Large Sample Comparison of Tests and Empirical Bayes Procedures

by

J.C. Kiefer and D.S. Moore
Purdue University

Technical Report #367

Department of Statistics
Purdue University

17.

LARGE SAMPLE COMPARISON OF TESTS AND EMPIRICAL BAYES PROCEDURES

Jack C. Kiefer

Mathematics Department
Cornell University
Ithaca, New York

David S. Moore

Department of Statistics
Purdue University
Lafayette, Indiana

The organizers of this conference have made a selection of recent influential statistical ideas, and have asked us to present an exposition of the two topics of the title. The emphasis of the first of these is to be on the use of limit theorems other than the central limit theorem in large sample comparison of tests, in contrast with the now more familiar "local" comparison treated by Pitman, Wilks, Wald, LeCam, Neyman, Weiss, and Wolfowitz, among others. The non-local comparison of tests was developed by Chernoff, Hodges and Lehmann, and Bahadur, producing a striking result in a paper of Hoeffding (1965). Empirical Bayes procedures were introduced by Robbins (1955).

LARGE SAMPLE COMPARISON OF TESTS

Introduction

We begin with a simple testing model: one observes independent and identically distributed random variables $X_1, \ldots, X_n$. The probability density function of X_i is unknown, but belongs to a known class

$\{f_\theta, \theta \in \Theta\}$ labeled in terms of an index set Θ. For example, Θ might be the upper half-plane and $f_{(\theta',\theta'')}$ the normal density with mean θ' and variance θ''. Or the class $\{f_\theta\}$ might consist of every symmetric density and $\theta = (\theta', g)$, where θ' is the median of f_θ and g the density of the "error" $X_1 - \theta'$, symmetric about 0.

It is desired to test the null hypothesis that $\theta \in \Theta_0$ for $\Theta_0 \subset \Theta$ against the alternative that $\theta \in \Theta - \Theta_0$. For simplicity, we shall assume throughout the first two sections of this paper that (1) the parameter space Θ is a subset of Euclidean k-dimensional space; (2) $\Theta_0 = \{\theta_0\}$, so that we are testing the simple null hypothesis $H_0: \theta = \theta_0$; (3) all critical regions considered are defined in terms of sums of iid random variables standardized to approach a normal distribution (under θ_0) by the central limit theorem. (The central limit theorem is not used in approaches that involve a computation like Eq. [2].) Given a sequence of critical regions $\{T_n\}$, we have two probabilities of error: the significance level or probability of erroneous rejection of H_0

$$\alpha_n = P_{\theta_0}[(X_1, \ldots, X_n) \in T_n]$$

and the probability of erroneous acceptance of H_0 when an alternative θ is true

$$\beta_n(\theta) = 1 - P_\theta[(X_1, \ldots, X_n) \in T_n] \qquad \theta \neq \theta_0$$

Suppose we have two competing families of critical regions for the same problem. (We say "family of critical regions" because the region actually used depends on the sample size n and the level α selected. Thus, one might compare--as Hoeffding did--the χ^2 and likelihood ratio families for a multinomial testing problem.) How shall we compare their performance? Given sequences $\{T_n\}$ and $\{T'_n\}$ of critical regions, if T_n has $\alpha_n = \alpha$ and $\beta_n(\theta) = \beta$ for a fixed

alternative $\theta \neq \theta_0$, we may ask how many observations m are required for the critical region T'_m to attain $\alpha'_m = \alpha$ and $\beta'_m(\theta) = \beta$. The ratio n/m is the *efficiency* of $\{T'_n\}$ relative to $\{T_n\}$. Unfortunately, this relative efficiency usually depends on several of α, β, θ_0, θ, and n. It is therefore natural to seek some large sample simplification by investigating the behavior of the error probabilities as the sample size n increases. In addition to identifying good statistical procedures for large samples, such studies may suggest the form of good procedures for small n.

Any reasonable sequence of tests has the property that as n increases and the level α remains fixed, the probability $\beta_n(\theta)$ of erroneous acceptance approaches zero for any alternative θ. Thus, some quantity (α, β, or θ) in addition to n must change as n increases, and various approaches to large sample comparison of tests can be distinguished by the constraints placed on these quantities.

