4-25-1974

Statistical Methods for Determining Quality of Communication Channels

Scott Guthery

Andres C. Salazar

STATISTICAL METHODS FOR DETERMINING QUALITY OF COMMUNICATION CHANNELS

S. B. GUTHERY and A. C. SALAZAR

Bell Telephone Laboratories

Abstract

The reliable transmission of information over communication channels depends immeasurably on the adequacy of a maintenance program provided for the transmission medium. An integral part of such a program is the procedure used for determining the quality of a communication link. In this paper, we discuss several statistical methods for ascertaining those impairments and their respective levels which reveal the most information about the fidelity of the channel in signal transmission. For the purpose of illustration, an example is presented which demonstrates how a testing program could implement each particular statistical method.

I. Introduction

The transmission medium of a communication system may consist of a variety of links - wire pairs, repeaters, multiplexors, switching equipment, microwave radio paths, etc. Thus, a communication channel or connection set up for the transmission of information between two points in the system may have to be evaluated according to a variety of impairments such as background noise, nonlinear distortion, impulsive noise, envelope delay distortion, etc. Each impairment, of course, affects the quality of the channel differently according to the type of modulation technique being used.

To clearly delineate the objectives of a maintenance program for communication channels we must first define what is meant by channel quality. Further, we must be told how the channel can be observed. That is, what equipment is available and how many measurements are possible. It is then the objective of the maintenance program engineer to establish the limits on the channel characteristics which permit adequate transmission. Further, he must specify what order the measurements must follow in order to minimize the effort required for ascertaining quality. Perhaps most important of all is the accuracy which the maintenance engineer feels his testing program has in ferreting out any transmission problem.

In this paper we will delve into many aspects of the construction of a channel testing program for maintenance purposes. The heart of the program consists of discovering relationships between channel fidelity and channel measurements. Analytical efforts often fail to lead to useful relationships simply because the channel models used are simplistic in nature and often ignore a large number of possible impairments. In addition, mathematical models often do not deal with interrelationships between various channel impairments and when they do, the mathematical problems which evolve are often intractable.

A useful approach to determining the relevant and statistically significant relationships between channel quality and measurements is through the statistical analysis of a large data base linking the two aspects of a communications system - performance and impairment levels.

There is a myriad of pattern recognition techniques or feature selection algorithms which could be used to advantage in such a statistical analysis. However, the objective here is to construct a channel testing program which requires little or no computation by a craftsman. This criterion delimits severely the number of statistical methods useful for our purpose. We have found two sequential nonparametric methods to be of practical value for channel maintenance programs. They will be discussed along with a parametric statistical procedure (offered as contrast) known as linear discriminant analysis.

II. Features of Channel Testing

A. Quality

How good a channel is in supporting the transmission of information is intimately related to the type of information being sent. Clearly subjective tests are called for in evaluating speech or image data. However, when well defined symbols are being transmitted, quality judgments become objective in nature. In this case the term symbol error rate or symbol block error rate is an unambiguous standard of channel quality.

Since symbol transmission allows a definition of channel quality why not use it as the sole test for maintenance purposes? Doubtless, such a test is simple to perform in many cases. However, no information as to what is wrong with the channel evolves from such tests. Further, a symbol transmission run can become a lengthy test if error rates determining channel quality are very small.

B. Observability

A channel can be observed or probed by a large repertoire of measurements available to a craftsman. Through experience and feasibility he may know which particular measurements reveal the most information for troubleshooting purposes. All the possible measurements, however, provide a parametrization of the channel. It is this description of the channel which can later be used to determine relationships between impairment level and performance.

C. Measurement Criteria

Precise numerical relationships between channel measurements and the quality associated with that channel is a desired objective in a maintenance program. A typical relationship might claim that in order to record tolerable performance on a given channel (i.e., symbol error rate $< 10^{-6}$) no more than 10 impulse noise counts can be observed in five minutes on an X type counter.

