A SUMMARY OF OBSERVATIONS CONCERNING THE INFORMATION IN THE SPECTRAL-TEMPORAL-ANCILLARY DATA AVAILABLE FOR ESTIMATING GROUND COVER CROP PROPORTIONS

Charles R. Perry, Jr., U.S. Department of Agriculture, Statistical Reporting Service*

ABSTRACT

In agricultural crop acreage surveys based on satellite-acquired data, the derived estimates ultimately depend on a probabilistic classification of small area units into a number of different ground cover classes. The goal is to assign the small area units to crop classes in such a way that the resulting crop proportion estimates are as accurate as possible. The statistical problem is that of assessing the quantity of information which the Landsat reflectance measurements supply about the unknown crop proportions, assuming the most efficient utilization of the data. This report provides a discussion of Fisher information theory and its application to crop proportion estimation using satellite-acquired data. The theoretical results are illustrated with examples based on data from 1978 Large Area Crop Inventory Experiment photointerpreter classifications of small grains segments. These examples indicate that a considerable information loss occurs during the photointerpretation process.

1. INTRODUCTION

In agricultural crop acreage surveys based on satellite-acquired data, the derived estimates ultimately depend on a probabilistic classification of small area units into a number of different ground cover classes (ref. 1). These classifications are usually made by examining the change through time of the Landsat reflectance measurements that are associated with the small area unit. In addition, other information concerning the practice of agriculture in the general region is important in the classification process; however, ground observations are assumed unavailable except for test purposes since applications are to foreign areas. These data sets are referred to as the spectral-temporal data and the spectral-temporal-ancillary data, respectively.

Historically, a great deal of effort and ingenuity has been devoted to the theory and methodology associated with making these probabilistic crop class determinations. Although not always clearly expressed, the goal has always been to assign the small area units to crop classes in such a way that the resulting crop proportion estimates are as accurate as possible. Latent in this goal is the question: Is the precision of the results limited by the classification techniques employed or by the intrinsic value of the satellite data used in the classification process? In other words, in obtaining an accurate estimate of the crop proportions for a specified area, are we limited by the procedures we employ or by the amount of information concerning the unknown crop proportions that is contained in the associated spectral-temporal-ancillary data?

From a purely statistical standpoint, the problem is that of assessing the quantity of information which the Landsat reflectance measurements supply about the unknown crop proportions, assuming the most efficient utilization of the data.

The measure of information used in this report was introduced by R. A. Fisher (ref. 2) and is generally referred to as Fisher information. As motivation for his information measure, a part of the introductory remarks given to the subject in reference 2 are included. Fisher introduces the definition of information with the sentence:

> "If, therefore, any such average is determined with a sampling variance V, we may define a quantity I such that I = 1/V, and I will measure the quantity of information supplied by the experiment in respect of the particular value to which the variance refers." (Ref. 2, p. 185)

Later in the context of estimation theory, he states:

> "The amount of information to be expected in respect of any unknown parameters, from a given number of observations of independent objects or events, the frequencies of which depend on that parameter, may be obtained by a simple application of the differential calculus." (ref. 2, pp. 215-216)

Fisher continues with a sequence of examples. The first of these examples is included for its instructive value and also to share with the reader the beauty in Fisher's expression of the basic idea.

> "Let us suppose that only two kinds of objects or events are to be distinguished, and that we are concerned to estimate the frequency, p, with which one of them occurs as a fraction of all occurrences; or, what comes to the same thing, the complementary frequency, q(= 1 - p), with which the alternative event occurs. We might, for example, be estimating the proportion of males in the aggregate of live births, or the proportion of sterile samples drawn from a bulk in which an unknown number of organisms are distributed, or the proportion of experimental animals which die under well-defined experimental conditions. The experimental or observational record will then give us the numbers of the two kinds of observations made, a of one kind and b of another, out of a total number of n cases examined. We wish to know how much information the examination of n cases may be expected to provide, concerning the values of p and q, which are to be estimated from the data.
>
> "A general procedure, which may be easily applied to many cases, is to set down the frequencies to be expected in each of the distinguishable classes in terms of the unknown parameter. For each class we then find the differential coefficient, with respect to p, of this expectation. The squares of these, divided by the corresponding expectations, and added together, supply the amount of information to

be anticipated from the observational record. That such a calculation will give a quantity of the kind we want, may be perceived at once by considering that the differential coefficients of the expectations, with respect to p, measure the rates at which these expectations will commence to be altered if p is gradually varied; and the greater these rates are, whether the expectations are increased or diminished as p is increased, or in other words, whether the differential coefficients are positive or negative, the more sensitively will the expectations respond to variations of p. Consequently, it might have been anticipated that the value of the observational record for our purpose would be simply related to the squares of these differential coefficients.