We wish to stress two themes in the development of this area: First, the use of tools other than the central limit theorem to compare tests. (This may be called the "mathematical front.") Second, establishment of the large sample optimality of procedures based on likelihood, in this case the likelihood ratio (LR) family of tests. (This is the "likelihood front," on which there have been significant advances in the theory of estimation as well as testing.) When the alternative as well as the hypothesis is simple, the Neyman-Pearson Fundamental Lemma, of course, states that any LR test is most powerful of its level for any sample size. Large sample optimality of LR tests (as of the analogous maximum likelihood estimators) has since been established with respect to a number of criteria. Hoeffding's contribution was to show that even families of tests that differ by little from LR tests and that are asymptotically equivalent to them in one sense (Pitman efficiency) may be inferior in another sense, in large samples.

Approaches to Large Sample Comparison

We will mention three approaches. The earliest of these was to study the relative efficiency of tests when α is fixed (or $\alpha_n \to \alpha$ and $0 < \alpha < 1$) and the alternative θ_n varies with n in such a way that $\beta(\theta_n) \to \beta$ for a fixed β, $0 < \beta < 1$. In the cases we are discussing, θ_n must approach θ_0 at rate $n^{-1/2}$ to obtain nontrivial α and β. This "local comparison" was systematized by Pitman in the one-dimensional case and bears his name. The central limit theorem is the essential mathematical tool in studying local alternatives. Early work in this setting is attributable to Wilks; Wald's definitive paper (1943) established several optimum properties of the LR and related families. Further work on local properties is contained in the work of LeCam, Neyman, and Weiss and Wolfowitz; this includes the more complex case of composite null hypotheses, various optimality criteria, and families of procedures other than LR tests. We omit details, as local comparisons are not our concern here.

[We remark that some of these last-mentioned developments are counterparts of the asymptotically efficient estimation results for Bayes and maximum likelihood estimators (LeCam, Wolfowitz), which are relevant to the discussion in the section on the standard Bayesian model (vide infra), and for Wolfowitz's maximum probability estimator.]

Other comparisons of tests leave the alternative θ fixed. Calculation of the probabilities of error can then no longer be handled by the central limit theorem, but requires results on probabilities of large deviations. (If t_n is a normalized sample mean, so that t_n converges in law to the standard normal distribution by the central limit theorem, $\{t_n \geq a_n\}$ is a *large deviation* of t_n if $n^{-1/2} a_n \to a$ for $0 < a < \infty$. In this case, the central limit theorem says only that $p_n = P[t_n \geq a_n]$ approaches 0, which is uninformative. We want to know the speed with which this probability approaches 0.) Crámer began the study of this probabilistic problem in 1938, and the literature now contains many sources for both order results (of the form $n^{-1} \log p_n \to c$) and asymptotic results (of the form $p_n/c_n \to 1$) for probabilities of large deviations. Only order results are required for the large sample comparisons of tests done to date.

Chernoff (1952) first used probabilities of large deviations to compare tests. We will mention two contrasting "fixed alternative" approaches. Hodges and Lehmann (1956) fixed θ and $\alpha (0 < \alpha < 1)$ and studied the rate of convergence to 0 of $\beta_n(\theta)$. In the usual cases

$$\beta_n(\theta) = e^{-nc(\theta)[1+o(1)]} \quad \text{for all } \alpha \qquad [1]$$

so that an asymptotic relative efficiency can be defined as the ratio of the indices $c(\theta)$ for competing families of tests.

Beginning in 1960, R. R. Bahadur produced an extensive theory of large sample properties of statistical procedures, which he recently summarized in a monograph, Bahadur (1971). His approach to tests can be stated as follows: fix θ and β and study the rate of convergence of α_n to 0. Again, one usually obtains

$$\alpha_n = e^{-nb(\theta)[1+o(1)]} \quad \text{for all } \beta \qquad [2]$$

so that an asymptotic relative efficiency can again be defined. Bahadur's approach has borne more fruit than has that of Hodges and Lehmann for two reasons. First, it is easier; Eq. [2] requires an order result for probabilities of large deviations under θ_0, whereas Eq. [1] requires a similar result under the alternative θ. The Bahadur index $b(\theta)$ has therefore been computed for many more families of tests than has the Hodges-Lehmann index $c(\theta)$. Second, we have done Bahadur an injustice to have described his work in this framework. His basic idea was to study the behavior of the actually attained level of the test as a random variable. This natural "stochastic comparison" of tests turns out to be equivalent to the nonstochastic comparison based on Eq. [2]. Bahadur has shown in some generality that LR tests have maximum $b(\theta)$ and are therefore asymptotically optimal by his criterion.

