STATISTICAL METHODS FOR MONITORING NUCLEAR TESTS

H.L. Gray and W.A. Woodward
Southern Methodist University

ABSTRACT

In previous papers, these authors have developed general methodology for detecting outliers. In particular, given a training set of, say, earthquake data, a test of hypothesis for testing whether or not a new observation should be classified as an earthquake was developed. The method was based on a generalized likelihood ratio test which did not require normality assumptions or even continuous data. In the most recent papers the technique was modified to allow missing data. In this paper the methodology is greatly extended to address a variety of issues that arise in a multistation environment. In particular, the method is generalized to allow the inclusion of expert opinion as part of the data and it is extended to allow an outlier to be an outlier to more than one population. Thus, for example, one might have training data from earthquakes and mining explosions and desire to know if an observed event should be classified as belonging to either of these groups or not. With the results of this paper, one can test, at a specified level, whether or not the new observation belongs to the joint population.

Alternative approaches to the generalized likelihood method are also considered for the problems addressed. However, in general, the generalized likelihood approach seems to be the best method of those considered.

Keywords: outlier detection, generalized likelihood ratio, multiple stations, discrimination.

19960624 187

869

OBJECTIVE:

The goal of this research is to establish the statistical methodology required to automatically detect the occurrence of a nuclear test anywhere in the world, given suitable discriminants.

PRELIMINARY RESEARCH RESULTS:

In previous papers these authors have developed automated methods for detecting outliers based on the generalized likelihood ratio (GLR). Although those methods are reasonably general and robust, they do not address all of the problems which can occur in a multistation, multisensor environment. In what follows, we list several of the problems and their partial solutions.

A)   Does adding m stations with p-discriminants create a computing problem?

Answer: It could, if, for example, $M = p \times m > 200$.

To be more specific, let $V = $ data vector,
where
$$V' = (X_{11}, \cdots, X_{p1}, X_{12}, \cdots, X_{p2}, X_{1m}, \cdots, X_{pm}),$$

where X_{ij} denotes the discriminant i measured at station j.

Using the generalized likelihood ratio approach involves inversion of an $M \times M$ matrix which could be near singular if care is not taken in selecting the discriminants and the stations.

Remark: It should be noted that neither stations nor discriminants need to be seismic. Thus, V can be a mixture of seismic and nonseismic data.

Since M could be large in the most general scenario, alternative approaches have and are being considered. Two alternatives which we have thoroughly investigated are based on inverse variance weighting and minimum variance weighting. The advantage of these approaches would be in the computing area, i.e., the data could be significantly compressed and hence reduce computing requirements substantially. Unfortunately extensive simulations show these methods to be unsatisfactory for reasons that are clearly demonstrated by the following example:

Setting: one discriminant, two stations.
Let
$$X_{1j} = \text{discriminant measured at station } j = 1, 2,$$
$$\mu_j^{(EQ)} = \text{mean of discriminant at station } j \text{ for earthquakes,}$$
$$\mu_j^{(EX)} = \text{mean of discriminant at station } j \text{ for explosions,}$$

and

$$Y = W_1 X_{11} + W_2 X_{12},$$

where $W_1 + W_2 = 1$ and W_1 and W_2 are chosen so that Y will have minimum variance. Y will be referred to as a compressed discriminant. The compressed discriminant, Y, would then be used in the GLR method rather than X_{11} and X_{12}.

Example: Suppose the earthquake population has

$$\mu_1^{(EQ)} = \mu_2^{(EQ)} = 0$$

$$\sigma_1^{(EQ)} = 1, \quad \sigma_2^{(EQ)} = 2, \quad \rho = .75.$$

(1)

Then it is easily shown that $W_1 = 1.25$, $W_2 = -.25$.

But suppose V is an observation vector from a population defined by

$$\mu_1^{(EX)} = 2, \quad \mu_2^{(EX)} = 10$$

$$\sigma_1^{(EX)} = 1, \quad \sigma_2^{(EX)} = 2, \quad \rho = .75. \tag{2}$$

The observation vector $V' = (X_{11}, X_{12})$ would almost surely be an obvious outlier by the GLR method based on V. However

$$E[Y] = \tfrac{5}{4} E[X_{11}] - \tfrac{1}{4} E[X_{12}]$$

$$= 0. \tag{3}$$

Consequently the compressed discriminant has a mean which is identical to the compressed earthquake mean and since it has the same variance, it is clear that Y cannot be used to discriminate between earthquakes and explosions. Thus the power of a test based on Y in this case would be virtually zero, even though it does have a much reduced variance. Of course, variance reduction is not the goal here, outlier detection is. The GLR approach based on the uncompressed data is effective here and in general because it, in essence, addresses the right problem. Essentially it classifies V as an outlier if it is highly unlikely that it belongs to the hypothesized population.

Having given this example, we should add that it is an extreme one and, in fact, for most cases of interest, the minimum variance method of compressing the data is an effective one. However, since such cases as the previous example can occur, and since at this time we believe the computing requirements for the GLR approach can be handled, we are presently opting to stay with the GLR method on the noncompressed data. We refer to this as the full vector approach. In the future if computing becomes a problem, we believe that it will be possible to develop a

hybrid method based on the Mahalanobis distance that will alleviate the problem described here. Another possible approach is based on principal components, but at the present we are not pursuing that method. Thus the short answer to question A is <u>no</u>.

