Unbiased Estimates of Class Proportions from Thematic Maps

Paul C. Van Deusen

Abstract

A statistical overview is presented for estimating various components related to map accuracy assessment. The emphasis is on estimation of the true proportions of each map class under several common sampling designs. A complete system is presented for relating alternative approaches and estimators using standard rules of probability theory. Covariance matrices for estimates of true class proportions are derived in the Appendices for each of the sampling designs discussed.

Introduction

The traditional approach to assessing map accuracy is based on the assumption that each pixel within the map has a correct classification, but errors in classification can be made. Unbiased estimation of true class proportions is based on a subsample where the true classification is obtained by an infallible and expensive method so that both the true and map classifications are known. The subsample results are then extrapolated to the entire map utilizing statistical procedures that are discussed below.

Tenenbein (1972) gives an early development based on a double sampling scheme. Card (1982) recognizes that applications to thematic maps are privy to complete knowledge of the map class marginal proportions, and he suggests appropriate alterations to Tenenbein's method. Grassia and Sundberg (1982) present a method for calibrating sorting machines that can also be applied to thematic maps. While Card's method requires a subsample after the map is made, Grassia and Sundberg's approach could utilize the training data acquired to calibrate the classification algorithm.

Bauer *et al.* (1978) and Hay (1988) are examples from the remote sensing literature that are similar to Grassia and Sundberg's method. Czaplewski and Catts (1992) compare Grassia and Sundberg's (1982) estimator and Tenenbein's (1972) estimator in a simulation study. While this comparison is valid, it does not recognize the improvement made to Tenenbein's method by Card (1982) for the special circumstances of thematic map accuracy assessment. Green *et al.* (1993) and Card (1982) present a Bayesian derivation of the producer's risk, a term used (Aronoff, 1982; Aronoff, 1985) to describe the conditional probability of the map class given the true class, say $p(m \mid t)$. Prisley and Smith (1987) and Story and Congalton (1986) both state that the user's accuracy, which is analogous to the consumer's risk and $p(t \mid m)$, is commonly estimated by dividing the number of sample observations correctly classified as category X by the total number of category X ground samples. Story and Congalton (1986) state that an alternative method is to divide the number of correctly classified samples of category X by the total number of samples classified as category X.

This paper also looks at alternative methods for computing producer's and consumers risk, but the emphasis is on how the sampling procedure used in deriving the confusion matrix determines the correct alternative to use. This paper lays out notation and a unifying statistical theory that encompasses both Card's (1982) approach and Grassia and Sundberg's (1982) method. Key points made are that estimators should depend on the sampling design and that known map marginal frequencies should be maintained by the estimators.

Notation and Probability Formulas

Assume there is a thematic map consisting of N pixels where each is assigned to 1 of K unordered classes. The information on the proportion assigned to class i but belonging in truth to class j can be summarized in a K by K table with the rows representing the map class and columns the true class. The number of entries in a cell for the population is denoted by N_{ij}, and these are easily converted to proportions because we know the total N. Let N_{i+} represent the row totals and N_{+j} represent the column totals with $N_{++} = N$. This two-way table can be referred to as the confusion or error matrix (Story and Congalton, 1986; Prisley and Smith, 1987). Corresponding sample value entries, n_{ij}, of the confusion matrix might be based on a simple random sample of size n.

It is convenient to work in terms of two-way discrete distributions to derive relevant statistical formulas for map accuracy assessment, and some basic probability formulas are now given. First, we have the joint distribution $p(m,t)$ whose entries are the same as the population level confusion matrix with each element divided by N, i.e., $p(m,t) = \{N_{ij}/N\}$. Now we have the marginal distributions, $p(m)$ and $p(t)$, which are K by 1 vectors containing the proportions of pixels in each class according to the map and in truth.

The usual situation is that $p(m)$ is known and $p(t)$ is unknown, but an estimate of $p(t)$ is desired. Using matrix notation and letting $\mathbf{1}$ be a column vector of ones, we have

$$p(m) = p(m, t) * \mathbf{1} \tag{1a}$$

and

$$p(t) = p(m, t)' * \mathbf{1} . \tag{1b}$$

Then we need the standard relations on conditional distributions: i.e.,

$$p(m, t) = p(m \mid t) * \text{diag}\,[p(t)] \tag{2a}$$

and

NCASI, Northeast Regional Center, Tufts University, Anderson Hall, Medford, MA 02155.

Photogrammetric Engineering & Remote Sensing, Vol. 62, No. 4, April 1996, pp. 409–412.

