0.0001
13	−0.0003	−0.0002	−0.0003	−0.0002	−0.0001	−0.0002	0.0014	−0.0001	−0.0002
23	−0.0002	−0.0001	−0.0002	−0.0001	−0.0001	−0.0001	−0.0001	0.0009	−0.0001
33	−0.0003	−0.0002	−0.0002	−0.0002	−0.0001	−0.0001	−0.0002	−0.0001	0.0013

[1] *The covariance matrix for the estimated joint probabilities ($\hat{p}_{ij}$) is estimated assuming the multinomial distribution (see Eqs. 46, 47, 128, 130, 131, and 132).*

[2] $\hat{\mathbf{Var}}(\hat{\kappa})$ *= 0.0101 in Bishop et al. (1975) is a computational error. The correct* $\hat{\mathrm{Var}}(\hat{\kappa})$ *is 0.008235 (Hudson and Ramm 1987).*

The approximate variance of κ_w expressed in matrix algebra is:

$$\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa}_w) = [\mathbf{d}_k' \hat{\mathbf{C}}\mathbf{ov}(vec\hat{\mathbf{P}})\, \mathbf{d}_k]/(1-\hat{p}_c)^4. \quad [43]$$

See Eqs. 104 and 105 for examples of $\hat{\mathbf{C}}\mathbf{ov}(vec\hat{\mathbf{P}})$. The variance estimator in Eq. 43 is equivalent to the estimator in Eq. 25. Tables 1 and 2 provide examples.

The structure of Eq. 43 reflects its origin as a linear approximation, in which $\hat{\kappa}_w \approx (vec\hat{\mathbf{P}})'\mathbf{d}_k$. This suggests different and more accurate variance approximations using higher-order terms in the multivariate Taylor series approximation for the $\mathbf{d}_k$ vector, and these types of approximations will be explored by the author in the future.

The variance of $\hat{\kappa}_w$ under the null hypothesis of chance agreement, i.e., $\hat{\mathrm{V}}\mathrm{ar}_o(\hat{\kappa}_w)$ in Eq. 28, is expressed by replacing p_o with p_c in Eq. 42:

$$\mathbf{d}_{k=0} = (1-p_c)\left\{vec\mathbf{W} - vec\begin{bmatrix}\mathbf{w}_{1\cdot}\\ \mathbf{w}_{2\cdot}\\ \vdots\\ \mathbf{w}_{k\cdot}\end{bmatrix}\right\} - vec\begin{bmatrix}\mathbf{w}'_{\cdot 1}\\ \mathbf{w}'_{\cdot 2}\\ \vdots\\ \mathbf{w}'_{\cdot k}\end{bmatrix}\Bigg\}, \quad [44]$$

$$\hat{\mathrm{V}}\mathrm{ar}_o(\hat{\kappa}_w) = [\mathbf{d}'_{k=o}\, \hat{\mathbf{C}}\mathbf{ov}_0(vec\hat{\mathbf{P}})\, \mathbf{d}_{k=0}]/(1-\hat{p}_c)^4. \quad [45]$$

Tables 1 and 2 provide examples of $\mathbf{d}_{k=0}$ and $\hat{\mathrm{V}}\mathrm{ar}_o(\hat{\kappa}_w)$ for the multinomial distribution. The covariance matrix $\hat{\mathbf{C}}\mathbf{ov}_o(vec\hat{\mathbf{P}})$ in Eq. 45 must be estimated under the conditions of the null hypothesis, namely that $E[p_{ij}] = p_{i\cdot}p_{\cdot j}$ (see Eqs. 113, 114, and 117).

The estimated variances for the unweighted $\hat{\kappa}$ statistics, e.g., $\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa})$ in Eq. 33 and $\hat{\mathrm{V}}\mathrm{ar}_o(\hat{\kappa})$ in Eq. 34, can be computed with matrix Eqs. 43 and 45 using $\mathbf{W} = \mathbf{I}$ in Eqs. 37, 38, 40, and 41, where I is the $k \times k$ identity matrix. Tables 1 and 2 provide examples.

Verification with Multinomial Distribution

The variance approximation $\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa}_w)$ in Eq. 25 has been derived by Everitt (1968) and Fleiss et al. (1969) for the special case of simple random sampling, in which each sample unit is independently classified into one and only one mutually exclusive category using each of two classifiers. In the case of simple random sampling, the multinomial or multivariate hypergeometric distributions provide the covariance matrix for $\hat{E}[\varepsilon_{ij}\varepsilon_{rs}]$ in Eq. 25. The purpose of this section is to verify that Eq. 25 includes the results of Fleiss et al. (1969) in the special case of the multinomial distribution.

The covariance matrix for the multinomial distribution is given by Ratnaparkhi (1985) as follows:

$$\hat{\mathrm{C}}\mathrm{ov}(\hat{p}_{ij}\hat{p}_{ij}) = \hat{\mathrm{V}}\mathrm{ar}(\hat{p}_{ij}) = E[\varepsilon_{ij}^2] - E^2[\varepsilon_{ij}] = \frac{\hat{p}_{ij}(1-\hat{p}_{ij})}{n} \quad [46]$$

$$\hat{\mathrm{C}}\mathrm{ov}(\hat{p}_{ij}\hat{p}_{rs|\{rs\}\neq\{ij\}}) = E[\varepsilon_{ij}\varepsilon_{rs|\{rs\}\neq\{ij\}}] - E[\varepsilon_{ij}]E[\varepsilon_{rs|\{rs\}\neq\{ij\}}] = -\frac{\hat{p}_{ij}\hat{p}_{rs}}{n}. \quad [47]$$

Equation 104 expresses these covariances in matrix form.

Replacing Eqs. 46 and 47 into the $\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa}_w)$ from Eqs. 13 and 24:

$$\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa}_w) = \left\{\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\left(\frac{\partial k_w}{\partial p_{ij}}\right)_{|p_{ij}=\hat{p}_{ij}}\right]^2 \frac{\hat{p}_{ij}(1-\hat{p}_{ij})}{n}\right\}$$

$$+\left\{\sum_{i=j}^{k}\sum_{j=1}^{k}\left[\left(\frac{\partial k_w}{\partial p_{ij}}\right)_{|p_{ij}=\hat{p}_{ij}}\right]\sum_{\substack{r=1\\ \{rs\}\neq\{ij\}}}^{k}\sum_{s=1}^{k}\left[\left(\frac{\partial k_w}{\partial p_{rs}}\right)_{|p_{rs}=\hat{p}_{rs}}\right]\frac{-\hat{p}_{ij}\hat{p}_{rs}}{n}\right\},$$

$$\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa}_w) = \left\{\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\left(\frac{\partial k_w}{\partial p_{ij}}\right)_{|p_{ij}=\hat{p}_{ij}}\right]^2 \frac{\hat{p}_{ij}}{n} - \sum_{i=1}^{k}\sum_{j=1}^{k}\left[\left(\frac{\partial k_w}{\partial p_{ij}}\right)_{|p_{ij}=\hat{p}_{ij}}\right]^2 \frac{\hat{p}_{ij}^2}{n}\right\}$$

$$-\left\{\begin{array}{l}\displaystyle\sum_{i=j}^{k}\sum_{j=1}^{k}\left[\left(\frac{\partial k_w}{\partial p_{ij}}\right)_{|p_{ij}=\hat{p}_{ij}}\right]\sum_{r=1}^{k}\sum_{s=1}^{k}\left[\left(\frac{\partial k_w}{\partial p_{rs}}\right)_{|p_{rs}=\hat{p}_{rs}}\right]\frac{\hat{p}_{ij}\hat{p}_{rs}}{n}\\ \displaystyle -\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\left(\frac{\partial k_w}{\partial p_{ij}}\right)_{|p_{ij}=\hat{p}_{ij}}\right]\frac{\hat{p}_{ij}^2}{n}\end{array}\right\},$$

$$\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa}_w) = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\left(\frac{\partial k_w}{\partial p_{ij}}\right)_{|p_{ij}=\hat{p}_{ij}}\right]^2 \hat{p}_{ij}$$

$$-\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\left(\frac{\partial k_w}{\partial p_{ij}}\right)_{|p_{ij}=\hat{p}_{ij}}\right] p_{ij} \sum_{r=1}^{k}\sum_{s=1}^{k}\left[\left(\frac{\partial k_w}{\partial p_{rs}}\right)_{|p_{rs}=\hat{p}_{rs}}\right]\hat{p}_{rs}. \quad [48]$$

The following term in $\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa}_w)$ from Eq. 48 can be simplified using the definition of p_o in Eq. 4, and the definition of p_{ij} and $p_{\cdot j}$ in Eqs. 2 and 3:

$$\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\left(\frac{\partial k_w}{\partial p_{ij}}\right)_{|p_{ij}=\hat{p}_{ij}}\right]\hat{p}_{ij} = \frac{\hat{p}_o - 1}{(1-\hat{p}_c)^2}\left[\sum_{i=1}^{k}\overline{W}_{i\cdot}\hat{p}_{i\cdot} + \sum_{j=1}^{k}\overline{w}_{\cdot j}\hat{p}_{\cdot j}\right] + \frac{\hat{p}_o}{1-\hat{p}_c}. \quad [49]$$

From Eqs. 5 and 22,

$$\sum_{i=1}^{k}\overline{w}_{i\cdot}\hat{p}_{i\cdot} = \sum_{i=1}^{k}\left(\sum_{j=1}^{k}w_{ij}\hat{p}_{\cdot j}\right)\hat{p}_{i\cdot} = \hat{p}_c, \quad [50]$$

$$\sum_{j=1}^{k} \overline{w}_{\cdot j} \hat{p}_{\cdot j} = \sum_{j=1}^{k} \left(\sum_{i=1}^{k} w_{ij} \hat{p}_{i\cdot} \right) \hat{p}_{\cdot j} = \hat{p}_c . \qquad [51]$$

Using the definition of p_c in Eqs. 50 and 51, Eq. 49 simplifies to:

$$\sum_{i=1}^{k} \sum_{j=1}^{k} \left[\left(\frac{\partial k_w}{\partial p_{ij}} \right)_{|p_{ij}=\hat{p}_{ij}} \right] \hat{p}_{ij} = \frac{\hat{p}_o \hat{p}_c - 2\hat{p}_c + \hat{p}_o}{(1-\hat{p}_c)} . \qquad [52]$$

Likewise, the following term in $\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa}_w)$ from Eq. 48 is derived directly from Eq. 24:

$$\sum_{i=1}^{k} \sum_{j=1}^{k} \left[\left(\frac{\partial k_w}{\partial p_{ij}} \right)_{|\hat{p}_{ij}=\hat{p}_{ij}} \right]^2 \hat{p}_{ij}$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{k} \hat{p}_{ij} \frac{[w_{ij}(1-\hat{p}_c) - (\overline{w}_{i\cdot} + \overline{w}_{\cdot j})(1-\hat{p}_o)]^2}{(1-\hat{p}_c)^4} . \qquad [53]$$

Substituting Eqs. 52 and 53 back into Eq. 48:

$$\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa}_w) = \frac{1}{n(1-\hat{p}_c)^4} \sum_{i=1}^{k} \sum_{j=1}^{k} \hat{p}_{ij} [w_{ij}(1-\hat{p}_c)$$

$$-(\overline{w}_i + \overline{w}_{\cdot j})(1-\hat{p}_o)]^2 - \frac{1}{n(1-\hat{p}_c)^4} (\hat{p}_o \hat{p}_c - 2\hat{p}_c + \hat{p}_o)^2 \qquad [54]$$

which agrees with the results of Fleiss et al. (1969). This partially validates the more general variance approximation $\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa}_w)$ in Eq. 24.

Likewise, $\hat{\mathrm{V}}\mathrm{ar}_o(\hat{\kappa}_w)$ in Eq. 28 can be shown to be equal to the results of Fleiss et al. (1969, Eq. 9) for the multinomial distribution using Eqs. 1 and 50 and the following identity:

$$\sum_{i=j}^{k} \sum_{j=1}^{k} \overline{w}_{i\cdot} \hat{p}_{i\cdot} \hat{p}_{\cdot j} - \sum_{i=1}^{k} \overline{w}_{i\cdot} \hat{p}_{i\cdot} \sum_{j=1}^{k} \hat{p}_{\cdot j} = \hat{p}_c . \qquad [55]$$

In the special case of the multinomial distribution, $\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa})$ in Eq. 33 agrees with Fleiss et al. (1969, Eq. 13), where Eqs. 29 and 30 and the following three identities are used in Eq. 33:

$$\sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{r=1}^{k} \sum_{s=1}^{k} \hat{p}_{ij} \hat{p}_{rs} (\hat{p}_{\cdot i} + \hat{p}_{j\cdot})(\hat{p}_{\cdot r} + \hat{p}_{s\cdot}) = 4\hat{p}_c^2$$

$$\sum_{i=1}^{k} \sum_{j=1}^{k} \hat{p}_{ii} \hat{p}_{jj} = \hat{p}_o^2$$

$$\left[\begin{array}{c} (1-\hat{p}_o)^2 \sum_{i=1}^{k} \hat{p}_{ii} (p_{\cdot i} + p_{i\cdot})^2 \\ -2(1-\hat{p}_o)(1-\hat{p}_c) \sum_{i=1}^{k} \hat{p}_{ii} (p_{\cdot i} + p_{i\cdot}) \end{array} \right]$$

$$= \left[\sum_{i=1}^{k} \hat{p}_{ii} \frac{\left[\begin{array}{c} (1-\hat{p}_c) \\ -(1-\hat{p}_o)(\hat{p}_{\cdot i} + \hat{p}_{i\cdot}) \end{array} \right]^2}{-\hat{p}_o (1-p_c)^2} \right] .$$

Examples given by Fleiss et al. (1969) and Bishop et al. (1975) were used to further validate the variance approximations, although this validation is limited by its empirical nature and use of the multinomial distribution. Results are in tables 1 and 2.

In a similar empirical evaluation, Eq. 43 for the unweighted kappa ($\hat{\kappa}_w$, $\mathbf{W} = \mathbf{I}$) agrees with the unpublished results of Stephen Stehman (personal communication) for stratified sampling in the 3×3 case when used with the covariance matrix in Eqs. 124 and 133 (after transpositions to change stratification to the column classifier as in Eq. 123 and using the finite population correction factor).

CONDITIONAL KAPPA ($\kappa_{i\cdot}$) FOR ROW *i*

Light (1971) considers the partition of the overall coefficient of agreement (κ) into a set of k partial κ statistics, each of which quantitatively describes the agreement for one category in the classification system. For example, assume that the rows of the contingency table represent the true reference classification. The "conditional kappa" ($\kappa_{i\cdot}$) is a coefficient of agreement given that the row classification is category i (Bishop et al. 1975, p. 397):

$$\kappa_{i\cdot} = \frac{p_{ii} - p_{i\cdot} p_{\cdot i}}{p_{i\cdot} - p_{i\cdot} p_{\cdot i}} . \qquad [56]$$

The Taylor series approximation of Eq. 56 is made using Eq. 10:

$$\kappa_{i\cdot} \approx \hat{\kappa}_{i\cdot} + \varepsilon_{ii} \left(\frac{\partial k_{i\cdot}}{\partial p_{ii}} \right)_{|p_{ij}=\hat{p}_{ij}}$$

$$+\varepsilon_{i1} \left(\frac{k_{i\cdot}}{\partial p_{i1}} \right)_{|p_{ij}=\hat{p}_{ij}} + \cdots + \varepsilon_{i(i-1)} \left(\frac{k_{i\cdot}}{\partial p_{i(i-1)}} \right)_{|p_{ij}=\hat{p}_{ij}}$$

$$+\varepsilon_{i(i+1)} \left(\frac{k_{i\cdot}}{\partial p_{i(i+1)}} \right)_{|p_{ij}=\hat{p}_{ij}} + \cdots + \varepsilon_{ik} \left(\frac{k_{i\cdot}}{\partial p_{ik}} \right)_{|p_{ij}=\hat{p}_{ij}}$$

$$+\varepsilon_{1j} \left(\frac{k_{i\cdot}}{\partial p_{1j}} \right)_{|p_{ij}=\hat{p}_{ij}} + \cdots + \varepsilon_{(i-1)j} \left(\frac{k_{i\cdot}}{\partial p_{(i-1)j}} \right)_{|p_{ij}=\hat{p}_{ij}}$$

$$+\varepsilon_{(i+1)j} \left(\frac{k_{i\cdot}}{\partial p_{(i+1)j}} \right)_{|p_{ij}=\hat{p}_{ij}} + \cdots + \varepsilon_{kj} \left(\frac{k_{i\cdot}}{\partial p_{kj}} \right)_{|p_{ij}=\hat{p}_{ij}} . \qquad [57]$$

p_{ij} is factored out of Eq. 56 to compute the partial derivatives in Eq. 57. First, define

$$a_{i|u} = \sum_{\substack{j=1 \\ j\neq u}}^{k} p_{ij} = p_{i\cdot} - p_{iu}, \qquad [58]$$

$$a_{\cdot i|u} = \sum_{\substack{j=1 \\ j\neq u}}^{k} p_{ji} = p_{\cdot i} - p_{ui}. \qquad [59]$$

