Don H. Card
NASA Ames Research Center
Moffett Field, CA 94035

Using Known Map Category Marginal Frequencies to Improve Estimates of Thematic Map Accuracy

Two statistical sampling schemes are discussed: simple random sampling of single points and random sampling stratified by map category.

Abstract: *It is generally recognized that estimating the accuracy of maps that are derived from remotely sensed data requires statistical sampling of photographs or ground plots to insure that the estimates are reliable and cost effective. The usual method is to cross-tabulate the categories identified for these plots with the categories associated with corresponding areas on the map in a table called a "contingency table." From this table, measures of map accuracy ("proportion-correct") are usually obtained by ratioing diagonal entries by marginal sums or by the total number of points in the table. Frequently, one has knowledge of the true map category marginal proportions, that is, the relative areas of each map category. These map category proportions can be used to improve estimates of "proportion-correct" for each map category. This paper derives these improved estimates with their asymptotic variances for two common sampling designs: simple random sampling of single points and sampling stratified by map category. A numerical example illustrates the computations.*

Introduction

The introduction of satellite-acquired imagery in the early 1970s and the concurrent advances in techniques of computer classification of digital remotely sensed data have made possible the rapid generation of land-use and land-cover maps. However, user acceptance of this kind of computer-generated map has lagged behind the technology and prevented realization of its full potential, principally because of difficulties inherent in the specification and statistical testing of accuracy. Users will not and should not take a map at face value without some associated estimate of error. As pointed out by Switzer (1969), a map without an accompanying statement of error is like a point estimate with no variance stated. Because in practice, it is impossible to check every point on the map, some type of statistical sampling procedure must be used to meet time and cost limitations.

Although various aspects of the assessment of the accuracy of maps derived from remotely sensed data have been dealt with in the literature (Ginevan, 1979); Hord and Brooner, 1976; Hay, 1979; van Genderen and Lock, 1977; van Genderen *et al.*, 1978), a unified treatment of the subject has been lacking. This writer suggests that "contingency-table" analysis is the most natural framework for accuracy assessment, both for the convenient display of empirical results and for ease of statistical analysis. This idea is not new—other authors have displayed their data in the form of a cross-tabulation of map category versus true category; however, the statistical treatment of these tables has frequently been either insufficient or incorrect. It has not always been appreciated that unbiased estimates of error probabilities derived from these tables are relative to the sampling scheme giving rise to the data.

This paper discusses two statistical sampling schemes for assessing map accuracy: simple random sampling of single points and random sampling stratified by map category. In the stratified sampling method, independent simple random samples of points are selected within each map category for identification by ground checking or by interpreting photographs.

Photogrammetric Engineering and Remote Sensing,
Vol. 48, No. 3, March 1982, pp. 431-439.

0099-1112/82/4803-0431$02.25/0

It will be seen that point estimates of "proportion-correct" are identical for both sampling schemes, but that variance formulas are slightly different. The idea of stratified sampling is quite attractive because a simple random or systematic sample drawn from the total map area tends to oversample categories of high frequency and to undersample categories of low frequency. Presumably, drawing samples from a low-frequency map category would give these points a better chance of being from the true category in question, assuming reasonably small commission error for that category. Even though this procedure is sometimes followed in practice, there seems to be general confusion about how to treat it statistically. Authors have understandably been cautious about interpreting data collected in this way, since the naive estimates of probabilities are not correct due to bias introduced by the stratified sampling. To quote van Genderen (1977): "Although several alternative methods have been used in other projects, few provide sufficient statistical justification for the allocation of sample points in each category of land use using remote sensing imagery." From another van Genderen paper (van Genderen, 1978): "It is considered that the above methods do not provide sufficient statistical justification for the allocation of sample points in each category of a classification scheme utilizing Landsat MSS imagery." Finally, in a paper by Hay (1979) in which he discusses a contingency table of ground observations versus classified data and in which the classified data served as a stratification: ". . . any effects in the rows are largely a consequence of the stratified sampling and should be treated with caution. . . ."

The present paper shows how to derive unbiased estimates of classification accuracy from data derived in this way, using known sampling rates in each stratum. The key to this approach is knowledge of the marginal proportions for each map category. For maps derived from Landsat digital multispectral data, these marginals are easily obtained from a routine by-product of computer classification—pixel counts by category. With a little more effort, the same kind of information can be derived for continuous maps; for example, by planimetering or by cutting and weighing.

