Exact Binomial Confidence Interval for Proportions

Jeffrey T. Morisette and Siamak Khorram

Introduction

In remote sensing accuracy assessment applications, the confidence interval is commonly used as a way to establish an appropriate sample size. However, confidence intervals are also informative when included in the accuracy assessment report. Many reports and papers give accuracy figures and leave out confidence intervals. In those cases where a confidence interval is constructed, the standard approach is to derive the interval through the use of a normal approximation of the binomial distribution or by referring to exact tables. This note briefly discusses the benefit of using confidence intervals. The main objective is to describe the calculation of an *exact* equal-tail confidence interval for the proportions of correctly classified pixels. The exact confidence interval is not based on a normal approximation but, instead, uses the relationship between the binomial and F distributions. While the derivation of the exact confidence interval relies on a somewhat involved mathematical relationship between the probability density functions, in practice the exact confidence interval is a relatively straightforward formula. With the percentiles of "F" Distributions now commonly available in hand calculators and spreadsheet (and other) programs, the exact confidence interval is easily calculated. After presenting the formula for the exact confidence interval, we will present an example using both the normal approximation and the exact confidence interval.

Confidence Intervals in Remote Sensing Accuracy Assessment

Traditional accuracy assessment of classified land-cover data is based primarily on data summarized in an "error matrix" or a "confusion matrix." This is a square matrix, with the number of rows and columns equal to the number of categories in the classification. The elements of the matrix compare the relationship between reference data and the corresponding satellite image classification (Lillesand and Kiefer, 1994, p. 612). The error matrix, once constructed, can be used to obtain estimates such as the overall accuracy, user's accuracy, producer's accuracy (Aronoff, 1982) and the Kappa Coefficient (Congalton *et al.*, 1983). All of these figures are estimates based on the accuracy assessment sample. Because it is only a sample of location on the thematic map, the accuracy assessment figures contain a stochastic or random element. That is, we do not know the true accuracy of the classified data but acquire *estimates* from the accuracy assessment sample. Because the figures are estimates, there is an associated margin of error. The magnitude of this margin of error is relayed through the statistical concept of a *confidence interval*. This interval has an associated confidence coefficient, which can be interpreted as the probability that the given interval contains the true parameter (Steel and Torrie, 1980, p. 63; see also most any introductory text on statistical inference).

The most common use of confidence intervals in remote sensing studies appears to be as a guide for sample size determinations (several references related to this point, as well as a relevant discussion, appear in Congalton (1991) under the "Sampling Considerations" section, p. 44). However, in addition to determining sample size, confidence intervals are also helpful in relaying the statistical uncertainty associated with the given estimate and can be informative and useful when presented along with the accuracy figure extracted from the error matrix (Card, 1982; Richards, 1993, p. 275).

Calculation of Exact Confidence Intervals for Thematic Raster Data

Consider the estimate of accuracy

$$\hat{p} = \frac{x}{n} \tag{1}$$

where x is the number of sample sites correctly classified and n is the number of samples. Note that the x and n could pertain to the overall accuracy or the accuracy for a given class. A common method used to construct a confidence interval found in many elementary statistics texts is to use a large scale, normal approximation (Bain and Engelhardt, 1987, p.345; Brockett and Levine, 1984, section 5.3; Casella and Berger, 1990, example 9.4.5; Cochran, 1977, pp. 57-58; Larson, 1982, pp. 294-297; Montgomery, 1991, section 2.4.3). Equation 2 gives the formula used to derive the confidence interval using the normal approximation: i.e.,

$$\hat{p} - z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} < p < \hat{p} + z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} \tag{2}$$

where z is the $100 \times (1 - \text{alpha})^{\text{th}}$ percentile from a standard normal distribution.

In the remote sensing literature, Hord and Brunner (1976) use a normal approximation to construct a table of 95 percent confidence intervals as a function of x and n while Card (1982) uses the normal approximation for his confidence intervals for marginal thematic map accuracies. Further work similar to these two studies could utilize the exact confidence interval given below by implementing the exact confidence interval in those places where the normal approximation has been used.

