



Taylor & Francis
Taylor & Francis Group



Sequential Trials, Sequential Analysis and the Likelihood Principle

Author(s): Jerome Cornfield

Source: *The American Statistician*, Vol. 20, No. 2 (Apr., 1966), pp. 18-23

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: <http://www.jstor.org/stable/2682711>

Accessed: 18-10-2016 12:19 UTC

REFERENCES

Linked references are available on JSTOR for this article:

http://www.jstor.org/stable/2682711?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://about.jstor.org/terms>



Taylor & Francis, Ltd., American Statistical Association are collaborating with JSTOR to digitize, preserve and extend access to *The American Statistician*

Sequential Trials, Sequential Analysis and the Likelihood Principle*

JEROME CORNFIELD
National Institutes of Health

1. I shall be concerned in this paper with a single question, the answer to which is of great importance to all those engaged in the sequential collection of data. The question is this: Do the conclusions to be drawn from any set of data depend only on the data or do they depend also on the stopping rule which led to the data? In discussing this question I shall draw heavily upon the theoretical work of others, particularly L. J. Savage, but also F. J. Anscombe, G. A. Barnard, A. Birnbaum, and D. V. Lindley. Biostatisticians have tended to regard the theoretical developments suggested by these names as unduly abstract and perhaps of no great relevance to statistical practice. I shall nevertheless review some of them in the hope that few of us would go so far as to say with Falstaff that because "Honour hath no skill in surgery . . . I'll none of it." These newer developments, abstract or not, are, in my opinion of great relevance to biostatistical practice and their absorption into thinking, teaching and consultation is becoming overdue.

At the outset it is useful to distinguish between sequential trials and sequential analysis. By a sequential trial I shall mean any form of data collection in which the decision to continue or discontinue further collection depends in some sense on the information previously obtained. By sequential analysis I mean any form of analysis in which the conclusion depends not only on the data, but also on the stopping rule. Although there are many unsettled questions in the construction of particular stopping rules for sequential trials, it is clear that data dependent stopping rules of some type are often appropriate. I shall however concern myself here only with the question of sequential analysis, i.e., whether the conclusion should depend on the stopping rule and not at all on how to choose among the rules.

2. To indicate that the question to be considered is a far from trivial one I should first like to give a simple example of sequential analysis in which the conclusion is far more dependent on the stopping rule than on the data. Consider the hypothesis that a normal mean has a particular value. This hypothesis might be tested either by observations gathered in a fixed sample size trial or in a Wald three-decision sequential trial [1]. In the former case the most powerful unbiased test would reject the hypothesis at, say the .05 level whenever the sample

mean differed from the hypothesized value by more than 1.96 standard errors, no matter what the value of n or the alternative assumed (for σ known). Proceeding sequentially, however, the number of standard errors by which the sample mean must differ from that hypothesized depends first of all on n , secondly on the value specified as a two-sided alternative, and thirdly, on the power required, given this alternative. Thus let the alternative differ from the hypothesized value by $\pm 0.25\sigma$, with power of 0.95. For $n = 1$ the hypothesis could not be rejected at the .05 level unless the observation departed from it by more than 14 standard errors! Even for $n = 10$ the hypothesis could not be rejected with a deviation of less than 5 standard errors [2]. This result cannot be explained by the assumption that σ is known, since if this assumption is relaxed and the sequential t -test [3] is used instead the situation becomes worse. The hypothesis then cannot be rejected at the .05 level until at least 14 observations have been accumulated, no matter what the values of the first 13, i.e., even if they all differ from the hypothesis value by 10^{10} . The conclusion is thus markedly, if not pathologically, dependent on the stopping rule, and the question with which we are concerned can hardly be considered a trivial one.