Substituting Eqs. 58 and 59 into Eq. 56, $\kappa_{i\cdot}$ can be rewritten as a function of p_{ii} and differentiated for the first term of the Taylor series approximation in Eq. 57:

$$k_{i\cdot} = \frac{p_{ii} - (a_{ii|i} + p_{ii})(a_{\cdot i|i} + p_{ii})}{(a_{i|i} + p_{ii}) - (a_{i|i} + p_{ii})(a_{\cdot i|i} + p_{ii})}$$

$$= \frac{(-a_{i|i}a_{\cdot i|i}) + (1 - a_{\cdot i|i} - a_{i|i})p_{ii} - p_{ii}^2}{(a_{i|i} - a_{i|i}a_{\cdot i|i}) + (1 - a_{\cdot i|i} - a_{i|i})p_{ii} - p_{ii}^2}, \qquad [60]$$

$$\frac{\partial k_{i\cdot}}{\partial p_{ii}} = \frac{\begin{bmatrix} (p_{i\cdot} - p_{i\cdot}p_{\cdot i})\left[(1 - a_{i|i} - a_{i|i}) - 2p_{ii}\right] \\ -(p_{ii} - p_{i\cdot}p_{\cdot i})\left[(1 - a_{\cdot i|i} - a_{i|i}) - 2p_{ii}\right] \end{bmatrix}}{(p_{i\cdot} - p_{i\cdot}p_{\cdot i})^2}$$

$$= \frac{(p_{i\cdot} - p_{ii})(1 - p_{i\cdot} - p_{\cdot i})}{(p_{i\cdot} - p_{i\cdot}p_{\cdot i})^2} \qquad [61]$$

Similarly, $\kappa_{i\cdot}$ can be rewritten as functions of p_{ij} or p_{ji}, $i \neq j$, then differentiated for the other terms in the Taylor series approximation in Eq. 57:

$$\kappa_{i\cdot} = \frac{p_{ii} - (a_{i|j} + p_{ij})p_{\cdot i}}{(a_{i|j} + p_{ij}) - (a_{i|j} + p_{ij})p_{\cdot i}}, \qquad [62]$$

$$\frac{\partial k_{i\cdot}}{\partial p^{ij}_{j\neq i}} = \frac{\begin{bmatrix} (p_{i\cdot} - p_{i\cdot}p_{\cdot i})[-p_{\cdot i}] \\ -(p_{ii} - p_{i\cdot}p_{\cdot i})[1 - p_{\cdot i}] \end{bmatrix}}{(p_{i\cdot} - p_{i\cdot}p_{\cdot i})^2} = \frac{-p_{ii}(1 - p_{\cdot i})}{p_{i\cdot} - p_{i\cdot}p_{\cdot i})^2}, \qquad [63]$$

$$\kappa_{i\cdot} = \frac{p_{ii} - p_{i\cdot}(a_{\cdot i|j} + p_{ji})}{p_{i\cdot} - p_{i\cdot}(a_{\cdot i|j} + p_{ji})}, \qquad [64]$$

$$\frac{\partial k_{i\cdot}}{\partial p^{ji}_{j\neq i}} = \frac{\begin{bmatrix} (p_{i\cdot} - p_{i\cdot}p_{\cdot i})[-p_{i\cdot}] \\ -(p_{ii} - p_{i\cdot}p_{\cdot i}[-p_{i\cdot}] \end{bmatrix}}{(p_{i\cdot} - p_{i\cdot}p_{\cdot i})^2} = \frac{-(p_{i\cdot} - p_{ii})p_{i\cdot}}{(p_{i\cdot} - p_{i\cdot}p_{\cdot i})^2}. \qquad [65]$$

Replacing the partial derivatives in Eqs. 61, 63, and 65 which are evaluated at $p_{ij} = \hat{p}_{ij}$, into the Taylor series approximation in Eq. 57:

$$(\kappa_{i\cdot} - \hat{\kappa}_{i\cdot}) \approx \varepsilon_{ii}\frac{(\hat{p}_{i\cdot} - \hat{p}_{ii})(1 - \hat{p}_{i\cdot} - \hat{p}_{\cdot i})}{(\hat{p}_{i\cdot} - \hat{p}_{i\cdot}\hat{p}_{\cdot i})^2}$$

$$- \frac{\hat{p}_{ii}(1 - \hat{p}_{\cdot i})}{(\hat{p}_{i\cdot} - \hat{p}_{i\cdot}\hat{p}_{\cdot i})^2}\sum_{\substack{j=1 \\ j\neq i}}^{k}\varepsilon_{ij} - \frac{(\hat{p}_{i\cdot} - \hat{p}_{ii})\hat{p}_{i\cdot}}{(\hat{p}_{i\cdot} - \hat{p}_{i\cdot}\hat{p}_{\cdot i})^2}\sum_{\substack{j=1 \\ j\neq i}}^{k}\varepsilon_{ji} \qquad [66]$$

$$(\kappa_{i\cdot} - \hat{\kappa}_{i\cdot})^2 \approx$$

$$\frac{(\hat{p}_{i\cdot} - \hat{p}_{ii})^2(1 - \hat{p}_{i\cdot} - \hat{p}_{\cdot i})^2}{(\hat{p}_{i\cdot} - \hat{p}_{i\cdot}\hat{p}_{\cdot i})^4}\varepsilon_{ii}^2$$

$$+ \frac{\hat{p}_{ii}^2(1 - \hat{p}_{\cdot i})^2}{(\hat{p}_{i\cdot} - \hat{p}_{i\cdot}\hat{p}_{\cdot i})^4}\sum_{\substack{j=1 \\ j\neq i}}^{k}\varepsilon_{ij}\sum_{\substack{s=1 \\ s\neq i}}^{k}\varepsilon_{is}$$

$$+ \frac{(\hat{p}_{i\cdot} - \hat{p}_{ii})^2\hat{p}_{i\cdot}^2}{(\hat{p}_{i\cdot} - \hat{p}_{i\cdot}\hat{p}_{\cdot i})^4}\sum_{\substack{j=1 \\ j\neq i}}^{k}\varepsilon_{ji}\sum_{\substack{s=1 \\ s\neq i}}^{k}\varepsilon_{si}$$

$$- \frac{(\hat{p}_{i\cdot} - \hat{p}_{ii})(1 - \hat{p}_{i\cdot} - \hat{p}_{\cdot i})\hat{p}_{ii}(1 - \hat{p}_{\cdot i})}{(\hat{p}_{i\cdot} - \hat{p}_{i\cdot}\hat{p}_{\cdot i})^4}\varepsilon_{ii}\left(\sum_{\substack{j=1 \\ j\neq i}}^{k}\varepsilon_{ij} + \sum_{\substack{s=1 \\ s\neq i}}^{k}\varepsilon_{is}\right)$$

$$- \frac{(1 - \hat{p}_{i\cdot} - \hat{p}_{\cdot i})(\hat{p}_{i\cdot} - \hat{p}_{ii})^2\hat{p}_{i\cdot}}{(\hat{p}_{i\cdot} - \hat{p}_{i\cdot}\hat{p}_{\cdot i})^4}\varepsilon_{ii}\left(\sum_{\substack{j=1 \\ j\neq i}}^{k}\varepsilon_{ji} + \sum_{\substack{s=1 \\ s\neq 1}}^{k}\varepsilon_{si}\right)$$

$$+ \frac{\hat{p}_{ii}(1 - \hat{p}_{\cdot i})(\hat{p}_{i\cdot} - \hat{p}_{ii})\hat{p}_{i\cdot}}{(\hat{p}_{i\cdot} - \hat{p}_{i\cdot}\hat{p}_{\cdot i})^4}\sum_{\substack{j=1 \\ j\neq i}}^{k}\varepsilon_{ij}\sum_{\substack{s=1 \\ s\neq i}}^{k}\varepsilon_{si}$$

$$+ \frac{\hat{p}_{ii}(1 - \hat{p}_{\cdot i})(\hat{p}_{i\cdot} - \hat{p}_{ii})\hat{p}_{i\cdot}}{(\hat{p}_{i\cdot} - \hat{p}_{i\cdot}\hat{p}_{\cdot i})^4}\sum_{\substack{j=1 \\ j\neq i}}^{k}\varepsilon_{ji}\sum_{\substack{s=1 \\ s\neq i}}^{k}\varepsilon_{is} \qquad [67]$$

[67] continued on next page

[67] *continued*

$$\hat{\mathrm{Var}}(\hat{\kappa}_{i\cdot}) = E\left[(\kappa_{i\cdot} - \hat{\kappa}_{i\cdot})^2\right] \approx \frac{(\hat{p}_{i\cdot} - \hat{p}_{ii})^2 (1 - \hat{p}_{i\cdot} - \hat{p}_{\cdot i})^2}{(\hat{p}_{i\cdot} - \hat{p}_{i\cdot}\hat{p}_{\cdot i})^4} \hat{E}[\varepsilon_{ii}^2]$$

$$+ \frac{\hat{p}_{ii}^2}{\hat{p}_{i\cdot}^4 (1 - \hat{p}_{\cdot i})^2} \sum_{\substack{j=1 \\ j \neq i}}^{k} \sum_{\substack{s=1 \\ s \neq i}}^{k} \hat{E}[\varepsilon_{ij}\varepsilon_{is}]$$

$$+ \frac{(\hat{p}_{i\cdot} - \hat{p}_{ii})^2}{\hat{p}_{i\cdot}^2 (1 - \hat{p}_{\cdot i})^4} \sum_{\substack{j=1 \\ j \neq i}}^{k} \sum_{\substack{s=1 \\ s \neq i}}^{k} \hat{E}[\varepsilon_{ij}\varepsilon_{si}]$$

$$-2 \frac{(\hat{p}_{i\cdot} - \hat{p}_{ii})(1 - \hat{p}_{i\cdot} - \hat{p}_{\cdot i})\hat{p}_{ii}}{\hat{p}_{i\cdot}^4 (1 - \hat{p}_{\cdot i})^3} \sum_{\substack{j=1 \\ j \neq i}}^{k} \hat{E}[\varepsilon_{ii}\varepsilon_{ij}]$$

$$-2 \frac{(1 - \hat{p}_{i\cdot} - \hat{p}_{\cdot i})(\hat{p}_{i\cdot} - \hat{p}_{ii})^2}{\hat{p}_{i\cdot}^3 (1 - \hat{p}_{\cdot i})^4} \sum_{\substack{j=1 \\ j \neq i}}^{k} \hat{E}[\varepsilon_{ii}\varepsilon_{ji}]$$

$$+2 \frac{\hat{p}_{ii}(\hat{p}_{i\cdot} - \hat{p}_{ii})}{\hat{p}_{i\cdot}^3 (1 - \hat{p}_{\cdot i})^3} \sum_{\substack{j=1 \\ j \neq i}}^{k} \sum_{\substack{s=1 \\ s \neq i}}^{k} \hat{E}[\varepsilon_{ij}\varepsilon_{si}].$$

The validity of the approximation in Eq. 67 was partially checked using the example provided by Bishop et al. (1975, p. 398); the results are in table 3. For example, $\hat{\kappa}_{3\cdot}$ is 0.2941 in table 3, which agrees with $\hat{\kappa}_3$ in Bishop et al. The 95% confidence interval is $0.2941 \pm (1.96 \times \sqrt{0.0122})$ in table 3, which agrees with the interval [0.078, 0.510] in Bishop et al.

The variance under the null hypothesis of independence between the row and column classifiers, $\hat{\mathrm{Var}}_o(\hat{\kappa}_{i\cdot})$, assumes that $p_{ii} = p_{i\cdot}p_{\cdot i}$. To compute $\hat{\mathrm{Var}}_o(\hat{\kappa}_{i\cdot})$, substitute $p_{ii} = p_{i\cdot}p_{\cdot i}$ into Eq. 67, and use the variance under the assumption $E[p_{ij}] = p_{i\cdot}p_{\cdot j}$ (see Eqs. 113, 114, and 117). An example of $\hat{\mathrm{Var}}_o(\hat{\kappa}_{i\cdot})$ is given in table 3.

Conditional Kappa ($\kappa_{\cdot i}$) for Column *i*

The kappa conditioned on the *i*th column rather than the *i*th row (Eq. 56) is defined as:

$$\kappa_{\cdot i} = \frac{p_{ii} - p_{i\cdot}p_{\cdot i}}{p_{\cdot i} - p_{i\cdot}p_{\cdot i}}. \qquad [68]$$

The Taylor series approximation of Eq. 68 is derived similar to Eq. 66:

$$(\kappa_{\cdot i} - \hat{\kappa}_{\cdot i}) \approx \varepsilon_{ii} \frac{(\hat{p}_{\cdot i} - \hat{p}_{ii})(1 - \hat{p}_{i\cdot} - \hat{p}_{\cdot i})}{(\hat{p}_{\cdot i} - \hat{p}_{i\cdot}\hat{p}_{\cdot i})}$$

$$- \frac{(\hat{p}_{\cdot i} - \hat{p}_{ii})\hat{p}_{\cdot i}}{(\hat{p}_{\cdot i} - \hat{p}_{i\cdot}\hat{p}_{\cdot i})^2} \sum_{\substack{j=1 \\ j \neq i}}^{k} \varepsilon_{ij} - \frac{\hat{p}_{ii}(1 - \hat{p}_{i\cdot})}{(\hat{p}_{\cdot i} - \hat{p}_{i\cdot}\hat{p}_{\cdot i})^2} \sum_{\substack{j=1 \\ j \neq i}}^{k} \varepsilon_{ji}. \qquad [69]$$

Equation 69 is used to derive $\hat{\mathrm{Var}}(\hat{\kappa}_{\cdot i})$ similar to Eq. 67:

$$\hat{\mathrm{Var}}(\hat{\kappa}_{\cdot i}) = E[(\kappa_{\cdot i} - \hat{\kappa}_{\cdot i})^2] \approx \frac{(\hat{p}_{\cdot i} - \hat{p}_{ii})^2 (1 - \hat{p}_{i\cdot} - \hat{p}_{\cdot i})^2}{(\hat{p}_{\cdot i} - \hat{p}_{i\cdot}\hat{p}_{\cdot i})^4} \hat{E}[\varepsilon_{ii}^2]$$

$$+ \frac{(\hat{p}_{\cdot i} - \hat{p}_{ii})^2}{\hat{p}_{\cdot i}^2 (1 - \hat{p}_{i\cdot})^4} \sum_{\substack{j=1 \\ j \neq i}}^{k} \sum_{\substack{s=1 \\ s \neq i}}^{k} \hat{E}[\varepsilon_{ij}\varepsilon_{is}]$$

$$+ \frac{\hat{p}_{ii}^2}{\hat{p}_{\cdot i}^4 (1 - \hat{p}_{i\cdot})^2} \sum_{\substack{j=1 \\ j \neq i}}^{k} \sum_{\substack{s=1 \\ s \neq i}}^{k} \hat{E}[\varepsilon_{ji}\varepsilon_{si}]$$

$$-2 \frac{(1 - \hat{p}_{i\cdot} - \hat{p}_{\cdot i})(\hat{p}_{\cdot i} - \hat{p}_{ii})^2}{\hat{p}_{\cdot i}^3 (1 - \hat{p}_{i\cdot})^4} \sum_{\substack{j=1 \\ j \neq i}}^{k} \hat{E}[\varepsilon_{ii}\varepsilon_{ij}]$$

$$-2 \frac{(\hat{p}_{\cdot i} - \hat{p}_{ii})(1 - \hat{p}_{i\cdot} - \hat{p}_{\cdot i})\hat{p}_{ii}}{\hat{p}_{\cdot i}^4 (1 - \hat{p}_{i\cdot})^3} \sum_{\substack{j=1 \\ j \neq i}}^{k} \hat{E}[\varepsilon_{ii}\varepsilon_{ji}]$$

$$+2 \frac{\hat{p}_{ii}(\hat{p}_{\cdot i} - \hat{p}_{ii})}{\hat{p}_{\cdot i}^3 (1 - \hat{p}_{i\cdot})^3} \sum_{\substack{j=1 \\ j \neq i}}^{k} \sum_{\substack{s=1 \\ s \neq i}}^{k} \hat{E}[\varepsilon_{ij}\varepsilon_{si}]. \qquad [70]$$

The variance under the null hypothesis of independence between the row and column classifiers, $\hat{\mathrm{Var}}_o(\hat{\kappa}_{\cdot i})$, assumes that $p_{ii} = p_{i\cdot}p_{\cdot i}$. To compute $\hat{\mathrm{Var}}_o(\hat{\kappa}_{\cdot i})$, substitute $p_{ii} = p_{i\cdot}p_{\cdot i}$ into Eq. 70, and use the variance under the assumption $E[p_{ij}] = p_{i\cdot}p_{\cdot j}$ (see Eqs. 113, 114, and 117). The validity of this approximation was partially checked by using $\hat{\mathbf{p}}'$ in the example provided by Bishop et al. (1975); the results are in table 4.

