Assessing Classification Probabilities for Thematic Maps

Abstract

In this paper we discuss the methods for estimating the probability that a pixel is actually in category i *given that it was classified into category* j, *and methods for calculating the probability that a pixel is classified into category* j, *given that it is actually in category* i. *We show that the proper method for calculating these probabilities depends upon how the sample pixels are selected. If they are selected from the entire map, without regard to classification, then both of the above probabilities can be computed using standard multinomial formulae. However, if separate random samples are taken from the pixels classified into each category, then the former probability can be estimated using multinomial formulae, while the latter should be estimated employing Bayes theorem. We show how to do this, present an estimate of the variance of the estimated probability, and conclude with two example applications; one with TM data from New Jersey, and one with simulated data.*

Introduction

Remotely sensed data have been used extensively to generate land-cover information for a variety of purposes using a wide range of classification approaches (Airola *et al.*, 1992). In this paper we discuss methods for estimating map accuracy. Although our methods are not exclusive to thematic maps constructed from remotely sensed data, we will focus on this case. We define a thematic map to be one in which all the areas on the map are categorized into one and only one of a discrete number of categories. A familiar example is a map which portrays land-cover type (e.g., agricultural, forested, water, residential, etc.). Accuracy testing is normally accomplished by comparing the classification of pixels on the map with reference data obtained from aerial photographs or ground checks. We will show that the appropriate method depends upon how the sample(s) are selected. For simplicity, we will consider two cases: (1) sample locations drawn randomly from the population of pixels on the map, without regard to the classification of the pixel or the category to which the pixel truly belongs, and (2) separate simple random samples of pixels drawn from each category on the map (here, the sample for category j is drawn from the population of pixels *classified* into category j).

Before proceeding further, it is important to distinguish two different types of errors. First, we may err in developing the map by wrongly classifying a pixel into category i when it is actually in category j. Let us denote the probability of such an error as Prob(Classified category = i|Actual category = j) = $p_{C_i|A_j}$. Secondly, once the map is constructed, it may be in error inasmuch as it may depict a certain pixel to be in category i whereas the pixel is actually in category j. Let us denote the probability of this type of error as Prob(Actual category = j|Classified category = i) = $p_{A_j|C_i}$. Aronoff (1985) and Story and Congalton (1986), among others, have discussed these two types of errors, and have termed the risk of the first type *producer* risk, and the risk of the second *consumer* risk.

Methods

Suppose we have a map in which, as usual, all pixels are classified into one and only one category. Further, suppose we are concerned with four categories: forest, agriculture, residential, and water. Now suppose a simple random sample of pixels is chosen from the map without regard to the classification of the pixels. These pixels are checked, and the result is an error matrix like that in Table 1. The reference data are the data representing *truth*, i.e., the data obtained from aerial photographs or ground checks. The classified data are the data obtained from the map, after it has been constructed.

Let n_{ij} = the number of pixels classified into category i and found to actually be in category j, $n_{\cdot j}$ = the total number of pixels actually in category j, $n_{i\cdot}$ = the total number of sample pixels in category i, $n_{\cdot\cdot}$ = the total number of pixels in the sample, and m = the number of categories on the map. Mathematically,

$$n_{\cdot j} = \sum_{i=1}^{m} n_{ij},\; n_{i\cdot} = \sum_{j=1}^{m} n_{ij},\; n_{\cdot\cdot} = \sum_{i=1}^{m}\sum_{j=1}^{m} n_{ij}$$

In this case it is usual and proper to estimate $p_{C_i|A_j}$ and $p_{A_j|C_i}$ as multinomial probabilities. When we are considering $p_{A_j|C_i}$, then we are concerned with the rows of the error matrix. For a given classification, the number of pixels *actually* in cate-

TABLE 1. ERROR MATRIX FOR HYPOTHETICAL THEMATIC MAP WITH FOUR CATEGORIES: FOREST (F), AGRICULTURE (A), RESIDENTIAL (R), AND WATER (W).