Contingency-Table Approach

The results of sampling are conveniently displayed in a manner such as that shown in Table 1, which is called a contingency table in statistical parlance. The cell entries show the number of sampled points whose map category is the column label, and whose true category is the row label ("true" in the context of this paper means having been derived from a presumably more accurate data source than the map; for example, from low-altitude aerial photography or ground visits). This framework for conceptualizing the problem is important from several standpoints. Two of the most important are: first, it gives a visual overview of the results, which may help to locate sources of observer bias very quickly; and second, it stresses the relationship between the two kinds of error inherent in the classification system.

Table 1. Contingency Table for Evaluating Map Accuracy

True Category*	Map Category A	B	C	D	E	Totals
A	142	1	5	1	—	149
B	3	65	—	—	5	73
C	—	2	104	—	1	107
D	1	—	2	172	—	175
E	—	—	1	10	64	75
Totals	146	68	112	183	70	579

* Categories identified by photographic interpretation or field checking.

These are (1) the error of identifying a class as A when in fact it is not A, and (2) the converse error of identifying it as something else when it is really A. These errors are sometimes called errors of commission and omission for class A, respectively. This way of looking at the problem is important because, in the view of the author, approaches described in the literature that have not used the contingency-table approach have implicitly ignored errors of omission. In other words, probabilities of correct classification were described simply as the number of points falling in a map category that were correct, divided by the total number of points in the category. This can be described, in terms of probability, as the probability of being correct given the map category. This ignores the points that were really of the class but were not so identified; that is, omission errors. The author believes that a case can be made that this is the most crucial type of error in the classification system, and that a user of the map is really more interested in the probability of correct classification given the true category. For example, of all the area that is really forest, what proportion of it is correctly called "forest" by the map? This paper will describe how estimates of these probabilities can be derived that will hold, not only for simple random sampling of test points, but for sampling stratified by map category as well.

Sample Design

A contingency table such as that shown in Table 1 cannot be analyzed until the sampling procedure which was followed in collecting the data is known. At least, intelligent inferences to the underlying population cannot be made. Different methods of sampling require different approaches to estimation. For instance, although the results of a haphazard sample can be compiled in a contingency table, such a sample allows no statistical statements at all to be made, unless special assumptions are made regarding accuracy as a function of location. For our purposes, let us consider two sampling plans: (1) simple random sampling of single points from a map and (2) stratified sampling with map categories as independent strata, that is, points located independently and randomly within map categories. Each sampled point is identified as to map category and ground- or photograph-checked as to "true" category. The verification data base, whether derived from ground visitation or photographic interpretation, is assumed to be substantially more accurate than the map; as a result, it makes sense to refer to the data base as "truth," at least informally.

Once the sampled data are compiled in a contingency table, we can make estimates of parameters of interest. Hay (1979) discusses what kind of questions a map user is likely to ask of the data in a contingency table. As he points out, for the stratified sampling case, the proportions-correct, given the true category, should not be estimated by the diagonal entry divided by the row sum, because of bias introduced by possible differential sampling rates within map categories. However, we shall demonstrate how one can use the known sampling rates within map categories to correct for this bias and give unbiased estimates for these proportions. Interestingly enough, the same estimators will be shown to be valid also for simple random sampling.

This paper follows quite closely the concepts and nomenclature given in Tenenbein (1972). Tenenbein discusses the effect of classification errors on the estimation of multinomial proportions and applies the results to problems in quality control. The problem that he addresses is this: one has a measurement technique that is very accurate but costly and also an alternative technique that is less accurate but relatively inexpensive. How does one use the inexpensive measurement technique to improve the precision of estimates of product quality? For example, suppose one can visually inspect a large number of units very quickly but with rather large error; on the other hand, one has at his disposal a very accurate measuring technique that destroys the product in the testing process. The approach is to visually inspect a large number of units and then to measure accurately a subset of these units to estimate classification errors and remove bias from the overall estimates. This is an application of two-phase or double sampling described by Cochran (1977). In the present context, the less accurate data source is the map whose accuracy is in question, and the more accurate but costly source is ground or photographic verification of map category. Whereas Tenenbein's initial, less accurate sample is partial, ours is complete, since the classification is known for the whole map. Also, he considers only simple random sampling, whereas we consider also stratified sampling. These differences in the problem formulation explain slight differences in final results.

Statistical Model

Suppose that n points are located on a map having r categories by either of two methods: (1) simple random sample or (2) stratified random sample, with map categories as the strata. Suppose also that an independent "ground truth" data source can be queried as to the "true" category for each sampled point. Then each point in the sample is identified as to "true" category and map category. Define random variables T and M (T refers to true category and M refers to map category) as

$T = i$ if the point belongs to category i ($i = 1, 2, \ldots, r$) according to "ground truth" or photographic interpretation, and

$M = j$ if the point is classified by the map as being in category j ($j = 1, 2, \ldots, r$).

