The Philippine Statistician

Vol. xxvi, Nos. 3 & 4 1977

"ON THE ASYMPTOTIC EFFICIENCY OF THE MAXIMUM LIKELIHOOD ESTIMATOR AND A FUZZY VERSION"[1]

by Antonio D. Baccay[2]

A. Introduction

This paper is divided into two parts. The first part presents the development in the proof of the asymptotic efficiency of the maximum likelihood estimator. It starts with the classical proof(1) which is valid only when a distribution function follows the so-called regularity conditions. This proof has been expanded by Wolfowitz(2) to accomodate more types of distribution functions as long as such distribution function satisfies the so-called Uniformity Condition. The work of Wolfowitz is valid only for one-dimensional parameter case. The extension of this to n-dimensional parameter space has been worked out by Kaufman(3).

The problem in the proof of the asymptotic efficiency of the maximum likelihood estimator is that there still exist some distribution functions that neither satisfy the regularity condition nor the uniformity conditions. Some examples of distribution functions have been cited by Daniel(4). The question now is; how does one show that the maximum likelihood estimator of a particular distribution function is asymptotically efficient? What Daniels did was to come up with the conditions that are satisfied by the so-called "non-regular" case distribution functions. Thus, if a given distribution function does not satisfy the regularity condition but satisfies the conditions worked out by Daniels, then the asymptotic efficiency of the maximum likelihood estimator of that particular distribution function is assured. Unfortunately, distribution functions do not fall simply in this two categories, there still exist some distribution functions that defy regularity condition or the non-regular case.

---

[1] Abstracted from M.S. Thesis of the same title.

[2] Assistant Professor, U.P. Statistical Center

Recent developments in the proof for the asymptotic efficiency of the maximum likelihood estimator have been in the direction of modification of the maximum likelihood estimator itself and showing it to hold only as special case for some of these modified estimators. For instance, Daniels has come up with a smoothed maximum likelihood estimator. Actually, the maximum likelihood estimator has been modified into this to suit certain type of distribution functions. This modification should not be taken as a generalization of the maximum likelihood estimator, unlike the next modification that is about to follow. Weiss and Wolfowitz(5) have also done some extensions by modifying the maximum likelihood estimator to make it asymptotically efficient. They first call it "Generalized Maximum Likelihood," but further refinement and extensions of the said estimator have prompted them to revise it into the so-called "Maximum Probability Estimator."

The second part of this paper is the writer's discussion of the problem of the asymptotic efficiency of the maximum likelihood extimator using the concept of fuzzy sets develop by Zadeh(6) in 1965. Fuzzy set may be viewed as a generalization of the ordinary set concept that we know of. In the set concept that almost every statistician is familiar with nowadays, a given element is either a member or not a member of a given set. We can define a function that will show membership of an element to a set. For instance, we can let the value of a function equal to 0 if a given element is not a member of a set, and the value 1 if a given element is a member of a set. Hence, the function takes on two values only, 0 or 1, depending as to whether a given element is not a member or a member of a given set. Fuzzy set is similarly defined by a "membership" function whose counter-domain is any value between 0 and 1 inclusive; not only 0 or 1 as in ordinary set concept. The value of the membership function of fuzzy set is interpreted as an indication of the "degree" or grade of membership of a given element to a given set. In effect, fuzzy sets have some flexibility in judging whether a given element is a member more or less of a set or not by assigning a vaue to its membership function intermediate between 0 and 1. In other words, ordinary set which give too rigid a criteria for its membership function can be generalize to fuzzy sets whose membership function is indicative of the degree of membership of a given element to a given set.

Fuzzy set becomes relevant to asymptotic efficiency because the word "asymptotic" in itself is a sign of fuzziness. The set of estimates generated by the maximum likelihood estimator, as n is varied, is a fuzzy set and for each estimate (for a given n), we assign a grade of membership to show how close it is to the true parameter. A solution to the problem of the asymptotic efficiency of the maximum likelihood estimator is attempted, using fuzzy sets, by first showing the possibility of defining a maximum likelihood estimator set that is fuzzy and then by establishing that such a set includes an element or a subset of elements whose membership function is the maximum among all other elements of the maximum likelihood estimator set. The concept of fuzzy sets, nevertheless, is still nascent and the establishment of it on firmer and more rigorous grounds might lead to a breakthrough for a comprehensive proof for the asymptotic efficiency of the maximum likelihood estimator.