Matrix Formulation of $\hat{\mathrm{Var}}(\hat{\kappa}_{i\cdot})$ and $\hat{\mathrm{Var}}(\hat{\kappa}_{\cdot i})$

The formulae above can be expressed in matrix algebra, which facilitates numerical implementation with matrix algebra software. The $p_{i\cdot}$ and $p_{\cdot i}$ terms that define $\hat{\kappa}_i$ in Eq. 56 are computed with Eqs. 2 and 3 or matrix Eqs. 35 and 36. The linear approximation of

$\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa}_{i\cdot})$ in Eq. 67 uses the following terms. First, define the diagonal $k \times k$ matrix $\mathbf{H}_{i\cdot}$ using the definitions of $p_{i\cdot}$ and $p_{\cdot i}$ in Eqs. 2 and 3, in which all elements equal zero except for the diagonal:

$$(H_{i\cdot})_{ii} = \frac{-\hat{p}_{ii}}{\hat{p}_{i\cdot}^2(1-\hat{p}_{\cdot i})}, \qquad [71]$$

which corresponds to the second term in Eq. 66. An example of $\mathbf{H}_{i\cdot}$ is given in table 3. Define the $k \times k$ matrix $\mathbf{M}_{i\cdot}$, in which all elements equal zero except for the ith column:

$$(M_{i\cdot})_{ji} = \frac{-(\hat{p}_{i\cdot} - \hat{p}_{ii})}{\hat{p}_{i\cdot}(1-\hat{p}_{\cdot i})^2}, 1 \le j \le n, \qquad [72]$$

which corresponds to the third term in Eq. 66. An example of $\mathbf{M}_{i\cdot}$ is given in table 3. Define the $k \times k$ matrix $\mathbf{G}_{i\cdot}$, in which all elements are zero except for the iith element:

$$(G_{i\cdot})_{ii} = \frac{1}{\hat{p}_{i\cdot}(1-\hat{p}_{\cdot i})}, \qquad [73]$$

which corresponds to the first term in Eq. 66 plus $abs(H_{i\cdot})_{ii}$ in Eq. 71 plus $abs(M_{i\cdot})_{ii}$ in Eq. 72. An example of $\mathbf{G}_{i\cdot}$ is given in table 3.

Table 3. — Example data[1] from Bishop et al. (1975) for conditional kappa ($\hat{\kappa}_{i\cdot}$), conditioned on the row classifier (i), including vectors used in matrix formulation. Contingency table is given in table 2.

Vectors used in matrix computations

$\mathbf{G}_{i\cdot}$ (Eq. 73)			$\mathbf{H}_{i\cdot}$ (Eq. 71)			$\mathbf{M}_{i\cdot}$ (Eq. 72)		
j=1	j=2	j=3	j=1	j=2	j=3	j=1	j=2	j=3
4.4690	0	0	−2.6197	0	0	−1.3407	0	0
0	0	0	0	−4.0614	0	−1.3407	0	0
0	0	0	0	0	−1.9548	−1.3407	0	0
0	0	0	−2.6197	0	0	0	−0.5428	0
0	5.7536	0	0	−4.0614	0	0	−0.5428	0
0	0	0	0	0	−1.9548	0	−0.5428	0
0	0	0	−2.6197	0	0	0	0	−0.9965
0	0	0	0	−4.0614	0	0	0	−0.9965
0	0	3.9095	0	0	−1.9548	0	0	−0.9965

Vectors used in matrix computations, null hypothesis $\hat{\kappa}_{i\cdot} = 0$

$\mathbf{G}_{i\cdot}$ (Eq. 73)			$\mathbf{H}_{i\cdot}$ (Eq. 71, $\hat{p}_{ii} = \hat{p}_{i\cdot}\hat{p}_{\cdot i}$)			$\mathbf{M}_{i\cdot}$ (Eq. 72, $\hat{p}_{ii} = \hat{p}_{i\cdot}\hat{p}_{\cdot i}$)		
j=1	j=2	j=3	j=1	j=2	j=3	j=1	j=2	j=3
4.4690	0	0	−1.9862	0	0	−1.8000	0	0
0	0	0	0	−1.5183	0	−1.8000	0	0
0	0	0	0	0	−1.1403	−1.8000	0	0
0	0	0	−1.9862	0	0	0	−1.3585	0
0	5.7536	0	0	−1.5183	0	0	−1.3585	0
0	0	0	0	0	−1.1403	0	−1.3585	0
0	0	0	−1.9862	0	0	0	0	−1.4118
0	0	0	0	−1.5183	0	0	0	−1.4118
0	0	3.9095	0	0	−1.1403	0	0	−1.4118

Resulting statistics

		$Cov(\hat{\kappa}_{i\cdot})$			$Cov(\hat{\kappa}_{i\cdot})$		
i	$\hat{\kappa}_{i\cdot}$	j=1	j=2	j=3	j=1	j=2	j=3
1	0.2551	0.0170	0.0052	0.0067	0.0165	0.0040	0.0046
2	0.6004	0.0052	0.0191	0.0000	0.0040	0.0161	0.0021
3	0.2941	0.0067	0.0000	0.0122	0.0046	0.0021	0.0101

[1] *The covariance matrix for the estimated joint probabilities ($\hat{p}_{ij}$) is estimated assuming the multinomial distribution (see Eqs. 46, 47, 128, 130, 131, and 132).*

The linear approximation of $\kappa_{i\cdot}$ equals $(vec\mathbf{P})'\mathbf{d}_{k_{i\cdot}}$, where the $k^2 \times k$ matrix $\mathbf{d}_{k_{i\cdot}}$ equals:

$$\mathbf{d}_{k_{i\cdot}} = \begin{bmatrix} \mathbf{G}_{1\cdot} \\ \mathbf{G}_{2\cdot} \\ \vdots \\ \mathbf{G}_{k\cdot} \end{bmatrix} + \begin{bmatrix} \mathbf{H}_{i\cdot} \\ \mathbf{H}_{i\cdot} \\ \vdots \\ \mathbf{H}_{i\cdot} \end{bmatrix} + \begin{bmatrix} \mathbf{M}_{1\cdot} \\ \mathbf{M}_{2\cdot} \\ \vdots \\ \mathbf{M}_{k\cdot} \end{bmatrix}. \quad [74]$$

An example of $\mathbf{d}_{k_{i\cdot}}$ is given in table 3. The $k \times k$ covariance matrix for the $k \times 1$ vector of conditional kappa statistics $\hat{\kappa}_{i\cdot}$ equals:

$$\hat{\mathbf{Cov}}(\hat{\kappa}_{i\cdot}) = \mathbf{d}'_{k_{i\cdot}} \hat{\mathbf{Cov}}(vec\hat{P})\, \mathbf{d}_{k_{i\cdot}}. \quad [75]$$

See Eqs. 104 and 105 for examples of $\hat{\mathbf{Cov}}(vec\hat{\mathbf{P}})$. An example of $\hat{\mathbf{Cov}}(\hat{\kappa}_{i\cdot})$ is given in table 3. The estimated variance for each $\hat{\kappa}_{i\cdot}$ equals the corresponding diagonal element of $\hat{\mathbf{Cov}}(\hat{\kappa}_{i\cdot})$ in Eq. 75. As in the case of $\hat{\kappa}_w$, better approximations of $\mathbf{d}_{k_{i\cdot}}$ in $\kappa_{i\cdot} \approx (vec\mathbf{P})'\mathbf{d}_{k_{i\cdot}}$ might lead to better approximations of $\hat{\mathbf{Cov}}(\hat{\kappa}_{i\cdot})$.

To compute the covariance matrix under the null hypothesis of independence between the row and column classifiers, $\mathbf{Cov}_o(\hat{\kappa}_{i\cdot})$, substitute $\hat{p}_{ii} = \hat{p}_{i\cdot}\hat{p}_{\cdot i}$ in Eqs. 71, 72, 74 and 75; and use the variance under the as-

Table 4. — Example data[1] from Bishop et al. (1975) for conditional kappa ($\hat{k}_{i}$), conditioned on the column classifier, including vectors used in matrix formulation. Contingency table is given in table .

Vectors used in matrix computations

$\mathbf{G}_{i\cdot}$ (Eq. 78)			$\mathbf{H}_{i\cdot}$ (Eq. 76)			$\mathbf{M}_{i\cdot}$ (Eq. 77)		
j=1	*j*=2	*j*=3	*j*=1	*j*=2	*j*=3	*j*=1	*j*=2	*j*=3
3.7674	0	0	−1.3142	0	0	−2.0015	0	0
0	0	0	0	−0.6314	0	−2.0015	0	0
0	0	0	0	0	−0.9333	−2.0015	0	0
0	0	0	−1.3142	0	0	0	−3.1331	0
0	4.9608	0	0	−0.6314	0	0	−3.1331	0
0	0	0	0	0	−0.9333	0	−3.1331	0
0	0	0	−1.3142	0	0	0	0	−3.3221
0	0	0	0	−0.6314	0	0	0	−3.3221
0	0	5.3665	0	0	−0.9333	0	0	−3.3221

Vectors used in matrix computations, null hypothesis $k_{i\cdot}=0$

$\mathbf{G}_{i\cdot}$ (Eq. 73)			$\mathbf{H}_{i\cdot}$ (Eq. 71, $\hat{p}_{ii} = \hat{p}_{i\cdot}\hat{p}_{\cdot i}$)			$\mathbf{M}_{i\cdot}$ (Eq. 72, $\hat{p}_{ii} = \hat{p}_{i\cdot}\hat{p}_{\cdot i}$)		
j=1	*j*=2	*j*=3	*j*=1	*j*=2	*j*=3	*j*=1	*j*=2	*j*=3
3.7674	0	0	−1.6744	0	0	−1.5174	0	0
0	0	0	0	−1.3091	0	−1.5174	0	0
0	0	0	0	0	−1.5652	−1.5174	0	0
0	0	0	−1.6744	0	0	0	−1.1713	0
0	4.9608	0	0	−1.3091	0	0	−1.1713	0
0	0	0	0	0	−1.5652	0	−1.1713	0
0	0	0	−1.6744	0	0	0	0	−1.9379
0	0	0	0	−1.3091	0	0	0	−1.9379
0	0	5.3665	0	0	−1.5652	0	0	−1.9379

Resulting statistics

		$Cov(\hat{k}_{i\cdot})$			$Cov_o(\hat{k}_{i\cdot})$		
i	$\hat{\kappa}_{i\cdot}$	*j*=1	*j*=2	*j*=3	*j*=1	*j*=2	*j*=3
1	0.2151	0.0124	0.0033	0.0079	0.0117	0.0029	0.0053
2	0.5177	0.0033	0.0166	0.0010	0.0029	0.0120	0.0024
3	0.4037	0.0079	0.0010	0.0207	0.0053	0.0024	0.0191

[1] *The covariance matrix for the estimated joint probabilities ($\hat{p}_{ij}$) is estimated assuming the multinomial distribution (see Eqs. 46, 47, 128, 130, 131, and 132).*

sumption $E[p_{ij}] = p_{i\cdot}p_{\cdot j}$ (see Eqs. 113, 114, and 117). An example of $\mathbf{Cov}_o(\hat{\kappa}_{\cdot i})$ is given in table 3.

The linear approximation of $\hat{\mathrm{Var}}(\hat{\kappa}_{\cdot i})$ in Eq. 70, which is used to estimate the precision of the kappa conditioned on the column classifier ($\hat{\kappa}_{\cdot i}$), can also be expressed in matrix algebra. As in Eq. 71, define the diagonal $k \times k$ matrix $\mathbf{H}_{\cdot i}$, in which all elements equal zero except for the diagonal:

$$(H_{\cdot i})_{ii} = -\frac{(\hat{p}_{\cdot i} - \hat{p}_{ii})}{\hat{p}_{\cdot i}(1-\hat{p}_{i\cdot})^2}, \qquad [76]$$

which corresponds to the second term in Eq. 70. An example of $\mathbf{H}_{\cdot i}$ is given in table 4. As in Eq. 72, define the $k \times k$ matrix $\mathbf{M}_{\cdot i}$, in which all elements equal zero except for the ith column:

$$(M_{\cdot i})_{ji} = -\frac{\hat{p}_{ii}}{\hat{p}_{\cdot i}^2(1-\hat{p}_{i\cdot})}, 1 \le j \le n, \qquad [77]$$

which corresponds to the third term in Eq. 70. An example of $\mathbf{M}_{\cdot i}$ is given in table 4. As in Eq. 73, define the $k \times k$ matrix $\mathbf{G}_{\cdot i}$, in which all elements are zero except for the iith element:

$$(G_{\cdot i})_{ii} = \frac{1}{\hat{p}_{\cdot i}(1-\hat{p}_{i\cdot})}, \qquad [78]$$

which corresponds to the first term in Eq. 70 plus $abs(H_{\cdot i})_{ii}$ in Eq. 76 plus $abs(M_{\cdot i})_{ii}$ in Eq. 77. An example of $\mathbf{G}_{\cdot i}$ is given in table 4.

The linear approximation of $\kappa_{\cdot i}$ equals $(vec\mathbf{P})'\mathbf{d}_{k\cdot i}$, where the $k^2 \times k$ matrix $\mathbf{d}_{k\cdot i}$ equals:

$$\mathbf{d}_{k\cdot i} = \begin{bmatrix} \mathbf{G}_{\cdot 1} \\ \mathbf{G}_{\cdot 2} \\ \vdots \\ \mathbf{G}_{\cdot k} \end{bmatrix} + \begin{bmatrix} \mathbf{H}_{\cdot i} \\ \mathbf{H}_{\cdot i} \\ \vdots \\ \mathbf{H}_{\cdot i} \end{bmatrix} + \begin{bmatrix} \mathbf{M}_{\cdot 1} \\ \mathbf{M}_{\cdot 2} \\ \vdots \\ \mathbf{M}_{\cdot 3} \end{bmatrix}. \qquad [79]$$

An example of $\mathbf{d}_{k\cdot i}$ is given in table 4. The $k \times k$ covariance matrix for the $k \times 1$ vector of conditional kappa statistics $\hat{\kappa}_{\cdot i}$ equals:

$$\mathbf{Cov}(\hat{\kappa}_{\cdot i}) = \mathbf{d}'_{k\cdot i}\left(\hat{\mathbf{Cov}}(vec\hat{\mathbf{P}})\right)\mathbf{d}_{k\cdot i}. \qquad [80]$$

See Eqs. 104 and 105 for examples of $\hat{\mathbf{Cov}}(vec\hat{\mathbf{P}})$. An example of $\mathbf{Cov}(\hat{\kappa}_{\cdot i})$ is given in table 4. The estimated variance for each $\hat{\kappa}_{\cdot i}$ equals the corresponding diagonal element of $\mathbf{Cov}(\hat{\kappa}_{\cdot i})$ in Eq. 80. To compute the covariance matrix under the null hypothesis of independence between the row and column classifiers, $\mathbf{Cov}_o(\hat{\kappa}_{\cdot i})$, substitute $\hat{p}_{ii} = \hat{p}_{i\cdot}\hat{p}_{\cdot i}$ into Eqs. 76, 77, 78, 79, and 80; and use the variance under the assumption $E[p_{ij}] = p_{i\cdot}p_{\cdot j}$ (see Eqs. 113, 114, and 117). An example of $\mathbf{Cov}_o(\hat{\kappa}_{\cdot i})$ is given in table 4.

The off-diagonal elements of $\mathbf{Cov}(\hat{\kappa}_{\cdot i})$ and $\mathbf{Cov}(\hat{\kappa}_{\cdot i})$ can be used to estimate precision of differences between partial kappa statistics from the same error matrix (**P**). The variance of this difference is used to test the hypothesis that the difference between $\hat{\kappa}_1$ and $\hat{\kappa}_2$ is zero, i.e., the conditional kappas are the same, and hence, the accuracy in classifying objects into categories i=1 and i=2 is the same. Table 3 provides an example. The difference between $\hat{\kappa}_1$ and $\hat{\kappa}_2$ is (0.2551–0.6004) = –0.3453; the variance of this estimated difference is 0.0170+0.0191+(2×0.0052) = 0.0465; the standard deviation is 0.2156; and the 95% confidence interval is $-0.3453 \pm (1.96 \times 0.2156) = [-0.7679, 0.0773]$. Since this interval contains zero, we fail to reject (at the 95% level) the null hypothesis that the two classifiers have the same agreement when the row classification is category i=1 or i=2. This test might have limited power to detect true differences in accuracy for specific categories.