Hoeffdings' Contribution

Hoeffding (1965) considered several testing problems involving the multinomial distribution with k cells and unknown vector $\theta = (\theta_1, \ldots, \theta_k)$ of cell probabilities. We will discuss only the problem of

testing the simple null hypothesis $\theta = \theta_0$ for a fixed probability vector θ_0. Hoeffding made an advance on both the mathematical front and the LR front. Mathematically, he built on work of Sanov to give an order result for probabilities of large deviations in this k-dimensional multinomial case. Previous comparisons of tests had used such results only for sums of univariate random variables.

On the LR front, Hoeffding succeeded in distinguishing the large sample performance of the LR family of tests for $\theta = \theta_0$ from that of the familiar Pearson χ^2 tests for this problem. If $\bar{x}_{in}$ for i = 1, ..., k is the proportion of n observations falling in the ith cell, the LR test is based on the information-distance statistic

$$L_n = \Sigma_i \bar{x}_{in} \log \frac{\bar{x}_{in}}{\theta_{0i}}$$

The χ^2 statistic is of course

$$Q_n^2 = \Sigma_i \frac{(\bar{x}_{in} - \theta_{0i})^2}{\theta_{0i}}$$

These tests had long been treated as being asymptotically equivalent because of their equivalence under local comparison. Q_n^2 is the dominant term in the Taylor's series expansion of L_n about θ_0; $2nL_n$ and nQ_n^2 have the same χ^2 limiting distribution under the null hypothesis; the two families of tests have the same large sample performance against local alternatives.

The spirit and nature of Hoeffding's comparison can be demonstrated with minimal mathematics in the two-cell (k = 2) case. This we do in the next section, which may be omitted without loss of continuity. Here, we content ourselves with observing that although Hoeffding's precise comparison was not one of those discussed above, he implicitly showed that the LR and χ^2 families have the same Hodges-Lehmann performance for all alternatives θ, but that *the LR family has strictly better Bahadur performance for "most" alternatives*. As Bahadur's theory has become a standard tool in the decade since Hoeffding's work, the

latter work is now most easily understood in Bahadur's framework. That it could be so understood was shown in detail by J. C. Gupta (1972).

Fixed-alternative comparisons ask more of the χ^2 test than its creators probably intended. The coincidence of power results for local alternatives is closer to the motivation for Q_n^2, which involves the relevance of the expected value of Q_n^2 and hence of normal theory. Nevertheless, this common test has been discredited for large samples and fixed alternatives by Hoeffding's result.

Progress on the LR front has, of course, continued. Brown (1971) has shown in considerable generality that appropriate tests of LR type (actually LR tests of possibly larger hypotheses) are at least as good as any given sequence of tests in both the Hodges-Lehmann and Bahadur senses. The more difficult task of analyzing what makes an apparently equivalent test strictly inferior to a LR test for large samples in the generality of Brown's setting awaits another advance on the mathematical front--more general large deviation results for multivariate problems. Herr (1967) has done this in certain multivariate normal cases, but much work remains. It would also be valuable to investigate the sample size required for LR tests to be close to optimal (or, alternatively, to be superior to χ^2 tests in the multinomial case).