A statistical approach to discovering relationships between channel quality and channel parameters requires observing many channels and the recording of large quantities of measurement data. The quality of each channel is ascertained by the transmission of a known sequence of symbols or a fixed period of an analog waveform depending on the kind of information format required for transmission. Next, parameterization of the channel is done by recording levels of various impairments. The objective then becomes one of statistically searching for relationships between the parameters and channel quality.

D. Order of Measurements

The frequency of finding certain intolerable levels of impairments in the communications medium is almost certain to differ widely between impairments. For example, our craftsman may know from experience that impairment Y is commonly found to be an objectionable channel feature. Hence, in his troubleshooting or maintenance work he checks out this aspect of the channel first. He proceeds with other measurements which test other, but less commonly objectionable, features of the channel. Statistical analysis can verify the order of measurements this craftsman uses and may lead to ranking the measurements in a more statistically precise manner. This ordering minimizes the time required to correct the substandard performance.

E. Accuracy

Any channel testing program can achieve only a certain degree of accuracy in detecting substandard channels. A statistical analysis of the accuracy of the testing program can be made before implementation is actually considered. The point here is that this important feature of a channel maintenance program is merely one inherent aspect of the statistical approach to determining relationships between channel quality and channel measurements.

III. Statistical Methods

We begin describing the statistical approach to solving the channel maintenance problems discussed in the previous section by first considering a conventional parametric statistical procedure. The purpose to be served by the first procedure is that of considering a large number of measurements simultaneously which then sets the stage for the sequential methods to follow.

A. Linear Discriminant Analysis (LDA)

1. The Linear Discriminant Method of Classification

Consider n channel classes, each defined by a specific grade of quality. Let Q denote the number of measurements $(x_1, x_2, \ldots, x_Q)$ which can be made on a channel. These measurements empirically lead to the concept of n Q-dimensional distribution functions each expressing the probability that a channel from its class has a vector value. $\vec{x} = (x_1, x_2, x_3, \ldots, x_Q)$. If R_j, $j = 1,2,\ldots n$ denotes that region in Q dimensional space in which a channel will be classified as belonging to class j then the probability of correct channel classification can be determined from the distribution functions.

If the elements of each of the n classes are normally distributed according to $N(\vec{\mu}_j, \Sigma)$, $j = 1,2,\ldots,n$, with the same covariance matrix but different means, then the minimization of the probability of misclassification occurs when R_j, $j = 1,2,\ldots,n$ is the set of points $\vec{x}$ where $u_j(\vec{x})$ is maximum over the set $\{u_k(\vec{x})\}_{k=1}^{n}$ and where

$$u_k(\vec{x}) = 2\vec{x}^T \Sigma^{-1} \vec{\mu} - \vec{\mu}_k^T \Sigma^{-1} \vec{\mu}_k. \qquad (1)$$

$u_j(\vec{x})$ is called the linear discriminant function for the class indexed by j.

The technique can be thought of as a search for a set of hyperplanes of dimension Q-1 (where n is the number of populations and Q is the number of variates, in our case, channel parameters) which partitions exhaustively a hypercube of dimension Q designated by the ranges of the values the variates can take, and whereby the partition separates the observations of any class from those of any other class. The partitioned hypercube then consists of n convex regions, each region being identified with a single population or class.

Equation (1) which leads to the definition of the hyperplanes separating the convex regions, is seen to depend upon the common covariance matrix Σ and the individual means of the populations. A rough measure of the population separation is then obtained by noting the difference in the means between the populations. An evaluation of the difference between the populations is obtained from the Mahalanobis statistic V given by

$$V = \sum_{i=1}^{Q} \sum_{j=1}^{Q} a^{ij} \sum_{r=1}^{n} N_r(\bar{x}_{ir}-\bar{x}_i)(\bar{x}_{jr}-\bar{x}_j)$$

where

a^{ij} are the entries of the matrix Σ^{-1}

N_r is the sample size of population r

$\bar{x}_{ir}$ is the sample mean of population r

$\bar{x}_i$ is the pooled sample mean of variate i given by

$$\bar{x}_i = \frac{\sum_{r=1}^{n} N_r x_{ir}}{\sum_{r=1}^{n} N_r}$$

The statistic can be used as χ^2 with $Q \cdot (n-1)$ degrees of freedom for testing the null hypothesis. In this case, the null hypothesis would be that the mean values of the n populations of channels designated by the error rates of specific data sets are really estimates of the same mean.