The marginal distributions of T and M are

$$p_i = \Pr[T = i], \tag{1}$$

$$\pi_j = \Pr[M = j], \tag{2}$$

for categories i and j and

$$\sum_{i=1}^{r} p_i = \sum_{j=1}^{r} \pi_j = 1.$$

The p_i can be interpreted as the true relative proportion of area in category i, and π_j is the area in category j according to the map.

The classification system is described by a function θ_{ij}:

$$\theta_{ij} = \Pr[M = j \mid T = i]. \tag{3}$$

This is the probability that a point, which is really in the ith category, is classified by the map as being in the jth category. It follows that

$$\sum_{j=1}^{r} \theta_{ij} = 1$$

and

$$\pi_j = \sum_{i=1}^{r} p_i\theta_{ij}. \tag{4}$$

The joint probability distribution of T and M is

$$p_{ij} = \Pr[T = i, M = j] = p_i\theta_{ij}. \tag{5}$$

Also of interest are the conditional probabilities λ_{ij}, where

$$\lambda_{ij} = \Pr[T = i \mid M = j]. \tag{6}$$

The relationship between θ_{ij} and λ_{ij} is

$$\lambda_{ij} = p_i\theta_{ij}/\pi_j. \tag{7}$$

Of special interest in the assessment of map accuracy are errors of omission and commission. These can be described as follows:

$$\begin{aligned}\theta_i &= \text{probability of misclassifying a point belonging to class } i\\ &= \text{error of omission for class } i \qquad (8)\\ &= \sum_{\substack{i \neq j\\ j=1}}^{r} \theta_{ij};\end{aligned}$$

$$\begin{aligned}\phi_j &= \text{probability of classifying a point as } j \text{ when it is not } j\\ &= \text{error of commission for class } j \qquad (9)\\ &= \sum_{\substack{i \neq j\\ i=1}}^{r} \theta_{ij}.\end{aligned}$$

Invoking an elementary theorem from probability theory, we have

$$\begin{aligned}\pi_j &= \Pr[M = j \mid T = j]\Pr[T = j] \\ &\quad + \Pr[M = j \mid T \neq j]\Pr[T \neq j] \qquad (10)\\ &= (1 - \theta_j)p_j + \phi_j(1 - p_j).\end{aligned}$$

This allows us to characterize the bias in estimating p_j by π_j:

$$\text{bias} = \pi_j - p_j = (1 - p_j)\phi_j - p_j\theta_j, \qquad (11)$$

and the relative bias is

$$\text{bias}/p_j = (1 - p_j)\phi_j/p_j - \theta_j. \qquad (12)$$

As pointed out by Tenenbein, this bias can be serious not only when ϕ_j is large and p_j small, but possibly even when θ_j and ϕ_j, the classification errors, are small. For example, he shows that the relative bias is 44 percent when $p_j = 0.1$, $\theta_j = 0.01$, and $\phi_j = 0.05$. This bias is especially interesting in the context of maps derived from Landsat multispectral scanner data. Estimates of percentage area for each category are sometimes routinely derived by taking raw pixel (picture element) counts as output by the computer. This is the estimate π_j and, as shown above, it can be seriously biased. If good estimates of the true proportions are important, then the π_j should not be used as direct estimates. Unbiased estimates of p_j can be derived using the π_j as weights, as shown in the next section.

Estimation Procedure

Suppose a land-use map has been generated from a source that is relatively inexpensive compared with low-altitude aerial photographs or ground survey, for example, Landsat data. However, it is presumably less accurate than these other sources, and so tends to generate a bias, characterized by the off-diagonal elements in the contingency table.

A sampling scheme designed to evaluate and correct for this bias is as follows: A sample of n points is located on the map and the true and map categories are determined (from some verification source, such as photographic interpretation or ground visit) for each point. The n points can be allocated in either of two ways:

(1) Locate a simple random sample of n points, or
(2) Locate random samples of n_j points independently in each map category j.

The results are tabulated in a two-way square contingency table, as shown in Table 2, where n_{ij} is the number of points in the sample whose true category is "i" and whose map category is "j."