Hoeffding's Result for Two Cells

To illuminate Hoeffding's discovery that Q_n^2 is inferior to L_n, we will consider the special case $k = 2$. This amounts to observing n independent Bernoulli random variables $X_1, \ldots, X_n$ with $\theta = P[X_i = 1]$ unknown. For $0 < p < 1$, the test statistics for the hypothesis $\theta = p$ are $L_n = I(\bar{X},p)$ and $Q_n^2 = Q^2(\bar{X},p)$, where $\bar{X}$ is the sample mean of $X_1, \ldots, X_n$ and

$$I(\theta,p) = \theta \log \frac{\theta}{p} + (1 - \theta)\log \frac{1 - \theta}{1 - p}$$

$$Q^2(\theta,p) = \frac{(\theta - p)^2}{p(1 - p)}$$

The information distance $I(\theta,p)$ between two probability vectors [here between $(\theta,1-\theta)$ and $(p,1-p)$] plays a central role in all large deviation comparisons of tests.

Analysis of this special case requires only one mathematical tool, an order result for probabilities of large deviations of binomial random variables, obtainable from Cramér's inequality. Specifically, if B_n is a binomial random variable with mean np, $q = 1 - p$ and

$$p_n = P\left[\frac{B_n - np}{(npq)^{1/2}} \geq bn^{1/2}\right] \qquad b > 0$$

then, for $p + b(pq)^{1/2} < 1$,

$$\frac{1}{n} \log p_n \to - I[p + b(pq)^{1/2}, p] \qquad [3]$$

From Eq. [3], one may first calculate that the Hodges-Lehmann index $c(\theta)$ defined in Eq. [1] is $I(p,\theta)$ for both L_n and Q_n. Thus, the Hodges-Lehmann approach also fails to distinguish Q_n from L_n.

The Bahadur index $b(\theta)$ defined in Eq. [2] depends on whether $p > 1/2$ or $p < 1/2$. (When $p = 1/2$, the tests based on L_n and Q_n are identical.) For the remainder of this discussion, we assume that the hypothesized p exceeds $1/2$. Another application of Eq. [3] then shows that for Q_n

$$\begin{aligned} b_Q(\theta) &= I(\theta,p) && 0 \leq \theta \leq p \\ &= I[\theta - 2(\theta - p), p] && p \leq \theta \leq 1 \end{aligned} \qquad [4]$$

whereas it is known that for the LR statistic L_n

$$b_L(\theta) = I(\theta,p) \qquad 0 \leq \theta \leq 1 \qquad [5]$$

The situation is illustrated in Fig. 1, where $Q^2(\theta,p)$ and $I(\theta,p)$ are drawn for $p = 3/4$. Note that $I(\theta,p)$ is not symmetric about p, but increases more slowly for $\theta < p$ when $p > 1/2$. Thus, Eqs. [4] and [5] say that $b_Q(\theta) < b_L(\theta)$ for $\theta > p$, so that Q_n is inferior to L_n against alternatives $\theta > p$.

Let us now look at this comparison as Hoeffding did. For sufficiently regular sets A, he showed (for general k) that

$$\frac{1}{n} \log P[\bar{X} \in A|\theta] \to - I(A,\theta) \qquad [6]$$

where

$$I(A,\theta) = \inf_{\omega \in A} I(\omega,\theta)$$

is the information distance of A from θ. Suppose, next, that for $\delta > 0$

$$A(\delta) = \{\theta: Q^2(\theta,p) \geq \delta\}$$

is a χ^2 critical region and

$$B(\delta) = \{\theta: I(\theta,p) \geq I(A(\delta),p)\}$$

is a corresponding LR critical region. These regions are illustrated in Fig. 1.

