**Multiparameter Hypothesis Testing and Acceptance Sampling**

**PERFORMING ORGANIZATION NAME AND ADDRESS**
Florida State University
Tallahassee, FL 32306

**CONTROLLING OFFICE NAME AND ADDRESS**
U. S. Army Research Office
Post Office Box 12211
Research Triangle Park, NC 27709

**MONITORING AGENCY NAME AND ADDRESS (if different from above)**

**DISTRIBUTION STATEMENT (of this Report)**
Approved for public release; distribution unlimited.

**ABSTRACT**
The quality of a product might be determined by several parameters, each of which must meet certain standards before the product is acceptable. In this paper, a method of determining whether all the parameters meet their respective standards is proposed. The method consists of testing each parameter individually and deciding that the product is acceptable only if each parameter passes its test. This simple method has some optimal properties including attaining exactly a pre-specified consumer's risk and uniformly minimizing the producer's risk. These results are obtained from more general hypothesis testing results concerning null hypothesis consisting of the unions of sets.
MULTIPARAMETER HYPOTHESIS TESTING AND
ACCEPTANCE SAMPLING

by

Roger L. Berger
The Florida State University

FSU Statistics Report M555
USARO Technical Report D48

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

July, 1980

This research was supported by the U.S. Army Research Office Grant No.
DAAG29-79-G-0158.

Roger L. Berger is Assistant Professor, Department of Statistics, The
Florida State University, Tallahassee, Florida 32306.
ABSTRACT

The quality of a product might be determined by several parameters, each of which must meet certain standards before the product is acceptable. In this paper, a method of determining whether all the parameters meet their respective standards is proposed. The method consists of testing each parameter individually and deciding that the product is acceptable only if each parameter passes its test. This simple method has some optimal properties including attaining exactly a prespecified consumer's risk and uniformly minimizing the producer's risk. These results are obtained from more general hypothesis testing results concerning null hypotheses consisting of the unions of sets.

Key words: consumer's risk, multiple inference, uniformly most powerful.
1. INTRODUCTION

In many situations, the quality of a product is determined by several parameters. The product is of acceptable quality to the consumer only if each of the parameters meets certain standards. For example, an upholstery fabric must meet standards for strength, colorfastness, and fire resistance. Based on some measurements on the product, the consumer must decide whether the product is acceptable, i.e., all of the parameters meet the standards, or unacceptable, i.e., one or more of the parameters do not meet the standards. In making this decision the consumer wishes to use a rule which controls the consumer's risk at a small level.

If there is only one parameter and only one kind of measurement, then a standard quality controls text such as Burr (1976) or Duncan (1974) gives methods for making this decision. Different methods are given depending on whether the parameter is a mean, variance or proportion of defectives and on whether the measurements are counts of defective units (sampling by attributes) or measurements on a continuous variable (sampling by measurements). But no text that the author has found deals with the situation in which there are multiple parameters of interest.

This problem will be formulated as a hypothesis testing problem in which the null hypotheses states that one or more of the parameters do not meet their standards and the alternative hypothesis states that all of the parameters do meet their standards. Then the probability of a Type I error will be the consumer's risk. Thus an \( \alpha \)-level test will be one which controls the consumer's risk at less than or equal to \( \alpha \).

The test proposed herein is so simple it must not be new, but the author has not been able to find it described in hypotheses testing or quality control literature. The test is the following. A hypothesis test is done on each parameter individually at level \( \alpha \). The overall test rejects the null hypothesis and decides that all of the parameters meet their standards if and only if each individual test decides that the individual parameter meets its standard.
This test has several interesting properties. First, the individual tests are performed at level \( \alpha \) and yet the overall test has level \( \alpha \). Usually when doing simultaneous inference about many parameters (see, e.g., Miller (1966)) inferences about individual parameters must be done with an error rate of less than \( \alpha \) to achieve an overall error rate of \( \alpha \). This, for example is the basis of the Bonferroni method of simultaneous confidence intervals. Second, under very mild conditions, the level of this test is exactly \( \alpha \). So the test is not being too conservative by requiring each of the individual tests to decide that the individual parameter meets its standard. Third, under more restrictive conditions a result of Lehmann (1952) can be used to prove that this test is uniformly most powerful in a reasonable class of tests. In terms of risks this says that this test uniformly minimizes the producer's risk. These properties indicate that not only is the test extremely easy to implement, since it deals with only one parameter at a time, but it also seems to be a reasonably good test.