2. Application of LDA

We have considered a data base consisting of 39 measurements made on each of 484 channels. Four grades (namely, Classes I, II, III, IV) of quality were established and each channel was assigned a grade according to the predetermined specifications regarding the quality of the transmission of information it would support. The measurement vector consisted of many different types of meter readings including those measuring ambient noise, impulse noise, envelope delay distortion, harmonic distortion, intermodulation distortion, etc. The identical form of information transmission was used on each channel.

The mean value of individual classes and the respective V statistic values were computed for the data base according to each grouping of the channel classes (I vs II vs III vs IV; I,II,III vs IV; I,II vs III, IV; and I vs II,III,IV). In every case the V statistic value was high. A distinct difference in the classes was indicated by the effects of its measure, i.e., the $1-\alpha$ level of the χ^2 value was always less than 10^{-3}.

Channels were classified by evaluating Equation (1). This was carried out by a linear discriminant computer program [1], using the channel data base. The results of this classification method (known as the resubstitution method) are given in Table 1. Several conclusions can be drawn from these results.

Table 1 shows that on this data base the linear discriminant classification procedure does only moderately well in distinguishing among all four classes simultaneously. When the classes are separated in other ways, i.e., simple two value combination of the four classes, as shown, the classification accuracy is improved. This indicates that either the classification technique or the data is capable of yielding coarse but not fine distinctions between classes.

In the resubstitution method of classification, the same data are used for evaluating the linear discriminant function and for finding the percentage of misclassification (the off-diagonal terms of the confusion matrix). The results are therefore optimistic. How optimistic they are is not clear, but since the data base is large, the bias introduced by resubstitution should not be great. The degree of bias was examined by another means of estimating the misclassification rate. A recent paper [2] considers the merit of using linear discriminant techniques for estimating misclassification rates, and reported that three less optimistic (and presumably more valid) estimates can be obtained from the V statistic. The three estimates are termed the D, DS and the D* and the actual estimates of the misclassification rate obtained are given by $\Phi\left(-\frac{D}{2}\right)$, $\Phi\left(-\frac{DS}{2}\right)$ $\Phi\left(-\frac{D^*}{2}\right)$

where Φ is the standardized normal distribution. The parameter ψ denotes the estimate of the probability of misclassification obtained through resubstitution.

Estimates of linear discriminant misclassification rates (two classes only) for the channel data obtained through the D, DS, D* and Ψ methods are given in Table 1. It can be seen that in many cases all estimates are nearly the same. In those cases where they do differ we can assume that the data was not of multivariate normal character or the population sizes (N_1, N_2) of the classes were rather small compared to the number of dimensions Q.

In all cases of channel classification, no a priori distributions were assigned to the individual channel classes nor were misclassification costs assumed.

B. Partition Analysis

The objective of partition analysis is to examine the effects of a set of interrelated independent variables on a dependent variable. The analysis employs a nonsymmetrical branching process based on variance analysis to subdivide the sample into a series of subgroups which maximize one's ability to predict values of the dependent variable. A complete description of the technique is given in [3] so we shall describe its utilization here by way of an example.