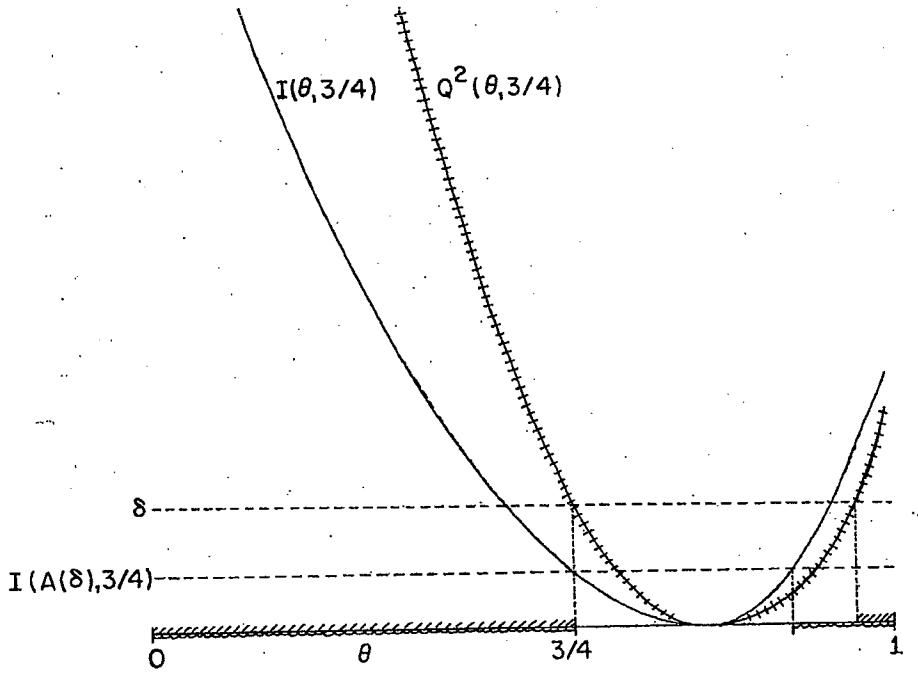

FIG. 1. $I(\theta,3/4)$ and $Q^2(\theta,3/4)$. The hatched region above the axis indicates the set $A(\delta)$ and the marked region below the axis indicates the set $B(\delta)$.

Applying Eq. [6] with $\theta = p$ and $A = A(\delta)$ or $B(\delta)$ shows that both critical regions have asymptotically the same log α_n, as both are the same information distance from p. The LR critical region includes the χ^2 region and is thus at least as powerful. More specifically, alternatives $\theta < p$ are the same information distance from both acceptance regions $A(\delta)^c$ and $B(\delta)^c$, and hence have asymptotically the same log β_n. [In this heuristic sketch, we consider only alternatives in $A(\delta)$; this includes any given $\theta \neq p$ for sufficiently small δ.] But alternatives $\theta > p$ are strictly closer to the χ^2 acceptance region and therefore Q_n has larger log β_n than does L_n for these alternatives.

The geometry of Fig. 1 is indicative of the general case. Hoeffding showed that for any k the analogs of $A(\delta)$ and $B(\delta)$ have only finitely many common boundary points, one on each line segment joining $p = (p_1, \ldots, p_k)$ to the unit vector in the direction of smallest components p_i. L_n is superior to Q_n for all θ not lying on these line segments, by arguments indicated above. When $k > 2$, the exceptional line segments form a small portion of the parameter space, so that the superiority of L_n is more striking in these cases.

THE EMPIRICAL BAYES MODEL

This interesting model was introduced and first studied by Robbins (1955). We shall depart slightly from the usual development of background material by summarizing not only the standard Bayesian model, but also the notions of structural parameter models and adaptive estimators. All of these possess features reflected in some of the empirical Bayes concepts, as well as important differences from the latter.

For simplicity, we shall describe the ideas only for estimation problems in the absolutely continuous case. Regularity conditions that are required will not be listed in detail.

The simplest estimation model is that introduced at the beginning of the first section, except that the object is now to *estimate* some function ϕ of the unknown θ governing the probability law of the X_i. A common example is $\phi(\theta) = \theta'$ in either of the examples of the first

section. An estimator t_n is a rule for guessing $\phi(\theta)$ on the basis of $X_1, \ldots, X_n$. As in the case of testing, it is often difficult to compute an estimator which is "optimal" in some prescribed sense for a given sample size n. It is again natural to study sequences $\{t_n\}$ of estimators as the sample size n increases in the hope of establishing desirable large sample properties.

The Standard Bayesian Model

The Bayesian model adds two assumptions to the estimation problem described above: (1) that the parameter θ can be regarded as a random variable, and (2) that the prior distribution G of this random variable is known. Note that the value of θ, once it is chosen according to G, remains the same in the density f_θ of each X_i.