3. To most scientists without previous exposure to statistics, as well as to most intelligent laymen, any dependence of conclusions on stopping rules, let alone the extreme dependence we have exhibited, seems like a violation of common sense. Those biostatisticians who defend sequential analysis on the other hand would argue that dependence of conclusions on stopping rules is required to preserve the critical level, i.e., the lowest significance level at which the hypothesis can be rejected for given data. If one accepts the importance of preserving the critical level, then clearly conclusions must depend on the stopping rule. But what is not immediately obvious is that the critical level provides an appropriate measure of the amount of evidence in the data for or against the hypothesis. Biostatisticians who accept sequential analysis would argue, if I understand them correctly, that the significance level is an index to or measure of weight of evidence, i.e., that if one is in possession of observations which (in the usual tail area sense) would occur rarely if the hypothesis is true, then one is also in possession of evidence against the hypothesis. The critical level is thus regarded as a universal yardstick. The fundamental postulate, on which

* A revision of remarks made at the round table on sequential clinical trials at the 1965 meeting of the American Public Health Association.

sequential analysis is based, thus appears to be, "All hypotheses rejected at the same critical level have equal amounts of evidence against them." I have never heard this postulate, which I shall call the α -postulate, explicitly stated, nor can I point to any statistician that I know for a fact to believe it, although many act as if they do.¹ But yet anyone who denies it seems to me to have denied the only reason known for believing in sequential analysis in the first place. Sequential analysis can be defended, in my opinion, if and only if something like the α -postulate is true.

I propose now to develop two lines of argument: 1. The α -postulate is not as reasonable as it might at first sight appear and 2. There is an alternative formulation of the idea of weight of evidence which is reasonable and which does lead to the conclusion that the stopping rule is irrelevant to the conclusion.

4. The first line of argument consists of citing three situations in which I believe everyone would agree that the critical level has no relation to the idea of weight of evidence, no matter how one chooses to define that elusive concept. (a) For most tests in common use a hypothesis which is rejected at a given significance level, say .01, for a given set of data will be rejected at the .02, .05, and all higher levels for that same set of data. Not all tests of hypotheses have this characteristic and then the α -postulate leads to contradictions. To see this, consider a hypothesis H_1 , which is rejected by data S_1 at the .02 and all higher levels but not the .01 level and hypothesis H_2 , which is rejected by the data S_2 at the .01, but not .05 level. If then we consider the .01 level S_2 supplies more evidence against H_2 than S_1 does against H_1 , but if we consider the .05 level the contrary is true. The universal applicability of the α -postulate therefore requires that if a most powerful test of a hypothesis against a simple alternative rejects at some significance level, α , for given data, that at any significance level $\alpha' > \alpha$ the hypothesis must also be rejected for the same data. But this is not always so. Lehman [7] has given a counter-example, and others are easily constructed. Whether or not such situations are often encountered in daily statistical practice, their mere existence

¹ My colleague Dr. Samuel W. Greenhouse points out, however, that numerous first cousins of the postulate have appeared in the literature, as for example,

"We shall describe two different tests T_1 and T_2 associated with critical regions w_1 and w_2 as equivalent when the probabilities $P_1(w_1)$ and $P_1(w_2)$ of making an error of Type I are equal." [4]

The critical level "gives an idea of how strongly the data contradict (or support) the hypothesis." [5]

"When a prediction is made, having a known low degree of probability, such as that a particular throw with four dice shall show four sixes, an event known to have a mathematical probability, in the strict sense, of 1 in 1296, the same reluctance will be felt towards accepting this assertion, and for just the same reason, indeed, that a similar reluctance is shown to accepting a hypothesis rejected at this level of significance . . . In general, tests of significance . . . lead . . . to a rational and well-defined measure of reluctance to the acceptance of the hypotheses they test." [6]

makes it impossible to believe that the critical level supplies a measure of weight of evidence under all circumstances.

(b) The following example will be recognized by statisticians with consulting experience as a simplified version of a very common situation. An experimenter, having made n observations in the expectation that they would permit the rejection of a particular hypothesis, at some predesignated significance level, say .05, finds that he has not quite attained this critical level. He still believes that the hypothesis is false and asks how many more observations would be required to have reasonable certainty of rejecting the hypothesis if the means observed after n observations are taken as the true values. He also makes it clear that had the original n observations permitted rejection he would simply have published his findings. Under these circumstances it is evident that there is no amount of additional observation, no matter how large, which would permit rejection at the .05 level. If the hypothesis being tested is true, there is a .05 chance of its having been rejected after the first round of observations. To this chance must be added the probability of rejecting after the second round, given failure to reject after the first, and this increases the total chance of erroneous rejection to above .05. In fact as the number of observations in the second round is indefinitely increased the significance approaches .0975 ($= .05 + .95 \times .05$) if the .05 criterion is retained. Thus no amount of additional evidence can be collected which would provide evidence against the hypothesis equivalent to rejection at the $P = .05$ level and adherents of the α -postulate would presumably advise him to turn his attention to other scientific fields. The reasonableness of this advice is perhaps questionable (as is the possibility that it would be accepted). In any event it does not seem possible to argue seriously in the face of this example that all hypotheses rejected at the .05 level have equal amounts of evidence against them.