The marginal sums are defined as follows:

$$n_{i\cdot} = \sum_{j=1}^{r} n_{ij}, \qquad (13)$$

$$n_{\cdot j} = \sum_{i=1}^{r} n_{ij}, \qquad (14)$$

$$n = \sum_{i=1}^{r}\sum_{j=1}^{r} n_{ij}. \qquad (15)$$

Maximum likelihood estimates for various probabilities of interest are derived in the Appendix. These estimates are

(1) Marginal proportion for true class i:

$$\hat{p}_i = \sum_{j=1}^{r} \pi_j n_{ij}/n_{\cdot j}. \qquad (16)$$

(2) Probability correct, given true class i:

$$\Pr[M = i \mid T = i] = \hat{\theta}_{ii} = \frac{\pi_i}{\hat{p}_i}\frac{n_{ii}}{n_{\cdot i}}. \qquad (17)$$

(3) Probability correct, given map class j:

$$\Pr[T = j \mid M = j] = \hat{\lambda}_{jj} = n_{jj}/n_{\cdot j}. \qquad (18)$$

(4) Overall probability correct:

$$\hat{P}_c = \sum_{j=1}^{r} \pi_j n_{jj}/n_{\cdot j}. \qquad (19)$$

Table 2. Contingency-Table Format

		M				
		1	2	. . .	r	
T	1	n_{11}	n_{12}	. . .	n_{1r}	$n_{1\cdot}$
	2	n_{21}	n_{22}		n_{2r}	$n_{2\cdot}$
	.					
	.					
	.					
	r	n_{r1}	n_{r2}		n_{rr}	$n_{r\cdot}$
		$n_{\cdot 1}$	$n_{\cdot 2}$		$n_{\cdot r}$	n

(5) Individual cell probabilities:

$$\hat{p}_{ij} = \Pr[T = i, M = j] = \hat{\theta}_{ij}\hat{p}_i = \pi_j n_{ij}/n_{\cdot j}. \quad (20)$$

The π_j are the known marginal proportions for each map category j. If totals are needed, rather than proportions, the modifications are straight forward. For example, if we are dealing with Landsat data, and $N_{\cdot j}$ are pixel counts for category j, and N is the total number of pixels, then

$$\pi_j = N_{\cdot j}/N, \quad (21)$$

and estimates of true marginal totals are

$$\hat{N}_{i\cdot} = \sum_{j=1}^{r} N_{\cdot j} n_{ij}/n_{\cdot j}. \quad (22)$$

Note that for the simple random sampling case, the $n_{\cdot j}$ are random variables, whereas for the stratified sampling, they are chosen by the experimenter. In each case, the maximum likelihood estimates (Equations 16 through 20) are the same, as is shown in the Appendix. However, asymptotic variances are slightly different.

As pointed out by Tenenbein, it is possible for $n_{\cdot j}$ to be zero for some map class j, if the true marginal proportion p_j for that class is small. In that case, define $n_{ij}/n_{\cdot j}$ to be 0. This has negligible effect on estimates involving class j if n is reasonably large.

Estimates of variances of Estimators 16 through 20 are derived in the Appendix. These allow confidence statements to be made for each estimate. A numerical example is given in the next section. In the simple random sampling case, the variance of $\hat{p}_i$ is

$$V(\hat{p}_i) = \sum_{j=1}^{r} p_{ij}(\pi_j - p_{ij})/n\pi_j. \quad (23)$$

For the stratified sampling case, the variance of $\hat{p}_i$ is

$$V(\hat{p}_i) = \sum_{j=1}^{r} p_{ij}(\pi_j - p_{ij})/n_{\cdot j}. \quad (24)$$

Variances for Estimators 17 through 19 are

(1) Stratified sampling:

$$V(\hat{\theta}_{ii}) = p_{ii}p_i^{-4}\left[p_{ii} \sum_{j\neq i}^{r} p_{ij}(\pi_j - p_{ij})/n_{\cdot j} + (\pi_i - p_{ii})(p_i - p_{ii})^2/n_{\cdot ji}\right], \quad (25)$$

$$V(\hat{\lambda}_{ii}) = p_{ii}(\pi_i - p_{ii})/(\pi_i^2 n_{\cdot i}), \quad (26)$$

$$V(\hat{P}_c) = \sum_{i=1}^{r} p_{ii}(\pi_i - p_{ii})/n_{\cdot i}. \quad (27)$$

(2) Simple random sampling:

$$V(\hat{\theta}_{ii}) = p_{ii}p_i^{-4}\left[p_{ii} \sum_{j\neq i}^{r} p_{ij}(\pi_j - p_{ij})/(\pi_j n) + (\pi_i - p_{ii})(p_i - p_{ii})^2/(\pi_i n)\right], \quad (28)$$

$$V(\hat{\lambda}_{ii}) = p_{ii}(\pi_i - p_{ii})/(\pi_i^3 n), \quad (29)$$

$$V(\hat{P}_c) = \sum_{i=1}^{r} p_{ii}(\pi_i - p_{ii})/(\pi_i n). \quad (30)$$

Numerical Example

Suppose a Landsat-derived map has five categories, A through E, and that random samples of 50 pixels have been selected from each category and ground checked, resulting in the summary shown in Table 3. Suppose, also, that the marginal proportions (sampling fractions) for each map category are known from a simple computer count of pixels (shown in the bottom line of the table).