Bayes' theorem combines G with the observed data to produce the posterior distribution of θ. Comparison of estimators in this model is based on the posterior expected loss $R(t_n,G)$ of an estimator t_n. A *Bayes estimator* $t^*_{G,n}$ of $\phi(\theta)$ is an estimator that minimizes this expected loss. For example, if (as in the examples of the first section) a real parameter $\phi(\theta)$ is to be estimated and loss is measured by squared error, then $t^*_{G,n}$ is the posterior expectation of $\phi(\theta)$.

A feature of interest to us in this Bayesian formulation is that the desired performance of the Bayes procedure is relatively insensitive to slight errors in the specification of G. More precisely

$$\frac{R(t^*_{G',n},G)}{R(t^*_{G,n},G)} \qquad [7]$$

is close to 1 when G is close to G' (under reasonable regularity conditions, as usual). Thus, using $t^*_{G',n}$ when the actual prior law is G (close to G') is almost as good as using the Bayes procedure $t^*_{G,n}$ relative to G. This is just the asymptotic optimality of Bayes estimators referred to in the section on large sample comparison of tests (vide supra).

If the Bayesian model as stated above is correct, there is no disagreement about using $t^*_{G,n}$. One source of controversy arises because doubt may be thrown on the simple-minded form assumed for $\{f_\theta\}$,

or for the assumed loss function, or on the stated aim of the inference [estimation of $\phi(\theta)$]. Another source of controversy lies in the Bayesian assumptions (1) and (2). Bayesian statisticians feel that a description of rational thought legitimizes the use of a subjective guess for G in the absence of knowledge of an actual G; others disagree strongly, but we need not discuss this controversy in detail here.

The Empirical Bayes Model

We now turn to Robbins' model. We are faced with a sequence of independent estimation problems, each of which must be acted on as it arises. These problems are, however, related as follows: the observed X_i in the ith problem has density f_{θ_i} once θ_i is given, but the θ_i are themselves iid with unknown distribution G. So at the nth inference, we have available $X_1, \ldots, X_n$ and we can hope that if n is large some information about the unknown prior law G can be wrung from the past observations $X_1, \ldots, X_{n-1}$. If we knew G exactly, we would estimate θ_n by $t^*_{G,1}(X_n)$, and in the absence of such exact knowledge it seems reasonable (as discussed below in the section on adaptive estimators) to use this estimator with G replaced by an estimator of G; this is Robbins' proposal, which we now describe in further detail.

One can construct an empirical Bayes estimator of θ_n by (1) finding an estimator $\hat{G}_{n-1}(X_1, \ldots, X_{n-1})$ of G; and (2) using the Bayes estimator $t'_n \stackrel{\text{def}}{=} t^*_{\hat{G}_{n-1},1}(X_n)$. One thus acts as if $\hat{G}_{n-1}$ were the known prior law. If the estimator $\hat{G}_{n-1}$ of G is a good one, then on the basis of Eq. [7] we expect that for large n

$$\frac{R(t'_n,G)}{R(t^*_{G,1},G)} \qquad [8]$$

is close to 1. Robbins found appropriate $\hat{G}_{n-1}$ for several problems and established Eq. [8] in these cases. Thus, we do as well asymp-

totically in estimating θ_n when G is unknown as we would if we knew the prior law G.

The proof of Eq. [8] in various settings and the study of the rate of convergence to 1 of the ratio in Eq. [8] has produced a body of literature by Hannan, Johns, van Ryzin, Samuel, Gilliland, and others. One can expect further research to yield reasonably efficient procedures for small n.

Of interest to many observers will be the extent to which Bayesians are able in practice to depart from the standard Bayesian model with a subjective guess of G, and can instead imbed the problem at hand as the nth one in the empirical Bayes model, to yield and use a more formally described guess $\hat{G}_{n-1}$. To non-Bayesians, Robbins' model will seem much more acceptable in many practical settings than the original Bayesian formulation. For example, X_i might be an observation of some biological characteristic of an organism at location i, governed by a parameter θ_i of which distribution G is characteristic of the species but is unknown. Or X_i might be the result of a diagnostic test on individual i made at a preventive medicine clinic run for workers in a large plant, and θ_i an index of the underlying condition having an unknown distribution characteristic of this population of workers.