B) In the previous example we posed the question of an outlier in the form "Is the new observation an earthquake or not?" or "Is the new observation an explosion or not?" That is, "Is the observation a member of a specified group or not?" Of course, the better question might be, "Is the new observation an earthquake or a mining explosion or not?" More generally, the question is, "Is the new observation a member of any one of several groups or not?"

In order to address this problem, we must first make some adjustment to the GLR test. In previous work, it was assumed that the covariance structure of the outlier was the same as the training set population. Even though the GLR method is robust to this assumption, inspection of equation 5 below should make it clear that this would be a poor assumption in this case. To avoid this, we simply make the noninformative assumption that the outlier has a constant distribution over the realistic support of a potential outlier, Z. The generalized likelihood ratio statistic then becomes

$$\lambda_1 = \frac{\sup_{\theta} L_0(\theta)}{\sup_{\theta} L_1(\theta)}, \tag{4}$$

where

$$L_0(\theta) = \prod_{i=1}^{n} f(V_i, \theta) f(Z, \theta)$$

and

$$L_1(\theta) = c \prod_{i=1}^{n} f(V_i, \theta),$$

where f is the pdf of the vector V, and c is a constant.

The distribution of λ_1 is then obtained by bootstrapping. Note that if n is sufficiently large, $\lambda_1 \approx f(Z, \widehat{\theta})$, where $\widehat{\theta} =$ maximum likelihood estimate of θ from the training sample. In this event, the bootstrapping can be greatly simplified by just resampling the Z from the training sample. The table below shows that for samples as large as 50 or 60, very little is lost in approximating the significance level α by this reduced bootstrapping and there seems to be a small gain in power.

Table 1. Simulation Results (nominal level $\alpha = .05$)

n	Approximate Level		Approximate Power	
	Full	One	Full	One
15	.065	.118	.568	.729
20	.063	.100	.588	.709
25	.048	.081	.601	.704
30	.051	.084	.609	.718
50	.050	.064	.645	.696
100	.057	.059	.657	.677
150	.061	.057	.664	.703

Suppose we have training data from several different populations. For simplicity, assume training samples from two populations (earthquakes and explosions, for example). The extension to more than two will be obvious. In this case the mixture distribution is given by

$$f(V, \theta) = p_1 g_1(V, \theta_1) + p_2 g_2(V, \theta_2), \tag{5}$$

where $\theta = (p_1, \theta_1, \theta_2)$, $p_1 + p_2 = 1$, and $p_1 \geq 0, p_2 \geq 0$.

In (5) p_1 and p_2 are the mixing proportions and g_1 and g_2 are the pdf's for populations Π_1 and Π_2, respectively. It follows that f is the pdf of the mixed population Π_{12}. Now denote the training sample by the two random samples $V_{11}, V_{12}, \cdots, V_{1n_1} \in \Pi_1$, and $V_{21}, V_{22}, \cdots,$ $V_{2n_2} \in \Pi_2$, where $n_1 + n_2 = n$. We wish to test the hypothesis

$$H_0 : Z \in \Pi_{12}$$

vs.

$$H_1 : Z \notin \Pi_{12},$$

given the random vectors $V_{11}, V_{12}, \cdots V_{1n_1} \in \Pi_1, V_{21}, V_{22}, \cdots V_{2n_2} \in \Pi_2$.

This can now be done by using λ_1 in equation (4) and f in equation (5), where p_1 and p_2 are either given or estimated by the training samples. This method is currently being coded for simulation runs and will be ready to apply to actual data sets by September 1, 1995.

C)   The GLR test is essentially a nonparametric test of the hypothesis that $\mu_1 = \mu_2$ vs $\mu_1 \neq \mu_2$. In practice it may be that $\mu_1 \neq \mu_2$, but the deviation is in a direction which doesn't concern us. To be more specific, suppose we wish to test the hypothesis that a new one dimensional observation, Z, is an earthquake. Then we test $\mu^{(EQ)} = \mu_Z$ against $\mu^{(EQ)} \neq \mu_Z$. However, it may be for this particular discriminant that $\mu_Z > \mu^{(EQ)}$ is of no interest, so we would not like to include this possibility in the rejection region. This is not a problem for a small number of events, but for a large number of events maintaining an acceptable false alarm rate could result in a substantial loss of power or detection capability. What is needed is a test which allows a more focused alternative, i.e., in this case the alternative $\mu_1 < \mu_2$. That is, $\mu_{i1} < \mu_{i2}$ for each i. Although this is a simple problem in one dimension, it is difficult in p-space. One approach which we are pursuing to solve this problem is a quasi- Bayesian approach which essentially limits the support of the distribution to regions which are physically plausible. Problems inhibiting this approach have been computational and appear to be solved now by a closed form solution. Confirmation of this will be forthcoming by the time of presentation of this paper. Other approaches to restricting the critical region are being considered but are not sufficiently underway to report on here.

CONCLUSIONS and RECOMMENDATIONS:

Results so far are encouraging. However, one thing is clear: new and better regional discriminants are needed. For regions in which such discriminants become available, the methods discussed and developed here make near optimal use of that data for outlier detection from one or several groups. For regions where such data are lacking, expert opinion may be used in the "Bayesian" GLR method to partially bridge the gap.