The Annals of Statistics
1977, Vol. 5, No. 6, 1183-1194

THE PERFORMANCE OF THE LIKELIHOOD RATIO TEST WHEN THE MODEL IS INCORRECT

BY ROBERT V. FOUTZ AND R. C. SRIVASTAVA

Virginia Polytechnic Institute and State University; Ohio State University

Let the random variable X have a distribution depending on a parameter $\theta \in \Theta$. Consider the problem of testing the hypothesis H: $\Theta_0 \subseteqq \Theta$ based on a sequence of observations on X. The likelihood ratio test for H is constructed by selecting a model for the unknown distribution of X. In this paper the asymptotic performance of the likelihood ratio test is studied when the model is incorrect, that is, when the probability distribution of X is not a member of the model from which the likelihood ratio test is constructed. Exact and approximate measures of the asymptotic efficiency of the likelihood ratio test when the model is incorrect are proposed.

1. Introduction. Let X be a random variable having a distribution depending on a parameter in the open subset Θ of Euclidean k-space, E^k. Let Θ_0 be a subspace of Θ, and consider the problem of testing the hypothesis H: Θ_0 based on a seqence $X_1, X_2, \cdots$ of independent observations of X. The likelihood ratio test for H is constructed by selecting a model for the unknown distribution of X. The model takes the form of a family, $\{P_\theta, \Theta\}$, of probability distributions dominated by a σ-finite measure, ν. Denote $f(x, \theta) = dP_\theta/d\nu$, then the likelihood ratio statistic for H is defined as

$$\lambda_n = \max\{\prod_{i=1}^n f(X_i, \theta); \theta \in \Theta_0\}/\max\{\prod_{i=1}^n f(X_i, \theta); \theta \in \Theta\}. \tag{1.1}$$

Standard results on the performance of λ_n for testing H depend on the assumption that the model is correct; that is, that the probability distribution of X is a member of the family, $\{P_\theta, \Theta\}$, from which λ_n is constructed.

In this paper, the asymptotic performance of the likelihood ratio test is studied when the probability distribution of X is not a member of the class of probability distributions from which λ_n is constructed. The following example illustrates the nature of the problem and motivates the problem's formulation in Section 2.

In this example, consider a random variable X having variance σ^2 and mean μ. For a specified σ_0^2 it is desired to test the composite hypothesis H: $\sigma^2 = \sigma_0^2$. For constructing a likelihood ratio test for H, let $P_{\sigma^2,\mu}$ be the normal distribution with variance σ^2 and mean μ. Let S_n^2 denote the sample variance of n independent observations on X, then the likelihood ratio statistic for H constructed from the

Received September 1975; revised November 1976.
AMS 1970 *subject classifications*. Primary 62A10, 62F20, 62F05.
Key words and phrases. Likelihood ratio, asymptotic distribution, model is incorrect, exact slope, approximate slope, Bahadur efficiency.

normal model is

$$\lambda_n\{S_n^2/\sigma_0^2\}^{n/2} \exp\{n(1 - S_n^2/\sigma_0^2)/2\}\,. \tag{1.2}$$

Now, consider a case where this model is incorrect. Suppose the distribution of X is a contaminated normal distribution of the form

$$Q_{\sigma^2,\mu,\gamma}(B) = .9P_{\sigma_1^2,\mu}(B) + .1P_{\sigma_2^2,\mu}(B)\,,$$

for measurable sets B. The variance and mean of the distribution $Q_{\sigma^2,\mu,\gamma}$ are $\sigma^2 = .9\sigma_1^2 + .1\sigma_2^2$ and μ respectively, and the parameter $\gamma = \sigma_2^2/\sigma_1^2$ is assumed known. Note that γ measures the departure of the above distribution from the normal distribution. For $\gamma = 1$, $Q_{\sigma^2,\mu,\gamma} = P_{\sigma^2,\mu}$. For $\gamma_1 > \gamma_2 \geqq 1$ or $\gamma_1 < \gamma_2 \leqq 1$, $Q_{\sigma^2,\mu,\gamma_1}$ gives heavier tail probabilities than $Q_{\sigma^2,\mu,\gamma_2}$. The "correct" model for the distribution of X, then, is

$$\{Q_{\sigma^2,\mu,\gamma};\, \sigma^2 > 0,\, -\infty < \mu < \infty\}\,. \tag{1.3}$$