A convenient way to lay out the calculations is to first form a table whose cell entries are the estimates $\hat{p}_{ij}$ (Table 4), using Equation 20. We can now compute the estimates for the true marginal proportions using Equation 16:

$$\hat{p}_1 = 0.393; \quad \hat{p}_2 = 0.403; \quad \hat{p}_3 = 0.126; \quad \hat{p}_4 = 0.047; \quad \hat{p}_5 = 0.031.$$

Table 3. Hypothetical Contingency Table for Numerical Example

True Category	Map Category A	B	C	D	E	Row Totals
A	48		2	5		55
B	1	49		4		54
C	1		47	3	3	54
D		1	1	34	12	48
E				4	35	39
Column Totals	50	50	50	50	50	250
Map Marginal Proportions	0.4	0.4	0.12	0.04	0.04	

TABLE 4. CONTINGENCY TABLE FOR ESTIMATES $\hat{p}_{ij}$ OF CELL PROBABILITIES p_{ij}

True Category	Map Category A	B	C	D	E
A	0.384		0.0048	0.004	
B	0.008	0.392		0.0032	
C	0.008		0.1128	0.0024	0.0024
D		0.008	0.0024	0.0272	0.0096
E				0.0032	0.028

These are just the row sums in Table 4. Estimates for probability correct, given the true category are given by Equation 17:

$$\hat{\theta}_{11} = 0.977; \quad \hat{\theta}_{22} = 0.973; \quad \hat{\theta}_{33} = 0.895; \quad \theta_{44} = 0.579; \ \hat{\theta}_{55} = 0.903.$$

These are the diagonal values divided by the row sums in Table 4.

It is interesting to compare the estimates of the true marginal proportions, $\hat{p}_i$, with estimates one would derive if one ignored the stratified sampling. It can be seen that the unbiased estimates $\hat{p}_i$ above are quite close to the map category marginals in Table 1, which is to be expected since there are few errors of classification. If one uncritically estimates the true row marginals by the marginal sums divided by 250, one obtains 0.22, 0.22, 0.22, 0.19, and 0.16, all of which are far from the values anticipated from knowledge of the map marginal proportions. This illustrates the importance of using the known π_j to remove bias.

Estimates of probability correct, given the map category, are straightforward using Equation 18:

$$\begin{aligned} \hat{\lambda}_{11} &= 48/50 = 0.96 \\ \hat{\lambda}_{22} &= 49/50 = 0.98 \\ \hat{\lambda}_{33} &= 47/50 = 0.94 \\ \hat{\lambda}_{44} &= 34/50 = 0.68 \\ \hat{\lambda}_{55} &= 35/50 = 0.70 \end{aligned}$$

The overall probability correct is estimated from Equation 19 as:

$$\hat{P}_c = 0.384 + 0.392 + 0.1128 + 0.0272 + 0.028 = 0.944.$$

The approximate variance of $\hat{P}_c$, overall probability correct, is given by Equation 27:

$$\begin{aligned} V(\hat{P}_c) = {} & [0.384(0.4 - 0.384) + 0.392(0.4 - 0.392) \\ & + 0.113(0.12 - 0.113) \\ & + 0.0272(0.04 - 0.0272) \\ & + 0.028(0.04 - 0.028)]/50 = 0.000215. \end{aligned}$$

Therefore, an approximate 95 percent confidence interval for $\hat{P}_c$ is (0.915, 0.973).

Example calculations for three typical variance formulas are now given:

$$\begin{aligned} V(\hat{\theta}_{11}) &= \text{variance of } \hat{\theta}_{11} = 0.384 \\ & \times 0.393^{-4}\{0.384[0.0048(0.12 - 0.0048)/50 \\ & + 0.004(0.04 - 0.004)/50] \\ & + (0.4 - 0.384)\,(0.393 - 0.384)^2/50\} \\ &= 0.0000866; \end{aligned}$$

$$\begin{aligned} V(\hat{\lambda}_{11}) &= 0.384(0.4 - 0.384)/0.4^2/50 \\ &= 0.000768; \end{aligned}$$

$$\begin{aligned} V(\hat{p}_1) &= [0.384(0.4 - 0.384) \\ & + 0.0048(0.12 - 0.0048) \\ & + 0.004(0.04 - 0.004)]/50 \\ &= 0.000137. \end{aligned}$$

We can now write down approximate confidence intervals for each estimate $\hat{\theta}_{ii}$ as

$$\hat{\theta}_{ii} \pm 2[V(\hat{\theta}_{ii})]^{1/2},$$

and similarly for the other estimators. If normality can be assumed, these will be nominal 95 percent confidence intervals; otherwise, by Chebyshev's inequality (Fowler, 1979), we are guaranteed at least 75 percent confidence. Table 5 gives the standard errors for $\hat{p}_i$, $\hat{\theta}_{ii}$, and $\hat{\lambda}_{jj}$; Table 6 gives a summary of each estimate with its approximate 95 percent confidence interval for each category.