(c) D. R. Cox [8] has constructed an example which suggests that the most powerful test of the hypothesis that a mean is zero against a particular alternative will sometimes reject the hypothesis when the observed mean is zero. Thus, the observation which would ordinarily be regarded as providing the strongest evidence for the null hypothesis is in this example treated as reason for rejecting it.

These examples are, I think, sufficient to suggest that the seemingly plausible assumption that critical level and weight of evidence are identical concepts is far from firmly established. I realize, of course, that practical people tend to become impatient with counter-examples of this type. Quite properly they regard principles as only approximate guides to practice, and not as prescriptions that must be literally followed even when they lead to absurdities. But if one is unwilling to be guided by the α -postulate in the examples given, why should he be any more willing to accept it when analyzing sequential trials? The biostatistician's responsibility for

providing biomedical scientists with a satisfactory explanation of inference cannot, in my opinion, be satisfied by applying certain principles when he agrees with their consequences and by disregarding them when he doesn't.

5. I turn now to my second line of argument—which is that there is a reasonable alternative explication of the idea of inference and one which leads to the rejection of sequential analysis. This explication is provided by the likelihood principle—which states that all observations leading to the same likelihood function should lead to the same conclusion.

I shall start by illustrating the idea of a likelihood function. Consider r successes in n independent trials, each with constant probability, p , of leading to a success. If n is a prespecified constant, r is a random variable whose distribution is given by the binomial distribution, namely

$$(1) \quad \binom{n}{r} p^r (1-p)^{n-r}.$$

If r is a prespecified constant, i.e., if we continue observation until the r^{th} success occurs and then stop, the number of trials, n , is a random variable whose distribution is given by the negative binomial, namely

$$(2) \quad \binom{n-1}{r-1} p^r (1-p)^{n-r}.$$

These two probabilities refer to different random variables, and even for given r and n will have different numerical values. Thus, for $p = 1/2$, $n = 2$, $r = 1$, the first probability is $1/2$, the second $1/4$. It will be observed, however, that each probability can be considered a product of two factors, one of which is a combinatorial term which does not depend on p , and differs for the two distributions, and the other of which depends on p and is the same for both distributions. The factor which depends on p , namely

$$(3) \quad p^r (1-p)^{n-r}$$

is an instance of a likelihood function. In general, given any probability or joint probability density function it is always possible to factor it into two parts, one of which depends on the unknown parameters, and provides the likelihood function, and the other which does not depend on unknown parameters.

As an additional example, the likelihood function for n independent normal observations, $x_1 \dots x_n$, with unknown mean and variance, is

$$(4) \quad \sigma^{-n} \exp - \left[\frac{1}{2\sigma^2} \sum (x_i - \mu)^2 \right],$$

while the second factor, independent of unknown parameters, is

$$(5) \quad (2\pi)^{-n/2}.$$

If σ were known, however, then the likelihood function

is

$$(6) \quad \exp - \left[\frac{n}{2\sigma^2} (\bar{x} - \mu)^2 \right] \text{ and the second factor is}$$

$$(7) \quad \sigma^{-n} (2\pi)^{-n/2} \exp - \left[\frac{\sum (x_i - \bar{x})^2}{2\sigma^2} \right].$$

The motivation for this definition, which goes back to Fisher, is that the likelihood function so defined represents that part of Bayes' formula which is dependent on the data. (See for example p. 326 of [9] and section 6 of Chapter III of [6].) If the entire probability or probability density function, rather than just the likelihood function is inserted in Bayes' formula, the second factor appears as a multiple of both numerator and denominator and cancels out, leaving just the likelihood function. No matter what its motivation, however, the likelihood function is a well-defined mathematical entity, and for the moment this is all we require.