2. Basic Results

Let \( X = (X_1, \ldots, X_n) \) be a random vector of observations whose distribution is determined by a vector parameter \( \theta = (\theta_1, \ldots, \theta_K) \). Let \( \Theta \) denote the parameter space. Let \( \Theta_i, \ i = 1, \ldots, K, \) be subsets of \( \Theta \). Let \( \Theta_0 = \bigcup_{i=1}^{K} \Theta_i \). Let \( A' \) denote the complement of the set \( A \). Note that \( \Theta_0' = \bigcap_{i=1}^{K} \Theta_i' \). The problem to be considered is that of testing \( H_0: \theta \in \Theta_0 \) vs \( H_1: \theta \in \Theta_0' \). In the example in the introduction, \( \Theta_i \) is the hypothesis that \( \theta_i \) does not meet its standard.
If \( \theta_i \) must be greater than \( c_i \) in order to meet its standard then \( C_i = \{ \theta : \theta_i \leq c_i \} \).

If \( \theta_i \) must be between \( c_i \) and \( d_i \) to meet its standard, then \( \theta_i = \{ \theta : c_i \leq \theta_i \leq d_i \} \).

With this formulation, \( H_0 \) is the hypothesis that at least one parameter does not meet its standard and \( H_1 \) is the hypothesis that every parameter meets its standard. Note that \( K \), the number of subsets, may be less than \( \ell \), the number of parameters. This will be the case if some of the parameters are nuisance parameters and do not have standards associated with them.

Let \( \alpha \), \( \alpha \leq 1 \) be fixed. For \( i = 1, \ldots, K \), let \( \psi_i(x) \) be an \( \alpha \)-level test of \( H_{0i} : \theta \in \theta_i \) vs \( H_{1i} : \theta \in \theta_i' \), i.e., \( E_\theta \psi_i(x) \leq \alpha \) for all \( \theta \in \theta_i \). Let \( \psi \) be the test of \( H_0 \) vs \( H_1 \) which rejects \( H_0 \) if and only if every \( \psi_i \) rejects \( H_{0i} \).

Other authors such as Birnbaum (1954), Birnbaum (1955), Lehmann (1955) and Spjotvoll (1972) have considered testing hypotheses \( H_0 \) and \( H_1 \). But in all these papers, except the one result of Lehmann to be discussed in Section 3, the null hypothesis is of the form \( E_1 \).

Tsutakawa and Hewett (1978) propose the test \( \psi \) for a problem comparing regression lines. Theorems 1 and 2, which follow, are broad generalizations of a result they prove about a test based on a bivariate \( t \) distribution.

Wilkinson (1951) has proposed a test like \( \psi \) but in a very different situation. Wilkinson assumes that the individual tests are \( \alpha \)-level tests for all of \( H_0 \), not just \( H_{0i} \). Wilkinson also assumes the individual tests are independent which typically will not be the case in the problems considered herein.

The facts that \( \psi \) is always an \( \alpha \)-level test and under mild conditions has size exactly equal to \( \alpha \) are presented in the following two theorems.

Theorem 1. \( \psi \) is an \( \alpha \)-level test of \( H_0 \) vs \( H_1 \), i.e., \( E_\theta \psi(x) \leq \alpha \) for all \( \theta \in \theta_0 \).

Proof: Let \( R_i \) be the event that \( \psi_i \) rejects \( H_{0i} \). Then \( R = \bigcap_{i=1}^{\ell} R_i \) is the event that \( \psi \) rejects \( H_0 \). Fix \( \theta \in \theta_0 \). Then \( \theta \in \theta_i \) for some \( i \) and

\[
E_\theta \psi(X) = \sum_{i=1}^{\ell} P_\theta(R_i) \leq \sum_{i=1}^{\ell} P_\theta(R_i) = E_\theta \psi_i(X) \leq \alpha \text{ since } \theta \in \theta_i \text{ and } \psi_i \text{ is an } \alpha \text{-level test of }
\]
Theorem 2: Suppose $G_1 = \{ \theta \mid \theta_1 \leq a_1 \}, \theta = 1, \ldots, K$. Suppose the power of $\psi_1$ depends only on $\theta_1$ and $\theta_{K+1}, \ldots, \theta_{K+L}$. Suppose $E_{n_{\psi_1}}(X) = \alpha$ if $\theta_1 = a_1$. Suppose there are upper bounds $b_1$ (possibly infinite) such that for any fixed values of $\theta_{K+1}, \ldots, \theta_{K+L}$,

$$
\lim_{\theta_1 \rightarrow b_1, \theta \in G_0} E_{\psi_1}(X) = 1.
$$

Then $\psi$ has size exactly $\alpha$, i.e., $\sup_{\theta \in G_0} E_{\theta}(X) = \alpha$.