B. DEVELOPMENTS IN THE PROOF

Given a sample of size n, denoted by $(x_1, \ldots\ldots x_n)$, the statistician will usually want to find the parameter of the distribution (usually known except for the parameter involved) to which the sample comes from. The method of maximum likelihood is usually used to estimate the parameter. However, how sure are we that the estimator based from a sample of size n is a "good" estimate of the parameter? We can never be sure, but knowledge of the asymptotic distribution of the maximum likelihood estimator, as the sample size n becomes large, makes us confident to act as if the asymptotic distribution is the actual distribution (which it is, to a close approximation). This assertion is the essence of the asymptotic efficiency of the maximum likelihood estimator.

The asymptotic efficiency of the maximum likelihood estimator then implies that the estimator is consistent and asymptotically normal as n becomes large, and that the variance of the asymptotic distribution should equal to the Cramer-Rao lower bounds.

R.A. Fisher(7), who popularized the method of maximum likelihood, proposed a method of proving the asymptotic efficiency which later become the classical proof and was incorporated in a book by Gramer(1). The method of proof basically answers the two propositions stated above, namely, normality is first established on its asymptotic distribution and then its consistency, as n becomes large, with the variance equalling the Cramer-Rao lower bounds.

The classical proof relies heavily on the differentiation methods of calculus in locating the maximum of the likelihood function. For this reason the regularity conditions are necessary prerequisites for the validity of the proof, for they provide for the existence, particularly, of the second-order derivative of the likelihood function. The regularity conditions are:

1. For almost all x, the partial derivatives $\frac{\partial \log f(x/\theta)}{\partial \theta}$, $\frac{\partial^2 \log f(x/\theta)}{\partial \theta^2}$, and $\frac{\partial^3 \log f(x/\theta)}{\partial \theta^3}$, exist for all θ which is an element of Θ

2. For every $\theta \epsilon \Theta$, we have $\left|\frac{\partial f(x/\theta)}{\partial \theta}\right| < A_1(x)$, $\left|\frac{\partial^2 f(x/\theta)}{\partial \theta^2}\right| < A_2(x)$, and $\left|\frac{\partial^3 \log f(x/\theta)}{\partial \theta^3}\right| < Z(x)$, the functions A_1 and A_2 being integrable over $(-\infty, +\infty)$ while $\int_{-\infty}^{\infty} z(x)\, f(x/\theta)\, dx < W$, where W is independent of θ.

3. For every θ in Θ, the integral $\int_{-\infty}^{\infty}\left\{\frac{\log f(x/\theta)}{\theta}\right\}^2 f(x/\theta)\,dx$ is finite and positive.

Cases are known, however, when the distribution function does not satisfy the regularity conditions, hence asymptotic efficiency cannot be ascertained using the classical approach. This case which is generally known as the non-regular case, has been the subject of research in the literature and a particular result by Daniels(4) is mentioned in this paper.

Daniels proposed two sets of sufficiency conditions to treat the non-regular case. These weaker conditions for the asymptotic efficiency are given which do not involve the second derivative of the likelihood function. Again the method of proof to show asymptotic efficiency is to show that the asymptotic distribution of the estimator is consistent and normally distributed wih variance equal to the Cramer-Rao lower bounds. The set of sufficiency conditions proposed by Daniels proved asymptotic efficiency without appeal to the Wald-Wolfowitz result but there is a convexity requirement imposed which is frequently not satisfied. The second set of conditions dispenses with the convexity requirement at the expense of some specialization elsewhere, but consistency has to be established by the Wald-Wolfowitz method. Nevertheless, these two sets of sufficiency conditions do not seem to be exhaustive enough to incorporate all non-regular cases. Of course, it can be argued that some non-regular cases which still do not fall under the two sets of sufficiency conditions seldom occur in practice and may be disregarded. However, a proposal which is to be accepted as a theory must incorporate all possible cases. A more general situation is considered where a modified maximum likelihood procedure is shown still to yield an asymptotically efficient estimator. Note now that a modified maximum likelihood is defined to establish precisely its asymptotic efficiency.