6. Consider now two investigators, one of whom decided to conduct n binomial trials and then stop, no matter what his number of successes, and another, who decided to continue until he had obtained r successes and then stop, no matter how many trials it took. Suppose further that the first investigator obtained exactly r successes, and the second obtained his r^{th} success on exactly the n^{th} trial. Then by (3) they are in possession of the same likelihood function, and if they both accept the likelihood principle they must come to the same conclusion about p , despite the use of quite different stopping rules. If they had adapted some different inferential principle, say that of unbiased estimation, however, the first investigator would have estimated p as r/n and the second as $(r-1)/(n-1)$. The fact that the numerical difference between these two estimates will often be small should not obscure the fact that we are dealing with an important difference in principle and that the conclusions yielded by the likelihood principle differ radically on this point from those yielded by more traditional principles.

For the binomial-negative binomial case the differing stopping rules had no effect upon the likelihood function. Can one be sure that this will always be the case? It is easy to show that this must always be so [10]. The likelihood function does not depend on the stopping rule and if one accepts the likelihood principle one must reject sequential analysis.

7. Why should anyone accept the likelihood principle? I remark first that it is an immediate consequence of Bayes' theorem, since the posterior probability density function of the unknown parameters is proportional to the product of the likelihood function and the prior probability density function. But acceptance of this justification requires acceptance of prior probabilities and hence a radical revision in the traditional objec-

tivistic outlook towards interpretation of data. A number of non-Bayesian arguments therefore also have been given [11,12]. I should like to sketch out one that owes a good deal to Savage and Lindley [13]. It seems to me to be particularly illuminating because it deduces the principle from concepts with which all of us raised in the Neyman-Pearson tradition are familiar.

We consider a null hypothesis, H_0 , a simple alternative, H_1 , and k different experimental designs. For each design we can divide the sample space into two parts, one containing points that lead to the acceptance of H_0 and a part that leads to its rejection. Each such division implies particular values for α_i and β_i ($i = 1, 2, \dots, k$), the errors of the first and second kind. We now ask how to select the "best" critical region for each design. If we accept the α -postulate "best" means that we set each α_i equal to some constant independent of i , and then by application of the Neyman-Pearson lemma (whose truth, of course, is not dependent on the postulate) find the rejection region for each design which minimizes β_i . The sample spaces for the different design need not be the same, but they may have certain points (i.e., possible observations like r successes in n trials) in common. As we have seen, for sample spaces so constructed, it is possible for a common point to be in the rejection region for some designs and outside it for others, i.e., for the inference to depend on the design as well as the data.

As an alternative definition of "best", consider the minimization of a linear function of the two types of errors, $\lambda \alpha_i + \beta_i$, where λ measures the undesirability or cost of an error of the first kind relative to one of the second kind and is the same for all k . We ask for a rejection region for each of k designs which will minimize this function. Consider any sample point, t , and the ratio of the likelihood of the alternative, H_1 , given t , to that of the null hypothesis, H_0 , given t and denote the ratio for the i^{th} design by $R_i(t)$. Then it is easy to show that the rejection region for the design must consist of all points for which $R_i(t) > \lambda$. But any common sample point which has the same likelihood function in each design will also have the same value of $R_i(t)$ in each of them. Since λ is the same for all designs, that sample point will then lead to the rejection of the null hypothesis in each of the designs if $R_i(t)$ exceeds λ and its acceptance if it does not. Thus, if instead of minimizing β for a given α , we minimize $\lambda \alpha + \beta$, we must come to the same conclusion for all sample points which have the same likelihood function, no matter what the design.