"We may now set out the process of calculation for the simple case of the estimation of the frequency of one of two classes.

TABLE 36

Observed Frequency, (x)	Expected Frequency, (m)	Differential Coefficient, dm/dp	$\frac{1}{m}\left(\frac{dm}{dp}\right)^2$
a	pn	n	n/p
b	qn	-n	n/q
n	n	o	n/pq

"The frequencies expected are found by multiplying the number of observations, n, by the theoretical frequency, p, which is the object of estimation, and by its complementary frequency, q. The differential coefficients of these expectations with respect to p are simply n and -n. The sum of these is zero, as must be the case whenever, as is usual, the number of observations made is independent of the parameter to be estimated. It is obviously, therefore, not the total of the differential coefficients which measures the value of the data, but effectively the extent to which these differ in the different distinguishable classes, as measured by their squares appropriately weighted, as shown in the last column.

"The total amount of information is found to be

$$I = \frac{n}{pq},$$

and we may now note the well-known fact that, if our sample of observations were indefinitely increased, the estimate of p, obtained from the data, tends in the limit to be distributed normally about the true value with variance $\frac{pq}{n}$." (ref. 2, pp. 216-218).

2. THE MATHEMATICAL FORMULATION OF FISHER INFORMATION AND THE CRAMER-RAO LOWER BOUND

A summary of the basic theoretical results utilized in the remainder of this report is presented in this section. The theorems are presented without proof. For a thorough discussion of the theory, see reference 3.

Even though redundant, the theory is divided into two cases: (1) the case where the underlying probability distribution depends on one real parameter Θ and (2) the case where the distribution depends on a vector of parameters $\vec{\Theta} = (\Theta_1, \Theta_2, \cdots, \Theta_n)$. Perhaps this approach will be helpful to those who are unfamiliar with the subject.

2.1 THE CASE OF ONE REAL PARAMETER

Throughout this section, we assume that the data X from an experiment are generated by an underlying probability process with a probability density function from the one parameter family of densities $\{f(X;\Theta) | \Theta$ is a real number$\}$; i.e., X is a random sample from a population having a probability density function $f(X;\Theta)$, where Θ is a real number.

Theorem (2-1)

a. Definition: The Fisher information regarding the parameter Θ in the experiment that yields the sample X is defined by

$$I_X(\Theta) = E\left\{\left[\frac{\partial}{\partial\Theta} \ln f(X;\Theta)\right]^2\right\}$$

b. Theorem: The information can be computed by the alternative formula

$$I_X(\Theta) = -E\left[\frac{\partial^2}{\partial\Theta^2} \ln f(X;\Theta)\right]$$

Theorem (2-2)

a. Theorem: Information is additive over independent experiments. Thus, if X and Y are independent, then

$$I_X(\Theta) + I_Y(\Theta) = I_{(X,Y)}(\Theta)$$

b. Corollary: The information in a random sample of size n is just n times the information in a single observation

$$I_Y(\Theta) = nI_X(\Theta)$$

where $Y = (X_1, X_2, \cdots, X_n)$

Theorem (2-3)

The information provided by a sufficient statistic $T = t(X)$ is the same as that in the sample X

$$I_T(\Theta) = I_X(\Theta)$$

Theorem (2-4)

If $T = t(X)$, then

$$I_T(\Theta) \leq I_X(\Theta)$$

with equality holding if and only if the statistic T is sufficient.