Another limitation to the classical approach is the requirement that other estimators have to be asympotically normal. This presents a constraint to the practicing statistician who is seeking an asymptocically efficient estimator that is reasonable in a satistically operational sense: Why should he be limited only to estimators that are asymptotically normal? A partial answer to this is the inadequacy of a basis for comparing the amount of condensation of a normal and a non-normal distribution, if estimators are not restricted to asymptotically normal ones. A better argument, however, has been proposed by Wolfowitz(2) and has got to do with generalizing the limiting distribution of the other estimators. The condition he imposed on competing estimators, aside from the usual regularity conditions similar to that of Cramer is called, the Uniformity Condition. The Uniformity Condition is stated as:

When $f(^{o}/\theta)$ is the density of the X_i, the distribution of $\sqrt{n}(B_n - \theta)$ approaches a limiting distribution of $D(^{o}/\theta)$ which may depend on θ, uniformly in both arguments of D(for $\theta \in \Theta$).

Wolfowitz's work may be regarded as an extension of Cramer's work for he assumes the density function to satisfy the regularity conditions also. As a matter of fact, the regularity conditions imposed by Wolfowitz is an expanded version of Cramer's. It incorporates already some established ideas that have been developed since Cramer's proof, like the works of Wald(8) on the consistency of the maximum likelihood estimator and other major theoretical results in statistics.

The proof of Wolfowitz consisted of six lemmas and one main theorem. The methods of proof is no longer the same as Cramer or Daniels where consistency and asymptotic normality with minimum variance imply asymptotic efficiency. The reason for this is that we no longer limit ourselves to asymptotically normal competing estimators. The lemmas are proven to establish the mathematical rigor of the limiting distribution

$D(x/\theta)$. Also we define

$$K(\theta_1) = \lim_{\delta \to o} \sup\{u(\theta)/\theta_1 < \theta < \theta_1 + \delta) \tag{B.1}$$

and

$$k(\theta_1) = \lim_{\delta \to o} \inf(r(\theta)/\theta_1 - \delta < \theta < \theta_1)^1 \tag{B.2}$$

With these definition the final result of Wolfowitz is stated as:

$$\lim_{n\to\infty} P(-m < \sqrt{n}(\theta_n - \theta) < h/\theta_o) \tag{B.3}$$

$$\lim_{n\to\infty} P(-m + k(\theta_o) < \sqrt{n}(B_n - \theta_o) < h + K(\theta)/\theta_o),$$

where (B_n) is a sequence of estimators that satisfy the regularity conditions and the uniformity condition. θ_o is the true parameter and n and h are arbitrary positive numbers.

It was left to Kaufman(3) to extend Wolfowitz result to the multidimensional parameter case. That is, for any analogously uniform (B_n) and any convex symmetric set $S < R^k$, we have; 1 $[r(\theta), u(\theta)]$ is the median interval of $D(\cdot/\cdot)$.

$$\lim_{n\to\infty} P(\sqrt{n}(\theta_n - \theta) \in S) < \lim_{n\to\infty} P(\sqrt{n}(B_n - \theta) \in S), \tag{B.4}$$

The uniformity condition imposed on estimator sequences is somewhat weakened. This leads to a corresponding weakening of the result, but such modification seems to be necessary if we wish not to exclude many reasonable estimators. The proof also assumes any regularity condition that implies the uniform asymptotic joint-normality of $\sqrt{n}(\theta_n\text{-}Q)$ and the asymptotic sufficiency of (θ_n).

The method of proof is based on a theorem by T.W. Anderson (9) which states that if H and I are independent random K-vectors and if I has a probability density characterized by convex symmetric levels, then for any convex symmetric set

$S \subset R^k$;

$$P(H + I \in S) < P(I \in S) \tag{B.5}$$

A symmetrized version of this in terms of our problem is,

$P_\theta(\sqrt{n}(B_n - \theta) \in S) = P_\theta(\sqrt{n}(B_n - \theta_n) + \sqrt{n}(\theta_n - \theta) \in S)$ (B.6)
and that asymptotically $\sqrt{n}(\theta_n - \theta)$ satisfies the hypothesis for I.

Hence, if itwere possible to show that $\sqrt{n}(B_n - \theta_n)$ and $\sqrt{n}(\theta_n - \theta)$ were in some sense "asymptotically independent" then it becomes reasonable to expect that the Anderson's thorem might apply asymptotically:

$$\lim_{n \to \infty} P(\sqrt{n}(\theta_n - \theta) \in S)$$

$$\lim_{n \to \infty} P(\sqrt{n}(B_n - \theta_n) + \sqrt{n}(\theta_n - \theta) \in S). \tag{B.6}$$

Actually, such a program cannot be carried out directly.