To extend this argument to designs which differ because of differences in stopping rules, it is sufficient to divide the sample space for each design into three parts, corresponding to rejection of H_0 , acceptance of H_0 and suspension of judgment. There are probabilities of falling into each of these three regions, under each hypothesis and appropriate costs. One attempts to define the regions for each design by minimizing a linear function

of the error probabilities, where the coefficients in the linear function depend on the costs of the corresponding errors. It is again easy to show that for each design the rejection region for H_0 consists of all points for which $R_i(t) > \lambda_2$, the acceptance region of all points for which $R_i(t) < \lambda_1$ and the region for suspending judgment of all points in which $\lambda_1 \leq R_i(t) \leq \lambda_2$. The constants λ_1 and λ_2 depend on costs and are the same for all k stopping rules.

Thus, the antagonism sometimes pointed to between the likelihood principle and the principle of minimizing errors is seen to depend entirely on a particular formulation of the idea of minimizing errors, namely one dependent on the α -postulate. If a linear function of the errors is minimized instead, one is led directly to the likelihood principle.

One might ask why to minimize a linear rather than some other function of the errors. Savage sketches out a reason for considering the linear function the only appropriate one to minimize. But even without this, it is clear that the entire basis for sequential analysis depends upon nothing more profound than a preference for minimizing β for given α rather than minimizing their linear combination. Rarely has so mighty a structure and one so surprising to scientific common sense, rested on so frail a distinction and so delicate a preference.

8. Earlier I remarked that the critical level is taken in current biostatistical practice as a universal yardstick. It will be observed that the likelihood ratio emerges from this argument as such a yardstick instead, at least for the comparison of given H_0 against a given H_1 . The operational sense in which the likelihood ratio supplies a yardstick is illuminated by considering consistent betting behavior. Suppose a statistician who, having observed the outcome of one of the k designs, were willing to bet that H_1 is true at odds of $f(t)$ to 1 or that H_0 is true at odds of 1 to $f(t)$, where $f(t)$ simply denotes the dependence of the odds on the outcome. How should he determine $f(t)$? If he accepts the likelihood principle, $f(t)$ will be constant for all outcomes for which $R(t)$ is constant, and in particular will not depend on the design. If he accepts the α -postulate $f(t)$ will be constant for all outcomes leading to a given critical level. Now it can be demonstrated that when averaging over all k designs one can realize an average gain whether H_0 or H_1 is true by betting against any statistician whose odds depend on the critical level. The strategy is simple. Bet on H_0 for all sample points for which $f(t)/R(t)$ is greater than a given constant and against it when less. If the statistician sets $f(t)/R(t)$ equal to a constant for all t , however, no system of bets can win both when H_0 is true and when H_1 is true. At best one can win when H_0 is true, but will lose when H_1 is true, or vice versa. The result is a direct consequence of deFinetti's demonstration under conditions more general than a simple choice between H_0 and H_1 that a coherent system of bet-

ting odds implies and is implied by Bayes' theorem [14].

It is not inappropriate to ask anyone who denies the relevance of betting odds to scientific inference but accepts critical levels to explain the operational sense in which it is possible for one set of data more strongly to contradict H_0 than another, at the same time that any system of bets based on this measure would lead to loss whether H_0 or H_1 is true.

9. In section 4 doubt was cast upon the α -postulate by considering special situations in which it leads to absurd results. Is a counter-attack possible in which something like this is done for the likelihood principle as well? In particular can one find so absurd a stopping rule, that no matter what the data, one is certain to come to the wrong conclusion if the inference is made without consideration of the stopping rule—as the likelihood principle says it should be. Such a counter-example has been proposed by Armitage [15,16]. It is worth considering, partly on its merits, and partly because it illuminates further the relation between the likelihood principle and Bayes' theorem.

The stopping rule is this: continue observations until a normal mean differs from the hypothesized value by k standard errors, at which point stop. It is certain, using the rule, that one will eventually differ from the hypothesized value by at least k standard errors even when the hypothesis is true. If one looks only at the data, i.e., the likelihood function, one would quite properly reject the hypothesis for reasonably large values of k , whereas if, in the light of the α -postulate, one looks at the stopping rule as well, with its implied α of unity, one would not. Thus, if one disregards the stopping rule and is guided only by the likelihood function one is certain to reject a true hypothesis. In Armitage's words this is a reason for "resisting immediate conversion" to the likelihood principle. Barnard [17] and Birnbaum [18], who accept the likelihood principle, but reject prior probabilities, have, as I understand them, admitted the anomalous nature of this result.