Theorem (2-5)

a. Theorem: For any statistic $T = t(X)$, the relationship between the information $I_X(\Theta)$ and the variance of the statistic T is given by the information inequality

$$\mathrm{Var}(T) \geq \frac{[\Psi_T'(\Theta)]^2}{I_X(\Theta)} = \frac{[1 + b_T'(\Theta)]^2}{I_X(\Theta)}$$

where $\Psi_T(\Theta) = E(T|\Theta)$ and $b_T(\Theta)$ is the bias in the estimator T at Θ. The information inequality is also known as the Cramer-Rao inequality or the Frêche inequality.

b. Corollary: The information inequality can also be expressed in terms of the mean squared error (MSE).

$$MSE(T) = var(T) + b_T^2(\theta) \geq \frac{[1 + b_T'(\theta)]^2}{I_X(\theta)} + b_T^2(\theta)$$

c. Corollary: In the class of unbiased estimators, the information inequality provides a lower bound that is independent of the estimator; this bound is just the reciprocal of the information in the data.

$$MSE(T) = Var(T) \geq \frac{1}{I_X(\theta)}$$

This lower bound is usually referred to as the Cramér-Rao lower bound. Generally, an estimator T is said to be efficient if it obtains this bound.

2.2 THE CASE OF SEVERAL PARAMETERS

As in the case of a single parameter, we assume that the data X are generated by a probability function from the family $\{f(X;\vec{\theta})|\vec{\theta}\}$. Now, however, $\vec{\theta}$ is a k-dimension real vector and hence no longer required to be a real number. In this setting, the Fisher information number becomes the information matrix.

Theorem (2-6)

a. Definition: The information matrix regarding the parameter $\vec{\theta} = (\theta_1, \theta_2, \cdots, \theta_k)$ in the experiment that yields the sample X is defined to be the matrix $I_X(\vec{\theta}) = [I_{ij}]$; $i,j = 1,2,\cdots,k$ where

$$I_{ij} = -E\left[\frac{\partial^2 \ln f(X;\vec{\theta})}{\partial\theta_i \partial\theta_j}\right]$$

If appropriate changes are made in the notation, most theorems for the single parameter case remain valid in the more general setting of several parameters. These changes usually involve:

1. Substitution of the information matrix $I_X(\vec{\theta})$ for the Fisher information number $I_X(\theta)$
2. Interpretation of the real inequality $A \leq B$ to mean the matrix $B - A$ is positive semidefinite
3. Replacement of the variance of T, var(T), with the variance covariance matrix V.
4. Replacement of the real number, $[\Psi_T'(\theta)]^2 \div I_X(\theta)$, by the analogous matrix expression, $\Delta\, I_X(\vec{\theta})^{-1}\, \Delta^T$.

The following theorems and corollaries summarize the major results in the several parameter setting. For a more complete discussion of the general theory, see reference 3, pages 326 to 331.

b. Theorem: Let $h_1(X), h_2(X), \cdots, h_r(X)$ be r statistics such that

1. $E(h_i) = g_i(\vec{\theta})$; $i = 1,2,\cdots,r$
2. $V = [V_{ij}]$
   where

$V_{ij} = E[(h_i - g_i)(h_j - g_j)]$ and $i,j = 1,2,\cdots,r$.

3. $$\frac{\partial g_i}{\partial\theta_j} = \frac{\partial}{\partial\theta_j}\int h_i f(X;\theta)dX = \int h_i \frac{\partial f(X,\theta)dX}{\partial\theta_j}$$

   where Δ is the r by k matrix

$$\frac{\partial g_i}{\partial\theta_j} \;;\; i = 1,2,\cdots,r \;;\; j = 1,2,\cdots,k$$

Then, the matrix

$$V - \Delta\, I_X(\vec{\theta})^{-1}\, \Delta^T$$

is positive semidefinite. If the information matrix is singular, then its inverse in the above expression is replaced with the generalized inverse.

c. Corollary: Suppose I^{mn} are the elements of the matrix inverse of the information matrix $I_X(\vec{\theta})$. Then

$$V_{ii} \geq \sum_{m=1}^{k}\sum_{n=1}^{k} I^{mn}\left(\frac{\partial g_i}{\partial\theta_m}\right)\left(\frac{\partial g_i}{\partial\theta_n}\right)$$

This is a generalization to the many parameter case of theorem (2-5). It shows that the variance of any estimator of $g_i(\vec{\theta})$ is greater than or equal to a quantity which is independent of the method of estimation. When $g_i(\theta) = \theta_i$, the above equation reduces to

$$V_{ii} \geq I^{ii} \geq (I_{ii})^{-1}$$

The last inequality

$$V_{ii} \geq (I_{ii})^{-1}$$

gives a lower limit for the Cramér-Rao lower bound of the second corollary in theorem (2-5).