CONDITIONAL PROBABILITIES

Fleiss (1981, p. 214) gives an example of assessing classification accuracy for individual categories with conditional probabilities. An example of a conditional probability is the probability of correctly classifying a member of the population (e.g., a pixel) as forest given that the pixel is classified as forest with remote sensing. Let $p_{(i|\cdot j)}$ represent the conditional probability that the row classification is category i given that the column classification is category j; in this case:

$$p_{(i|\cdot j)} = \frac{p_{ij}}{p_{\cdot j}}. \qquad [81]$$

The variance for an estimate of $p_{(i|\cdot j)}$ can be approximated with the Taylor series expansion as in Eq. 57. First, p_{ri} is factored out of Eq. 81 using Eq. 59 so that the partial derivatives can be computed:

$$p_{(i|\cdot j)} = \frac{p_{ij}}{a_{\cdot j|r} + p_{ri}}, 1 \le r \le k. \qquad [82]$$

The partial derivative of Eq. 82 with respect to p_{ri} is:

$$\left(\frac{\partial p_{(i|\cdot j)}}{\partial p_{ri}}\right)_{|p_{ri}=\hat{p}_{ri}} = \frac{\hat{p}_{\cdot j} - \hat{p}_{ij}}{(\hat{a}_{\cdot j|i} + \hat{p}_{ij})^2} = \frac{\hat{p}_{\cdot j} - \hat{p}_{ij}}{\hat{p}_{\cdot j}^2}, r = i,$$

$$\left(\frac{\partial p_{(i|\cdot j)}}{\partial p_{ri}}\right)_{|p_{ri}=\hat{p}_{ri}} = \frac{\hat{p}_{\cdot j}}{(\hat{a}_{\cdot j|r} + \hat{p}_{ri})^2} = -\frac{\hat{p}_{ij}}{\hat{p}_{\cdot j}^2}, r \ne i. \qquad [83]$$

The Taylor series approximation of Var ($p_{(i|\cdot j)}$) is made similar to Eq. 57 for $\hat{\mathrm{Var}}(\hat{\kappa}_{\cdot i})$:

$$\varepsilon_{p(i|\cdot j)} \approx \sum_{r=1}^{k} \varepsilon_{ri}\left(\frac{\partial p_{(i|\cdot j)}}{\partial p_{ri}}\right)_{|p_{ri}=\hat{p}_{ri}} = \varepsilon_{ij}\left(\frac{\hat{p}_{\cdot j} - \hat{p}_{ij}}{\hat{p}_{\cdot j}^2}\right) + \sum_{\substack{r=1 \\ r \ne i}}^{k} \varepsilon_{ri}\left(\frac{-\hat{p}_{ij}}{\hat{p}_{\cdot j}^2}\right) \qquad [84]$$

$$\varepsilon^2_{p_{(i|\cdot j)}} \approx \left(\frac{\hat{p}_{\cdot j} - \hat{p}_{ij}}{\hat{p}^2_{\cdot j}} \varepsilon_{ij} - \frac{\hat{p}_{ij}}{\hat{p}^2_{\cdot j}} \sum_{\substack{r=1 \\ r \neq i}}^{k} \varepsilon_{rj} \right) \left(\frac{\hat{p}_{\cdot j} - \hat{p}_{ij}}{\hat{p}^2_{\cdot j}} \varepsilon_{ij} - \frac{\hat{p}_{ij}}{\hat{p}^2_{\cdot j}} \sum_{\substack{s=1 \\ s \neq i}}^{k} \varepsilon_{sj} \right)$$

$$\hat{\mathrm{Var}}(\hat{p}_{i|\cdot j}) = \frac{(\hat{p}_{\cdot j} - \hat{p}_{ij})^2}{\hat{p}^4_{\cdot j}} \hat{E}[\varepsilon^2_{ij}] - 2\frac{(\hat{p}_{\cdot j} - _{ij})\hat{p}_{ij}}{\hat{p}^4_{\cdot j}} \sum_{\substack{r=1 \\ r \neq i}}^{k} \hat{E}[\varepsilon_{ij}\varepsilon_{rj}] + \frac{\hat{p}^2_{ij}}{\hat{p}^4_{\cdot j}} \sum_{\substack{r=1 \\ r \neq i}}^{k} \sum_{\substack{s=1 \\ s \neq i}}^{k} \hat{E}[\varepsilon_{rj}\varepsilon_{sj}]. \quad [85]$$

Conditional probabilities that are conditioned on the row classification, rather than the column classification, are also useful. Let $p_{(i|j\cdot)}$ represent the conditional probability that the column classification is category i given that the row classification is category j; in this case:

$$p_{(i|j\cdot)} = \frac{p_{ji}}{p_{j\cdot}}. \quad [86]$$

The variance for an estimate of $p_{(i|\cdot j)}$ can be approximated with the Taylor series expansion as in Eqs. 82 to 85:

$$\hat{\mathrm{Var}}(\hat{p}_{i|j\cdot}) = \frac{(\hat{p}_{j\cdot} - \hat{p}_{ji})^2}{\hat{p}^4_{j\cdot}} \hat{E}[\varepsilon^2_{ji}] - 2\frac{(\hat{p}_{j\cdot} - \hat{p}_{ji})\hat{p}_{ji}}{\hat{p}^4_{j\cdot}} \sum_{\substack{r=1 \\ r \neq i}}^{k} \hat{E}[\varepsilon_{ji}\varepsilon_{jr}] + \frac{\hat{p}^2_{ji}}{\hat{p}^4_{\cdot i}} \sum_{\substack{r=1 \\ r \neq i}}^{k} \sum_{\substack{s=1 \\ s \neq i}}^{k} \hat{E}[\varepsilon_{jr}\varepsilon_{js}]. \quad [87]$$

Of special interest is the case in which $i = j$, i.e., the conditional probabilities on the diagonal of the error matrix ($\hat{\mathbf{P}}$). In this case,

$$\hat{\mathrm{Var}}(\hat{p}_{i|\cdot i}) = \frac{(\hat{p}_{\cdot i} - \hat{p}_{ii})^2}{\hat{p}^4_{\cdot i}} \hat{E}[\varepsilon^2_{ii}] - 2\frac{(\hat{p}_{\cdot i} - \hat{p}_{ii})\hat{p}_{ii}}{\hat{p}^4_{\cdot i}} \sum_{\substack{r=1 \\ r \neq i}}^{k} \hat{E}[\varepsilon_{ii}\varepsilon_{ri}] + \frac{\hat{p}^2_{ii}}{\hat{p}^4_{i\cdot}} \sum_{\substack{r=1 \\ r \neq i}}^{k} \sum_{\substack{s=1 \\ s \neq i}}^{k} \hat{E}[\varepsilon_{ir}\varepsilon_{is}], \quad [88]$$

$$\hat{\mathrm{Var}}(\hat{p}_{i|i\cdot}) = \frac{(\hat{p}_{i\cdot} - \hat{p}_{ii})^2}{\hat{p}^4_{i\cdot}} \hat{E}[\varepsilon^2_{ii}] - 2\frac{(\hat{p}_{i\cdot} - \hat{p}_{ii})\hat{p}_{ii}}{\hat{p}^4_{i\cdot}} \sum_{\substack{r=1 \\ r \neq i}}^{k} \hat{E}[\varepsilon_{ii}\varepsilon_{ir}] + \frac{\hat{p}^2_{ii}}{\hat{p}^4_{i\cdot}} \sum_{\substack{r=1 \\ r \neq i}}^{k} \sum_{\substack{s=1 \\ s \neq i}}^{k} \hat{E}[\varepsilon_{ir}\varepsilon_{is}]. \quad [89]$$

Matrix Formulation for $\hat{\mathrm{Var}}(\hat{p}_{i|\cdot i})$ and $\hat{\mathrm{Var}}(\hat{p}_{i|i\cdot})$

Equation 88 can be expressed in matrix algebra as follows. First, define the $k \times k$ matrix $\mathbf{H}_{p(\cdot i)}$, in which all elements equal zero except for the ith column:

$$(H_{p(\cdot i)})_{ri} = \frac{\hat{p}_{ii}}{\hat{p}^2_{\cdot i}}, 1 \leq r \leq k. \quad [90]$$

Equation 90 corresponds to the second term in Eq. 84, and an example is given in table 5. Define the $k \times k$ matrix $\mathbf{G}_{p(\cdot i)'}$, in which all elements are zero except for the iith element:

$$(G_{p(\cdot i)})_{ii} = \frac{1}{\hat{p}_{\cdot i}}. \quad [91]$$

Equation 91 corresponds to the first term in Eq. 84, and an example is given in table 5. The linear approximation of $\mathbf{p}_{(i|\cdot i)}$ equals $(vec\mathbf{P})'\mathbf{D}_{(i|\cdot i)'}$, where $\mathbf{p}_{(i|\cdot i)}$ is the $k \times 1$ vector of diagonal conditional probabilities with its ith element equal to $\mathbf{p}_{(i|\cdot i)}$. The $k^2 \times k$ matrix $\mathbf{D}_{(i|\cdot i)}$ equals:

$$\mathbf{D}_{(i|\cdot i)} = \begin{bmatrix} \mathbf{G}_{p(\cdot 1)} \\ \mathbf{G}_{p(\cdot 2)} \\ \vdots \\ \mathbf{G}_{p(\cdot k)} \end{bmatrix} + \begin{bmatrix} \mathbf{H}_{p(\cdot 1)} \\ \mathbf{H}_{p(\cdot 2)} \\ \vdots \\ \mathbf{H}_{p(\cdot k)} \end{bmatrix}. \quad [92]$$

An example of $\mathbf{D}_{(i|\cdot i)}$ (Eq. 92) is given in table 5. The $k \times k$ covariance matrix for the $k \times 1$ vector of estimated conditional probabilities $\mathbf{p}_{(i|\cdot i)}$ on the diagonal of the error matrix (conditioned on the column classification) equals:

$$\mathbf{Cov}(\mathbf{p}_{(i|\cdot i)}) = \mathbf{D}'_{(i|\cdot i)} \hat{\mathbf{C}}\mathbf{ov}(vec\hat{\mathbf{P}})\, \mathbf{D}_{(i|\cdot i)}. \quad [93]$$

See Eqs. 104 and 105 for examples of $\hat{\mathbf{C}}\mathbf{ov}(vec\hat{\mathbf{P}})$. An example of Eq. 93 is given in table 5.

The variance of the estimated conditional probabilities that are conditioned on the column classifications ($\hat{p}_{(i|j\cdot)}$ in Eq. 89) can similarly be expressed in matrix form. First, define the $k \times k$ diagonal matrix $\mathbf{H}_{p(i\cdot)'}$ in which all elements equal zero except for the diagonal elements (ii):

$$(H_{p(i\cdot)})_{ii} = -\frac{\hat{p}_{ii}}{\hat{p}^2_{i\cdot}}, 1 \leq i \leq k. \quad [94]$$

An example is given in table 5. Define the $k \times k$ matrix $\mathbf{G}_{p(i\cdot)}$, in which all elements are zero except for the iith element:

$$(G_{p(i\cdot)})_{ii} = \frac{1}{\hat{p}_{i\cdot}}. \quad [95]$$

An example is given in table 5. The linear approximation of $\mathbf{p}_{(i|i\cdot)}$ equals $(vec\mathbf{P})'\mathbf{D}_{(i|i\cdot)}$, where $\mathbf{p}_{(i|i\cdot)}$ is the $k \times 1$ vector of diagonal probabilities conditioned on the row classification. The $k^2 \times k$ matrix $\mathbf{D}_{(i|i\cdot)}$ equals:

$$\mathbf{D}_{(i|i\cdot)} = \begin{bmatrix} \mathbf{G}_{p(1\cdot)} \\ \mathbf{G}_{p(2\cdot)} \\ \vdots \\ \mathbf{G}_{p(k\cdot)} \end{bmatrix} + \begin{bmatrix} \mathbf{H}_{p(1\cdot)} \\ \mathbf{H}_{p(2\cdot)} \\ \vdots \\ \mathbf{H}_{p(k\cdot)} \end{bmatrix}. \qquad [96]$$

An example is given in table 5. The $k \times k$ covariance matrix for the $k \times 1$ vector of estimated conditional probabilities ($\mathbf{p}_{(i|i\cdot)}$) on the diagonal of the error matrix (conditioned on the row classification) equals:

$$\mathbf{Cov}(\mathbf{p}_{(i|i\cdot)}) = \mathbf{D}'_{(i|i\cdot)} \left(\hat{\mathbf{Cov}}(vec\hat{\mathbf{P}})\right) \mathbf{D}_{(i|i\cdot)}. \qquad [97]$$

See Eqs. 104 and 105 for examples of $\hat{\mathbf{Cov}}(vec\hat{\mathbf{P}})$. An example of Eq. 97 is given in table 5.

Test for Conditional Probabilities Greater Than Chance

It is possible to test whether an observed conditional probability is greater than that expected by chance. In

Table 5. — Examples of conditional probabilities[1] and intermediate matrices using contingency table in table 2.

Conditional probabilities, conditioned on columns ($p_{(i|\cdot i)}$)

i	p_{ii}	$p_{\cdot i}$	$p_{(i\|\cdot i)}$	$diag[\mathbf{Cov}(P_{(i/\cdot i)})]$ (Eq. 93)	Approximate 95% confidence bounds
1	0.2361	0.4444	0.5313	0.0078	[0.3583, 0.7042]
2	0.1667	0.2639	0.6316	0.0122	[0.4147, 0.8485]
3	0.1806	0.2917	0.6190	0.0112	[0.4113, 0.8268]

$\mathbf{G}_{p(\cdot i)}$ (Eq. 91)			$\mathbf{H}_{p(\cdot i)}$ (Eq. 90)			$D_{(i/\cdot i)}$ (Eq. 92)		
j=1	j=2	j=3	j=1	j=2	j=3	j=1	j=2	j=3
2.2500	0	0	−1.1953	0	0	1.0547	0	0
0	0	0	−1.1953	0	0	−1.1953	0	0
0	0	0	−1.1953	0	0	−1.1953	0	0
0	0	0	0	−2.3934	0	0	−2.3934	0
0	3.7895	0	0	−2.3934	0	0	1.3961	0
0	0	0	0	−2.3934	0	0	−2.3934	0
0	0	0	0	0	−2.1224	0	0	−2.1224
0	0	0	0	0	−2.1224	0	0	−2.1224
0	0	3.4286	0	0	−2.1224	0	0	1.3062

Conditional probabilities, conditioned on rows ($p_{(i|i\cdot)}$)

$i\cdot$	p_{ii}	$p_{i\cdot}$	$p_{(i\|i\cdot)}$	$diag[\mathbf{Cov}(p_{(i\|i\cdot)})]$ (Eq. 97)	Approximate 95% confidence bounds
1	0.2361	0.4028	0.5862	0.0084	[0.4070, 0.7655]
2	0.1667	0.2361	0.7059	0.0122	[0.4893, 0.9225]
3	0.1806	0.3611	0.5000	0.0096	[0.3078, 0.6922]

$\mathbf{G}_{p(i\cdot)}$ (Eq. 95)			$\mathbf{H}_{p(i\cdot)}$ (Eq. 94)			$\mathbf{D}_{(i\|i\cdot)}$ (Eq. 96)		
j=1	j=2	j=3	j=1	j=2	j=3	j=1	j=2	j=3
2.4828	0	0	−1.4554	0	0	1.0273	0	0
0	0	0	0	−2.9896	0	0	−2.9896	0
0	0	0	0	0	−1.3846	0	0	−1.3846
0	0	0	−1.4554	0	0	−1.4554	0	0
0	4.2353	0	0	−2.9896	0	0	1.2457	0
0	0	0	0	0	−1.3846	0	0	−1.3846
0	0	0	−1.4554	0	0	−1.4554	0	0
0	0	0	0	−2.9896	0	0	−2.9896	0
0	0	2.7692	0	0	−1.3846	0	0	1.3846

[1]*The covariance matrix for the estimated joint probabilities ($\hat{p}_{ij}$) is estimated assuming the multinomial distribution (see Eqs. 46, 47, 128, 130, 131, and 132).*

most cases, practical interest is confined to the conditional probabilities on the diagonal ($p_{(i|i)}$ or $p_{(i|i)}$). This is closely related to the hypothesis that the conditional kappa (κ_{i} or κ_{i}) is no greater than that expected by chance (see $\hat{\mathrm{Var}}_o(\hat{\kappa}_{i})$ and $\hat{\mathrm{Var}}_o(\hat{\kappa}_{i})$ following Eqs. 67 and 70).

The proposed test is based on the null hypothesis that the difference between an observed conditional probability and its corresponding conditional probability expected under independence between the row and column classifiers is not greater than zero. First, consider the conditional probability on the diagonal of the ith row that is expected if classifiers are independent:

$$E[p_{(i|i)}] = \frac{p_{i\cdot}p_{\cdot i}}{p_{\cdot i}} = p_{i\cdot}. \qquad [98]$$

$p_{i\cdot}$ in Eq. 98 is defined in Eq. 2, but the Taylor series approximation of $p_{i\cdot}$ can be expressed differently in matrix algebra as:

$$\hat{\mathbf{p}}'_{i\cdot} = vec\hat{\mathbf{P}}'\,\mathbf{D}_{i\cdot}. \qquad [99]$$

Recall that $\hat{\mathbf{p}}_{i\cdot}$ in Eq. 99 is the $k\times1$ vector in which the ith element is $\hat{p}_{i\cdot}$ (Eq. 35), *and* $vec\hat{\mathbf{P}}'$ is the transpose of the $k^2\times1$ vector version of the $k\times k$ error matrix $\hat{\mathbf{P}}$ (Eqs. 35 and 36). In Eq. 99, $\mathbf{D}_{i\cdot}$ is a $k^2\times k$ matrix of zeros and ones, where $\mathbf{D}_{i\cdot} = (\mathbf{I}\,|\,\mathbb{L}\,|\,\mathbf{I})'$ and $\mathbf{I}$ is the $k\times k$ identity matrix. Let the $2k\times1$ vector $\hat{\mathbf{p}}_{\cdot i-}$ equal $(\hat{\mathbf{p}}_{(i|\cdot i)}|\hat{\mathbf{p}}'_{i\cdot})'$, where $\hat{\mathbf{p}}_{(i|\cdot i)}$ is the $k\times1$ vector of observed conditional probabilities (Eq. 92) and $\hat{\mathbf{p}}_{i\cdot}$ is the $k\times1$ vector of expected conditional probabilities under the independence hypothesis (Eq. 99). The covariance matrix for $\hat{\mathbf{p}}_{\cdot i-}$ using the Taylor series approximation is:

$$\hat{\mathbf{Cov}}_o(\mathbf{p}_{\cdot i}) = \mathbf{D}'_{\cdot i-}\left[\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}})\right]\mathbf{D}_{\cdot i-}, \qquad [100]$$

where $\mathbf{D}_{\cdot i-}$ is the $k^2\times2k$ matrix equal to $[\mathbf{D}_{(i|\cdot i)}|\mathbf{D}_{i\cdot}]$, $\mathbf{D}_{(i|\cdot i)}$ is defined in Eq. 92, and $\mathbf{D}_{i\cdot}$ is defined following Eq. 99. An example of $\mathbf{D}_{\cdot i-}$ is given in table 6. The covariance matrix expected under the null hypothesis, $\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}})$, is used in Eq. 100 (see Eqs. 113, 114, and 117).