The Bayesian viewpoint of the example is as follows [2]. If one is seriously concerned about the probability that a stopping rule will certainly result in the rejection of a true hypothesis, it must be because some possibility of the truth of the hypothesis is being entertained. In that case it is appropriate to assign a non-zero prior probability to the hypothesis. If this is done, differing from the hypothesized value by k standard errors will not result in the same posterior probability for the hypothesis for all values of n . In fact for fixed k the posterior probability of the hypothesis monotonically approaches unity as n increases, no matter how small the prior probability assigned, so long as it is non-zero, and how large the k , so long as it is finite. Differing by k standard errors does not therefore necessarily provide any evidence against the hypothesis and disregarding the stopping rule does not lead to an absurd conclusion. The Bayesian viewpoint thus indicates that the hypothesis is

certain to be erroneously rejected—not because the stopping rule was disregarded—but because the hypothesis was assigned zero prior probability and that such assignment is inconsistent with concern over the possibility that the hypothesis will certainly be rejected when true.

10. The previous remarks have been confined to tests of hypotheses, because this is the only form of sequential analysis for which a general mathematical basis now exists. This should not be interpreted to mean that the difficulties with the α -postulate would disappear if ever a firm mathematical basis for sequential confidence limits were found. The confidence set yielded by a given body of data is the set of all hypotheses not rejected by the data, so that the relation between hypothesis tests and confidence limits is close. In fact the confidence limit equivalent of the α -postulate is, "All statements made with the same confidence coefficient have equal amounts of evidence in their favor." That this may be no more reasonable than the α -postulate is suggested by the very common problem of inference about the ratio of two normal means. The most selective unbiased confidence set for the unknown ratio has the following curious characteristic: for every sample point there exists an $\alpha > 0$ such that all confidence limits with coefficients $\geq 1 - \alpha$ are plus to minus infinity [19]. But to assert that the unknown ratio lies between plus and minus infinity with confidence coefficient of only $1 - \alpha$ is surely being over-cautious. Even worse, the postulate asserts that there is less evidence for such an infinite interval than there is for a finite interval about a normal mean, but made with confidence coefficient $1 - \alpha'$, where $\alpha' < \alpha$. The α -postulate cannot therefore be considered any more reasonable for confidence limits than it is for hypothesis testing.

It has been proposed by proponents of confidence limits that this clearly undesirable characteristic of the limits on a ratio be avoided by redefining the sample space so as to exclude all sample points that lead to infinite limits for given α . This is equivalent to saying that if the application of a principle to given evidence leads to an absurdity then the evidence must be discarded. It is reminiscent of the heavy smoker, who, worried by the literature relating smoking to lung cancer, decided to give up reading.

REFERENCES

- [1] Wald, A., *Sequential Analysis*, (1947) New York, John Wiley and Sons.
- [2] Cornfield, J., "A Bayesian test of some classical hypotheses—with applications to sequential clinical trials," Submitted for publication.
- [3] U.S. National Bureau of Standards, *Tables to Facilitate Sequential t -Tests*, (1951), Applied Mathematics Series (NBS-AMS-7) Washington, D. C.: Government Printing Office.
- [4] Neyman, J. and Pearson, E. S., "The testing of statistical hypotheses in relation to probabilities a priori," *Proceedings of the Cambridge Philosophical Society*, 29 (1933), 492-510.
- [5] Lehmann, E. L., *Testing Statistical Hypotheses*, (1959), New York, John Wiley and Sons, p. 62.