		Reference Data				
		F	A	R	W	Total
Classified Data	F	20	2	3	0	25
	A	1	21	2	1	25
	R	7	8	10	0	25
	W	0	2	0	23	25
	TOTAL	28	33	15	24	100

Edwin J. Green
Department of Natural Resources, Remote Sensing, Cook College—Rutgers University, New Brunswick, NJ 08903-0231 pp. 000-000.

William E. Strawderman
Department of Statistics, Rutgers University, New Brunswick, NY 08903-0231

Teuvo M. Airola
Cook College Remote Sensing Center, Cook College–Rutgers University, New Brunswick, NJ 08903-0231

Photogrammetric Engineering & Remote Sensing,
Vol. 59, No. 5, May 1993, pp. 635–639.

0099-1112/93/5905-635$03.00/0

gory j is a multinomial random variable, i.e., $n_{ij}|C_i \sim$ multinomial$(n_{i\cdot}, p_{A_1|C_i}, p_{A_2|C_i}, \ldots, p_{A_m|C_i})$, where the first term in the parentheses denotes sample size and the remaining terms are the multinomial probabilities. From standard statistical theory (e.g., see Bickel and Doksum (1977)), we know that the maximum-likelihood estimate for $p_{A_j|C_i}$ is

$$\hat{p}_{A_j|C_i} = n_{ij}/n_{i\cdot}$$

and the maximum-likelihood estimate for the variance of $\hat{p}_{A_j|C_i}$ is

$$\hat{v}(\hat{p}_{A_j|C_i}) = [n_{ij}/(n_{i\cdot}^2)](1 - n_{ij}/n_{i\cdot}).$$

The usual 95 percent confidence interval (justified by the Central Limit Theorem; e.g., see Mendenhall and Scheaffer (1973)) for $p_{A_j|C_i}$ is

$$\hat{p}_{A_j|C_i} \pm 2\sqrt{\hat{v}(\hat{p}_{A_j|C_i})}.$$

Note that as a consequence of the definition of $\hat{p}_{A_j|C_i}$, we have $\Sigma_j(\hat{p}_{A_j|C_i}) = 1$.

Because there was no restriction on the randomization, i.e., the sample was randomly drawn from the entire population of pixels on the map, $p_{C_i|A_j}$ is also a multinomial probability, and can be estimated in the same way as $p_{A_j|C_i}$, making obvious changes.

Now let us consider the case where separate random samples are drawn from the pixels classified into each category. One common manifestation of this is stratified random sampling (e.g., see Cochran (1977)). When we look at the rows of the error matrix, we are looking at the data conditioned on (or *given*) the classification. Because stratified random sampling amounts to selecting a simple random sample in each stratum, it is clear that we have a multinomial random variable in each row, and that the sum of the row probabilities *must* be one. Hence, the above estimates for $p_{A_j|C_i}$ and $v(p_{A_j|C_i})$ are still valid. However, when we look at the columns, we are conditioning on the actual category. But in stratified sampling, the randomization is only applied within classified categories. Hence, there is no randomization within actual categories and the above multinomial method is invalid for $p_{C_i|A_j}$. However, a valid estimate of $p_{C_i|A_j}$ is available by an application of Bayes' theorem.

The celebrated Bayes' Theorem states simply that if X and Y are two events, then $P(X|Y) = P(Y|X)P(X)/P(Y)$ (see, among others, Box and Tiao (1973)). The theorem is a simple, undisputable consequence of standard probability axioms and is not controversial. However, when used as the basis for an inference system, the theorem has been the source of much disagreement among statisticians (see, e.g., Berger (1985)). It is important to note that the latter is not what we propose here. We use Bayes theorem simply to obtain one conditional probability from another. This is in keeping with the traditional use of the theorem in classical statistics.