DISCUSSION

Using two simple sampling plans suggested in the accuracy assessment literature, we have shown how one can use knowledge of map-category relative sizes to improve estimates of various probabilities of interest to map users. The fact that maximum likelihood estimates of cell probabilities for the two sampling schemes (simple random sampling and sampling stratified by map category) were identical allowed a unified treatment of the contingency-table analysis.

Results for the stratified case allow a rigorous analysis of the effect of sampling independently within map categories, a procedure often recommended in the literature but heretofore not backed

TABLE 5. APPROXIMATE STANDARD ERRORS FOR ESTIMATES $\hat{p}_i$, $\hat{\theta}_{ii}$, AND $\hat{\lambda}_{jj}$

True Category i	Approximate Standard Error $\hat{p}_i$	$\hat{\theta}_{ii}$
A	0.0117	0.00931
B	0.0113	0.0195
C	0.00908	0.0331
D	0.00901	0.109
E	0.00301	0.0447

Map Category j	Approximate Standard Error of $\hat{\lambda}_{jj}$
A	0.0277
B	0.0198
C	0.0336
D	0.0660
E	0.0648

TABLE 6. APPROXIMATE 95 PERCENT CONFIDENCE INTERVALS FOR $\hat{p}_i$, $\hat{\theta}_{ii}$, AND $\hat{\lambda}_{jj}$

True Category	Confidence Interval		Confidence Interval		Confidence Interval	
	$\hat{p}_i$	Interval	$\hat{\theta}_{ii}$	Interval	$\hat{\lambda}_{jj}$	Interval
A	0.393	(0.369, 0.416)	0.978	(0.959, 0.997)	0.960	(0.905, 1.0)
B	0.403	(0.381, 0.476)	0.972	(0.933, 1.0)	0.980	(0.940, 1.0)
C	0.126	(0.107, 0.144)	0.898	(0.832, 0.964)	0.940	(0.873, 1.0)
D	0.047	(0.0292, 0.0652)	0.576	(0.358, 0.794)	0.680	(0.548, 0.812)
E	0.031	(0.0252, 0.0372)	0.897	(0.808, 0.986)	0.700	(0.570, 0.830)

up by statistical analysis. Whether this procedure is helpful or not will not be known until actual accuracy assessment studies are made. Given the results of an accuracy study, one can estimate the gains from stratification as described in Cochran (1977). This is impossible to determine *a priori* since it is relative to the actual map generation process and the actual classification error structure in any particular case.

It should be mentioned that certain critical questions of statistical design, such as optimal sample size selection to achieve desired precision in various estimators, have not been discussed here. These considerations, though important in actual applications, are not central to the estimation problem considered in this paper. The estimators derived are valid regardless of how sample sizes are chosen, including sample sizes arrived at by guessing.

APPENDIX

To derive maximum likelihood estimates of cell probabilities, it should be noted that the joint likelihood of the p_{ij} for both sampling schemes (simple random sampling and stratified random sampling) is proportional to

$$L = \prod_{i=1}^{r} \prod_{j=1}^{r} p_{ij}^{n_{ij}}, \qquad \text{(A-1)}$$

and therefore the estimates of p_{ij} that maximize log L will maximize the likelihood for both sampling schemes. Since the marginal proportions π_j are known, we have the problem

$$\text{Maximize } \log L \text{ subject to} \sum_{i=1}^{r} p_{ij} = \pi_j$$

$$\text{for } j = 1, 2, \ldots, r \text{ and} \sum_i \sum_j p_{ij} = 1. \qquad \text{(A-2)}$$

Define Lagrange multipliers λ_j such that

$$F = \log L - \sum_{j=1}^{r} \lambda_j \left(\sum_{i=1}^{r} p_{ij} \right) - \lambda \sum_i \sum_j p_{ij}.$$

Taking partials of F with respect to p_{ij} and setting them equal to 0,

$$\frac{\partial F}{\partial p_{ij}} = \sum_i \sum_j (\log p_{ij}) n_{ij} - \lambda_i - \lambda = 0$$

$$= n_{ij}/p_{ij} - \lambda_j - \lambda = 0.$$

Solving for p_{ij},

$$p_{ij} = n_{ij}/(\lambda_j + \lambda). \qquad \text{(A-3)}$$

Summing both sides over i:

$$\sum_i p_{ij} = \sum_i n_{ij}/(\lambda_j + \lambda) = n_{\cdot j}/(\lambda + \lambda_j)$$

$$= \pi_j,$$

by the constraint in the statement of the problem.