Structural Parameter Models

Robbins' model may be compared with one which had already been the subject of a large body of literature by 1955, estimation of structural parameters. This is a non-Bayesian framework.

Here again the observations $X_1, \ldots, X_n$ are independent, but the density of X_i is indexed by (α,θ_i). The structural parameter α is to be estimated, while θ_i is an incidental parameter which varies from observation to observation. In the most common example, the X_i are points in the plane derived from an unknown line α by adding independent error vectors with zero means to points on α with abscissas θ_i. This simplest line-fitting situation is illustrated in Fig. 2. One approach to the structural parameter problem is to consider the θ_i to

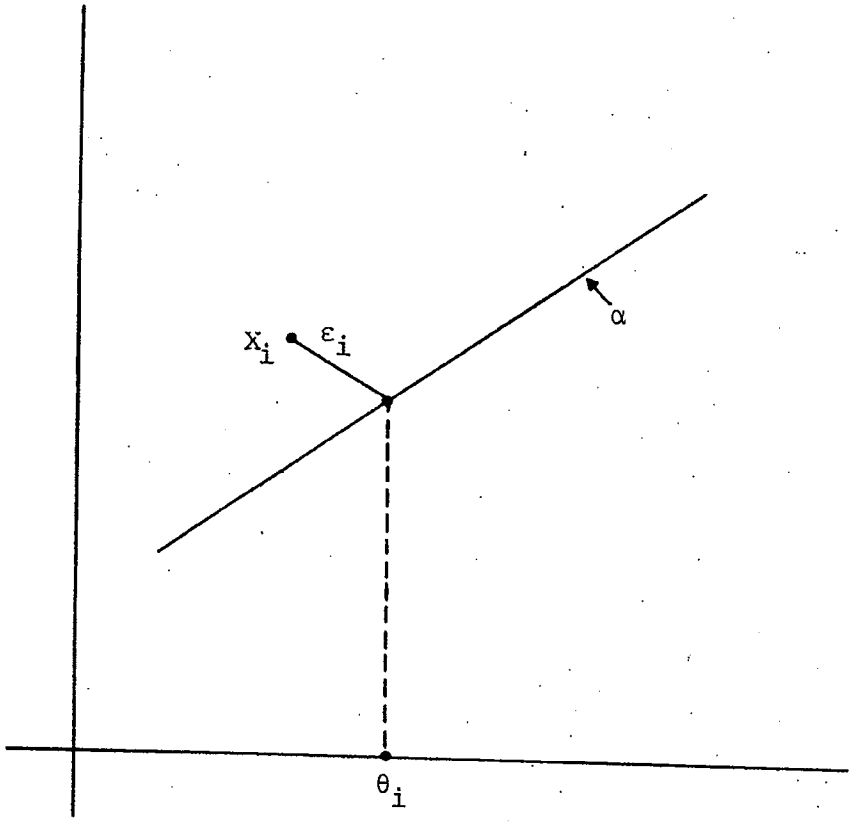

FIG. 2. The structural parameter model with α an unknown line, and observation X_i determined by adding a random error vector ε_i to the point on α with abscissa θ_i.

be independent random variables with the same law G. Indeed, the term "structural model" is sometimes reserved for this case. Such models have been studied by Geary, Reiersol, Wald, Neyman and Scott, Wolfowitz, and others. A survey of their work is given by Moran (1971).

The frameworks of the empirical Bayes model and the structural parameter model have been exhibited in Fig. 3 to point out their considerable similarities. The difference between the models lies primarily in the inference to be made.

EMPIRICAL BAYES MODEL

$X_1, \ldots, X_n$ independent observations
X_i has parameter θ_i
$\theta_1, \ldots, \theta_n$ iid with law G unknown
X_i has marginal density

$$f_G(X) = \int f_\theta(x)\, dG(\theta)$$

Use $X_1, \ldots, X_n$ to estimate θ_n

STRUCTURAL PARAMETER MODEL

$X_1, \ldots, X_n$ independent observations
X_i has parameter (α, θ_i)
$\theta_1, \ldots, \theta_n$ iid with law G unknown
X_i has marginal density

$$f_{\alpha,G}(x) = \int f(x|\alpha,\theta)\, dG(\theta)$$

Use $X_1, \ldots, X_n$ to estimate α

FIG. 3. Comparison of the empirical Bayes and structural parameter models for estimation.