In the present notation, we write $p_{C_i|A_j} = p_{A_j|C_i}(p_{C_i}/p_{A_j})$. As mentioned above, we can use the usual estimator for $p_{A_j|C_i}$. In addition, there is an obvious method to compute p_{C_i}. The latter would simply be the percentage of all pixels in the map classified into category i (note that we may then regard p_{C_i} as *known*, and not as an estimate subject to error).

All that remains is to determine a method for estimating p_{A_j}. Assume the samples within each category are independent, i.e., separate random samples are selected in each. Then, because the categories are mutually exclusive, we can estimate p_{A_j} with

$$\hat{p}_{A_j} = \sum_{k=1}^{m} (\hat{p}_{A_j|C_k})p_{C_k}.$$

Hence, we have

$$\hat{p}_{C_i|A_j} = (\hat{p}_{A_j|C_i} \cdot p_{C_i})/\hat{p}_{A_j} = \frac{(n_{ij}/n_{i\cdot})p_{C_i}}{\sum_{k=1}^{m}(n_{kj}/n_{k\cdot})p_{C_k}} \quad (1)$$

It is clear from the definition of p_{C_i} that $\Sigma_i(p_{C_i}) = 1$, and, because $p_{A_j|C_i}$ is an estimate of a multinomial probability, $\Sigma_j(\hat{p}_{A_j|C_i}) = 1$. It can be shown that, as one would expect, $\Sigma_i(\hat{p}_{C_i|A_j}) = 1$ and $\Sigma_j(\hat{p}_{A_j}) = 1$.

The variance of $\hat{p}_{C_i|A_j}$ is derived in the Appendix. There we show that a reasonable estimate for the variance of $p_{C_i|A_j}$ is

$$\hat{v}(\hat{p}_{C_i|A_j}) = \left[a_i\left(\frac{1}{\sum_{k=1}^{m} w_k} - \frac{w_i}{\left(\sum_{k=1}^{m} w_k\right)^2}\right)\right]^2 \times n_{ii}\left(1 - \frac{n_{ii}}{n_{i\cdot}}\right) + \sum_{r \neq i}\left[\left(\frac{a_r w_r}{\left(\sum_{k=1}^{m} w_k\right)^2}\right)^2 \times n_{ir}\left(1 - \frac{n_{ir}}{n_{i\cdot}}\right)\right] \quad (2)$$

where $a_k = p_{C_k}$ and $w_k = p_{C_k}(n_{ik}/n_{i\cdot})$.

An approximate 95 percent confidence interval for $p_{C_i|A_j}$ may then be obtained as

$$\hat{p}_{C_i|A_j} \pm \sqrt{\hat{v}(\hat{p}_{C_i|A_j})}$$

Application 1

We use the artificial data in Table 1 for illustration purposes. Here, $m = 4$. First, we compute $\hat{p}_{A_j|C_i}$ and $\sqrt{\hat{v}(\hat{p}_{A_j|C_i})}$, using the multinomial formulae described above. The results are presented in Tables 2 and 3.

Now, suppose the data were obtained from four simple random samples (one in each classified category). Further, suppose that, over the entire map, 25, 35, 35, and 5 percent of the pixels were classified as forest, agriculture, residential, and water, respectively. Hence, $p_{C_1} = 0.25$, $p_{C_2} = 0.35$, $p_{C_3} = 0.35$, and $p_{C_4} = 0.05$. Then, our estimates of $\hat{p}_{C_i|A_j}$ and

TABLE 2. ESTIMATES OF $p_{A_j|C_i}$ FOR DATA IN APPLICATION 1, USING MULTINOMIAL FORMULA.

		Reference Data			
		F	A	R	W
Classified Data	F	0.80	0.08	0.12	0.00
	A	0.04	0.84	0.08	0.04
	R	0.28	0.32	0.40	0.00
	W	0.00	0.08	0.00	0.92

TABLE 3. ESTIMATES OF $\sqrt{v(\hat{p}_{A_j|C_i})}$ FOR DATA IN APPLICATION 1, USING MULTINOMIAL FORMULA (VALUES SHOWN ARE $10^3 \times$ ACTUAL VALUE).