Theorem (2-7)

a. Theorem: In addition to conditions 1 and 2 of theorem (2-6), suppose $\vec{T}$ is a sufficient statistic for the vector $\vec{\theta}$. Then, there exist r functions $k_1, k_2, \cdots, k_r$ of the sufficient statistic $\vec{T}$ such that

$$E(k_i) = g_i(\vec{\theta}) \;;\; i = 1,2,\cdots,r$$

Furthermore, if

$$U = [U_{ij}]$$

where $U_{ij} = E[k_i - g_i)(k_j - g_j)]$ and $i,j = 1,2,\cdots,r$, then the matrix $V - U$ is positive semidefinite.

b. Corollary: Since the matrix $(V - U)$ is positive semidefinite, it follows that

$$V_{ii} \geq U_{ii} \;;\; i = 1,2,\cdots,r$$

Hence, estimators with minimum variance are explicit functions of a sufficient statistic.

3. INFORMATION AND MIXTURES OF TWO DENSITIES

This section presents a review of B. M. Hill's 1963 paper, "Information for Estimating the Proportion in Mixtures of Exponential and Normal Distributions" (ref. 4), followed by a discussion of the implications of Hill's results to agricultural surveys based on remotely sensed data. Hill's basic expansion of the information $I_X(p)$

in the mixture density $f(X) = pf_1(X) + (1 - p) f_2(X)$ motivates similar result for the general mixture model. Familiarity with the derivation of basic expansion will facilitate of the theoretical development in the general setting.

Hill's paper is concerned with the information $I_X(p)$ for estimating the proportion p in a mixture $f(X) = pf_1(X) + (1 - p)f_2(X)$ of two densities $f_1(X)$ and $f_2(X)$. The main qualitative result is that, even in very simple situations, such as a mixture of two exponential or of two normal distributions with all parameters except p known, the expected precision in estimating p is very low unless the distributions in the mixture are well separated. Quantitative results include approximations for the information in mixtures of exponential and normal distributions. For example, in the case of two normal distributions, $N(\mu_1,\sigma)$ and $N(\mu_2,\sigma)$, where $[(\mu_1 - \mu_2) \div \sigma]$ is small, one may use the approximation

$$I_X(p) \approx \left(\frac{\mu_1 - \mu_2}{\sigma}\right)^2$$

<u>Theorem (3-1)</u>

Let X be a random variable and let $f(X), = pf_1(X) + (1 - p)f_2(X)$ be a mixture density associated with the probability density functions $f_1(X)$ and $f_2(X)$. Then the Fisher information regarding the proportion p in a random sample of size n from a population with probability density function f(X) is given by

$$\frac{m}{p(1 - p)}\left[1 - \int \frac{f_1(x)f_2(x)dx}{pf_1(x) + (1 - p)f_2(x)}\right]$$

Proof: The Fisher information in a single sample regarding p is

$$I_X(p) = -E\left\{\frac{\partial^2 \ln[f(X)]}{\partial p^2}\right\}$$

Computing the derivatives yields

$$I_X(p) = E\left\{\left[\frac{f_1(X) - f_2(X)}{f(X)}\right]^2\right\}$$

Taking the expected value and simplifying gives

$$I_X(p) = \int\left(\frac{[f_1(x) - f_2(x)]^2}{[pf_1(x) + (1 - p)f_2(x)]^2}\right)[pf_1(x) + (1 - p)f_2(x)]dx$$

$$= -\frac{1}{p(1 - p)}\int \frac{[f(x) - f_1(x)]\,[f(x) - f_2(x)]}{f(x)}\,dx$$

$$= \frac{1}{p(1 - p)}\left[1 - \int \frac{f_1(x)\ f_2(x)}{f(x)}\,dx\right]$$

where integration is replaced with summation for discrete distributions. Applying the corollary to theorem (2-1) establishes the theorem.