Proof: Let $R_i$ and $R$ be defined as in the proof of Theorem 1. Let $\theta_1 = (\theta_{11}, \ldots, \theta_{n_1}), j = 1, 2, \ldots\ldots$ be a sequence of parameter points satisfying $\theta_{11} = a_1, \theta_{1j} = b_j, j = 2, \ldots, \ldots$, and $\theta_{(K+1)}i, \ldots, \theta_{K+L}$ are fixed for all $i$. Then $\theta_1 \in G_1$ for all $i$ and $P_{\theta_1}(R_i) = 1 - E_{\theta_1}(X) = 1 - \alpha$ for all $i$. Also, for $j = 2, \ldots, K, \lim_{\theta_1 \rightarrow b_1} P_{\theta_1}(R_i) = 1 - \lim_{i \rightarrow b_1} P_{\theta_1}(R_i) = 1 - 1 = 0$.

Therefore,

$$
\sup_{\theta \in G_0} E_{\theta}(X) \geq \lim_{i \rightarrow b_1} E_{\theta_1}(X)
$$

$$
= \lim_{i \rightarrow b_1} P_{\theta_1}(\cap_{j=1}^K R_j)
$$

$$
= \lim_{i \rightarrow b_1} (1 - P_{\theta_1}(\cup_{j=1}^K R_j))
$$

$$
= 1 - \lim_{i \rightarrow b_1} \sum_{j=1}^K P_{\theta_1}(R_j)
$$

$$
= 1 - \lim_{i \rightarrow b_1} \left(1 - \alpha + \sum_{j=2}^K P_{\theta_1}(R_j)\right)
$$

$$
= 1 - (1 - \alpha) - 0 = \alpha.
$$

From Theorem 1, $\sup_{\theta \in G_0} E_{\theta}(X) \leq \alpha$.

Thus the size of $\psi$ is exactly $\alpha$. ||

The proof of Theorem 2 shows that for the test $\psi$ the maximum of the consumer's risk, the maximum of the probability of a Type I error, occurs when one parameter just fails the standard, $\theta_1 = a_1$, and all the other parameters are very good (large). This is exactly what would be expected. If many of the parameters do not meet their standards, it should be easy to decide that the product is unacceptable.
But if all the parameters but one are very good and the one exception nearly meets the standard, it will be most difficult to decide that the product is unacceptable.

The conditions of Theorem 2 will hold if $\theta_i$ is a normal mean and $\psi_i$ is a one tailed t-test or if $\theta_i$ is the proportion of non-defective items and $\psi_i$ is a one tailed binomial test. An important case in which the conditions of Theorem 2 will not hold is the case in which $\theta_i$ is a normal mean, the standard is $c_i < \theta_i < d_i$ and $c_i$ and $d_i$ are finite numbers. No $\alpha$-level test will have a power approaching one on this set of alternatives.

3. An optimality Result.

Lehmann (1952) considered multiparameter hypothesis tests. He was primarily concerned with testing the null hypothesis $H_1$ vs the alternative $H_0$. But he proved one result, Theorem 4.2, for the situation in which $H'_0$ is the null hypothesis. Although Lehmann did not speak in terms of combining individual tests to get an overall test, his Theorem 4.2 says that under certain conditions the test $\psi$ we have proposed is uniformly most powerful in a certain class. The remainder of this section is a review of Lehmann's result in terms of our notation.