The Taylor series approximation of the $k\times1$ vector of differences between the observed conditional probabilities on the diagonal ($\hat{\mathbf{p}}_{(i|i)}$) and their expected values under the independence hypothesis ($\hat{\mathbf{p}}_{i\cdot}$) equals $\mathbf{p}'_{\cdot i-}[\mathbf{I}|-\mathbf{I}]'$, where $[\mathbf{I}|-\mathbf{I}]'$ is a $2k\times k$ matrix of ones and zeros (table 6), and $\mathbf{I}$ is the $k\times k$ identity matrix. Since this represents a simple linear transformation, the Taylor series approximation of its covariance matrix is:

$$\hat{\mathbf{Cov}}_o(\hat{\mathbf{p}}_{(i|\cdot i)} - \hat{\mathbf{p}}_{i\cdot}) = [\mathbf{I}|-\mathbf{I}]\left[\hat{\mathbf{Cov}}_o(\mathbf{p}_{\cdot i})\right][\mathbf{I}|-\mathbf{I}]. \qquad [101]$$

Equation 101 can be combined with Eq. 100 for $\hat{\mathbf{Cov}}_o(\mathbf{p}_{\cdot i-})$ to make the expression more succinct with respect to $\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}})$:

$$\begin{aligned}&\hat{\mathbf{Cov}}_o(\hat{\mathbf{p}}_{(i|\cdot i)} - \hat{\mathbf{p}}_{i\cdot})\\ &= \mathbf{D}'_{(i|\cdot i)-i\cdot}\left[\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}})\right]\mathbf{D}_{(i|\cdot i)-i\cdot},\end{aligned} \qquad [102]$$

where $\mathbf{D}_{(i|\cdot i)-i\cdot}$, $\mathbf{D}_{(i|\cdot i)-i\cdot} = \mathbf{D}_{\cdot i} - [I\,|-I]'$ in Eq. 102. An example of $\mathbf{D}_{(i|\cdot i)-i\cdot}$ is given in table 6.

A similar test can be constructed for the diagonal probabilities conditioned on the row classification, in which the null hypothesis is independence between classifiers given the row classification is category i, i.e., $E[p_{(i|i\cdot)}] = p_{\cdot i}$ (see Eq. 98). Define $\mathbf{D}_{\cdot i*}$ as a $k^2\times k$ matrix of zeros and ones defined as follows. Let $\mathbf{D}_{\cdot 1}$ be the $k\times k$ matrix with ones in the first column and zeros in all other elements, let $\mathbf{D}_{\cdot 2}$ be the $k\times k$ matrix with ones in the second column and zeros in all other elements, and so forth; then, $\mathbf{D}_{\cdot i*}$ equals $(\mathbf{D}'_{\cdot 1}|\mathbf{D}'_{\cdot 2}|\mathbb{L}\,|\mathbf{D}'_{\cdot k})'$. As in Eq. 100, define $\mathbf{D}_{i-}$ as the $k^2\times2k$ matrix equal to $[\mathbf{D}_{(i|i\cdot)}|\mathbf{D}_{i*}]$, where $\mathbf{D}_{(i|i\cdot)}$ is given in Eq. 96. The approximate covariance matrix for the $k\times1$ vector of differences between the observed and expected conditional probabilities is derived as in Eq. 102:

$$\hat{\mathbf{Cov}}_o(\hat{\mathbf{p}}_{(i|i\cdot)} - \hat{\mathbf{p}}_{\cdot i}) = \mathbf{D}'_{(i|i\cdot)-i}\left[\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}})\right]\mathbf{D}_{(i|i\cdot)-i}, \qquad [103]$$

where $\mathbf{D}_{(i|i\cdot)-i} = \mathbf{D}_{i-}[\mathbf{I}|-\mathbf{I}]'$. The covariance matrix expected under the null hypothesis, $\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}})$, is used in Eq. 103 (see Eqs. 113, 114, and 117). An example of Eq. 103 is given in table 6.

The variances on the diagonal of $\hat{\mathbf{Cov}}_o(\hat{\mathbf{p}}_{(i|i\cdot)} - \hat{\mathbf{p}}_{i\cdot})$ in Eqs. 101 or 102, $\hat{\mathbf{Cov}}_o(\hat{\mathbf{p}}_{(i|i\cdot)} - \hat{\mathbf{p}}_{\cdot i})$ in Eq. 103, can be used to estimate an approximate probability of the null hypothesis being true. It is assumed that the distribution of random errors is normal in the estimate of $(\hat{\mathbf{p}}_{(i|i\cdot)} - \hat{\mathbf{p}}_{i\cdot})$ or $(\hat{\mathbf{p}}_{(i|i\cdot)} - \hat{\mathbf{p}}_{\cdot i})$, and $\hat{\mathbf{Cov}}_o(\hat{\mathbf{p}}_{(i|\cdot i)} - \hat{\mathbf{p}}_{i\cdot})$ and $\hat{\mathbf{Cov}}_o(\hat{\mathbf{p}}_{(i|i\cdot)} - \hat{\mathbf{p}}_{\cdot i})$ are accurate estimates. A one-tail test is used because practical interest is confined to testing whether the observed conditional probabilities are *greater* than those expected by chance. An example of these tests is given in table 6. Tests on conditional probabilities might be more powerful than tests with the conditional kappa statistics because the covariance matrix for conditional probabilities use fewer estimates. This will be tested with Monte Carlo simulations in the future.

Examples given by Green et al. (1993) were used to partially validate the variance approximations in Eqs. 93 and 97. This validation is limited by its empirical nature and use of the multinomial distribution for stratified sampling.

COVARIANCE MATRICES FOR $\hat{E}[\varepsilon_{ij}\varepsilon_{rs}]$ AND $vec\hat{\mathbf{P}}$

Estimated variances of accuracy assessment statistics require estimates of the covariances of random errors between estimated cells in the contingency table ($\hat{\mathbf{p}}$). These are denoted $\hat{E}[\varepsilon_{ij}\varepsilon_{rs}]$ for the covariance between cells $\{i,j\}$ and $\{r,s\}$ in $\hat{\mathbf{p}}$, or $\hat{\mathbf{Cov}}(vec\hat{\mathbf{P}})$ for the $k^2\times k^2$ covariance matrix of all covariances associated with the vector version of the estimated contingency table ($vec\hat{\mathbf{P}}$). Key examples of the need for these covariance estimates are in Eqs. 25 and 43 for the weighted kappa statistic ($\hat{\kappa}_w$); Eq. 33 for the unweighted kappa statistic ($\hat{\kappa}$); Eqs. 67, 70, 75, and 80 for the conditional kappa statistics ($\hat{\kappa}_{i\cdot}$ and $\hat{\kappa}_{\cdot i}$); Eqs. 85, 87, 88, 89, 93, and 97 for conditional probabilities ($\hat{\mathbf{p}}_{i|j}$ and $\hat{\mathbf{p}}_{j|i}$); and Eqs. 101, 102, and 103 for differences between diagonal conditional probabilities and their expected values under the independence assumption.

Table 6. — Examples of tests with conditional probabilities[1] and intermediate matrices using contingency table in table 2 and conditional probabilities in table 5.

Conditional probabilities, conditioned on columns $p_{(i|\cdot i)}$

$diag[\hat{\mathbf{C}}\mathbf{ov}(\hat{\mathbf{p}}_{(i|\cdot i)} - \hat{\mathbf{p}}_{i\cdot})]$
$p_{(i|i\cdot)}$ **(Eqs. 101 and 102)**

i	p_{ii}	$p_{\cdot i}$	$p_{(i|\cdot i)}$	$p_{i\cdot}$	$-p_{i\cdot}$	Variance[2]	Std. Dev.	*z*–value[3]	*p*–value[4]
1	0.2361	0.4444	0.5313	0.4028	0.1285	0.0045	0.0668	1.9231	0.0272
2	0.1667	0.2639	0.6316	0.2361	0.3955	0.0130	0.1142	3.4622	0.0003
3	0.1806	0.2917	0.6190	0.3611	0.2579	0.0100	0.1001	2.5760	0.0050

$\mathbf{D}_{\cdot j-} = [\mathbf{D}_{(j|\cdot j)} \mid \mathbf{D}_{j\cdot}]$
(Eqs. 92, 99, 100)

j=1	*j*=2	*j*=3	*j*=4	*j*=5	*j*=6
1.0547	0	0	1	0	0
−1.1953	0	0	0	1	0
−1.1953	0	0	0	0	1
0	−2.3934	0	1	0	0
0	1.3961	0	0	1	0
0	−2.3934	0	0	0	1
0	0	−2.1224	1	0	0
0	0	−2.1224	0	1	0
0	0	1.3061	0	0	1

$\mathbf{D}_{(i|\cdot i)-i\cdot} = \mathbf{D}_{\cdot i-}[\mathrm{I} \mid -\mathrm{I}]'$
(Eqs. 100, 101, 102)

j=1	*j*=2	*j*=3
0.0547	0	0
−1.1953	−1	0
−1.1953	0	−1
−1	−2.3934	0
0	0.3961	0
0	−2.3934	−1
−1	0	−2.1224
0	−1	−2.1224
0	0	0.3061

$[\mathrm{I} \mid -\mathrm{I}]'$ **(Eq. 101)**

j=1	*j*=2	*j*=3
1	0	0
0	1	0
0	0	1
−1	0	0
0	−1	0
0	0	−1

Conditional probabilities, conditioned on rows ($p_{(i|i\cdot)}$)

$diag[\hat{\mathbf{C}}\mathbf{ov}(\hat{\mathbf{p}}_{(i|i\cdot)} - \hat{\mathbf{p}}_{\cdot i})]$
$p_{(i|i\cdot)}$ **(Eq. 103)**

i	p_{ii}	$p_{i\cdot}$	$p_{(i|i\cdot)}$	$p_{\cdot i}$	$-p_{i\cdot}$	Variance[1]	Std. Dev.	*z*–value	*p*–value
1	0.2361	0.4028	0.5862	0.4444	0.1418	0.0055	0.0742	1.9117	0.0280
2	0.1667	0.2361	0.7059	0.2639	0.4420	0.0175	0.1323	3.3405	0.0004
3	0.1806	0.3611	0.5000	0.2917	0.2083	0.0061	0.0784	2.6580	0.0039

$\mathbf{D}_{i\cdot-} = [\mathbf{D}_{(i|i\cdot)} \mid \mathbf{D}_{\cdot i}]$
(Eq. 103)

j=1	*j*=2	*j*=3	*j*=4	*j*=5	*j*=6
1.0273	0	0	1	0	0
0	−2.9896	0	1	0	0
0	0	−1.3846	1	0	0
−1.4554	0	0	0	1	0
0	1.2457	0	0	1	0
0	0	−1.3846	0	1	0
−1.4554	0	0	0	0	1
0	−2.9896	0	0	0	1
0	0	1.3846	0	0	1

$\mathbf{D}_{(i|i\cdot)-\cdot i} = \mathbf{D}_{i\cdot-}[\mathrm{I} \mid -\mathrm{I}]'$
(Eq. 103)

j=1	*j*=2	*j*=3
0.0273	0	0
−1	−2.9896	0
−1	0	−1.3846
−1.4554	−1	0
0	0.2457	0
0	−1	−1.3846
−1.4554	0	−1
0	−2.9896	−1
0	0	0.3846

[1] *The covariance matrix for the estimated joint probabilities ($\hat{p}_{ij}$) is estimated assuming the multinomial distribution (see Eqs. 46, 47, 128, 130, 131, and 132).*

[2] *The variance of ($\hat{\mathbf{p}}_{(i|i\cdot)-}\hat{\mathbf{p}}_{i\cdot}$) equals the diagonal elements of $\hat{\mathbf{C}}\mathbf{ov}[(\hat{p}_{(i|i\cdot)} - \hat{p}_{i\cdot})]$.*

[3] *The z–value is the difference ($\hat{\mathbf{p}}_{(i|\cdot i)-}\hat{\mathbf{p}}_{i\cdot}$) divided by its standard deviation.*

[4] *The* p*–value is the approximate probability that the null hypothesis is true. The null hypothesis is that the observed conditional probability is not greater than the conditional probability expected if the two classifiers are independent. This is a one-tail test that assumes the estimation errors are normally distributed.*

The multinomial distribution pertains to the special case of simple random sampling, in which each sample unit is independently classified into one and only one mutually exclusive category using each of two classifiers. Up until recently, variance estimators for accuracy assessment statistics have been developed only for this special case.

Covariances for the multinomial distribution are given in Eqs. 46 and 47, where they were used to verify that $\hat{\mathrm{Var}}(\hat{\kappa}_w)$ in Eq. 25 agrees with the results of Everitt (1968) and Fleiss et al. (1969). These can also be expressed in matrix form as:

$$\hat{\mathbf{Cov}}(vec\hat{\mathbf{P}}) = (1-F)\left[diag(vec\hat{\mathbf{P}}) - vec\hat{\mathbf{P}}(vec\hat{\mathbf{P}})'\right]/n, \quad [104]$$

where n is the sample size of units that are classified into one and only one category by each of the two classifiers; and $diag(vec\hat{\mathbf{P}})$ is the $k^2 \times k^2$ matrix with $vec\hat{\mathbf{P}}$ on its main diagonal, with all other elements equal to zero (i.e., $diag(vec\mathbf{P})_{rr} = vec\,\mathrm{P}_r$ for all r, and $diag(vec\mathbf{P})_{rs} = 0$ for all r≠s). (1-F) in Eq. 104 is the finite population correction factor, which represents the proportional difference between the multinominal and multivariate hypergeometric distributions. F equals zero if sampling is with replacement or really zero if the population size is large relative to the sample size, which is the usual case in remote sensing (e.g., the number of randomly selected pixels for reference data is an insignificant proportion of all classified pixels). An example of this type of covariance matrix is $\hat{\mathbf{Cov}}(vec\hat{\mathbf{P}})$ in table 2.

However, there are many other types of reference data that do not fit the multinomial or hypergeometric models. $\hat{\mathbf{Cov}}(vec\mathbf{P})$ might be the following sample covariance matrix for a simple random sample of cluster-plots:

$$\hat{\mathbf{Cov}}(vec\mathbf{P}) = \frac{\sum_{r=1}^{n}(vec\mathbf{P}_r - \overline{vec\mathbf{P}})(vec\mathbf{P}_r - \overline{vec\mathbf{P}})'}{n}, \quad [105]$$

$$\overline{vec\mathbf{P}} = \frac{\sum_{r=1}^{n} vec\mathbf{P}_r}{n},$$

where n is the sample size of cluster-plots and $vec\mathbf{P}_r$ is the $k^2 \times 1$ vector version of the $k \times k$ contingency table or "error matrix" for the rth cluster plot. Czaplewski (1992) gives another example of $\hat{\mathbf{Cov}}(vec\mathbf{P})$, in which the multivariate composite estimator is used with a two-phase sample of plots (i.e., the first-phase plots are classified with less-expensive aerial photography, and a subsample of second-phase plots is classified by more-expensive field crews).