0099-1112/96/6204–409$3.00/0

$$p(m, t) = \text{diag}\,[p(m)] * p(t \mid m) \quad (2b)$$

where the conditional distributions, $p(m \mid t)$ and $p(t \mid m)$, have been called the producer's and consumer's risk (Aronoff, 1982; Aronoff, 1985). Both $p(m \mid t)$ and $p(t \mid m)$ can be displayed as matrices with the rows representing the levels of m amd the columns representing the levels of t. Rearranging Equation 2a, we get the producer's risk formula

$$p(m \mid t) = p(m, t) * \text{diag}\,[p(t)]^{-1} \,. \quad (3a)$$

From Equation 2b, we get the consumer's risk formula

$$p(t \mid m) = \text{diag}\,[p(m)]^{-1} * p(m, t) \,. \quad (3b)$$

By applying Equation 1b to Equation 2b, we get a second useful formula for $p(t)$,

$$p(t) = p(t \mid m)' * p(m), \quad (4a)$$

and likewise from Equations 1a and 2a we get

$$p(m) = p(m \mid t) * p(t) \,. \quad (4b)$$

By setting Equation 2a equal to Equation 2b and rearranging terms, we can relate the producer's risk to the consumer's risk as

$$p(m \mid t) = \text{diag}\,[p(m)] * p(t \mid m) * \text{diag}\,[p(t)]^{-1} \,. \quad (5)$$

Equation 5 is also known as Baye's rule, utilized by Green *et al.* (1993) and Card (1982) in their derivations of the producer's risk.

The final result given for this section is the basic formula used by Grassia and Sundberg (1982) for estimating $p(t)$ from calibration data. This is simply derived from Equation 4b as

$$p(t) = p(m \mid t)^{-1} * p(m) \,. \quad (6)$$

In practice, $p(m \mid t)$ for application of Equation 6 could come from the training data that were used to calibrate the classifier that produced the map.

Whenever it is necessary to refer to specific elements in a marginal, joint, or conditional distribution in this paper, it is done with the convention that a map class is denoted by i and the true class by j. Thus, $p(m{=}i \mid t{=}j)$ is the conditional probability of map class i given that the true class is j.

Sampling and Point Estimation

A number of *ad-hoc* estimators for the various components of thematic map accuracy assessment could be devised. In order to limit the list of such estimators, several criteria are given that an estimator should meet:

(1) The estimator should depend on the sampling design,
(2) The estimator should be compatible with the known map-class marginal distribution, and
(3) The estimator should yield estimates that are compatible with Equations 1a, and 1b, 2a, and 2b.

Estimators that do not satisfy the above criteria will be called incompatible and should not be used without special justification. Criteria 2 and 3 require specifically that $\hat{p}(m,t)*1 = p(m)$ and that the equalities in Equations 2a and 2b hold when the estimates are substituted in. Criterion 1 requires that the randomization process used in sampling pixels for their true class (or map class) be considered. The implications of this are discussed next. Maximum-likelihood estimators will satisfy the above criteria, but they are not necessarily the only estimators that will.

Simple Random Sampling after the Map

This discussion relates to the common situation where a map has been produced and one wishes to make estimates of $p(t)$, $p(t \mid m)$, $p(m \mid t)$, and/or $p(m,t)$. With simple random sampling (SRS), pixels are chosen from the map at random for determination of their true class membership. The K by K matrix that summarizes these results with elements $\{n_{ij}/n\}$ is an unbiased estimate of $p(m,t)$, say $\hat{p}(m,t)$. We have already met compatibility Criterion 1 by recognizing that the estimate of $p(m,t)$ comes directly from the sample data with SRS. The second compatibility criterion is not met by SRS alone, because $\hat{p}(m,t)*1 = \hat{p}(m)$ is not necessarily equal to the known $p(m)$. The marginal distribution of $\hat{p}(m,t)$ can be corrected to equal the known marginal, $p(m)$, as follows:

$$\hat{p}_c\,(m, t) = \text{diag}\,[p(m)/\hat{p}(m)] * \hat{p}(m, t) \quad (7)$$

$$\hat{p}(t) = \hat{p}_c\,(m, t)'\,\mathbf{1} = \left\{ \sum_{i=1}^{K} \frac{n_{ij}}{n_{i+}}\; p(m = i) \right\} \quad (8a)$$