Since $[p(1 - p)]^{-1}$ is the information regarding p in a pure binomial situation, the additional uncertainty as to the population from which an observation comes in the mixture $pf_1(X) + (1 - p)f_2(X)$ is reflected in the factor

$$S(p,f_1,f_2) = \left[1 - \int \frac{f_1(x)\ f_2(x)}{f(x)}\,dx\right]$$

Clearly, $0 \le S(p,f_1,f_2) \le 1$. Thus, if the densities f_1 and f_2 do not overlap, then the full binomial information is obtained, whereas if f_1 and f_2 are identical, the information in the mixture concerning p is zero.

Theorem (3-1) can be applied at two different levels in agricultural surveys based on satellite-acquired data. One level of application is to the spectral-temporal data which is discussed in reference 5. The other level is to the classification results and the discussion follows.

Suppose that a random sample of n pixels is selected from a sampling cluster. (The cluster can be a segment, a full frame, or any other specified collection of pixels.) Further, suppose the crop class of each sampled pixel can be ascertained only according to the following probability functions.

$$P_s(x) = \begin{cases}\alpha \text{ for } x = s\\ 1 - \alpha \text{ for } x = o\end{cases}$$

and

$$P_o(x) = \begin{cases}1 - \beta \text{ for } x = s\\ \beta \text{ for } x = o\end{cases}$$

where s denotes the crop of interest, and o denotes all other crops. Setting

$$P(x) = pP_s(x) + (1 - p)P_o(x)$$

and applying theorem (3-1) yields

$$I(p) = n\,[p(1 - p)]^{-1}\,S(p,P_s,P_o)$$

where

$$S(p,P_s,P_o) = 1 - \sum_{x\epsilon\{s,o\}} \frac{P_s(x)P_o(x)}{P(x)}$$

$$= \left\{1 - \left[\frac{\alpha(1 - \beta)}{p\alpha + (1 - p)(1 - \beta)} + \frac{\beta(1 - \alpha)}{p(1 - \alpha) + (1 - p)\beta}\right]\right\}$$

This expresses the information available in the classified sample regarding the crop proportion p in terms of the classification probabilities. Equivalently, the Cramér-Rao lower bound for any estimator T of p based on the classified sample is given by

$$Var(T) \ge p(1 - p)[\psi_T'(p)]^2[nS(p,P_s,P_o)]^{-1}$$

As a numerical example, we consider the results from an analysis of the Large Area Crop Inventory Experiment (LACIE) sample segment 1664, located in North Dakota. The analysis was performed at the National Aeronautics and Space Administration (NASA), Lyndon B. Johnson Space Center (JSC), on December 20, 1978, using spectral-temporal-ancillary data from the 1978 crop year. In the analysis, 142 pixels (good type-2 dots) were labeled by a photointerpreter. Table 3-1 shows a two-way classification of the photointerpreter's attached pixel labels versus the true ground cover crop classes.

Combining the entries of table 3-1 for wheat and barley, we obtain

$$\alpha = 0.70,\ \beta = 0.83,\ \text{and}\ p = 0.33$$

and

$$S(p,P_s,P_o) = 0.28$$

where s indicates wheat or barley and o indicates neither wheat nor barley.

TABLE 3-1.- PHOTOINTERPRETER PIXEL LABELS VERSUS THE THREE GROUND COVER CROP CLASSES

True ground cover classes	Photointerpreter classifier labels		
	Wheat	Barley	Other
Wheat	22	4	13
Barley	0	7	1
Other	12	4	79

This means that, for this set of data, there is, on the average, only 28 percent as much information in a photointerpreter label as there is in a true ground cover label for use in estimating the crop proportion p. Furthermore, the variance of any estimator T of p must be greater than or equal to $3.58[\psi_T'(p)]^2$ times the corresponding binomial variance.

Figures 3-1 and 3-2 display the relationship (or lack of a relationship) between the average field size and the quantity $S(p,P_S,P_O)$ for all 1978 LACIE segments located in Montana, South Dakota, North Dakota, and Minnesota, for which ground cover data were collected. A simple count reflects an information loss of more than 80 percent in slightly over one-half of the results from these segments.

4. INFORMATION AND MIXTURES OF SEVERAL DENSITIES

Results similar to those for mixtures of two densities can be developed in the general setting (ref. 5).

The information matrix for a mixture of several densities furnishes an example (ref. 6). It is analogous to the second equation in the proof of theorem (3-1).