1. An Example of Partition Analysis

The dependent variable under consideration is

$$\text{Channel Grade } Y = \begin{cases} 1 & \text{Excellent (Class I)} \\ 2 & \text{Good (Class II)} \\ 3 & \text{Fair (Class III)} \\ 4 & \text{Poor (Class IV)} \end{cases}$$

(Code / Actual Range)

and two exemplary independent variables might be

$$\text{Measurement } X_1 = \begin{cases} 1 & 0 \\ 2 & \text{1-5 units} \\ 3 & 6+ \end{cases}$$

(Code / Actual Range)

and

$$\text{Measurement } X_2 = \begin{cases} 1 & \text{0-19 units} \\ 2 & 20+ \end{cases}$$

(Code / Actual Range)

Suppose that the data consists of the following four observations:

Obs. No.	Y	X_1	X_2
1	1	1	1
2	1	2	1
3	2	3	1
4	4	2	2

Now there are four ways we can partition the observations using the variables X_1 and X_2 individually; namely,

$X_1 = 1$ vs. $X_1 = 2,3$,

$X_1 = 2$ vs. $X_1 = 1,3$,

$X_1 = 3$ vs. $X_1 = 1,2$,

and

$X_2 = 1$ vs. $X_2 = 2$.

In order to determine which partition is best we must have some measure of how well a partition discriminates between values of the dependent variable. The measure we shall consider is between sum of squares which may be described as follows:

Suppose $Y_1,\ldots,Y_N$ is a sample of dependent values which has been partitioned into two subgroups; $Y_{11},\ldots,Y_{1N_1}$, and $Y_{21},\ldots,Y_{2N_2}$. The total sum of squares in the sample is given by

$$TSS = \sum_{j=1}^{N} (Y_j-\bar{Y})^2$$

where $\bar{Y}$ is the overall mean and the sum of squares within each subgroup is given by

$$WSS_i = \sum_{j=1}^{N_i} (Y_{ij}-\bar{Y}_i)^2$$

where $\bar{Y}_i$ is the mean within each subgroup. The between sum of squares is given by

$$BSS = N_1(\bar{Y}_1-\bar{Y})^2 + N_2(\bar{Y}_2-\bar{Y})^2$$

and it is easily seen that

$$TSS = WSS_1 + WSS_2 + BSS$$

The effect of picking that partition which maximizes BSS is therefore to maximize the amount of variation accounted for by the partition while at the same time minimizing the amount of residual variation within the resulting subgroups.

In our example TSS = 6 and the BSS for each partition is given by

Partition	Part 1	Part 2	WSS_1	BSS	WSS_2
$X_1 = 1$ vs $X_1 = 2,3$	1	1,2,4	0	1.3	4.7
$X_1 = 2$ vs $X_1 = 1,3$	1,4	1,2	4,5	1.0	0.5
$X_1 = 3$ vs $X_1 = 1,2$	2	1,1,4	0	0	6
$X_2 = 1$ vs $X_2 = 2$	1,1,2	4	0.7	5.3	0

so that the partition on X_2 best discriminates between values of the dependent variable and accounts for about 88 percent of its variation.

Once a sample has been partitioned into two subgroups the same analysis can be applied to each of them thereby generating a succession of bipartite partitions of the original sample. In the discrimination of channel quality, we see that each partition leads to a group of channels whose quality is nearly the same. Thus, channel measurements and ranges used to generate the partitions lead in an obvious way to the identification of a population of channels whose quality for the transmission of information is nearly uniform.

C. Partition Analysis of a Channel Data Base

Forty independent variables (not identical to the 39 used in the LDA analysis) were considered at each stage in the partition analysis of a data base consisting of 501 channels. The few instances of missing data were coded as a value of zero with negligible effect. The results of the partition analysis appear in Tables 2-6 where we note that only partitions which accounted for more than three percent of the total sum of squares and lead to different error classes were considered (although at times not all channel classes are represented at the end points). The program used to perform these analyses was written by J. A. Sonquist and J. N. Morgan of the Institute for Social Research, the University of Michigan.[3] Tables 3-6 show the independent variable that was used to partition each group along with the range of the variable associated with the high quality channel group. The quality class distribution within each final subgroup is also given along with the maximum likelihood estimate of the channel class therein. Program limitations sometimes prevent each final subgroup from being associated with a different channel class and can also preclude a channel class from being identified with some final subgroup. The percent of correct classification figures and the expected number of tests needed to attain this classification rate is given in Table 2.