Computer Graphic Center, North Carolina State University, Raleigh, NC 27695-7106.

J.T. Morisette is presently with Earth Satellite Corp., 6011 Executive Blvd., Suite 400, Rockville, MD 20852-3804 (jmoriset@earthsat.com).

Photogrammetric Engineering & Remote Sensing, Vol. 64, No. 4, April 1998, pp. 281–283.

0099-1112/98/6304-281$3.00/0

TABLE 1. APPROXIMATE AND EXACT CONFIDENCE INTERVALS FOR THREE SITUATIONS

				Normal Approximation		Exact Interval	
	x	n	p	lower limit	upper limit	lower limit	upper limit
Situation 1	24	25	0.96	0.8832	1.0368	0.7965	0.9990
Situation 2	48	50	0.96	0.9057	1.0143	0.8629	0.9951
Situation 3	96	100	0.96	0.9216	0.9984	0.9007	0.9890

The normal approximation assumes that the "margin of error" associated with the estimated parameter is symmetric. However, for a confidence interval on a proportion near one, say 0.98, the confidence interval if not symmetric. That is, the lower limit on the confidence interval is further from 0.98 than is the upper limit. In general, if the sample is relatively small or the proportion is near either 0 or 1, the normal approximation can be inappropriate. The common approach is to say that, for n greater than a number between 30 and 60, it is safe to use the normal approximation (Richards, 1993, p. 612). Samuels and Lu (1992) give more involved guidelines for when the normal approximation is acceptable. In the event that a normal approximation is not appropriate, one should turn to exact confidence intervals based on the binomial distribution itself (Cochran, 1977, p. 59). Blyth and Still (1983) present a detailed discussion of binomial confidence intervals and normal approximations; they also give a table for exact binomial 95 percent and 99 percent confidence intervals. The trouble with using binomial tables is that they may not be readily available and they may be limited to only a few confidence coefficients.

There is an exact confidence interval based on the relationship between the binomial distribution and the F distribution (Blyth, 1986). The confidence interval has the form

$$\frac{1}{1+\frac{n-x+1}{x}F_{2(n-x+1),\ 2x,\alpha/2}} \le p \le \frac{\frac{x+1}{n-x}F_{2(x+1),2(n-x),\alpha/2}}{1+\frac{x+1}{n-x}F_{2(x+1),2(n-x),\alpha/2}} \quad (3)$$

where $F_{v1,\ v2,\ \text{alpha}}$ is the upper 100 $\times$(1 $-$ alpha)$^{\text{th}}$ percentile from an F distribution with $v1$ and $v2$ degrees of freedom. Most introductory statistical texts contain a description of the F distribution and the associated degrees of freedom; for example, see Casella and Berger (1990, p. 449). Equation 3 requires that, in the endpoint adjustment, the lower endpoint is 0 if $x = 0$ and the upper endpoint is 1 when $x = n$. That is, if $x = 0$, the lower confidence limit is set to zero; if $x = n$, the upper confidence limit is set to 1. The derivation of the formula in Equation 3 is given in Blyth (1986). There are several steps required in the derivation, and the resulting formula is not intuitively obvious. However, the following example will hopefully show how the formula is easy to implement and provides more statistically sound confidence intervals.

Some Example Confidence Intervals

In these examples, we will present the upper and lower confidence intervals for both the normal approximation and the exact interval. Each of the confidence intervals will be for a 95 percent confidence coefficient. We will do this for three different situations. In all three cases, the estimated proportions are the same while the sample size differs.

First, consider the situation where 25 samples were collected. From these, 24 point were correctly classified. That is,

$$x = 24 \text{ and } n = 25, \text{ which results in } p = 0.96.$$

Using these numbers with the normal approximation (given in Equation 2), we get

$$0.96 - 1.96\sqrt{0.96(1-0.96)/25} < p < 0.96 + 1.96\sqrt{0.96(1-0.96)/25},$$

which results in the interval [0.883186, 1.036814].