Covariances Under Independence Hypothesis

Under the hypothesis that the two classifiers are independent, and any agreement between the two classifiers is a chance event, $E[p_{ij}] = p_{i\cdot}p_{\cdot j}$. This effects $\hat{E}_o[\varepsilon_{ij}\varepsilon_{rs}]$ and $\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}})$ for $\hat{\mathrm{Var}}_o(\hat{\kappa}_w)$ in Eqs. 28 and 45; $\hat{E}_o[\varepsilon_{ii}\varepsilon_{rr}]$ for $\hat{\mathrm{Var}}_o(\hat{\kappa})$ in Eq. 34; $\hat{\mathrm{Var}}_o(\hat{\kappa}_{i\cdot})$, $\hat{\mathrm{Var}}_o(\hat{\kappa}_{\cdot i})$, $\hat{\mathbf{Cov}}_o(\hat{\kappa}_i)$, and $\hat{\mathbf{Cov}}(\hat{\kappa}_{\cdot i})$ for certain tests with Eqs. 67, 70, 71, 72, 74, 75, and 80; $\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}})$ for $\hat{\mathbf{Cov}}_o(\mathbf{p}_{\cdot i-})$ in Eq. 100, $\hat{\mathbf{Cov}}_o(\hat{\mathbf{p}}_{(i|\cdot i)} - \hat{\mathbf{p}}_{i\cdot})$ in Eq. 101, and $\hat{\mathbf{Cov}}_o(\hat{\mathbf{p}}_{(i|j\cdot)} - \mathbf{p}_{\cdot i})$ in Eq. 103. The true $p_{i\cdot}$ and $p_{\cdot j}$ are unknown, but the following estimates are available: $\hat{E}[\hat{p}_{ij}] = \hat{p}_{i\cdot}\hat{p}_{\cdot j}$

In the special case of the multinomial distribution, $\hat{E}_o[\varepsilon_{ij}\varepsilon_{rs}]$ is readily estimated as follows, using Eqs. 46 and 47:

$$\hat{E}_o[\varepsilon_{ij}\varepsilon_{ij}] = \frac{(\hat{p}_{i\cdot}\hat{p}_{\cdot j})[1-(\hat{p}_{i\cdot}\hat{p}_{\cdot j})]}{n}, \quad [106]$$

$$\hat{E}_o[\varepsilon_{ij}\varepsilon_{rs}] = \frac{-(\hat{p}_{i\cdot}\hat{p}_{\cdot j})(\hat{p}_{r\cdot}\hat{p}_{\cdot s})}{n}, \{i,j\} \neq \{r,s\}. \quad [107]$$

In matrix form, this is equivalent to:

$$\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}}) = \left[diag(vec\hat{\mathbf{P}}_c) - vec\hat{\mathbf{P}}_c(vec\hat{\mathbf{P}}_c)'\right]/n, \quad [108]$$

where $\hat{\mathbf{P}}_c = \hat{p}_{i\cdot}\hat{p}_{\cdot j}$ is the expected contingency table under the null hypothesis. For example, Eqs. 106 and 107 are used with Eq. 55 to show that $\hat{\mathrm{Var}}_o(\hat{\kappa}_w)$ in Eq. 28 agrees with the results of Fleiss et al. (1969, Eq. 9).

$\hat{E}_o[\varepsilon_{ij}\varepsilon_{rs}]$ is more difficult to estimate in the more general case. Using the first two terms of the multivariate Taylor series expansion (Eq. 10):

$$\begin{aligned}
p_{i\cdot}p_{\cdot j} \approx{}& \hat{p}_{i\cdot}\hat{p}_{\cdot j} + \varepsilon_{ij}\left(\frac{\partial p_{i\cdot}p_{\cdot j}}{\partial p_{ij}}\right)_{|p_{ij}=\hat{p}_{ij}} \\
&+\varepsilon_{i1}\left(\frac{\partial p_{i\cdot}p_{\cdot j}}{\partial p_{i1}}\right)_{|p_{ij}=\hat{p}_{ij}} + \cdots + \varepsilon_{i(i-1)}\left(\frac{\partial p_{i\cdot}p_{\cdot j}}{\partial p_{i(j-1)}}\right)_{|p_{ij}=\hat{p}_{ij}} \\
&+\varepsilon_{i(j+1)}\left(\frac{\partial p_{i\cdot}p_{\cdot j}}{\partial p_{i(j+1)}}\right)_{|p_{ij}=\hat{p}_{ij}} + \cdots + \varepsilon_{ik}\left(\frac{\partial p_{i\cdot}p_{\cdot j}}{\partial p_{ik}}\right)_{|p_{ij}=\hat{p}_{ij}} \\
&+\varepsilon_{1j}\left(\frac{\partial p_{i\cdot}p_{\cdot j}}{\partial p_{1j}}\right)_{|p_{ij}=\hat{p}_{ij}} + \cdots + \varepsilon_{(i-1)j}\left(\frac{\partial p_{i\cdot}p_{\cdot j}}{\partial p_{(i-1)j}}\right)_{|p_{ij}=\hat{p}_{ij}} \\
&+\varepsilon_{(i+1)j}\left(\frac{\partial p_{i\cdot}p_{\cdot j}}{\partial p_{(i+1)j}}\right)_{|p_{ij}=\hat{p}_{ij}} + \cdots + \varepsilon_{kj}\left(\frac{\partial p_{i\cdot}p_{\cdot j}}{\partial p_{kj}}\right)_{|p_{ij}=\hat{p}_{ij}}
\end{aligned} \quad [109]$$

Using Eqs. 58 and 59, the partial derivatives in the first-order Taylor series approximation are solved as follows:

$$\begin{aligned}
p_{i\cdot}p_{\cdot j} &= (a_{i|j} + p_{ij})(a_{\cdot j|i} + p_{ij}) \\
&= a_{i|j}a_{\cdot j|i} + p_{ij}(a_{\cdot j|i} + a_{i|j}) + p_{ij}^2,
\end{aligned} \quad [110]$$

$$\left(\frac{\partial p_{i\cdot}p_{\cdot j}}{\partial p_{ij}}\right)_{|p_{ij}=\hat{p}_{ij}} = (a_{\cdot j|i} + a_{i|j}) + 2p_{ij} = (p_{i\cdot} + p_{\cdot j}),$$

$$p_{i\cdot}p_{\cdot j} = (a_{i|s} + p_{is})p_{\cdot j},\quad 1 \le s \le k,\ s \ne j,$$

$$\left(\frac{\partial p_{i\cdot}p_{\cdot j}}{\partial p_{is}}\right)_{\substack{s\ne j \\ |p_{ij}=\hat{p}_{ij}}} = p_{\cdot j}, \qquad [111]$$

$$p_{i\cdot}p_{\cdot j} = p_{i\cdot}(a_{\cdot j|r} + p_{rj}),\ 1 \le r \le k,\ r \ne i,$$

$$\left(\frac{\partial p_{\cdot i}p_{\cdot j}}{\partial p_{rj}}\right)_{\substack{r\ne i \\ |p_{ij}=\hat{p}_{ij}}} = p_{i\cdot}. \qquad [112]$$

Substituting Eqs. 110, 111, and 112 into Eq. 109:

$$(p_{i\cdot}p_{\cdot j} - \hat{p}_{i\cdot}\hat{p}_{\cdot j}) = \varepsilon_{o|ij}$$

$$\varepsilon_{o|ij} \approx \varepsilon_{ij}(p_{i\cdot} + p_{\cdot j}) + \sum_{\substack{s=1 \\ s\ne j}}^{k}\varepsilon_{is}p_{\cdot j} + \sum_{\substack{r=1 \\ r\ne i}}^{k}\varepsilon_{rj}p_{i\cdot}$$

$$= p_{\cdot j}\sum_{s=1}^{k}\varepsilon_{is} + p_{i\cdot}\sum_{r=1}^{k}\varepsilon_{rj}$$

$$\varepsilon_{o|ij}^2 \approx \left(p_{i\cdot}\sum_{r=1}^{k}\varepsilon_{rj} + p_{\cdot j}\sum_{s=1}^{k}\varepsilon_{is}\right)\left(p_{i\cdot}\sum_{u=1}^{k}\varepsilon_{uj} + p_{\cdot j}\sum_{v=1}^{k}\varepsilon_{iv}\right)$$

$$= p_{i\cdot}^2\sum_{r=1}^{k}\sum_{u=1}^{k}\varepsilon_{rj}\varepsilon_{uj} + p_{i\cdot}p_{\cdot j}\left(\sum_{r=1}^{k}\sum_{v=1}^{k}\varepsilon_{iv}\varepsilon_{rj} + \sum_{u=1}^{k}\sum_{s=1}^{k}\varepsilon_{is}\varepsilon_{uj}\right)$$
$$+ p_{\cdot j}^2\sum_{s=1}^{k}\sum_{v=1}^{k}\varepsilon_{is}\varepsilon_{iv}$$

$$\hat{\mathrm{V}}\mathrm{ar}(\hat{p}_{i\cdot}\hat{p}_{\cdot j}) =$$
$$p_{i\cdot}^2\sum_{r=1}^{k}\sum_{s=1}^{k}\hat{E}[\varepsilon_{rj}\varepsilon_{si}] + 2p_{i\cdot}p_{\cdot j}\sum_{r=1}^{k}\sum_{s=1}^{k}\hat{E}[\varepsilon_{is}\varepsilon_{rj}]$$
$$+ p_{\cdot j}^2\sum_{r=1}^{k}\sum_{s=1}^{k}\hat{E}[\varepsilon_{ir}\varepsilon_{is}]. \qquad [113]$$

Equation 113 provides an estimate of the $\hat{E}_o[\varepsilon_{o|ij}^2]$ under the null hypothesis of independence between classifiers, which is the diagonal of the $k^2 \times k^2$ covariance matrix $\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}})$. The off-diagonal elements are estimated with the Taylor series approximation as follows:

$$\varepsilon_{o|ij}\,\varepsilon_{o|rs} \approx \left(p_{i\cdot}\sum_{u=1}^{k}\varepsilon_{uj} + p_{\cdot j}\sum_{v=1}^{k}\varepsilon_{iv}\right)\left(p_{r\cdot}\sum_{x=1}^{k}\varepsilon_{xs} + p_{\cdot s}\sum_{y=1}^{k}\varepsilon_{ry}\right),$$

$$\hat{E}_o[\varepsilon_{ij}\,\varepsilon_{rs}] = p_{i\cdot}p_{r\cdot}\sum_{u=1}^{k}\sum_{v=1}^{k}\hat{E}[\varepsilon_{uj}\varepsilon_{vs}] + p_{\cdot j}p_{r\cdot}\sum_{u=1}^{k}\sum_{v=1}^{k}\hat{E}[\varepsilon_{iv}\varepsilon_{us}]$$
$$+ p_{i\cdot}p_{\cdot s}\sum_{u=1}^{k}\sum_{v=1}^{k}\hat{E}[\varepsilon_{uj}\varepsilon_{rv}] + p_{\cdot j}p_{\cdot s}\sum_{u=1}^{k}\sum_{v=1}^{k}\hat{E}[\varepsilon_{iu}\varepsilon_{rv}]. \qquad [114]$$

Matrix Formulation for $\hat{E}_o[\varepsilon_{ij}\varepsilon_{rs}]$

Equations 113 and 114 can be expressed in matrix algebra. First, define the $k^2 \times k^2$ matrix $\hat{\mathbf{P}}_{i\cdot}$ as follows:

$$\hat{\mathbf{P}}_{i\cdot} = \begin{bmatrix} \hat{\mathbf{P}}_i & 0 & \cdots & 0 \\ 0 & \hat{\mathbf{P}}_i & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\mathbf{P}}_i \end{bmatrix}, \qquad [115]$$

where $\mathbf{0}$ is a $k \times k$ matrix of zeros, and $\hat{\mathbf{P}}_i$ is the $k \times k$ matrix in which the ith column vector equals $\hat{p}_{i\cdot}$ as defined in Eqs. 2 or 33. Table 7 includes an example of $\hat{\mathbf{P}}_{i\cdot}$ in Eq. 115. Next, define $\hat{\mathbf{p}}_{\cdot j}$ as the following $k^2 \times k^2$ matrix:

$$\hat{\mathbf{P}}_{\cdot j} = \begin{bmatrix} \mathbf{I}\hat{p}_{\cdot 1} & \mathbf{I}\hat{p}_{\cdot 2} & \cdots & \mathbf{I}\hat{p}_{\cdot k} \\ \mathbf{I}\hat{p}_{\cdot 1} & \mathbf{I}\hat{p}_{\cdot 2} & \cdots & \mathbf{I}\hat{p}_{\cdot k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{I}\hat{p}_{\cdot 1} & \mathbf{I}\hat{p}_{\cdot 2} & \cdots & \mathbf{I}\hat{p}_{\cdot k} \end{bmatrix}, \qquad [116]$$

where $\mathbf{I}$ is the $k \times k$ identity matrix, and $\hat{p}_{\cdot j}$ is the scalar marginal for the jth column of the error matrix (Eqs. 3, 31, or 120). Table 7 includes an example of $\hat{\mathbf{P}}_{\cdot j}$ in Eq. 116.

The $k^2 \times k^2$ covariance matrix for the estimated vector version of the error matrix, $\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}})$, expected under the null hypothesis of independence between classifiers, equals:

$$\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}}) = (\hat{\mathbf{P}}_{i\cdot} + \hat{\mathbf{P}}_{\cdot j})'\,\hat{\mathbf{Cov}}(vec\hat{\mathbf{P}})\,(\hat{\mathbf{P}}_{i\cdot} + \hat{\mathbf{P}}_{\cdot j}), \qquad [117]$$

where $\hat{\mathbf{P}}_{i\cdot}$ and $\hat{\mathbf{P}}_{\cdot j}$ are defined in Eqs. 115 and 116; and $\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}})$ is the $k^2 \times k^2$ covariance matrix for the estimated vector version of the error matrix ($vec\hat{\mathbf{P}}$), examples of which are given in Eqs. 104 and 105. Table 7 includes an example of $\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}})$ in Eq. 117. Equation 117 is merely a different expression of $\hat{E}_o[\varepsilon_{ij}\varepsilon_{rs}]$ in Eqs. 113 and 114.

STRATIFIED SAMPLE OF REFERENCE DATA

Stratified random sampling can be more efficient than simple random sampling when some classes are substantially less prevalent or important than others

(Campbell 1987, p. 358; Congalton 1991). This section considers strata that are defined by remotely sensed classifications, and reference data that are a separate random sample of pixels (with replacement) within each stratum. This concept includes not only pre-stratification (e.g., Green et al. 1993), but also post-stratification of a simple random sample based on the remotely sensed classification. Since the stratum size in the total population is known without error for each remotely sensed category (through a computer census of classified pixels), pre- and post-stratification could potentially improve estimation precision in accuracy assessments and estimates of area in each category as defined by the protocol used for the reference data.

The current section assumes that each sample unit is classified into only one category by each classifier, which precludes reference data from cluster plots (Congalton 1991), such as photo-interpreted maps of sample areas (e.g., Czaplewski et al. 1987). The covariance matrix for the multinomial distribution, which is given in Eqs. 46, 47, and 104 is appropriate for simple random sampling, but must be used differently for stratified random sampling since sampling errors are independent among strata.

Table 7. — Example of the covariance matrix assuming the null hypothesis of independence between classifiers and intermediate matrices. The contingency table in table 2 is used.

$\hat{\mathbf{C}}\mathbf{ov}_0(vec\hat{\mathbf{P}})$ (Eq. 117)

i	j	i=1 j=1	i=2 j=1	i=3 j=1	i=1 j=2	i=2 j=2	i=3 j=2	i=1 j=3	i=2 j=3	i=3 j=3
1	1	0.0015	0.0001	0.0002	0.0001	-0.0004	-0.0006	0.0002	-0.0004	-0.0006
2	1	0.0001	0.0006	-0.0001	-0.0000	0.0003	-0.0001	-0.0005	0.0001	-0.0005
3	1	0.0002	-0.0001	0.0010	-0.0005	-0.0004	0.0000	-0.0003	-0.0002	0.0003
1	2	0.0001	-0.0000	-0.0005	0.0005	0.0003	0.0001	-0.0000	-0.0000	-0.0004
2	2	-0.0004	0.0003	-0.0004	0.0003	0.0005	0.0002	-0.0004	0.0002	-0.0003
3	2	-0.0006	-0.0001	0.0000	0.0001	0.0002	0.0004	-0.0003	0.0000	0.0001
1	3	0.0002	-0.0005	-0.0003	-0.0000	-0.0004	-0.0003	0.0007	0.0000	0.0004
2	3	-0.0004	0.0001	-0.0002	-0.0000	0.0002	0.0000	0.0000	0.0002	0.0001
3	3	-0.0006	-0.0005	0.0003	-0.0004	-0.0003	0.0001	0.0004	0.0001	0.0009

$\hat{\mathbf{P}}_{i\cdot}$ (Eq. 115)

i	j	i=1 j=1	i=2 j=1	i=3 j=1	i=1 j=2	i=2 j=2	i=3 j=2	i=1 j=3	i=2 j=3	i=3 j=3
1	1	0.4028	0.2361	0.3611	0	0	0	0	0	0
2	1	0.4028	0.2361	0.3611	0	0	0	0	0	0
3	1	0.4028	0.2361	0.3611	0	0	0	0	0	0
1	2	0	0	0	0.4028	0.2361	0.3611	0	0	0
2	2	0	0	0	0.4028	0.2361	0.3611	0	0	0
3	2	0	0	0	0.4028	0.2361	0.3611	0	0	0
1	3	0	0	0	0	0	0	0.4028	0.2361	0.3611
2	3	0	0	0	0	0	0	0.4028	0.2361	0.3611
3	3	0	0	0	0	0	0	0.4028	0.2361	0.3611

$\hat{\mathbf{P}}_{\cdot j}$ (Eq. 116)

i	j	i=1 j=1	i=2 j=1	i=3 j=1	i=1 j=2	i=2 j=2	i=3 j=2	i=1 j=3	i=2 j=3	i=3 j=3
1	1	0.4444	0	0	0.2639	0	0	0.2917	0	0
2	1	0	0.4444	0	0	0.2639	0	0	0.2917	0
3	1	0	0	0.4444	0	0	0.2639	0	0	0.2917
1	2	0.4444	0	0	0.2639	0	0	0.2917	0	0
2	2	0	0.4444	0	0	0.2639	0	0	0.2917	0
3	2	0	0	0.4444	0	0	0.2639	0	0	0.2917
1	3	0.4444	0	0	0.2639	0	0	0.2917	0	0
2	3	0	0.4444	0	0	0.2639	0	0	0.2917	0
3	3	0	0	0.4444	0	0	0.2639	0	0	0.2917