1. Application of Partition Analysis to the Identification of Data Channel Quality

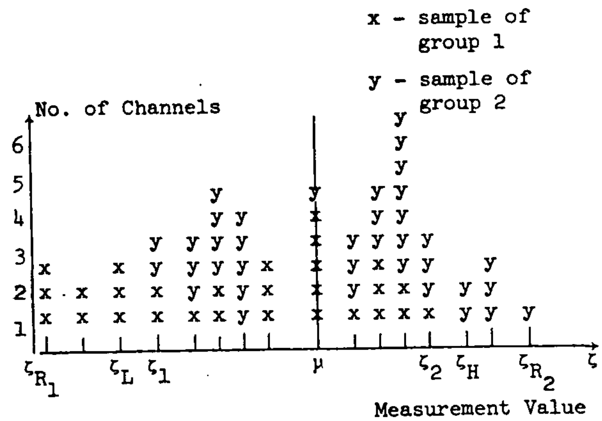
The advantage of partition analysis as a channel classifier lies in the fact that it sequentially "selects" the channel feature and associated range which best distinguishes quality in a data channel. The sequential aspect of the technique minimizes the number of tests a craftsman, for example, may need to make to ascertain channel quality. The time and economic considerations of this advantage demonstrate its importance in possible troubleshooting applications.

Of course, no feature or small number of channel features is seriously expected to "explain" the error rates observed in a channel. It is more likely that a great many causes of substandard performances are at work in a channel. Thus, in general, the more measurements we make in a channel, the better we can assess its quality. Further evaluation of any channel quality selector would then require consideration of both the economy and accuracy aspects of using many channel measurements to identify channel quality.

D. Nonparametric Test by Kendall

Channel quality identification has been found to be a difficult classification problem especially since channel measurement distribution forms are unknown. Hence a need for nonparametric testing procedures has been demonstrated. Nonparametric tests (or distribution-free methods) usually involve rank comparisons and it is this idea [4] which will form the nucleus of the channel classification procedure we will now discuss.

Each measurement has a sample distribution associated with it formed from a channel data base. Let the following distribution be a representative one.



DISTRIBUTION FOR MEASUREMENT Z

We have indicated several points on the ζ axis which are of special interest. The point μ is merely the sample mean. The other points ζ_L, ζ_1, ζ_2, ζ_H are demarcation points on the individual distribution samples of group 1 (x's) and group 2 (y's). The groups are the familiar observed error Classes I, II, III, IV as we wish to combine them. To consider them individually in the distribution we would of course need four symbols instead of the x and y we have chosen for this example. We observe that up to point ζ_L measurement z has a distribution portion formed by only members of group 1, namely x's. This means intuitively that no channel of group y can have a measurement z level that low. A similar situation occurs for the ζ_H point. There is no group 1 channel above ζ_2.

Thus, ζ_L and ζ_H can be used to identify correctly eight channels on the left and six channels on the right. Our testing procedure for measurement z then becomes

(a) Declare channel to be in group 1 if $\zeta < \zeta_L$

(b) Declare channel to be in group 2 if $\zeta \geq \zeta_H$

(c) Make no decision if $\zeta_L < \zeta < \zeta_H$

Notice that it is possible that $\zeta_L < \zeta_{R_1}$, namely that there is a mixture of x' and y's for every ζ value all the way to the left end of the distribution. Also, we envision the possibility of $\zeta_2 = \zeta_{R_2}$. When either of these cases occur (but not both) it is still possible to gain from our testing procedure with measurement z. Of course, the measurement is useless for identification purposes if $\zeta_1 = \zeta_{R_1}$ and $\zeta_2 = \zeta_{R_2}$.