Therefore,

$$\lambda + \lambda_j = n_{\cdot j}/\pi_j, \qquad \text{(A-4)}$$

and substituting this result in Equation A-3, we have for the estimate $\hat{p}_{ij}$ of p_{ij}:

$$\hat{p}_{ij} = \frac{n_{ij}}{n_{\cdot j}} \pi_j. \qquad \text{(A-5)}$$

From the invariance property of maximum likelihood estimates, estimates of functions of the p_{ij}, such as p_i and P_c, can be obtained by a direct substitution of $\hat{p}_{ij}$ for p_{ij} in the function. For example,

$$\hat{p}_i = \sum_{j=1}^{r} \hat{p}_{ij},$$

since

$$p_i = \sum_{j=1}^{r} p_{ij}$$

for the true values.

Asymptotic variances of the estimators $\hat{p}_{ij}$ are most easily obtained by inverting the matrix whose k, jth entry (for fixed i) is $-E[\partial^2 F/\partial p_{ik} \partial p_{ij}]$ (Kendall and Stuart, 1967, p. 55). The calculations are simplest if we fix i, the row index, and collapse all other rows to a category that we can call not-i. Since we then have n_{ij} entries for each cell in row i with probabilities p_{ij}, and $n_{\cdot j} - n_{ij}$ entries in "not-i" with probabilities $\pi_j - p_{ij}$, in the contingency table, then the likelihood function is proportional to

$$L = \prod_{j=1}^{r} p_{ij}^{n_{ij}} (\pi_j - p_{ij})^{(n_{\cdot j} - n_{ij})} \quad \text{(A-6)}$$

for both simple random sampling and stratified sampling. As before, define

$$F = \log L = \sum_{j=1}^{r} n_{ij} \log p_{ij} + \sum_{j=1}^{r} (n_{\cdot j} - n_{ij}) \log(\pi_j - p_{ij}) \quad \text{(A-7)}$$

For fixed i,

$$\frac{\partial^2 F}{\partial p_{ij} \partial p_{ik}} = -n_{ij}/p_{ij}^2 - \frac{(n_{\cdot j} - n_{ij})}{(\pi_j - p_{ij})^2} \text{ for } j = k$$
$$= 0 \qquad \text{for } j \neq k \quad \text{(A-8)}$$

Therefore,

$$E\left(\frac{-\partial^2 F}{\partial p_{ij} \partial p_{ik}}\right) = E\left]\frac{n_{ij}}{p_{ij}^2} + \frac{n_{\cdot j} - n_{ij}}{(\pi_j - p_{ij})^2}\right] \quad \text{(A-9)}$$

Under simple random sampling, $E(n_{ij}) = p_{ij}n$ and $E(n_{\cdot j}) = n\pi_j$; therefore,

$$E\left[\frac{n_{ij}}{p_{ij}^2} + \frac{n_{\cdot j} - n_{ij}}{(\pi_j - p_{ij})^2}\right] = n\pi_j/[p_{ij}(\pi_j - p_{ij})] \quad \text{(A-10)}$$

This is the diagonal term of the information matrix all of whose off-diagonal terms are zero by Equation A-8. Therefore, the inverse of the information matrix is the diagonal matrix whose jth diagonal term is the reciprocal of the expression in Equation A-10, and the variance of $\hat{p}_{ij}$ is given by the jth diagonal term of the inverse:

$$\text{Var}(\hat{p}_{ij}) = \frac{p_{ij}(\pi_j - p_{ij})}{n\pi_j} \quad \text{(A-11)}$$

Similarly, for stratified sampling, since

$$E(n_{ij}) = p_{ij} n_{\cdot j}/\pi_j$$

we have,

$$\text{Var}(\hat{p}_{ij}) = \frac{p_{ij}(\pi_j - p_{ij})}{n_{\cdot j}}. \quad \text{(A-12)}$$

For the variance of $\hat{p}_i$, we use the well known formula for the covariance of a sum of random variables. Since

$$\hat{p}_i = \sum_{j=1}^{r} p_{ij},$$

we have

$$\text{Var}(\hat{p}_i) = \sum_{j=1}^{r} \sum_{k=1}^{r} \text{Cov}(p_{ij}, p_{ik}).$$

From Equation A-8, $\text{Cov}(p_{ij}, p_{ik}) = 0$ for $j \neq k$, and so

$$\text{Var}(\hat{p}_i) = \sum_{j=1}^{r} \text{Var}(p_{ij}) = \sum_{j=1}^{r} p_{ij}(\pi_j - p_{ij})/(n\pi_j) \quad \text{(A-13)}$$

from Equation A-11 for simple random sampling. Similarly, for stratified sampling,

$$\text{Var}(\hat{p}_i) = \sum_{j=1}^{r} p_{ij}(\pi_j - p_{ij})/n_{\cdot j}. \quad \text{(A-14)}$$