The problem is to examine the performance of the test based on (1.2) for testing H. The performance is to be examined, however, when the distribution of X is in the model (1.3). This example is studied further in Section 5.

The main purpose of this paper is to propose exact and approximate measures of the asymptotic efficiency of the likelihood ratio test when the model is incorrect. Following Bahadur (1960) and (1967), we propose measures that are based on the exact slope and on the approximate slope of the test statistic. After some preliminaries in Section 2, the approximate slope of the likelihood ratio when the model is incorrect is obtained in Section 3. Measures of asymptotic efficiency are proposed in Section 4 and are illustrated with an example in Section 5.

2. Preliminaries. The problem of examining the performance of the likelihood ratio test when then model is incorrect may be formulated as in the example of Section 1. Let a random variable X have a distribution depending on an unknown parameter in the open subset Θ of E^k. Let an assumed model for the distribution of X be $\{P_\theta, \Theta\}$, and take $\{f(x, \theta), \Theta\}$ to be a corresponding family of probability density functions. Let $\Theta_0 \subset \Theta$, and let λ_n of (1.1) be the likelihood ratio statistic constructed from this model for testing H: Θ_0.

The problem is to examine the performance of λ_n for testing H: Θ_0 when the distribution of X is a member of another family of distributions, $\{Q_{\theta,\gamma}, \Theta\}$ for some $\gamma \in \Gamma$. As in the example of Section 1, the additional known parameter γ measures "in some sense" the departure of the correct model $\{Q_{\theta,\gamma}, \Theta\}$ from the assumed model $\{P_\theta, \Theta\}$.

The study of the above problem will require the asymptotic distribution of $-2\log\lambda_n$ when X has distribution in $\{Q_{\theta,\gamma}, \Theta\}$. Under the assumption of a correct model Wilks (1938) and Roy (1957) derive the limiting chi-square distribution of $-2\log\lambda_n$ using the consistency and asymptotic normality of maximum likelihood estimates. These arguments may be extended to find the limiting distribution of $-2\log\lambda_n$ when the model is incorrect as well—provided the consistency and limiting distribution of the maximum likelihood estimate have been established.

Maximum likelihood estimates have, in fact, been shown to be consistent and asymptotically normal when the model is incorrect under various assumptions: in particular, Huber (1965) obtains the results without assuming second order derivatives of the log likelihood. Additional related references are Pfanzagl (1969), Perlman (1971), Berk (1972), and Foutz and Srivastava (1974).

In A2 below we assume these previously established properties of maximum likelihood estimates. The remaining A1, A3 and A4 permit the limiting distribution of $-2\log\lambda_n$ to be established.

For fixed $\theta^* \in \Theta$, $\gamma \in \Gamma$, the model $\{P_\theta, \Theta\}$ is defined to be regular with respect to the distribution $Q_{\theta^*,\gamma}$ if the following conditions are satisfied. (All expectations are taken under the distribution $Q_{\theta^*,\gamma}$. The notation $g(x, \theta)$ is used for $\log f(x, \theta)$. Subscript notation is used for partial derivatives: $g_p(x, \theta) = dg(x, \theta)/d\theta_p$, $E_{pq}g(X, \theta) = d^2 \int g(t, \theta)\, dQ_{\theta^*,\gamma}(t)/d\theta_p\, d\theta_q$, etc.)

A1. For $\theta \neq \theta^*$, $Eg(X, \theta) < Eg(X, \theta^*)$.