A subset of $A$ of $n^K$ is called monotone if $x \in A$ and $y_i \geq x_i$, $i = 1, \ldots, K$, implies $y \in A$. Suppose $K = L$. The parameter space is the finite or infinite open rectangle $a_i < \theta_i < b_i$, $i = 1, \ldots, K$. Let $Y_1, \ldots, Y_K$ denote $K$ statistics. Suppose $p_{\theta}(Y)$ is the joint density of the $Y$'s and the positive sample space of the $Y$'s is the open rectangle $u_i < y_i < v_i$ independent of the $\theta$'s. Suppose the marginal distribution of $Y_i$ depends only on $\theta_i$ and $Y_i$ converges in probability to $v_i$ as $\theta_i \rightarrow b_i$. Suppose the joint distribution of the $Y$'s has the property that if $\theta$ and $\theta'$ are two parameter points with $\theta_i \leq \theta'_i$, $i = 1, \ldots, K$, then
Theorem 3. (Lehmann's Theorem 4.2).

Under the above assumptions, the test \( \psi \) which rejects \( H_0: \theta < \theta_0 \) if and only if each \( \psi_i \) rejects \( H_{0i}: \theta < \theta_i \) is uniformly most powerful among all non-randomized tests which have monotone rejection regions.

Although Lehmann did not discuss nuisance parameters, the results of Theorem 3 will continue to hold if the assumptions hold for each fixed set of values for \( \theta_{K+1}, \ldots, \theta_L \).

Lehmann gives some techniques for verifying the monotonicity property assumed about the joint distribution of \( Y_1, \ldots, Y_k \). It will hold, for example, if the \( Y \)'s are independent t-statistics and the \( \theta_i \)'s are normal means.

4. AN EXAMPLE

An example of specifications given in terms of many parameters may be found in the textile industry. Table 1 lists specifications for upholstery fabric from the American Society for Testing and Materials. The specifications give standards for nine parameters related to strength, dimensional stability, colorfastness and flammability.

The first four parameters might be assumed to be normal means; mean breaking strength, mean tear strength, etc. The first three standards say the mean must be greater than some value. The dimensional change standard gives an upper and lower bound for the mean. These four hypotheses might be tested using t-tests. An upper bound on the variance of the dimensional change measurements will have to be assumed in order to construct an \( \alpha \)-level test based on a t-statistics because
of the hypothesis' two sided form.

The last five parameters might be measured by binomial variables, each variable counting the number of units in a sample which pass the corresponding test. Each parameter would then be the proportion of units in the population (a particular manufacturer's output) which achieve one of the standards. The usual binomial test could be used to test each of these five, one-sided hypotheses.

It would be very difficult to posit a realistic multivariate model for these nine variables. Some are discrete and some are continuous. Some are likely to be correlated. Yet it is relatively easy to construct an \( \alpha \)-level test for each parameter individually.

These tests can be combined into the overall \( \Psi \) to test the hypothesis \( H_0 \) with a consumer's risk of \( \alpha \). Theorems 2 and 3 indicate that, in certain situations, the resulting acceptance plan is fairly efficient in terms of both producer's and consumer's risks.

5. CONCLUSIONS

Acceptance sampling procedures for individual parameters are well known. This paper proposes a way of combining these procedures in the situation in which the quality of a product is measured by standards on several parameters. Not only is the method easy to implement but it controls the consumer's risk at exactly a pre-assigned level in typical situations when the standards are one-sided (either upper or lower bounds). Under slightly more restrictive conditions this method also uniformly minimizes the producer's risk. This method can be used in hypothesis testing problems other than acceptance sampling if the null hypothesis is a union of sets.
### TABLE I

<table>
<thead>
<tr>
<th>Test</th>
<th>Minimum Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breaking Strength</td>
<td>50 pounds</td>
</tr>
<tr>
<td>Tongue Tear Strength</td>
<td>6 pounds</td>
</tr>
<tr>
<td>Surface Abrasion (heavy duty)</td>
<td>15,000 cycles</td>
</tr>
<tr>
<td>Dimensional Change</td>
<td>5% shrinkage, 2% gain</td>
</tr>
<tr>
<td>Colorfastness to:</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>class 4</td>
</tr>
<tr>
<td>Crocking</td>
<td></td>
</tr>
<tr>
<td>Dry</td>
<td>class 4</td>
</tr>
<tr>
<td>Wet</td>
<td>class 3</td>
</tr>
<tr>
<td>Light-40 AATCCF Fading Units</td>
<td>class 4</td>
</tr>
<tr>
<td>Flammability</td>
<td>Pass</td>
</tr>
</tbody>
</table>

REFERENCES