- [6] Fisher, R. A., *Statistical Methods and Scientific Inference*, (1956), London, Oliver and Boyd, p. 43.
- [7] *Op. cit.*, Ex. 29, p. 116.
- [8] Cox, D. R., "Some problems connected with statistical inference," *Annals of Mathematical Statistics*, 29 (1958), 357-372.
- [9] Fisher, R. A., *Contributions to Mathematical Statistics* (1950), New York, John Wiley and Sons.
- [10] Anscombe, F. J., Discussion to D. V. Lindley, "Statistical Inference," *Journal of the Royal Statistical Society* (B), 15 (1953), 30-76.
- [11] Barnard, G. A., "Statistical Inference," *Journal of the Royal Statistical Society* (B), 11 (1949), 115-159.
- [12] Birnbaum, A., "On the foundations of statistical inference," *Journal of the American Statistical Association*, 57 (1962), 269-326.
- [13] Savage, L. J., "The foundations of statistics reconsidered," in *Studies in Subjective Probability*, ed. by H. E. Kyburg, Jr. and H. E. Smokler, (1964), New York, John Wiley and Sons.
- [14] deFinetti, B., "Foresight: its logical laws, its subjective sources," in *Studies in Subjective Probability*, *op. cit.*
- [15] Armitage, P., Discussion to Smith, C.A.B., "Consistency in statistical inference and decision," *Journal of the Royal Statistical Society*, (B), 23 (1961), 1-37.
- [16] Armitage, P., "Sequential medical trials: some comments on F. J. Anscombe's paper," *Journal of the American Statistical Association*, 58, (1963), 384-387.
- [17] Barnard, G. A., "Comment on Stein's 'A remark on the likelihood principle'," *Journal of the Royal Statistical Society, Series A* (1962), 569-573.
- [18] Birnbaum, A., *The Anomalous Concept of Statistical Evidence, Axioms, Interpretations in Elementary Exposition* (invited paper presented to the Joint European Conference of Statistical Societies, Berne, Switzerland, September 14, 1964).
- [19] Lehmann, *op. cit.*, Ex. 11, p. 182.

The Teacher's Corner

PROBABILITY AND STATISTICS TEACHING IN WESTERN EUROPEAN SECONDARY SCHOOLS

A. L. O'TOOLE
Drake University

During the summers of 1964 and 1965, the writer visited thirteen countries of Western Europe and talked informally with university professors of mathematics, other teachers, students, and officials in ministries of national education, about the amount of probability and statistics taught now in the secondary schools and about their attitudes toward the teaching of these subjects in the secondary schools of the future. The countries visited were Austria, Belgium, Denmark, England, France, West Germany, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, and Switzerland.

Although considerable experimental work with new ideas in the teaching of mathematics at the secondary-school level is being done in many of these countries, in most of them it seems that no significant amount of probability and statistical inference is taught in the secondary schools at present. The standard curricula in mathematics and the experimental curricula that are being tried in Italy may be cited as illustrative of what is taking place now in the secondary schools of Western Europe [1, 2, 3].

Spain seems to have been the first country in Western Europe to attempt to teach any probability and statistics in the secondary schools. As early as 1953, some work in this area was required for the school-leaving certificate [4, pp. 15, 150]. Pupils 12 to 13 years of age are introduced to some elementary notions of statistics, such as histograms and pictographs, in the third course [5]. In fact, some experimental material dealing with stochastic processes is being tried at present in the third course in a few schools. In the fifth course (14 to 15 year old pupils), the last six of the forty-one lessons comprising

the course are devoted to concepts in probability and statistics [6]. The topics studied are: permutations and combinations, the binomial theorem and binomial probabilities, the concept of frequency and the concept of probability, simple probabilities and compound probabilities, mathematical expectation, graphs of frequency distributions and relative frequency distributions, numerical characterization of distributions (central tendency, dispersion, and properties of the deviations from the arithmetic mean), sigma notation for summation, objectives of the mathematical theory of statistics, the binomial distribution and the normal distribution [7]. The writer was told that all pupils except those going into law and letters are required to take the fifth course in mathematics.

In France, there does not seem to be any probability and statistics taught in the classical and modern lyceums. The programs for classes preparing for the baccalauréat series "technical and economic" in the technical lyceums contain some probability and statistics. The Classe de Première (T') studies the common types of elementary descriptive statistics plus topics such as the method of least squares, linear transformation of variables, analysis of time series, linear correlation, and linear regression [8, pp. 123-124]. The Classe de Technique Economique (T'E) studies simple probabilities, total probabilities, compound probabilities, mathematical expectation, the law of large numbers, Tchebicheff's inequality, Bernoulli's theorem, mean deviation, mean square deviation, equally probable deviation, moments, characteristics of the binomial distribution, the distribution of means of samples, the normal distribution, and the normal curve.