Theorem (4-1)

Let X be a random variable and let

$$f(X) = \theta_1 f_1(X) + \theta_2 f_2(X) + \cdots + \theta_{m-1} f_{m-1}(X) + (1 - \theta_1 - \cdots - \theta_{m-1}) f_m(X)$$

be a mixture density associated with the probability density functions $f_1(X), f_2(X), \cdots, f_m(X)$. Then the information matrix regarding the proportions $\theta_1, \theta_2, \cdots, \theta_{m-1}$ for a random sample of size n from a population with probability density function f(X) is given by the (m - 1) by (m - 1) matrix $I = n[I_{ij}]$, where

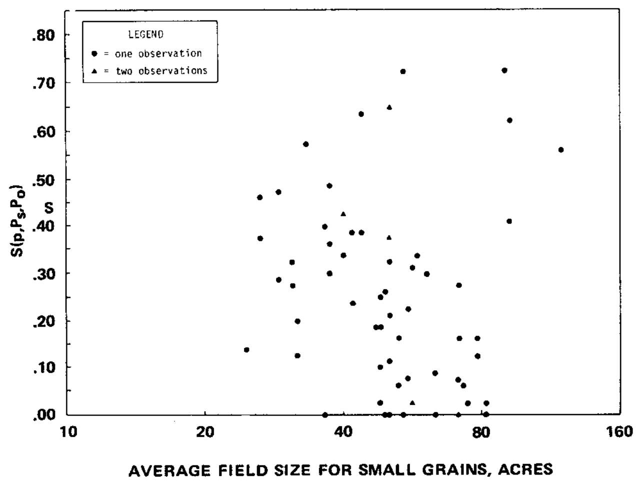

Figure 3-1.- Plot of field size versus quantity using type 1 dot data.

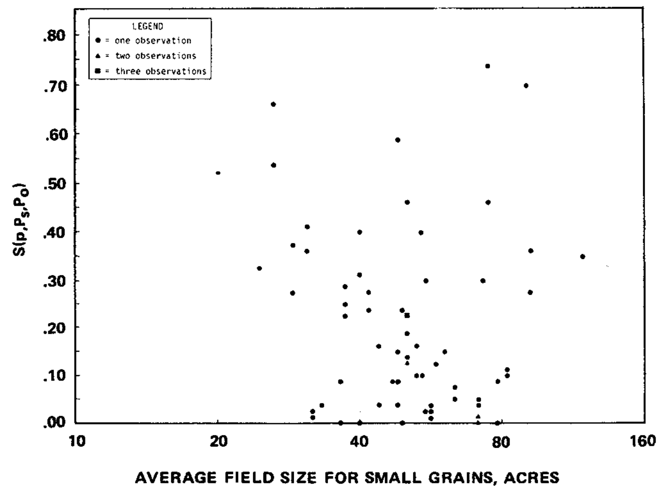

Figure 3-2.- Plot of field size versus quantity using type 2 dot data.

$$I_{ij} = \int \frac{[f_i(x) - f_m(x)][f_j(x) - f_m(x)]}{f(x)}\,dx$$

for $i,j = 1,2,\cdots,m - 1$.

The proof is similar to the one given for the Theorem (3-1)

Theorem (4-1) is a partial generalization of theorem (3-1); and, like theorem (3-1), it has two applications. To demonstrate the basic ideas, the application to the classification data is developed.

Suppose a random sample of n pixels is selected from a cluster. Further, suppose that the crop class of a pixel can be ascertained only according to the probability functions

$$P_\ell(x) = \begin{cases} P_{\ell 1} & \text{for } x = 1 \\ P_{\ell 2} & \text{for } x = 2 \\ \vdots & \\ P_{\ell m} & \text{for } x = m \end{cases}$$

where $\ell = 1,2,\cdots,m$ denotes the m crop classes.