It is possible to go through this procedure for every measurement of the Q available to us $z_j, j = 1,2,3,\ldots Q$. So we have a pair $(\zeta_L(K),\zeta_H(K))$ and a total $N(K)$ of channels identified correctly associated with each measurement, $k_1 = 1,2,\ldots,Q$. Suppose we choose as our first channel measurement that z_{K_1} for which $N(K_1)$ is the maximum, i.e., $N(K_1) = \max_{1\leq j\leq Q} \{N(j)\}$. Having identified $N(K_1)$ channels correctly with a $(\zeta_L(K_1),\zeta_H(K_1))$ pair we now turn to the set of channels for which no identity decision was made in this step. We know that z_{K_1} is useless as a possible identifier for the next step so we don't consider it. But we do redraw all our other distributions for $z_j \neq z_{K_1}$, $j = 1,2,\ldots,Q$ without those x's and y's (channels) which have been identified by z_{K_1}. Again we find that z_{K_2} which has a maximum $N(K_2) = \max_{\substack{1\leq t\leq Q\\ t\neq K_1}} \{N(t)\}$ and has $(\zeta_L(K_2), \zeta_H(K_2))$ among all redrawn distributions. Now we have $N(K_1) + N(K_2)$ channels correctly identified. Again we discard these identified channels from our set of data and redraw the distributions. However, this time we can consider z_{K_1} as a possible identifier but not z_{K_2} since a different set of data is now used for drawing the distributions. Our iterative procedure continues until either we classify all channels correctly (i.e., no more work to be done) or we come to an impasse formed by the situation in which <u>all</u> channel measurements have $\zeta_1(\ell) = \zeta_{R_1}(\ell)$ <u>and</u> $\zeta_2(\ell) = \zeta_{R_2}(\ell)$, $\ell = 1,2,\ldots,Q$. The latter event is unlikely to occur if we are using the smallest cell size possible for the distribution and we have a large number of measurements (Q).

Thus, by listing enough measurements with appropriate thresholds we can expect to identify all channels correctly. However, a reasonable testing procedure should involve only a few measurements, for example, a maximum of say 20. Hence, if our sequential extreme value testing scheme does not identify all the channels in the data base with a few measurements then little has been gained.

We list in Table 7 the sequence of measurements found for a data base consisting of 380 channels with 39 measurements for grouping I, II, III vs IV (the other groupings are not shown due to space limitations). A sequence of 89 measurements were required to identify the four classes individually while the other class groupings required about half this number or fewer. It is clear that there is a slow convergence of the algorithm to correctly identify all the channels.

IV. SUMMARY

Three statistical methods have been described for usage in a channel testing program. Many facets of such a program have been described in order to give proper perspective to the channel classification techniques presented. We have applied these techniques to dozens of sets of data bases and have observed their value in distinguishing channel quality from various characterizing measurements made on circuits constituting the transmission path. The application of these techniques (or slight variants of them) can aid in the construction of a channel maintenance program involving a great variety of possible measurements. A final program schedule, resulting from statistical analysis, would specify the type of measurement to be made, the levels of that measurement delineating channel fidelity, the order of importance for the measurement repertoire and finally an estimate of accuracy the program would have in distinguishing quality.

V. Acknowledgments

The authors express their gratitude to John Tischendorf for the many helpful discussions we have had with him concerning the project reported here. In addition, we thank Mrs. Barbara Butler and F. J. Brophy for their programming assistance.

<u>References</u>

[1] Biomedical Computer Programs, (W. J. Dixon, ed.) Health Sciences Computing Facility, UCLA, Nos. BMD05M and BMD07M, September 1, 1965.

[2] P. Langebruch, M. Mickey, "Estimation of Error Rates in Discriminant Analysis," Technometrics, February 1968, p. 715.

[3] J. A. Sonquist, J. N. Morgan, "Detection of Interaction Effects," Monograph 35, Institute for Social Research, U. of Michigan, 1964.

[4] M. G. Kendall, "Discrimination and Classification," from <u>Multivariate Analysis</u>, P. R. Krishnaiah, (ed.), Academic Press, 1966, pp. 165-185.