By making use of the concept of asymptotic sufficiency of the maximum likelihood estimator (thus, the necessity of incorporating this as part of the regularity condition), we can define a "modified maximum likelihood" and competing estimators in place of $\hat{\theta}_n$ and B_n in such a way that asymptotic probabilities in convex symmetric sets are preserved while at the same time the above mentioned properties of asymptotic independence is attained. This modification process constitutes the bulk of the lemmas proven by Kaufman before his main theorem on the asymptotic efficiency of the maximum likelihood estimator for multidimensional parameter space.

Like Daniels, Weiss & Wolfowitz (5) have defined a new estimator which is asymptotically efficient even for the non-regular case. This new estimator is said to be a generalization of the maximum likelihood estimator or rather, the maximum likelihood estimator is a special case when the regularity conditions are assumed. This new estimator is called **Maximum** Probability Estimator.

G. A Fuzzy Interpretation

One new concept that might perhaps lead to a comprehensive proof of the asymptotic efficiency of the maximum likelihood estimator without even modifying the same is the concept of "fuzzy sets" introduced by Zadeh(6) in 1965. As the world implies, fuzzy sets deal with "classes" or "sets" that do not have precisely defined criteria of membership. It is this type of imprecision that arises when we say, for example, that a class of integer is much greater than n since the set of integers cannot be divided dichotomously into those that are much greater than n and those that are not for just how much is much greater. In general, what distinguishes such classes from classes that are well-defined in the conventional mathematical sense is the fuzziness of their boundaries. In effect, in the case with a fuzzy boundary, an object may have a grade of membership in it that lies somewhere between full membership and non-membership. Thus, a class that admits of the possibility of partial membership in it is called a fuzzy set.

We make a fuzzy statement or assertion when some of the words appearing in the statement or assertion in question are indicative of fuzzy sets. For example:

1. Juan de la Cruz married **young.** The class of men who married "young" is a fuzzy set because "young" to some may no longer look "young" to others.

2. Maria is **beautiful**. The class of "beautiful" women is fuzzy because what is beautiful to some may not be beautiful to others and the criteria are oftentimes subjective;

3. "y is **approximately** equal to 10" is fuzzy because we do not exactly know what value should y take to qualify as "approximately" equal to 10;

4. "x is **much larger** than 30," is also fuzzy because we do not know exactly what x should be in order for it to be called "much larger" than 30.

In these statements, the source of fuzziness are the underlined phases, which in effect define fuzzy sets.

Fuzzy sets become relevant to asymptotic efficiency because the word "asymptotic" in itself is a sign of fuzziness.

In this case we take the sets of estimates generated by the maximum likelihood estimator (we shall assume unidimentional parameter space for simplicity), as $n\phi$ is varied, as a fuzzy set and for each estimates (for a given sample size), we assign a grade of membership to show how close it is to the (unknown) true parameter.

Before proceeding any further, however, let us first formally define a fuzzy set as:

Definition C.1: Let Z = (z) denote a set of points (objects) with z denoting a generic element of Z. Then a fuzzy set A in Z is a set of ordered pairs,

$$A = [\langle Z, \lambda_A(z)\rangle], \quad z \in Z, \tag{C.1}$$

where $\lambda_A(z)$ is called the "grade of membership" of z in A. Thus, if $\lambda_A(z)$ takes value in a space π - called the membership space - then A is basically a function from z to π. The function $\lambda_a : z \to \pi$, that defines A, is called the membership function of A. (When π contains only two points 0 and 1, then A is not fuzzy and its membership function is the same as the usual characteristic function.)

A fuzzy set A in Z is a class without sharply defined boundaries, that is, a class in which a point or object z may have a grade of membership somewhere between full membership and nonmembership. The important point to consider is that such a fuzzy set can be defined precisely by associating each z its grade of membership in A. We shall assume for simplicity that π is the interval (0,1), with the grade 1 representing full membership on a fuzzy set. Therefore, a fuzzy set A in Z, although lacking in sharpy defined boundaries can be precisely characterized by a membership function that associates with each z in Z a number in the interval (0,1) representing the grade of membership of z in A.

It is important to consider also that in the case of a fuzzy set, it does not make sense to say that an object belongs or does not belong to a particular set, except for objects whose grade of membership in the set is O or 1/. Hence, if A is the fuzzy set of beautiful women, then the statement; "Maria is beautiful" should not be interpreted to mean that Maria belongs to A. Such a statement should rather be interpreted as an association of Maria to a fuzzy set A, as associaton which will be denoted by,

$$\text{Maria} \underset{n}{\in} A, \qquad \langle C.2a\rangle$$

to distinguish if from an assertion of belonging in the usual non-fuzzy set, that is,

$$\text{Maria} \in A, \qquad \langle C.2b\rangle$$

which is meaningful only when A is not fuzzy.