		Reference Data			
		F	A	R	W
Classified Data	F	80	54	65	0
	A	39	73	54	39
	R	90	93	98	0
	W	0	54	0	54

TABLE 4. ESTIMATES OF $p_{A_j|C_i}$ FOR DATA IN APPLICATION 1, USING BAYES THEOREM.

		Reference Data			
		F	A	R	W
Classified Data	F	0.64	0.05	0.15	0.00
	A	0.05	0.68	0.14	0.23
	R	0.31	0.26	0.71	0.00
	W	0.00	0.01	0.00	0.77

TABLE 6. ESTIMATES OF $p_{C_i|A_j}$ FOR DATA IN APPLICATION 1, USING MULTINOMIAL FORMULA.

		Reference Data			
		F	A	R	W
Classified Data	F	0.71	0.06	0.20	0.00
	A	0.04	0.64	0.13	0.04
	R	0.25	0.24	0.67	0.00
	W	0.00	0.06	0.00	0.96

$\sqrt{\hat{v}(\hat{p}_{C_i|A_j})}$ should be obtained with Equations 1 and 2. These results are given in Tables 4 and 5.

Finally, suppose we fail to recognize the fact that, due to the randomization within classified categories, we do not have random samples from the classified categories within actual categories, and we mistakenly compute $\hat{p}_{C_i|A_j}$ using the usual multinomial formula. Then we obtain the results given in Table 6.

Evidently, one could get very misleading answers if the randomization structure is not recognized and taken into account. For instance, the probability of a pixel being classified into category 4 given that it actually is in category 4 is 0.77. However, if we naively use multinomial formulae to compute $\hat{p}_{C_i|A_j}$, we would estimate the same probability as 0.96.

Application 2

In New Jersey, the use of Landsat Thematic Mapper (TM) data for land-cover classification has been evaluated for improving the accuracy of the estimates derived as part of the mandated Soil Conservation Service National Resource Inventory (NRI) process (Airola and Vogel, 1988). The initial work, undertaken as part of the 1987 NRI, focused on a quadrangle sized study site. The data used in the present study were the result of a land-cover classification that was developed for the entire state of New Jersey as part of the ongoing 1992 NRI. This work has been funded through a cooperative agreement designating both the New Jersey SCS and the Cook College Remote Sensing Center as one of four Inventory Development Field Sites , charged with evaluating technologies designed to streamline the NRI process.

Current land-cover information for the state of New Jersey was derived from TM digital data. Three relatively cloud-free TM digital data sets providing coverage of the entire state were acquired from EOSAT for 17 March 1991. Analysis of the data was accomplished using the ERDAS image processing software package running on both SUN workstations and PCs using a local area network for data storage and file transfer.

Following receipt of the digital image data from EOSAT (path 14, rows 31, 32, and 33), the data were preprocessed, merged, and subdivided to include only coverage of the state. In order to facilitate processing and to differentiate between the physiographic provinces of the state, the image data set was further subdivided to approximate the boundary between the southern coastal plain and the northern (piedmont, highlands, and ridge and valley) provinces. As a consequence of the early spring date on which the data were acquired, a significant area in the northern portion of the state was obscured by snow from artificial snow-making at a number of recreational facilities. Using a masking procedure, cloud-free TM imagery from 15 August 1988 was merged into the data set to eliminate those areas that were snow covered at the later date.

Land-cover data were then generated for the masked area, the northern, and the southern regions of the state using a hybrid unsupervised-supervised classification approach. Each of the data sets was initially classified using an unsupervised clustering approach to generate between 50 and 100 clusters. The results were then used as input to a maximum-likelihood classification algorithm. Visual interpretation of the image data, color infrared aerial photography, and existing USGS quadrangle maps and orthophotos were used to collapse the resulting land-cover categories into six possible land-cover types: forest, nonforest vegetation, built-up/developed, bare or barren, water, and cloud. In order to reduce noise in the classified data sets, a 3 by 3 majority smoothing filter was used prior to evaluating the results.