Setting

$$P(x) = \Theta_1 P_1(x) + \Theta_2 P_2(x) + \cdots + \Theta_{m-1} P_{m-1}(x) + (1 - \Theta_1 - \cdots - \Theta_{m-1}) P_m(x)$$

and applying theorem (4-1) yields

$$I(\vec{\Theta}) = n\,[I_{ij}] \quad ; \quad i,j = 1,2,\cdots,m-1$$

where

$$I_{ij} = \sum_{x=1}^{m} \frac{[P_i(x) - P_m(x)][P_j(x) - P_m(x)]}{P(x)}$$

$$= \sum_{x=1}^{m} \frac{[P_{ix} - P_{mx}][P_{jx} - P_{mx}]}{\Theta_1 P_{1x} + \cdots + (1 - \Theta_1 - \cdots - \Theta_{m-1}) P_{mx}}$$

This expresses the information matrix in terms of the probability matrix P and the crop proportions $\Theta_1, \Theta_2, \cdots, \Theta_{m-1}$. To complete the analogy, suppose $T = (T_1, T_2, \cdots, T_{m-1})$ is an estimator of the crop proportions $\vec{\Theta} = (\Theta_1, \Theta_2, \cdots, \Theta_{m-1})$ based solely on pixel labels. Further, assume that $\vec{T}$ satisfies the normality condition of theorem (2-1) and denote the expected value of T_i by $\psi_i(\vec{\Theta})$. Then, an application of the second corollary in theorem (2-6) shows

$$\mathrm{Var}(T_i) \geq \sum_{\ell=1}^{m-1} \sum_{k=1}^{m-1} I^{\ell k} \left(\frac{\partial \psi_i}{\partial \Theta_\ell}\right) \left(\frac{\partial \psi_i}{\partial \Theta_k}\right)$$

for $i = 1,2,\cdots,m-1$. This result and other generalizations are interpreted numerically in reference 5.

5. CONCLUSIONS

a. The theory of Fisher information applied to mixture densities provides an appropriate and useful measure of the information in the Landsat data available for estimating crop proportions.

b. The examples indicate that considerable loss of information concerning the crop proportions occurs during the photointerpretation process.

c. If there is a relationship between average field size and information loss (photointerpreter labeling), then it is not apparent using the data set consisting of the 1978 LACIE segments located in Montana, South Dakota, North Dakota, and Minnesota.

6. REFERENCES

1. Feiveson, A. H.: Estimating Crop Proportion from Remotely Sensed Data. Proceedings of Technical Sessions - the LACIE Symposium, JSC-16015, Volume II, July 1979, pp. 633-646.
2. Fisher, R. A.: The Design of Experiments. Hafner Publishing Company, Inc. (New York) 1960.
3. Rao, C. R.: Linear Statistical Inference and Its Applications. John Wiley & Sons (New York) 1973.
4. Hill, B. M.: Information for Estimating the Proportions in Mixtures of Exponential and Normal Distributions. Journal of American Statistical Association, Volume 58, No. 305, pp. 918-932.
5. Perry, C. R.: A Summary of Observation Concerning the Information in the Spectral-Temporal-Ancillary Data Available for Estimating Ground Crop Proportions. AgRISTARS FC-J0-00486, JSC-16815, February 1981.
6. Odell, P. L.: The Cramer-Rao Lower Bound as a Criteria for Evaluating a Large Complicated Data Reduction System Such as LACIE. NASA-JSC (to be published).
7. Walker, H. F.: Estimating the Proportions of Two Populations in a Mixture Using Linear Maps. Comm. Stat., Theories and Methodology, A9(8), 1980, pp. 837-849.

ACKNOWLEDGEMENT

The author wishes to thank Dr. P. L. Odell of the University of Texas at Dallas for helpful suggestions in preparing this report.

---

*The research described herein was performed while the author was a National Research Council Senior Resident Associate with the Earth Resources Applications Division, NASA/JSC. The author is now assigned to the Early Warning/Corp Condition Assessment project within the AgRISTARS program, located at 1050 Bay Area Blvd., Houston, TX 77058.

* * *

REPLY TO DISCUSSANT'S COMMENTS

This work was developed only for the case where ground observations are unavailable, and applications are to foreign regions where ground collection of data is impossible. There is no "ground truth" except for small historical test data sets in the United States. Hence, the discussant's remarks regarding double sampling completely missed the point. This fact is pointed out on the second page of the discussant's reference: "When it is impossible to obtain the true classifications of the units in the sample, the above sampling scheme is not applicable." Thus in this setting, the discussant's quantity K (the square of the correlation coefficient between the true and the classified unit labels) suffers from the same questions of appropriateness as other measures of information.