The elements of Equation 7 are $\{p(m{=}i)*n_{ij}/n_{i+}\}$, which are identical to the estimates proposed by Card (1982). Now with $\hat{p}_c(m,t)$ being compatible with the known $p(m)$, the other estimates follow directly from the equations given previously; e.g., from Equations 1b, 3b, and 3a, we have

$$\hat{p}(t \mid m) = \text{diag}\,[p(m)]^{-1} * \hat{p}_c\,(m, t) = \left\{ \frac{n_{ij}}{n_{i+}} \right\} \quad (8b)$$

$$\hat{p}(m \mid t) = \hat{p}_c(m, t) * \text{diag}\,[\hat{p}(t)]^{-1} = \left\{ \frac{\frac{n_{ij}}{n_{i+}}\, p(m{=}i)}{\hat{p}(t{=}j)} \right\} \quad (8c)$$

Variance approximations for these formulas under SRS and stratified sampling can be found in Card (1982). However, in Appendix A it is shown that an approximate covariance matrix for $\hat{p}(t)$ is

$$\mathbf{V}[\hat{p}(t)] = \Big[\text{diag}[p(t)] - p(t) * p\,(t)' \Big] / n \,. \quad (8d)$$

A complete covariance matrix is necessary for placing confidence intervals on linear functions of the elements of $p(t)$, e.g., $p(t{=}j){+}p(t{=}h)$. Equation 8d is used by substituting current estimates for $p(t)$.

Stratified Sampling after the Map

After the thematic map is produced, the map class of each pixel is known. Therefore, it makes sense to perform sampling independently within each of the known map strata. This allows one to control how much effort is devoted to each map class and avoids the possibility that some classes are missed entirely, which might happen with SRS. This process involves drawing n_{i+} samples from stratum i, which produces an unbiased estimate of $p(t \mid m)$ with elements $\{n_{ij}/n_{i+}\}$ and each row statistically independent.

Now we simply follow the formulas, as was done for SRS, but recognizing that the stratified sample produces an unbiased estimate of $p(t \mid m)$ rather that $p(m,t)$ as with SRS. It is instructive to look at the result for $\hat{p}(m, t)$ under stratified sampling: i.e.,

$$\hat{p}(m, t) = \text{diag}\,[p(m)] * \hat{p}(t \mid m) = \left\{ p(m = i) * \frac{n_{ij}}{n_{i+}} \right\} . \quad (9a)$$

The elements of $\hat{p}(m,t)$ are the same as the constrained SRS estimates as Card (1982) shows. In fact, all of the point estimators are the same for simple random or stratified sampling due to the constraint imposed by the known $p(m)$. Alternative variance approximations to those of Card (1982) for $p(m \mid t)$ can be found in Green *et al.* (1993). A complete covariance matrix for $\hat{p}(t)$ is derived in Appendix B. Because no asymptotic approximations are used in Appendix B, this result is preferable to previously derived variance approximations.

Stratified Sampling before the Map

The discussion here relates to the use of training data for obtaining improved estimates of the true proportions, $p(t)$, from Equation 6. Grassia and Sundberg (1982) give variance estimates for the case where the map marginals, $p(m)$, are unknown. The appropriate modifications to their covariance estimators are given in Appendix C, because $p(m)$ is known for thematic maps.

Training data are usually collected in advance of applying a statistical classifier in order to provide estimates for the classifier's unknown parameters. For example, the standard maximum-likelihood classifier requires estimates of the mean and covariance matrix for the spectral bands representing each class. The training data are obtained from pixels where the true classification is known in advance. After classifying the map, the training data provide an estimate of $p(m \mid t)$ that can be used with Equation 6 to estimate $p(t)$.

Conclusions

A thematic map requires an associated accuracy assessment to be fully useful to its users. The purpose of this paper is to show how the sampling process affects estimation of the components of accuracy assessment, e.g., the producers risk, the consumers risk, and the map and true marginal proportions. It is hoped that this systematic approach will aid in future applications of map accuracy assessment. Related methods can be used for statistically assessing change in true class proportions over time (Van Deusen, 1994).

Full covariance matrices, which were not previously available in the literature, are derived for estimates of the true map proportions $p(t)$, and are given in the Appendices for each sampling scheme. The results in Appendix B for stratified sampling require no asymptotic approximations and can be corrected for bias according to a result from Cochran (1977).

The final section discussed the use of training data in correcting for misclassification bias. This is an approach that has seen little use in the remote sensing literature, but should be widely applicable. Training data are usually acquired before a map can be made, and this approach provides a low-cost method to obtain improved estimates of the true class proportions. The caveat here is that the training data should provide an unbiased estimate of the producer's risk $p(m \mid t)$. If for some reason the training data are not representative of the entire map, the approach should be used with caution. However, if this were the case, then the entire map should be used with caution.