The results of the classification procedure for the entire state are presented in Table 7. The majority of the state was classified as being either forest or in the non-forest vegetation class (which includes agricultural land cover categories). Built-up land was estimated to include approximately 11 percent of the state, while water surfaces, including coastal embayments, accounted for approximately 16 percent. Both cloud and barren categories accounted for less than 1 percent.

In order to assess the accuracy of the results, an ERDAS subroutine was used to select 100 random points from each of the three classified sub-images (southern provinces, northern provinces, and masked areas). The strata were the land-cover classes, and the sample points were distributed among the strata using proportional allocation (i.e., if 40 percent of the pixels in a sub-image were classified as forest, then 40 percent of the sample would consist of pixels classified as forest). Each point was then displayed over selected bands of the raw image data and the land cover class determined by an interpreter using the image data and the available ancillary land cover information. The results of the interpretation were then combined and the error matrix in Table 8 was produced. Using the multinomial formulae described earlier, we

TABLE 5. ESTIMATES OF $\sqrt{v(\hat{p}_{A_j|C_i})}$ DATA IN APPLICATION 1, USING ESTIMATOR DERIVED IN APPENDIX 1 (VALUES SHOWN ARE $10^3 \times$ ACTUAL VALUE)

		Reference Data			
		F	A	R	W
Classified Data	F	74	30	76	0
	A	42	60	87	176
	R	73	59	103	0
	W	0	6	0	176

TABLE 7. RESULTS OF THE CLASSIFICATION PROCEDURE

Land cover Class	Percentage
Forest	37.62
Nonforest	34.36
Built-up	11.41
Barren	0.47
Water	16.06
Cloud	0.08

TABLE 8. ERROR MATRIX FOR THEMATIC MAP OF LAND COVER IN NEW JERSEY, WITH SIX CATEGORIES: FOREST (F), NONFOREST (N), BUILT-UP/DEVELOPED (D), BARE/BARREN (B), WATER (W), AND CLOUD (C).

		Reference Data						
		F	N	D	B	W	C	TOTAL
	F	129	11	6	0	6	0	146
	N	8	71	9	0	0	0	88
Classified	D	5	2	25	0	0	0	32
Data	B	0	0	0	1	0	0	1
	W	0	0	0	0	32	0	32
	C	0	0	0	0	0	1	1
	Total	142	84	40	1	32	1	300

compute $\hat{p}_{A_j|C_i}$ and $\sqrt{\hat{v}(\hat{p}_{A_j|C_i})}$; these values are shown in Tables 9 and 10.

Our estimates of $\hat{p}_{C_i|A_j}$ and $\sqrt{\hat{v}(\hat{p}_{C_i|A_j})}$ obtained with Equations 1 and 2 (using the above percentages of the state classified into each cover class for p_{C_i}) are presented in Tables 11 and 12.

Again, suppose we fail to recognize the fact that, due to the randomization within classified categories, we do not have random samples within actual categories, and we mistakenly compute $\hat{p}_{C_i|A_j}$ using the usual multinomial formulae. Then we obtain the results presented in Table 13.

While the differences between the estimates of $p_{C_i|A_j}$ obtained using Equation 1 and those obtained by incorrectly using the multinomial formulae are not as great as in Application 1, it is clear that the two methods yield different estimates. For example, the correct estimate (obtained using Equation 1) of the probability that a pixel which is truly forested will be classified as such is 0.87. If we mistakenly use the multinomial formulae, we arrive at an estimate of 0.91.