It should be noted also that the imprecision due to fuzziness does not stem from randomness but from a lack of sharp transition from membership in a class to nonmembership in it. Although the membership function of a fuzzy set has some semblance to a probability function, they differ essentially from each other; the notion of a fuzzy set is nonstatistical in nature. Correspondingly, the mathematical techniques for dealing with fuzziness are quite different from those of probability theory. The notion of probability measure in probability theory corresponds the simpler notion of membership function in the theory of fuzziness [1b].

We mention above that the statement "as n approaches infinity," is fuzzy. Recalling that n here refers to the sample size, we can find the maximum likelihood estimate $\hat{\theta}_n$, given n, as long as it satisfy the condition that $L(\theta, S) \leq L(\hat{\theta}, S)$. Of course, given n, we can generate a set of values for $\hat{\theta}_n$ by simply varying the composition of the sample. This set of values is random, by nature, and contains the maximum likelihood estimator. Consider only, however, a representative value of the maximum likelihood estimator, say, the mode. In general, the criterion for the representative value would be to select that particular maximum likelihood estimate that closely resemble the (unknown) true parameter θ_o. Actually, however, we do not know the true parameter θ_o. The set of values of the maximum likelihood estimate, given n, will have a unimodal distribution and possibly symmetric, so that we can take the mode as the representative value because that value occurs the most number of times and not much computation is involved in locating it. Hence, by definition of the maximum likelihood estimator, the mode must resemble quite closely the true parameter θ_o than any other maximum likelihood estimate for a given n. Let us designate the mode of the maximum likelihood estimator as $\hat{\theta}_{on}$.

Consider now the set of modes of the maximum likelihood estimator, called it the maximum likelihood estimator set, M, as n varies from one to infinity. This may be designated also as

$$M = [\theta_{o1}, \theta_{o2}, ..., \theta_{on}, ...] \qquad (C.3)$$

Obviously, M is finite if the population is finite, otherwise it is infinite. Moreover, M is a subset of the parameter space Θ.

Let us examine the elements of M. Each of these elements estimates that (unknown) true parameter θ_o. Intuitively, as we increase the sample size, we expect the estimate to closely resemble the true parameter θ_o. That is $\hat{\theta}_{o2}$, would be less efficient that $\hat{\theta}_{o15}$, and becomes more efficient as n becomes larger. For a finite population N, it will be feasible to compute the true parameter θ_o, and we can see that the maximum likelihood estimator set , M, will contain a sequence of elements of estimators whose value will tend to the computed true parameter θ_o. For infinite population, the same tendency will also hold but now we have no idea what the limit is because we do not exactly know the value of the true parameter θ_o. This would imply that the elements of the maximum likelihood estimator set is not random and have a tendency to cluster about a particular point, which we know as the true parameter θ_o. The non-randomness of the maximum likelihood estimator set is further strengthen by the fact that since the elements of this set are the modes of any representative value, that means that we are preassigning the values of the maximum likelihood estimator set so as to give it an estimate that is most likely to occur.

Moreover, it is possible to associate a membership function, called it $\lambda_M(\Theta)$, that specifies the degree or grade of membership or closedness of a particular $\hat{\theta}_on$ to the true parameter θ_o. For finite population, this will be computationally possible for θ_o can be computed precisely, but for infinite population or for a finite population with a very large N, the definition of the membership function might be quite difficult to specify. Nevertheless, it is still logically possible.

Hence, we can define the maximum likelihood estimator set, M, as a fuzzy set.

Definition C.2: Let $\Theta = [\theta]$ denote a space of points with θ_{on} denoting a generic element of Θ. Then a fuzzy set M in Θ is a set of ordered pairs

$$M = [\langle\hat{\theta}_{on}, \lambda_M(\Theta)\rangle], \quad \hat{\theta}_{on} \in \Theta, \qquad \langle C.4\rangle$$

where $\lambda_M(\theta_{on})$ is called the grade of membership of $\hat{\theta}_{on}$ in M.

For simplicity, let $\lambda_M(\theta)$ take value in the space (0,1). That is, if particular $\hat{\theta}_{on}$, deviates considerably from the true parameter θ_o, then the value of its membership function is close to zero, otherwise it becomes close to 1.