Using these same $n = 25$ and $x = 24$ with the exact formula (given in Equation 3), we get

$$\frac{1}{1+\frac{2}{24}F_{4,48,.05/2}} \le p \le \frac{\frac{25}{1}F_{50,2,.05/2}}{1+\frac{25}{1}F_{50,2,.05/2}}$$

which results in the interval [0.7965, 0.9990]. This confidence interval and confidence intervals for two other situations are given in Table 1.

Note the desirable feature that the exact intervals is bounded by zero and one whereas, for $n = 25$ and 50, the normal approximation puts an upper limit on p beyond the range of the parameter. A common practice is to simply cut off the confidence interval at 1. However, this changes the width of the interval and thus changes the confidence coefficient associated with the interval. By construction, the exact confidence interval will not have a lower limit less than zero nor have an upper limit greater than one. This means that there is no need to adjust the confidence interval and, so, the given confidence coefficients can be maintained.

Conclusion

With the value of the F distribution built into statistical software packages, many hand calculators, and most current spreadsheet programs, the exact confidence interval is easy enough to compute. Because this was not always the case, the normal approximation was often the easiest way to estimate the desired confidence interval. When the normal approximation was not appropriate, tables were the easiest way to get exact confidence intervals. Because the percentiles of the F distribution built into calculators and software are now generally more accessible (and certainly less bulky and time consuming) than statistical tables, and given the relatively simple formula in Equation 3, it is as easy to construct the exact confidence interval as it is to use the normal approximation and it is easier than using tables.

We hope that the simple formula presented in Equation 3 will encourage those reporting overall accuracy assessment figures to include confidence intervals for the estimated proportion and use the exact confidence interval.

Acknowledgments

The authors would like to thank Drs. Russell Congalton, Dave Colby, Randy Ferguson, and Casson Stalling for helpful comments. We would also like to thank the anonymous reviewers who also provided suggestions that enhanced this note. The work of Jeff Morisette was supported by the North Carolina Space Grant Consortium and NASA's Earth System Science Graduate Student Fellowship.

References

Aronoff, S., 1982. Classification Accuracy: A User Approach/View, *Photogrammetric Engineering & Remote Sensing*, 48(8):1299–1312.

Bain, L.J., and M. Engelhardt, 1987. *Introduction to Probability and Mathematical Statistics*, Doxbury Press, Boston.

Blyth, C.R., 1986. Approximate Binomial Confidence Limits, *Journal of the American Statistical Association*, 81(395):843–855.

Blyth, C.R., and H.A. Still, 1983. Binomial Confidence Intervals, *Journal of the American Statistical Association*, 78(381):108–116.

Brockett, P., and A. Levine, 1984. *Statistics and Probability and Their Applications*, Saunders College Publishing, Philadelphia.

Casella, G., and R.L. Berger, 1990. *Statistical Inference*, Wadsworth and Brooks/Cole, Pacific Grove, California.

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

Congalton, R.G., 1991. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data, *Remote Sensing of Environment*, 37:35–46.

Congalton, R.G., R.G. Oderwald, and R.A. Mead, 1983. Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques, *Photogrammetric Engineering & Remote Sensing*, 49(12):1671–1678.

Card, Don H., 1982. Using Map Category Marginal Frequencies to Improve Estimates of Thematic Map Accuracy, *Photogrammetric Engineering & Remote Sensing*, 49(12):431–439.

Hord, R.M., and W. Brooner, 1976. Land-Use Map Accuracy Criteria, *Photogrammetric Engineering & Remote Sensing*, 42(5):671–677.

Larson, H.J., 1982. *Introduction to Probability Theory and Statistical Inference, Third Edition*, John Wiley & Sons, New York.

Lillesand, T.M., and R.W. Kiefer, 1994. *Remote Sensing and Image Interpretation*, John Wiley and Sons, Inc., New York.

Montgomery, D.C., 1991. *Introduction to Statistical Quality Control, Second Edition*, John Wiley & Sons, New York.

Richards, J.A., 1993. *Remote Sensing Digital Image Analysis: An Introduction, Second Edition*, Springer-Verlag, New York.

Samuels, N.L., and T.-F. Lu, 1992. Sample Size Requirements for the Back-of-the-Envelope Binomial Confidence Interval, *The American Statistician*, 46(3):228–231.