Conclusions

We have seen that it is necessary to recognize any restrictions on the randomization when pixels are selected to check a mapping classification. If one random sample is drawn, without regard to the map classification, then it is appropriate to use the usual multinomial formulae for both $p_{C_i|A_j}$ and $p_{A_j|C_i}$. If, however, separate random samples are drawn from the populations of pixels classified into each category, then it is necessary to use Bayes' theorem to recover estimates of $p_{C_i|A_j}$ from estimates of $p_{A_j|C_i}$. We have shown how to do this, and presented a method for approximating the variance of the resulting estimates. In the latter case, multinomial formulae are still appropriate for $p_{A_j|C_i}$.

Finally, if somehow one selected separate random samples from each category in the reference data, then the above procedure would be reversed. Multinomial formulae would be used to estimate $p_{C_i|A_j}$ and Bayes theorem would be used to estimate $p_{A_j|C_i}$.

TABLE 9. ESTIMATES OF $p_{A_j|C_i}$ FOR DATA IN APPLICATION 2, USING MULTINOMIAL FORMULA.

		Reference Data					
		F	N	D	B	W	C
	F	0.88	0.08	0.04	0	0	0
	N	0.09	0.81	0.10	0	0	0
Classified	D	0.16	0.06	0.78	0	0	0
Data	B	0	0	0	1.00	0	0
	W	0	0	0	0	1.00	0
	C	0	0	0	0	0	1.00

TABLE 10. ESTIMATES OF $\sqrt{\hat{v}(\hat{p}_{A_j|C_i})}$ FOR DATA IN APPLICATION 2, USING MULTINOMIAL FORMULA (VALUES SHOWN ARE $10^4 \times$ ACTUAL VALUE).

		Reference Data					
		F	N	D	B	W	C
	F	265	218	164	0	0	0
	N	306	421	323	0	0	0
Classified	D	642	428	731	0	0	0
Data	B	0	0	0	0	0	0
	W	0	0	0	0	0	0
C	0	0	0	0	0	0	0

Acknowledgments

This is New Jersey Agricultural Experiment Station Publication No. D-155033-1-92, supported by state funds, the New Jersey Council on Affordable Housing, and through a cooperative research agreement with the Soil Conservation Service.

References

Airola, T.M., and J. Vogel, 1988. Use of thematic mapper digital data for updating the New Jersey land cover component of the 1987 National Resources Inventory. *Journal of Soil and Water Conservation* 43:425-428.

Airola, T.M., J. Gasprich, E.J. Green, and W.E. Strawderman, 1992. Estimating the undeveloped land in New Jersey using thematic mapper data and the US Census Tiger files (manuscript in preparation.

Aronoff, S., 1985. The minimum accuracy value as an index of classification accuracy. *Photogrammetric Engineering & Remote Sensing* 51:99-111.

Bickel, P.J., and K.A. Doksum, 1977. *Mathematical Statistics: Basic Ideas and Selected Topics.* Holden-Day, San Francisco, California.

Berger, J.O., 1985. *Statistical Decision Theory and Bayesian Analysis*, Second Edition. Springer-Verlag, New York, N.Y.

Box, G.E.P., and G.C. Tiao, 1973. *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, Massachusetts.

Cochran, W.G., 1977. *Sampling Techniques*, Third Edition. John Wiley and Sons, New York, N.Y.

Lehmann, E.L., 1983. *Theory of Point Estimation.* John Wiley and Sons, New York, N.Y.

Mendenhall, W., and R.L. Scheaffer, 1973. *Mathematical Statistics with Applications.* Duxbury Press, North Scituate, Massachusetts.

Story, M., and R.G. Congalton, 1986. Accuracy assessment: A user's perspective. *Photogrammetric Engineering & Remote Sensing* 52:397-399.

(Received 27 April 1992; revised and accepted 20 October 1992)

Appendix

Here we derive an approximate variance for $\hat{p}_{C_i|A_j}$ when this probability is obtained by means of Bayes theorem. From

TABLE 11. ESTIMATES OF $p_{C_i|A_j}$ FOR DATA IN APPLICATION 2, USING BAYES THEOREM.