The choice of the mode as the representative value for the maximum likelihood estimator, given n, may not be operationally feasible. It means that for a particular sample of size n, we have to exhaustively compute all the maximum likelihood estimates, as we vary the composition of the sample. Hence, even with the use of computers, for each particular n and as we vary n, the computation of the maximum likelihood estimates, in order to identify that which occur the most number of times, would indeed be very tedious and sometimes impractical. Perhaps an alternative choice would be to select a particular sample of size n, and then compute the maximum likelihood estimate, $\hat{\theta}_{Hn}$. Let that be representative value when $n = n_1$. Then to compute a representative value when $n = n_1 + 1 = n_2$, we retain the sample of size n_1, and simply pick one more element to increase the size to $n_1 + 1 = N_2$ and then compute the maximum likelihood estimate based on this particular sample size n_2. This process may be continued to generate the elements of the set.

Having defined the maximum likelihood estimator set as fuzzy, we now proceed to show the procedure that could lead to its asymptotic efficiency. Recall that the asymptotic efficiency of the maximum likelihood estimator is the following: For large sample size, n, the maximum likelihood estimate, $\hat{\theta}_n$, approximates closely the unknown true parameter θ_o.

Let α be any real number near 1 but never greater than 1. The asymptotic efficiency of the maximum likelihood estimator set would be equivalent to showing that $\alpha = \text{Max } \lambda_M(\hat{\theta}_{oM})$. More generally, let (C(M) be the subset of all points in M at which α is essentially attained, then this subset is called the core of M.

The idea behind the proposed equivalence above is that the maximum likelihood estimator set is an increasing sequence of points leading to θ_O. Now, since α is very close to 1 (but never greater than 1), and since α is the maximum of all the values of the membership function for M, then the maximum likelihood estimate with the highest sample size must approximate closely the true parameter θ. For finite population we expect α to equal 1 when the sample size equals the population size, for in this case the true parameter θ_O can be computed precisely; but for infinite population we expect it to tend to 1.

In effect, the numerical value of a may be taken as a quantitative measure of the asymptotic efficiency of any estimator. That is, an estimator with an a closer to 1 would be more asymptotically efficient than one further from it. Thus, we can also show that the maximum likelihood estimator set contains the most asymptotically efficient estimates among all other estimators. In this case, we have to estabish that the "parameter space" generated by the different types of estimators are fuzzy. This can be done, logically, in the same way that the maximum likelihood estimator set was established.

The concept of fuzzy sets is still nascent and perhaps will take sometime for it to be well-established. The establishment of the concept on firmer and more rigorous grounds could be one breakthrough into a unification of the proof for the asymptotic efficiency of the maximum likelihood estimator. More rigorous development in the concept of fuzzy sets itself is also awaited. Until such an appropriate time, a detailed and rigorous proof will be in order which will definitely constitute an interesting topic for further research.

REFERENCES

(1) Harold Cramer, "Mathematical Methods of Statistics," Princeton: Princeton University Press, 1946.

(2) Jacob Wolfowitz, "Asymptotic Efficiency of the Maximum Likelihood Estimator," Theory of Probability & Application, Vol. 10, No. 2, 1965.

(3) Sol Kaufman, "Asymptotic Efficiency of the Maximum Estimator," Annals of Institute of Statistical Mathematics, Vol. 18, 1966.

(4) H. Daniels, "The Asymptotic Efficiency of a Maximum Likelihood Estimator," Proceedings Fourth Symposium of Mathematical Statistics & Probability, Vol. 1, Berkeley: University of California Press, 1961.

(5) L. Weiss & J. Wolfowitz, "Maximum Probability Estimators," Annals of Institute of Statistical Mathematics, Vol. 19 No. 2, 1967.

(6) L. A. Zadeh, "Fuzzy Sets," Journal of Information & Control, Vol. 8, 1965.

(7) Ronald A. Fisher, "Contribution to Mathematical Statistics," Index prepared by John Tukey, New York: Wiley, 1950.

(8) Abraham Wald, "Selected Papers in Statistics & Probability," Edited by T.W. Anderson, New York: McGraw-Hill, 1955.

(9) T.W. Anderson, "The Integral of a Symmetric Unimodal Function," Proceedings American Mathematical Society, Vol. 6, 1955.

(10) R.E. Bellman & L.A. Zadeh, "Decision Making in a Fuzzy Environment," Management Science, Vol. 17, No. 4, 1970.