To find the variance of

$$\hat{P}_c = \sum_{j=1}^{r} \hat{p}_{jj}$$

we can proceed in a manner similar to the technique in the previous paragraph. It can be shown that $\text{Cov}(p_{jj}, p_{kk}) = 0$ for $j \neq k$ by considering the diagonal cells in the contingency table as the first row of a new $2 \times r$ table, with the second row as summed off-diagonal elements in each column, as follows:

		Map Category				
		1	2		r	
True Category	diagonal	n_{11}	n_{22}	$\cdots$	n_{rr}	$\sum_{i=1}^{r} n_{ii}$
	non-diagonal	$n_{\cdot 1} - n_{11}$	$n_{\cdot 2} - n_{22}$	$\cdots$	$n_{\cdot r} - n_{rr}$	$n - \sum_{i=1}^{r} n_{ii}$
		$n_{\cdot 1}$	$n_{\cdot 2}$		$n_{\cdot r}$	

The result is (using Equation A-14):

$$V(\hat{P}_c) = \sum_{j=1}^{r} \sum_{k=1}^{r} \mathrm{Cov}(p_{jj}, p_{kk})$$
$$= \sum_{j=1}^{r} p_{jj}(\pi_j - p_{jj})/n_{.j} \qquad \text{(A-15)}$$

for stratified sampling, and (using Equation A-13):

$$V(\hat{P}_c) = \sum_{j=1}^{r} p_{jj}(\pi_j - p_{jj})/(\pi_j n) \qquad \text{(A-16)}$$

for simple random sampling.

By Equation 20, $\hat{\lambda}_{jj} = \hat{p}_{jj}/\pi_j$, and so by Equations A-11 and A-12,

$$\mathrm{Var}(\hat{\lambda}_{jj}) = \frac{1}{\pi_j^2}\mathrm{Var}(\hat{p}_{jj})$$
$$= \frac{p_{jj}(\pi_j - p_{jj})}{n\pi_j^3}$$

for simple random sampling, and

$$\mathrm{Var}(\hat{\lambda}_{jj}) = \frac{p_{jj}(\pi_j - p_{jj})}{n_{.j}\pi_j^2}$$

for stratified sampling.

The asymptotic variance of $\hat{\theta}_{ii}$ can be obtained by expanding $\hat{\theta}_{ii}$ in a Taylor's series about the true value $\theta_{ii} = p_{ii}/p_{i.}$. The results are given by Equations 25 and 28. The details will not be given here.

References

Cochran, W. G., 1977. *Sampling Techniques*, third edition, John Wiley & Sons.

Fowler, G. W., and D. Hauke, 1979. A Distribution-Free Method for Interval Estimation and Sample Size Determination, *Resource Inventory Notes*, USDI, Bureau of Land Management, Denver, Colorado.

Ginevan, M. E., 1979. Testing Land-use Map Accuracy: Another Look, *Photogrammetric Engineering and Remote Sensing*, v. 45, no. 10, pp. 1371-1377.

Hay, A. M., 1979. Sampling Designs to Test Land Use Map Accuracy, *Photogrammetric Engineering and Remote Sensing*, v. 45, no. 4, pp. 529-533.

Hord, R. M., and W. Brooner, 1976. Land-use Map Accuracy Criteria, *Photogrammetric Engineering and Remote Sensing*, v. 42, no. 5, pp. 671-677.

Kendall, M. G., and A. Stuart, 1967. *The Advanced Theory of Statistics*, second edition, Charles Griffin & Company Limited, London.

Switzer, P., 1969. *Mapping a Geographically Correlated Environment*, Technical Report No. 145, Department of Statistics, Stanford University, Stanford, Calif.

Tenenbein, A., 1972. A Double Sampling Scheme for Estimating from Misclassified Multinomial Data with Applications to Sampling Inspection, *Technometrics*, v. 14, no. 1, pp. 187-202.

van Genderen, J. L., and B. F. Lock, 1977. Testing Land Use Map Accuracy, *Photogrammetric Engineering and Remote Sensing*, v. 43, no. 9, pp. 1135-1137.

van Genderen, J. L., B. F. Lock, and P. A. Vass, 1978. Remote Sensing: Statistical Testing of Thematic Map Accuracy, *Remote Sensing of Environment*, v. 7, pp. 3-14.

(Received 9 March 1981; accepted 17 July 1981; revised 15 October 1981)