Adaptive Estimators

The methodology of estimation used in the empirical Bayes setting is related to a methodology arising in the example $\theta = (\theta', g)$ of the first section and which can also be employed in the structural parameter model. If we knew g in the nonparametric problem given in the first section, we could use some known good estimator, e.g., Pitman's location parameter estimator $t_{g,n}$ (say), for estimating θ'. The form

of this estimator of course depends on g. Because we do not know g, we find an appropriate estimator $\hat{g}_n$ of it, based on $X_1, X_2, \ldots, X_n$ and then use $t_{\hat{g}_n,n}$ to estimate θ'. This rough recipe requires care in its execution, but such an approach has been carried out in various settings by Weiss and Wolfowitz, LeCam, Hajek, and others. Under suitable assumptions on the unknown g, one can find an estimator of the form $t_{\hat{g}_n,n}$ or something similar, whose accuracy is asymptotically the same as that of the $t_{g,n}$ we would use if we *knew* g.

The spirit of this approach, of constructing procedures by *adapting* their form to what the data seems to say about the error law, is also used in a number of small sample-size studies of "robust" estimators.

Empirical Bayes estimators make use of adaptive estimation in the estimation of G by $\hat{G}_n$. It is also clear that a possible approach to the construction of "good" estimators of α in the structural model is to first estimate G by some $\hat{G}_n(X_1, \ldots, X_n)$, then substitute $\hat{G}_n$ into the estimator $t_{G,n}$ of α that we would use were G known. In some of the work in this setting, G is only estimated implicity; in other work, explicit estimates are given (e.g., Wolfowitz's minimum distance estimator).

Thus, the empirical Bayes model is not only connected with the Bayesian formulation of inference problems but is tied in spirit to structural models and adaptive estimators. One may even ask if there are some practical structural parameter problems in which the successive θ_i are of enough interest to be estimated along with α. Empirical Bayes methods can then be used.

ACKNOWLEDGMENT

The work of the first author was written under National Science Foundation Grant 35816 GPX. The work of the second author was sponsored, in part, by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR-72-2350. The U.S. government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

REFERENCES

Bahadur, R. R. (1971). *Some limit theorems in statistics. Reg. Conf. Ser. Appl. Math. 4,* Soc. Ind. Appl. Math., Philadelphia.

Brown, L. D. (1971). Non-local asymptotic optimality of appropriate likelihood ratio tests. *Ann. Math. Statist. 42,* 1206-1240.

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist. 23,* 493-507.

Gupta, J. C. (1972). Probabilities of medium and large deviations with statistical applications. Ph.D. thesis, University of Chicago.

Herr, D. G. (1967). Asymptotically optimal tests for multivariate normal distributions. *Ann. Math. Statist. 38,* 1829-1844.

Hodges, Jr., J. L. and Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the t-test. *Ann. Math. Statist. 27,* 324-335.

Hoeffding, W. (1965). Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist. 36,* 369-408.

LeCam, L. (1956). On the asymptotic theory of estimation and testing hypotheses. *Proc. 3rd Berkeley Symp. Math. Statist. Prob. 1,* 129-156.

Moran, P. A. P. (1971). Estimating structural and functional relationships. *J. Multivariate Anal. 1,* 232-255.

Neyman, J. (1959). Optimal tests of composite statistical hypotheses. In *The Harold Cramér Volume,* pp. 213-234. Almquist & Wiksell, Stockholm.

Robbins, H (1955). An empirical Bayes approach to statistics. *Proc. 3rd Berkeley Symp. Math. Statist. Prob. 1,* 157-163

Wald, A. (1943). Tests of hypotheses concerning many parameters when the number of observations is large. *Trans. Amer. Math. Soc. 54,* 426-482.

Weiss, L., and Wolfowitz, J. (1969). Asymptotically minimax tests of composite hypotheses. *Z. Wahrschein. Verw. Geb. 14,* 161-168.