		Reference Data					
		F	N	D	B	W	C
	F	0.87	0.09	0.11	0	0	0
	N	0.08	0.89	0.25	0	0	0
Classified	D	0.05	0.02	0.64	0	0	0
Data	B	0	0	0	1.00	0	0
	W	0	0	0	0	1.00	0
	C	0	0	0	0	0	1.00

TABLE 12. ESTIMATES OF $\sqrt{v(\hat{p}_{C_i|A_j})}$ FOR DATA IN APPLICATION 2, USING ESTIMATOR DERIVED IN APPENDIX (VALUES SHOWN ARE 10^4 × ACTUAL VALUE).

		Reference Data					
		F	N	D	B	W	C
Classified Data	F	295	243	408	0	0	0
	N	255	276	623	0	0	0
	D	184	153	619	0	0	0
	B	0	0	0	0	0	0
	W	0	0	0	0	0	0
	C	0	0	0	0	0	0

TABLE 13. ESTIMATES OF $p_{C_i|A_j}$ FOR DATA IN APPLICATION 2, USING MULTINOMIAL FORMULA.

		Reference Data					
		F	N	D	B	W	C
Classified Data	F	0.91	0.13	0.15	0	0	0
	N	0.06	0.85	0.23	0	0	0
	D	0.03	0.02	0.62	0	0	0
	B	0	0	0	1.00	0	0
	W	0	0	0	0	1.00	0
	C	0	0	0	0	0	1.00

Equation 1 in the text we have

$$\hat{p}_{C_i|A_j} = (\hat{p}_{A_j|C_i} \cdot p_{C_i})/\hat{p}_{A_j} = \frac{(n_{ij}/n_{i\cdot})p_{C_i}}{\sum_{k=1}^{m} (n_{kj}/n_{k\cdot})p_{C_k}}$$

As mentioned in the text, we may regard p_{C_i} as known. For notational convenience, we shall write $a_i = p_{C_i}$. Thus, we have

$$\hat{p}_{C_i|A_j} = \frac{(n_{ij}/n_{i\cdot})a_i}{\sum_{k=1}^{m} (n_{kj}/n_{k\cdot})a_k} = f(n_{1j}, n_{2j}, \ldots, n_{mj})$$

Now, we estimate the variance of $\hat{p}_{C_i|A_j}$ by a Taylor Series expansion about n_{ij} (see, e.g., Lehmann (1983), Theorem 5.1, p 106; or Bickel and Doksum (1977), p. 31):

$$v(\hat{p}_{A_j|C_i}) \approx \sum_{k=1}^{m} \left[\left(\frac{\partial f}{\partial n_{kj}} \right)^2 \Big|_{p_{C_k|A_j}} \times p_{C_k|A_j}(1 - p_{C_k|A_j})n_{kj} \right]$$

Taking derivatives, evaluating at $\hat{p}_{C_i|A_j}$, letting $a_k = p_{C_k}$ and $w_k = p_{C_k}n_{ik}/n_{i\cdot}$), and rearranging terms, we have

$$\hat{v}(\hat{p}_{C_i|A_j}) = \left[a_i \left(\frac{1}{\sum_{k=1}^{m} w_k} - \frac{w_i}{\left(\sum_{k=1}^{m} w_k\right)^2} \right) \right]^2 \times n_{ii}\left(1 - \frac{n_{ii}}{n_{i\cdot}}\right) + \sum_{r \neq i} \left[\left(\frac{a_r w_r}{\left(\sum_{k=1}^{m} w_k\right)^2} \right)^2 \times n_{ir}\left(1 - \frac{n_{ir}}{n_{i\cdot}}\right) \right]$$

Note that this is a large sample approximation. Caution is advised for applications where $n_{\cdot\cdot}$ is small, say $n_{\cdot\cdot} < 30$. Furthermore, note that this approximation is based on the normal approximation to a multinomial probability. The latter is justified by the Central Limit Theorem, and should be a close approximation for moderate sample sizes, say $n_{i\cdot} \geq 15$, $\forall i$.