Let the rows (i.e., i or r subscripts) of the contingency table represent the true reference classifications, and the columns (i.e., j or s subscripts) represent the less-accurate classifications (e.g., remotely sensed categorizations). Assume pre-stratification of reference data is based on the remotely sensed classifications, which are available for all members of the population (e.g., all pixels in an image) before the sample of reference data is selected. In stratified random sampling, sampling errors between all pairs of strata (i.e., columns in contingency table) are assumed to be mutually independent:

$$\hat{C}ov(\hat{p}_{ij}\hat{p}_{rs}) = 0 \text{ for } j \neq s. \quad [118]$$

Assume the size of each stratum j (i.e., $\hat{p}_{\cdot j}$) is known without error (e.g., a proportion based on a complete enumeration or census of all pixels for each remotely sensed classification). Let $n_{\cdot j}$ be the sample size of reference plots in the jth stratum, and n_{ij} be the number of reference plots classified as category i in the jth stratum. In this case,

$$n_{\cdot j} = \sum_{i=1}^{k} n_{ij} \quad [119]$$

$$\hat{p}_{ij} = p_{\cdot j}\left(\frac{n_{ij}}{n_{\cdot j}}\right). \quad [120]$$

The multinomial distribution provides the covariance matrix for sampling errors within each independent stratum j (see Eqs. 46 and 47). This distribution with Eq. 120 produces:

$$\hat{C}ov(\hat{p}_{ij}\hat{p}_{ij}) = p_{\cdot j}^2\left[\frac{\frac{n_{ij}}{n_{\cdot j}}\left(1-\frac{n_{ij}}{n_{\cdot j}}\right)}{n_{\cdot j}}\right] = \frac{p_{\cdot j}^2(n_{ij}n_{\cdot j} - n_{ij}^2)}{n_{\cdot j}^3} \quad [121]$$

$$\hat{C}ov(\hat{p}_{ij}\hat{p}_{rj}) = -p_{\cdot j}^2\left[\frac{\left(\frac{n_{ij}}{n_{\cdot j}}\right)\left(\frac{n_{rj}}{n_{\cdot j}}\right)}{n_{\cdot j}}\right] = -\left(\frac{p_{\cdot j}^2 n_{ij} n_{rj}}{n_{\cdot j}^3}\right), r \neq j. \quad [122]$$

The general variance approximation for $\hat{\kappa}_w$ is given in Eq. 25. Replacing Eqs. 118, 121, and 122 into Eq. 25, and noting that the fourth summation disappears from Eq. 25 because of the independence of sampling errors across strata:

$$\hat{V}ar(\hat{\kappa}_w) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\begin{matrix}(\overline{w}_{i\cdot}+\overline{w}_{\cdot j})(\hat{p}_o-1)\\ +w_{ij}(1-\hat{p}_c)\end{matrix}\right]\sum_{r=1}^{k}\hat{E}[\varepsilon_{ij}\varepsilon_{rj}]\left[\begin{matrix}(\overline{w}_{r\cdot}+\overline{w}_{\cdot j})(\hat{p}_o-1)\\ +w_{rj}(1-\hat{p}_c)\end{matrix}\right]}{(1-\hat{p}_c)^4}$$

$$\hat{V}ar(\hat{\kappa}_w) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\begin{matrix}(\overline{w}_{i\cdot}+\overline{w}_{\cdot j})(\hat{p}_o-1)\\ +w_{ij}(1-\hat{p}_c)\end{matrix}\right]\left(\frac{p_{\cdot j}^2}{n_{\cdot j}^3}\right)\left\{\begin{matrix}(n_{ij}n_{\cdot j}-n_{ij}^2)\left[\begin{matrix}(\overline{w}_{i\cdot}+\overline{w}_{\cdot j})(\hat{p}_o-1)\\ +w_{ij}(1-\hat{p}_c)\end{matrix}\right]\\ -\sum_{r=1}^{k} n_{ij}n_{rj}\left[\begin{matrix}(\overline{w}_{r\cdot}+\overline{w}_{\cdot j})(\hat{p}_o-1)\\ +w_{rj}(1-\hat{p}_c)\end{matrix}\right]\\ +n_{ij}^2\left[\begin{matrix}(\overline{w}_{i\cdot}+\overline{w}_{\cdot j})(\hat{p}_o-1)\\ +w_{ij}(1-\hat{p}_c)\end{matrix}\right]\end{matrix}\right\}}{(1-\hat{p}_c)^4}$$

$$\hat{V}ar(\hat{\kappa}_w) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\begin{matrix}(\overline{w}_{i\cdot}+\overline{w}_{\cdot j})(\hat{p}_o-1)\\ +w_{ij}(1-\hat{p}_c)\end{matrix}\right]^2\frac{p_{\cdot j}^2 n_{ij}}{n_{\cdot j}^2}}{(1-\hat{p}_c)^4}$$

$$\frac{\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\begin{matrix}(\overline{w}_{i\cdot}+\overline{w}_{\cdot j})(\hat{p}_o-1)\\ +w_{ij}(1-\hat{p}_c)\end{matrix}\right]\left(\frac{p_{\cdot j}^2 n_{ij}}{n_{\cdot j}^3}\right)\sum_{r=1}^{k} n_{rj}\left[\begin{matrix}(\overline{w}_{r\cdot}+\overline{w}_{\cdot j})(\hat{p}_o-1)\\ +w_{rj}(1-\hat{p}_c)\end{matrix}\right]}{(1-\hat{p}_c)^4}. \quad [123]$$

Accuracy Assessment Statistics Other Than $\hat{\kappa}_w$

The covariance matrix $\hat{\mathbf{Cov}}_s$ ($vec\mathbf{P}$) for the covariances $\hat{\mathbf{Cov}}(\hat{p}_{i\cdot}\hat{p}_{rj})$ for stratified random sampling (Eqs. 121 and 122) can be expressed in matrix algebra for use with the matrix formulations of accuracy assessment statistics in this paper (e.g., Eqs. 75, 80, 93, 97, 101, 102, and 103).

First, the $k \times k$ matrix of estimated joint probabilities ($\hat{\mathbf{p}}$) must be estimated from the $k \times k$ matrix of estimated conditional probabilities $\hat{\mathbf{P}}_s$ from the stratified sample. The strata are defined by the classifications on the columns of $\hat{\mathbf{P}}_s$; therefore, the column marginals all equal 1 (i.e., $\hat{\mathbf{P}}_s'\mathbf{1} = \mathbf{1}$). The strata sizes are assumed known without error (e.g., pixel count, where remotely sensed classifications are on the column and are used for pre-stratification of sample), and are represented by the $k \times 1$ vector of proportions of the domain that are in each stratum ($\mathbf{n}_s$, where $\mathbf{n}_s'\mathbf{1} = 1$). $\hat{\mathbf{P}}$ is estimated from $\hat{\mathbf{P}}_s$ by dividing each element in the jth column of $\hat{\mathbf{P}}_s$ by the jth element of $\mathbf{n}_s$, and then is used to define the $k^2 \times 1$ vector version ($vec\hat{\mathbf{P}}$) of the $k \times k$ matrix ($\hat{\mathbf{P}}$).

Next, compute the covariance matrix for this estimate $vec\hat{\mathbf{P}}$. Let $\hat{\mathbf{p}}_j$ represent the $k\times 1$ vector in which the ith element is the observed proportion of category i in stratum j. The $k\times k$ covariance matrix for the estimated proportions in the jth stratum, assuming the multinomial distribution, is:

$$\hat{\mathbf{Cov}}(\hat{\mathbf{p}}_j) = (1 - F_j)\left[diag(\hat{\mathbf{p}}_j) - \hat{\mathbf{p}}_j\hat{\mathbf{p}}_j'\right]/n_j, \qquad [124]$$

where n_j is the sample size of units that are classified into one and only one category by each of the two classifiers in the jth stratum; $diag(\hat{\mathbf{p}}_j)$ is the $k\times k$ matrix with $\hat{\mathbf{p}}_j$ on its main diagonal, and all other elements are equal to zero. $(1-F_j)$ in Eq. 124 is the finite population correction factor for stratum j. F_j equals zero if sampling is with replacement or the population size is large relative to the sample size, which is the usual case in remote sensing. Equation 124 is closely related to Eq. 104 for simple random sampling. The joint probabilities in the jth column of the contingency table ($\hat{\mathbf{p}}$) equal $\hat{\mathbf{p}}_j$ divided by the stratum size $\mathbf{p}_{\cdot j}$. Since $\mathbf{p}_{\cdot j}$ is known without error in the type of stratified random sampling being considered here,

$$\hat{\mathbf{Cov}}_s(vec\mathbf{P}) = \begin{bmatrix} (\hat{\mathbf{Cov}}(\hat{\mathbf{p}}_1)p_{\cdot 1}^2 & 0 & \cdots & 0 \\ 0 & (\hat{\mathbf{Cov}}(\hat{\mathbf{p}}_2)p_{\cdot 2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\hat{\mathbf{Cov}}(\hat{\mathbf{p}}_k)p_{\cdot k}^2 \end{bmatrix}, \qquad [125]$$

where $\hat{\mathbf{Cov}}(\hat{\mathbf{p}}_j)$ is defined in Eq. 124 and $\mathbf{0}$ is the $k\times k$ matrix of zeros.

Equations 124 and 125, when used with Eqs. 93 and 97 for conditional probabilities on the diagonal, agree with the results of Green et al. (1993) after transpositions to change stratification to the column classifier rather than the row classifier. Equations 124 and 125, when used with Eq. 43 for unweighted kappa ($(\hat{\kappa}_w, \mathbf{W} = \mathbf{I})$), agree with the unpublished results of Stephen Stehman (personal communication) after similar transpositions to change stratification to the column classifier. Congalton (1991) suggested testing the effect of stratified sampling with the variance estimator for simple random sampling, and Eq. 125 permits this comparison.

SUMMARY

Tests of hypotheses with the estimated accuracy assessment statistics can require a variance estimate. Most existing variance estimators for accuracy assessment statistics assume that the multinomial distribution applies to the sampling design used to gather reference data. The multinomial distribution implies that this design is a simple random sample where each sample unit (e.g., a pixel) is separately classified into a single category by each classifier. This assumption is overly restrictive for many, perhaps most, accuracy assessments in remote sensing, where more complex sampling designs and different sample units are more practical or efficient. Unfortunately, variance estimators for simple random sampling are naively applied when other sampling designs are used (e.g., Stenback and Congalton, 1990; Gong and Howarth, 1990). This improper use of published variance estimators surely affects tests of hypotheses, although the typical magnitude of the problem is unknown (Stehman 1992).

The variance estimators for the weighted kappa statistic [Eqs. 24 and 43 for $\hat{\mathrm{Var}}(\hat{\kappa}_w)$ and Eqs. 28 and 45 for $\hat{\mathrm{Var}}_o(\hat{\kappa}_w)$]; the unweighted kappa statistic [Eq. 33 for $\hat{\mathrm{Var}}(\hat{\kappa})$ and Eq. 34 for $\hat{\mathrm{Var}}(\hat{\kappa})$]; the conditional kappa statistic [Eqs. 67 and 75 for $\hat{\mathrm{Var}}_o(\hat{\kappa}_i)$ and Eqs. 70 and 80 for $\hat{\mathrm{Var}}(\hat{\kappa}_i)$]; conditional probabilities [(Eqs. 85, 87, 88, 89, 93, and 97 for $\hat{\mathrm{Var}}(\hat{p}_{i|j})$ and $\hat{\mathrm{Var}}(\hat{p}_{i|j})$]; and differences between diagonal conditional probabilities and their expected values under the independence assumption (Eqs. 101, 102, and 103) are the first step in correcting this problem. These equations form the basis for approximate variance estimators for other sampling situations, such as cluster sampling, systematic sampling (Wolter 1985 in Stehmam 1992), and more complex designs (e.g., Czaplewski 1992). Stratified random sampling is an important design in accuracy assessments, and the more general variance estimators in this paper were used to construct the appropriate $\hat{\mathrm{Var}}(\hat{\kappa}_w)$ in Eq. 123 and other accuracy assessment statistics using $\hat{\mathbf{Cov}}_s(vec\mathbf{P})$ in Eq. 125. Rapid progress in assessments of classification accuracy with more complex sampling and estimation situations is expected based on the foundation provided in this paper.

ACKNOWLEDGMENTS

The author would like to thank Mike Williams, C. Y. Ueng, Oliver Schabenberger, Robin Reich, Rudy King, and Steve Stehman for their assistance, comments, and time spent evaluating drafts of this manuscript. Any errors remain the author's responsibility. This work was supported by the Forest Inventory and Analysis and the Forest Health Monitoring Programs, USDA Forest Service, and the Forest Group of the Environmental Monitoring and Assessment Program, U.S. Environmental Protection Agency (EPA). This work has not been subjected to EPA's peer and policy review, and does not necessarily reflect the views of EPA.

LITERATURE CITED

Bishop, Y. M. M.; Fienberg, S. E.; Holland, P. W. 1975. Discrete multivariate analysis: Theory and practice. Cambridge, Massachusetts: MIT Press. 557 p.

Campbell, J. B. 1987. Introduction to remote sensing. New York: The Guilford Press. 551 p.

Christensen, R. C. 1991. Linear models for multivariate, time series, and spatial data. New York: Springer-Verlag. 317 p.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement. 20: 37–46.

Cohen, J. 1968. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin. 70: 213–220.

Congalton, R. G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment 37: 35–46.

Congalton, R. G.; R. A. Mead. 1983. A quantitative method to test for consistency and correctness in photointerpretation. Photogrammetric Engineering and Remote Sensing. 49: 69–74.

Czaplewski, R. L. 1992. Accuracy assessment of remotely sensed classifications with multi-phase sampling and the multivariate composite estimator. In: Proceedings 16th International Biometrics Conference, Hamilton, New Zealand. 2: 22.

Czaplewski, R. L.; Catts, G. P.; Snook, P. W. 1987. National land cover monitoring using large, permanent photo plots. In: Proceedings of International Conference on Land and Resource Evaluation for National Planning in the Tropics; 1987 Chetumal, Mexico. U.S. Department of Agriculture, Gen. Tech. Rep. GTR WO–39 Washington, DC: Forest Service: 197–202.

Deutch, R. 1965. Estimation theory. London: Prentice-Hall, Inc.

Everitt, B. S. 1968. Moments of the statistics kappa and weighted kappa. British Journal of Mathematical and Statistical Psychology. 21: 97–103.

Fleiss, J. L. 1981. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley & Sons, Inc.

Fleiss, J. L.; Cohen, J; Everitt, B. S. 1969. Large sample standard errors of kappa and weighted kappa. Psychological Bulletin. 72: 323–327.

Gong, P.; Howarth, P. J. 1990. An assessment of some factors influencing multispectral land-cover classification. Photogrammetric Engineering and Remote Sensing. 56: 597–603.

Goodman, L. A. 1968. The analysis of cross-classified data: independence, quasi-independence, and interaction in contingency tables with and without missing cells. Journal of the American Statistical Association. 63: 1091–1131.

Green, E. J.; Strawderman, W. E.; Airola, T. M. 1993. Assessing classification probabilities for thematic maps. Photogrammetric Engineering and Remote Sensing. 59: 635–639.

Hudson, W. D.; Ramm, C. W. 1987. Correct formulation of the kappa coefficient of agreement. Photogrammetric Engineering and Remote Sensing. 53: 421–422.

Landis, J. R.; Koch, G. G. 1977. The measurement of observer agreement for categorical data. Biometrics. 33: 159–174.

Light R. J. 1971. Measures of response agreement for qualitative data: some generalizations and alternatives. Psychological Bulletin. 76: 363–377.

Monserud, R. A.; Leemans, R. 1992. Comparing global vegetation maps with the kappa statistic. Ecological Modelling. 62: 275–293.

Mood, A. M.; Greybull, F. A.; Boes, D. C. 1963. Introduction to the theory of statistics. 3rd. ed. New York: McGraw-Hill.

Rao, C. R. 1965. Linear statistical inference and its applications. New York: John Wiley & Sons, Inc. 522 p.

Ratnaparkhi, M. V. 1985. Multinominal distributions. In: Encyclopedia of statistical sciences, vol. 5. Kotz, S. and Johnson, N. L., eds. New York: John Wiley & Sons, Inc. 741 p.

Rosenfield, G. H.; Fitzpatrick-Lins, K. 1986. A coefficient of agreement as a measure of thematic classification accuracy. Photogrammetric Engineering and Remote Sensing. 52: 223–227.

Stehman, S. V. 1992. Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data. Photogrammetric Engineering and Remote Sensing. 58: 1343–1350.

Stenback, J. M.; Congalton, R. G. 1990. Using thematic mapper imagery to examine forest understory. Photogrammetric Engineering and Remote Sensing. 56: 1285–1290.

Story, M.; Congalton, R. G. 1986. Accuracy assessment: a user's perspective. Photogrammetric Engineering and Remote Sensing. 52: 397–399.

Wolter, K. 1985. Introduction to variance estimation. New York: Springer-Verlag. 427 p.