References

Aronoff, S., 1982. Classification accuracy: A user approach, *Photogrammetric Engineering & Remote Sensing*, 48:1299–1307.

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

Bauer, M.E., M.M. Dixon, B.J. Davis, and J.B. Etheridge, 1978. Area estimation of crops by digital analysis of Landsat data, *Photogrammetric Engineering & Remote Sensing*, 44:1033–1043.

Card, D.H., 1982. Using known map category marginal frequencies to improve estimates of thematic map accuracy, *Photogrammetric Engineering & Remote Sensing*, 48:431–439.

Cochran, W.G., 1977. *Sampling Techniques*, Third edition, John Wiley & Sons, New York, 427 p.

Czaplewski, R.L., and G.P. Catts, 1992. Calibration of remotely sensed proportion or area estimates for misclassification error, *Remote Sensing of Environment*, 39:29–43.

Grassia, A., and R. Sundberg, 1982. Statistical precision in the calibration and use of sorting machines and other classifiers, *Technometrics*, 24:117–121.

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

Hay, A.M., 1988. The derivation of global estimates from a confusion matrix, *International Journal of Remote Sensing*, 9:1395–1398.

Prisley, S.P., and J.L. Smith, 1987. Using classification error matrices to improve the accuracy of weighted land–cover models, *Photogrammetric Engineering & Remote Sensing*, 53:1259–1263.

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

Tenenbein, A., 1972. A double sampling scheme for estimating from misclassified multinomial data with applications to sampling inspection, *Technometrics*, 14(1):187–202.

Van Deusen, P.C., 1994. Correcting bias in change estimates from thematic maps, *Remote Sensing of Environment*, 50:67–73.

(Received 28 October 1993; revised and accepted 20 December 1994)

Appendix A

An asymptotic approximation for the covariance matrix of $\hat{p}(t)$ under simple random sampling is derived. This result is derived as if the constraint on the map marginal, $p(m)$, is irrelevant. Although this is true asymptotically, the effect of this constraint is to make the actual variance of $\hat{p}(t)$ smaller than it would be under SRS. Thus, the derivation here provides a conservative estimate.

The elements of $\hat{p}(m,t)$ have a multinomial distribution with parameter vector $\pi = \text{VEC}(p(m,t))$ where VEC denotes stacking the columns of $p(m,t)$ so that π is a K^2 by 1 vector. In general, the variance of an element in $\hat{p}(m,t)$ is

$$\mathbf{V}[\hat{p}(m = i, t = j)] = \frac{1}{n}\, p\,(m = i, t = j) * \Big(1 - p(m = i, t = j)\Big) \quad \text{(A1)}$$

and the covariance between the two elements in $\hat{p}(m,t)$ is

$$\mathbf{C}[\hat{p}(m = i, t = j), \hat{p}(m = h, t = j)] = -\frac{1}{n}\, p(m = i, t = j) * p(m = h, t = j)\,. \quad \text{(A2)}$$

The entire covariance matrix of $\text{VEC}(\hat{p}(m,t))$ is therefore

$$\mathbf{V}(\hat{\pi}) = \Big[\text{diag}\,(\pi) - \pi * \pi'\Big]/n\,. \quad \text{(A3)}$$

We will use Equation A3 as an asymptotic approximation to the covariance of $\text{VEC}(\hat{p}_c(m,t))$.

Now write $\hat{p}(t)$ as

$$\hat{p}(t) = \mathbf{J} * \hat{\pi} \quad \text{(A4)}$$

where $\mathbf{J}$ is a K by K^2 matrix with row j having 1's in positions $(jK\text{-}1)+1$ through jK and 0's elsewhere. So row j of $\mathbf{J}$ picks off the jth column of $p_c(m,t)$ from π and sums it to get $p(t{=}j)$. Now the covariance matrix we seek is

$$\mathbf{V}(\hat{p}(t)) = \mathbf{J} * \mathbf{V}(\hat{\pi}) * \mathbf{J}' = \frac{1}{n} * \Big[\mathbf{J} * \text{diag}\,(\pi) * \mathbf{J}' - \mathbf{J} * \pi * \pi' * \mathbf{J}'\Big]\,. \quad \text{(A5)}$$

After some algebraic manipulation this result can be shown to equal

$$\mathbf{V}[\hat{p}(t)] = \left[\text{diag}[p(t)] - p(t) * p(t)'\right]/n . \tag{A6}$$

The resulting Equation A6 is the same as Equation 8d and is used by replacing unknown values with current estimates.