Steel, R.G.D., and J.H. Torrie, 1980. *Principles and Procedures of Statistics: A Biometric Approach, Second Edition*, McGraw-Hill, Inc., New York.

(Received 29 January 1996; accepted 18 June 1996; revised 18 June 1997)

Forthcoming Articles

Michel Arnaud and *Albert Flori*, **Bias and Precision of Different Sampling Methods for GPS Positions.**

Edward A. Ashton and *Alan Schaum*, **Algorithms for the Detection of Sub-Pixel Targets in Multispectral Imagery.**

Stéphane Chalifoux, François Cavayas, and *James T. Gray*, **Map-Guided Approach for the Automatic Detection on Landsat TM Images of Forest Stands Damaged by the Spruce Budworm.**

F.M. Danson, **Teaching the Physical Principles of Vegetation Canopy Reflectance Using the SAIL Model.**

Sheldon D. Drobot and *David G. Barber*, **Towards Development of a Snow Water Equivalence (SWE) Algorithm Using Microwave Radiometry over Snow Covered First-Year Sea Ice.**

George F. Hepner, Bijan Houshmand, Igor Kulikov, and *Nevin Bryant*, **Investigation of the Integration of AVIRIS and IFSAR for Urban Analysis.**

Stanley R. Herwitz, Robert E. Slye, and *Stephen M. Turton*, **Co-Registered Aerial Stereopairs from Low-Flying Aircraft for the Analysis of Long-Term Tropical Rainforest Canopy Dynamics.**

Michael E. Hodgson, **What Size Window for Image Classification? – A Cognitive Perspective.**

Ross S. Lunetta, John G. Lyon, Bert Guindon, and *Christopher D. Elvidge*, **North American Landscape Characterization Dataset Development and Data Fusion Issues.**

Magaly Koch and *Farouk El-Baz*, **Identifying the Effects of the Gulf War on the Geomorphic Features of Kuwait by Remote Sensing and GIS.**

Kenneth C. McGwire, **Mosaicking Airborne Scanner Data with the Multiquadric Rectification Technique.**

Kenneth C. McGwire, **Improving Landsat Scene Selection Systems.**

Victor Mesev, **The Use of Census Data in Urban Image Classification.**

S.V. Muller, S.A. Walker, F.E. Nelson, N.A. Auerbach, J.G. Bockheim, S. Guyer, and D. Sherba, **Accuracy Assessment of a Land-Cover Map of the Kuparuk River Basin Alaska: Considerations for Remote Sensing.**

Ram M. Narayanan and *Brian D. Guenther*, **Effects of Emergent Grass on Mid-Infrared Laser Reflectance of Soil.**

Elijah W. Ramsey III, Dal K. Chappell, Dennis Jacobs, Sijan K. Sapkota, and *Dan G. Baldwin*, **Resource Management of Forested Wetlands: Hurricane Impact and Recovery Mapped by Combining Landsat TM and NOAA AVHRR Data.**

E. Terrence Slonecker, Denice M. Shaw, and *Thomas M. Lillesand*, **Emerging Legal and Ethical Issues in Advanced Remote Sensing Technology.**

M. Stojic, J. Chandler, P. Ashmore, and *J. Luce*, **The Assessment of Sediment Transport Rates by Automated Digital Photogrammetry.**

David M. Stoms, Michael J. Bueno, Frank W. Davis, Kelly M. Cassidy, Kenneth L. Driese, and James S. Kagan, **Map-Guided Classification of Regional Land-Cover with Multi-Temporal AVHRR Data.**

Chuang Tao, Rongxing Li, and *Michael A. Chapman*, **Automatic Reconstruction of Road Centerlines from Mobile Mapping Image Sequences.**

Randolph H. Wynne, Thomas M. Lillesand, Murray K. Clayton, and *John J. Magnuson*, **Satellite Monitoring of Lake Ice Breakup on the Laurentian Shield (1980-1994).**

David A. Yocky and *Benjamin F. Johnson*, Repeat-Pass **Dual-Antenna Synthetic Aperture Radar Interferometric Change Detection Post-Processing.**