APPENDIX A: Notation

Variable	Definition	Equation
$a_{i\|u}$	$p_{i\cdot} - p_{iu}$	58
$a_{\cdot i\|u}$	$p_{\cdot i} - p_{ui}$	59
b_{ij}	coefficients containing p_{ij} in p_c	17, 21
c_{ij}	coefficients not containing p_{ij} in p_c	18
$\hat{\mathrm{C}}\mathrm{ov}(\hat{p}_{ij}\hat{p}_{ij})$	estimated variance for $\hat{p}_{ij}$ $\hat{\mathrm{V}}\mathrm{ar}(\hat{p}_{ij})$	46
$\hat{\mathrm{C}}\mathrm{ov}(\hat{p}_{ij}\hat{p}_{rs})$	estimated covariance between $\hat{p}_{ij}$ and $\hat{p}_{rs}$	47, 118
$\mathbf{Cov}(\mathbf{p}_{(i\|i\cdot)})$	$k\times k$ covariance matrix for the $k\times 1$ vector of estimated conditional probabilities $(\mathbf{p}_{(i\|i\cdot)})$ on the diagonal of the error matrix (conditioned on the row classification)	97
$\mathbf{Cov}(\mathbf{p}_{(i\|\cdot i)})$	$k\times k$ covariance matrix for the $k\times 1$ vector of estimated conditional probabilities $(\mathbf{p}_{(i\|\cdot i)})$ on the diagonal of the error matrix (conditioned on the column classification)	93
$\hat{\mathbf{C}}\mathbf{ov}(vec\mathbf{P})$	$k^2\times k^2$ covariance matrix for the estimate $vec\hat{\mathbf{P}}$	104, 105, 125
$\mathbf{Cov}(\hat{\kappa}_{i\cdot})$	covariance matrix for the conditional kappa statistics $\hat{\kappa}_{i\cdot}$	75
$\mathbf{Cov}(\hat{\kappa}_{\cdot i})$	covariance matrix for the conditional kappa statistics $\hat{\kappa}_{\cdot i}$	80
$\hat{\mathbf{C}}\mathbf{ov}_o(p_{\cdot i-})$	$2k\times 2k$ covariance matrix for $\hat{\mathbf{p}}_{\cdot i-}$	100
$\hat{\mathbf{C}}\mathbf{ov}_o(\hat{\mathbf{p}}_{(i\|i\cdot)} - \hat{\mathbf{p}}_{\cdot i})$	$k\times k$ covariance matrix for differences between the observed conditional probabilities on the diagonal $(\hat{\mathbf{p}}_{(i\|i\cdot)})$ and their expected values under the independence hypothesis	103
$\hat{\mathbf{C}}\mathbf{ov}_o(\hat{\mathbf{p}}_{(i\|\cdot i)} - \hat{\mathbf{p}}_{i\cdot})$	$k\times k$ covariance matrix for differences between the observed conditional probabilities on the diagonal $(\hat{\mathbf{p}}_{(i\|\cdot i)})$ and their expected values under the independence hypothesis	101, 102
$\hat{\mathbf{C}}\mathbf{ov}_o(vec\hat{\mathbf{P}})$	$k^2\times k^2$ covariance matrix for the estimate $vec\hat{\mathbf{P}}$ under the null hypothesis of independence between the row and column classifiers	45
$\hat{\mathbf{C}}\mathbf{ov}_s(vec\hat{\mathbf{P}})$	$k^2\times k^2$ covariance matrix for the estimate $vec\hat{\mathbf{P}}$ for stratified random sampling	125
$\mathbf{d}_k$	$k^2\times 1$ vector containing the first-order Taylor series approximation of $\hat{\kappa}_w$ or $\hat{\kappa}$	42, 43
$\mathbf{D}_i$	$k^2\times k$ matrix of zeros and ones defined in Eq. 99	99
$\mathbf{D}_{(i\|i\cdot)}$	$k\times k$ intermediate matrix used to compute $\hat{\mathrm{V}}\mathrm{ar}(\hat{p}_{(i\|i\cdot)})$, where $(vec\mathbf{P})'\mathbf{D}_{(i\|i\cdot)}$ is the linear approximation of $\mathbf{p}_{(i\|i\cdot)}$	96
$\mathbf{D}_{(i\|\cdot i)}$	$k\times k$ intermediate matrix used to compute $\hat{\mathrm{V}}\mathrm{ar}(\hat{p}_{i\|\cdot i})$, where $(vec\mathbf{P})'\mathbf{D}_{(i\|\cdot i)}$ is the linear approximation of $\mathbf{p}_{(i\|\cdot i)}$	92
$\mathbf{D}_{(i\|i\cdot)-i}$	$k^2\times k$ intermediate matrix, equal to $\mathbf{D}_{i-}[\mathrm{I}\|-\mathrm{I}]'$, used to compute $\hat{\mathbf{C}}\mathbf{ov}_o(\hat{\mathbf{p}}_{(i\|i\cdot)} - \hat{\mathbf{p}}_{i\cdot})$	103
$\mathbf{D}_{(i\|\cdot i)-i\cdot}$	$k^2\times k$ intermediate matrix, equal to $\mathbf{D}_{i-}[\mathrm{I}\|-\mathrm{I}]'$, used to compute $\hat{\mathbf{C}}\mathbf{ov}_o(\hat{\mathbf{p}}_{(i\|\cdot i)} - \hat{\mathbf{p}}_{i\cdot})$	102

APPENDIX A: Notation (Continued)

Variable	Definition	Equation
$\mathbf{D}_{\kappa=0}$	$k^2 \times 1$ vector containing the first-order Taylor series approximation of $\hat{\kappa}_w$ or $\hat{\kappa}$ under the null hypothesis of independence between the row and column classifiers	44
$\mathbf{D}_{i-}$	$k^2 \times 2k$ matrix equal to $[\mathbf{D}_{(i\|i)} \mid \mathbf{D}_{i\cdot}]$, used to compute $\hat{\mathbf{Cov}}_o(\mathbf{p}_{i-})$	100
$\mathbf{d}_{\kappa_{i\cdot}}$	$k^2 \times k$ matrix used in the matrix computation of $\mathbf{Cov}(\hat{\kappa}_{i\cdot})$	74
$\mathbf{d}_{\kappa_{\cdot i}}$	$k^2 \times k$ matrix used in the matrix computation of $\mathbf{Cov}(\hat{\kappa}_{\cdot i})$	79
$diagA$	$k \times 1$ vector containing the diagonal of the $k \times k$ matrix $\mathbf{A}$.	
$E[\cdot]$	expectation operator.	
$\hat{E}_o[\varepsilon_{ij}\varepsilon_{rs}]$	expected covariance between cells ij and rs in the contingency table under the null hypothesis of independence between classifiers	113, 114, 117
F	finite population correction factor for covariance matrix under simple random sampling	104
F_j	finite population correction factor for stratum j in covariance matrix for stratified random sampling	124
$\mathbf{G}_{i\cdot}$	$k \times k$ matrix used in the matrix computation of $\mathbf{Cov}(\hat{\kappa}_{i\cdot})$	73
$\mathbf{G}_{\cdot i}$	$k \times k$ matrix used in the matrix computation of $\mathbf{Cov}(\hat{\kappa}_{\cdot i})$	78
$\mathbf{G}_{p(\cdot i)}$	$k \times k$ intermediate matrix used to compute $\hat{\mathrm{V}}\mathrm{ar}(\hat{p}_{i\|\cdot i})$	91
$\mathbf{G}_{p(i\cdot)}$	$k \times k$ intermediate matrix used to compute $\hat{\mathrm{V}}\mathrm{ar}(\hat{p}_{i\|i\cdot})$	95
$\mathbf{H}_{i\cdot}$	$k \times k$ matrix used in the matrix computation of $\mathbf{Cov}(\hat{\kappa}_{i\cdot})$	71
$\mathbf{H}_{p(\cdot i)}$	$k \times k$ intermediate matrix used to compute $\hat{\mathrm{V}}\mathrm{ar}(\hat{p}_{i\|\cdot i})$	90
$\mathbf{H}_{p(i\cdot)}$	$k \times k$ intermediate matrix used to compute $\hat{\mathrm{V}}\mathrm{ar}(\hat{p}_{i\|i\cdot})$	94
$\mathbf{H}_{\cdot i}$	$k \times k$ matrix used in the matrix computation of $\mathbf{Cov}(\hat{\kappa}_{\cdot i})$	76
$\mathbf{I}$	the $k \times k$ identity matrix in which $\mathbf{I}_{ij} = 1$ for $i \neq j$ and $\mathbf{I}_{ij} = 0$ otherwise.	
i	row subscript for contingency table	6
j	column subscript for contingency table	6
k	number of categories in the classification system	6
$\mathbf{M}_{i\cdot}$	$k \times k$ matrix used in the matrix computation of $\mathbf{Cov}(\hat{\kappa}_{i\cdot})$	72
$\mathbf{M}_{\cdot i}$	$k \times k$ matrix used in the matrix computation of $\mathbf{Cov}(\hat{\kappa}_{\cdot i})$	77
$n_{\cdot j}$	sample size of reference plots in the jth stratum	119
p_c	matching proportion expected assuming independence between the row and column classifiers	5

APPENDIX A: Notation (Continued)

Variable	Definition	Equation
$\hat{p}_c$	matching proportion expected assuming independence between the row and column classifiers (estimated)	7, 30
p_{ij}	ijth proportion in contingency table	6, 26
$\hat{p}_{ij}$	estimated ijth proportion in contingency table	7
$p_{i\cdot}$	row marginal of contingency table	2, 32
$\mathbf{p}_{i\cdot}$	$k\times 1$ vector in which the ith element is $p_{i\cdot}$	35, 99
$\hat{\mathbf{p}}_j$	$k\times 1$ vector in which the ith element is the observed proportion of category i in stratum j (used for stratified random sampling example)	125
$\hat{\mathbf{p}}_{i-}$	$2k\times 1$ vector $(\hat{\mathbf{p}}'_{(i\|\cdot i)} \vert \hat{\mathbf{p}}'_{i\cdot})'$ containing the observed and expected conditional probabilities on diagonal of error matrix, conditioned on column classification	100
p_o	proportion matching classifications	4, 27
$\hat{p}_o$	estimated proportion matching classifications.	7, 29
$p_{\cdot j}$	column marginal of contingency table	3, 31, 120, 125
$p_{\cdot i}$	$k\times 1$ vector in which the ith element is $p_{\cdot i}$	36
$p_{(i\|j\cdot)}$	conditional probability that the column classification is category i given that the row classification is category j	86
$p_{(i\|\cdot j)}$	conditional probability that the row classification is category i given that the column classification is category j	81
$p_{(i\|\cdot i)}$	$k\times 1$ vector of diagonal conditional probabilities with its ith element equal to $p_{(i\|\cdot i)}$	92
$\mathbf{P}$	$k\times k$ matrix (i.e., the error matrix in remote sensing jargon) in which the ijth element of $\mathbf{P}$ is the scalar p_{ij}	35, 36
$\mathbf{P}_c$	$E[\mathbf{P}]$ under null hypothesis of independence between the row and column classifiers	37
$\hat{\mathbf{P}}_{i\cdot}$	$k^2\times k^2$ intermediate matrix used to compute $\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}})$	115, 117
$\hat{\mathbf{P}}_i$	$k\times k$ intermediate matrix used to compute $\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}})$	115
$\hat{\mathbf{P}}_{\cdot j}$	$k\times k$ intermediate matrix used to compute $\hat{\mathbf{Cov}}_o(vec\hat{\mathbf{P}})$	116, 117
R	remainder in Taylor series expansion	10
$\hat{\mathrm{Var}}(\hat{p}_{ij})$	estimated variance for $\hat{p}_{ij}$ $\hat{\mathbf{Cov}}(\hat{p}_{ij}\hat{p}_{ij})$	46
$\hat{\mathrm{Var}}(\hat{p}_{i\|j\cdot})$	estimated variance of random errors for estimating conditional probability $\hat{p}_{i\|j\cdot}$ (conditioned on row j)	87, 89
$\hat{\mathrm{Var}}(\hat{p}_{i\|\cdot j})$	estimated variance of random errors for estimating conditional probability $\hat{p}_{i\|\cdot j}$ (conditioned on column j)	85, 88

APPENDIX A: Notation (Continued)

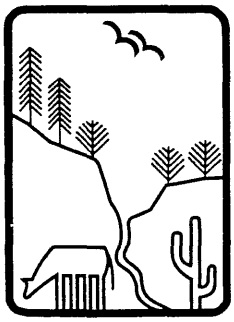
Variable	Definition	Equation
$\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa})$	estimated variance of random errors for kappa	33
$\hat{\mathrm{V}}\mathrm{ar}(\kappa_{i.})$	estimated variance of random errors for conditional kappa, conditioned on row classifier	67
$\hat{\mathrm{V}}\mathrm{ar}_o(\kappa_{i.})$	estimated variance of random errors for conditional kappa, conditioned on row classifier, under the null hypothesis of independence between classifiers	67
$\hat{\mathrm{V}}\mathrm{ar}(\kappa_{.i})$	estimated variance of random errors for conditional kappa, conditioned on column classifier	70
$\hat{\mathrm{V}}\mathrm{ar}_o(\kappa_{.i})$	estimated variance of random errors for conditional kappa, conditioned on column classifier, under the null hypothesis of independence between classifiers	70
$\mathrm{Var}(\hat{\kappa}_w)$	variance of random estimation errors for weighted kappa	9
$\hat{\mathrm{V}}\mathrm{ar}(\hat{\kappa}_w)$	estimated variance of random errors for weighted kappa	25
$\hat{\mathrm{V}}\mathrm{ar}_o(\hat{\kappa}_w)$	estimated variance of random errors for weighted kappa under the null hypothesis of independence between the row and column classifiers (i.e., κ_w)	28
$\hat{\mathrm{V}}\mathrm{ar}_o(\hat{\kappa})$	estimated variance of random errors for unweighted kappa under the null hypothesis of independence between the row and column classifiers (i.e., κ)	28
$vec\mathbf{A}$	the $k^2\times1$ vector version of the $k\times k$ matrix $\mathbf{A}$. If $\mathbf{a}_j$ is the $k\times1$ column vector in which the ith element equals a_{ij}, then $\mathbf{A}\,[\mathbf{a}_1\vert\mathbf{a}_2\vert\cdots\vert\mathbf{a}_k]$, and $vec\mathbf{A}\,[\mathbf{a}_1'\vert\mathbf{a}_2'\vert\cdots\vert\mathbf{a}_k']'$	42, 104
w_{ij}	weight placed on agreement between category i under the first classification protocol, and category j under the second protocol	6
$\overline{w}_{i.}$	weighted average of the weights in the ith row	22, 31
$\mathbf{w}_i$	$k\times1$ vector used in the matrix computation of $\hat{\kappa}$	40
$\overline{w}_{.j}$	weighted average of the weights in the jth column	23, 32
$w_{.j}$	$k\times1$ vector used in the matrix computation of $\hat{\kappa}$	41
$\mathbf{W}$	$k\times k$ matrix in which the ijth element is w_{ij}	38, 37
ε_κ^2	squared error in estimated kappa $(\kappa_w-\hat{\kappa}_w)^2$	12
ε_κ	random error in estimated kappa $(\kappa_w-\hat{\kappa}_w)$	8, 11
ε_{ij}	$(p_{ij}-\hat{p}_{ij})$	10
$\kappa_{i.}$	conditional kappa for row category i	56
$\kappa_{.i}$	conditional kappa for column category i	68
κ_w	weighted kappa statistic	6
$\hat{\kappa}_w$	weighted kappa statistic (estimated)	7
$\mathbf{0}$	$k\times k$ matrix of zeros	115

APPENDIX A: Notation (Continued)

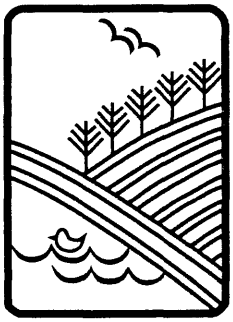
Variable	Definition	Equation
$\left(\frac{\partial \kappa}{\partial p_{ij}}\right)_{\mid p_{ij}=\hat{p}_{ij}}$	partial derivative of κ with respect to p_{ij} evaluated at p_{ij} $\hat{p}_{ij}$ for all i, j	20, 83
$\otimes$	element-by-element multiplication, where the ijth element of $(\mathbf{A} \otimes \mathbf{B})$ is $a_{ij}b_{ij}$, and matrices **A** and **B** have the same dimensions	38, 37

*U.S. GOVERNMENT PRINTING OFFICE:1994-576-749/05145

Rocky Mountains



Southwest



Great Plains

U.S. Department of Agriculture
Forest Service

Rocky Mountain Forest and Range Experiment Station

The Rocky Mountain Station is one of eight regional experiment stations, plus the Forest Products Laboratory and the Washington Office Staff, that make up the Forest Service research organization.

RESEARCH FOCUS

Research programs at the Rocky Mountain Station are coordinated with area universities and with other institutions. Many studies are conducted on a cooperative basis to accelerate solutions to problems involving range, water, wildlife and fish habitat, human and community development, timber, recreation, protection, and multiresource evaluation.

RESEARCH LOCATIONS

Research Work Units of the Rocky Mountain Station are operated in cooperation with universities in the following cities:

Albuquerque, New Mexico
Flagstaff, Arizona
Fort Collins, Colorado*
Laramie, Wyoming
Lincoln, Nebraska
Rapid City, South Dakota

*Station Headquarters: 240 W. Prospect Rd., Fort Collins, CO 80526