Appendix B

The covariance matrix is derived here for $\hat{p}(t)$ under stratified sampling. Each row of the estimated $p(t \mid m)$ matrix has an independent multinomial distribution with parameter vector $p(t \mid m = i)$ under stratified sampling. Therefore, each row of $\hat{p}(t \mid m)$ has its own K by K covariance matrix, $\mathbf{V}_i[\hat{p}(t \mid m = i)]$. By well known results on multinomial covariances

$$\mathbf{V}_i = \left[\text{diag}(p(t \mid m = i)) - p(t \mid m = i) * p\,(t \mid m = i)'\right]/n_{i+} . \tag{B1}$$

The appropriate formula here for $\hat{p}(t)$ is

$$\hat{p}(t) = \hat{p}(t \mid m)' * p(m) . \tag{B2}$$

The fact that each row of $\hat{p}(t \mid m)$ is independent leads to the following result:

$$\mathbf{V}(\hat{p}(t)) = \sum_{i=1}^{K} p\,(m = i)^2\, \mathbf{V}_i . \tag{B3}$$

So the covariance matrix of $\hat{p}(t)$ under stratified sampling is a weighted sum of the individual covariance matrices of the independent rows of $\hat{p}(m \mid t)$. Cochran (1977) shows that an unbiased estimate of multinomial variances results by dividing by 1 less than the number of sample observations to correct for degrees of freedom. Therefore, if n_{i+} is replaced by (n_{i+} -1) in Equation B1, the overall result in Equation B3 is unbiased. Because the derivations in Appendices A and C involve asymptotic approximations, there is no point in correcting for degrees of freedom for those cases.

Appendix C

The covariance matrix of $\hat{p}(t)$ is derived here closely following the procedures in Grassia and Sundberg (1982). The formula for $\hat{p}(t)$ is

$$\hat{p}(t) = \hat{p}(m \mid t)^{-1} * p(m) . \tag{C1}$$

This formula is useful when training data are available to provide an estimate of the producer's risk, $p(m \mid t)$. To abbreviate the notation used here, we rewrite Equation C1 as

$$\hat{\mathbf{T}} = \hat{\mathbf{A}}^{-1} * \mathbf{B} . \tag{C2}$$

Following Grassia and Sundberg, approximate $\hat{\mathbf{A}}^{-1}$ by

$$\hat{\mathbf{A}}^{-1} - \hat{\mathbf{A}}^{-1} \cong -\mathbf{A}^{-1}\,(\hat{\mathbf{A}} - \mathbf{A})\,\mathbf{A}^{-1} . \tag{C3}$$

Now substitute Equation C3 into Equation C2 to get a linearized approximation to Equation C2: i.e.,

$$\hat{\mathbf{T}} \approx \mathbf{T} - \mathbf{A}^{-1}\,(\hat{\mathbf{A}} - \mathbf{A})\,\mathbf{T} . \tag{C4}$$

It is clear that $\hat{\mathbf{T}}$ is unbiased and its covariance matrix is approximately

$$\mathbf{V}(\hat{\mathbf{T}}) \approx \mathbf{A}^{-1}\,\mathbf{V}\,(\hat{\mathbf{A}}\mathbf{T})\,\mathbf{A}^{-1'} \tag{C5}$$

Now remembering that $\mathbf{A} = p(m \mid t)$ and $\mathbf{T} = p(t)$, we need to compute $\mathbf{V}(\hat{\mathbf{A}}\mathbf{T})$.

The training data are collected from areas where the true class is known, and therefore each column of the estimated $p(m \mid t)$ has an independent multinomial distribution with parameter vector $p(m \mid t = j)$. Therefore, each column of $\hat{p}(m \mid t)$ has its own K by K covariance matrix, $\mathbf{V}_j[\hat{p}(m \mid t = j)]$, which is derived analogously to covariance matrix $\mathbf{V}_i$ in Appendix B. Now the variance we need for Equation C5 is

$$\mathbf{V}[\hat{p}(m \mid t) * p(t)] = \sum_{j=1}^{K} p(t = j)^2 * \mathbf{V}_j\,[\hat{p}(m \mid t = j)] . \tag{C8}$$

Equation C8 gives the intuitively appealing result that the overall variance will involve a weighted sum of the variances of the individual covariance matrices of the columns of $\hat{p}(m \mid t)$.

Rewriting Equation C5 using the full notation gives the final result

$$\mathbf{V}[\hat{p}(t)] = p(m \mid t)^{-1} * \mathbf{V}[\hat{p}(m \mid t) * p(t)] * p(m \mid t)^{-1'} . \tag{C9}$$

In order to apply Equation C9 in practice, all of the unknown values are replaced with their current estimates.