A2. Let $\hat{\theta}_n$ be the maximum likelihood estimate of θ constructed from the model $\{P_\theta, \Theta\}$; that is, $\hat{\theta}_n$ is defined to maximize $\prod_{i=1}^n f(X_i, \theta)$ over Θ. Assume the $k \times k$ dimensional matrix $\Lambda = \Lambda(\theta^*)$ with pqth element $E_p g_q(X, \theta^*)$ exists and is nonsingular at $\theta = \theta^*$, and let $C = C(\theta^*)$ be the covariance matrix of $(g_1(X, \theta^*), \cdots, g_k(X, \theta^*))'$. Assume $\hat{\theta}_n$ converges almost surely to θ^*, and assume $n^{\frac{1}{2}}(\hat{\theta}_n - \theta^*)$ is asymptotically normal with mean 0 and covariance matrix $\Lambda^{-1}C(\Lambda')^{-1}$ (where Λ' is the transpose of Λ).

A3. Assume $Eg_p(X, \theta^*) = 0, p = 1, 2, \cdots, k$ and $E_p g_q(X, \theta^*) = Eg_{pq}(X, \theta^*)$ $p, q = 1, 2, \cdots, k$.

A4. Assume

$$\sup_{\theta \in \Theta^*}\left|\frac{1}{n}\sum_{i=1}^n g_{pq}(X_i, \theta) - Eg_{pq}(X, \theta)\right| \to 0\,,$$

$$\sup_{\theta \in \Theta^*}\left|\frac{1}{n}\sum_{i=1}^n g(X_i, \theta) - Eg(X, \theta)\right| \to 0$$

almost surely, for some open neighborhood Θ^* about θ^*.

The model $\{P_\theta, \Theta\}$ is now said to be regular with respect to $\{Q_{\theta,\gamma}, \Theta\}$ if the model is regular with respect to each $Q_{\theta^*,\gamma}$ in $\{Q_{\theta,\gamma}, \Theta\}$.

REMARK 2.1. Note that when the model is correct A1 holds if P_{θ^*} is the probability distribution of X and the distributions in the model $\{P_\theta, \Theta\}$ are distinct. (See, for example, Wald 1949, Lemma 1.) When the model is incorrect, A1 imposes the notational convention that each distribution $Q_{(\cdot),\gamma}$ in the family $\{Q_{(\cdot),\gamma}, \cdot\}$ be labeled by the assumed unique θ^* maximizing $Eg(X, \theta)$ over Θ when the expectation is taken under $Q_{(\cdot),\gamma}$. This convention is adopted for convenience.

Let Θ_0 be a subspace of Θ of the form

$$\Theta_0 = \{(\theta_1, \cdots, \theta_r, \theta_{r+1}, \cdots, \theta_k)' : \theta_j = \theta_{0j}, j = 1, \cdots, r\},$$

where θ_{0j} ($j = 1, \cdots, r$) are specified parameter values, and let λ_n be the likelihood ratio statistic for H: Θ_0 constructed from the model $\{P_\theta, \Theta\}$. The asymptotic distribution of λ_n is given in

THEOREM 2.1. *Let $X_1, X_2, \cdots$ be independent with common probability distribution $Q_{\theta^*,\gamma}$, and assume the model $\{P_\theta, \Theta\}$ to be regular with respect to $Q_{\theta^*\gamma}$. Denote by $M = M(\theta^*)$ the upper $r \times r$ dimensional diagonal block of the matrix $\Lambda^{-1}C(\Lambda')^{-1}$ of condition* A2. *Partition Λ in a form having upper $r \times r$ dimensional diagonal block Λ_1:*

$$\Lambda = \begin{pmatrix} \Lambda_1 & \Lambda_2' \\ \Lambda_2 & \Lambda_3 \end{pmatrix},$$

and write $W = W(\theta^*) = -\Lambda_1 + \Lambda_2'\Lambda_3^{-1}\Lambda_2$. Then

(1) *If $\theta^* \in \Theta_0$ then $-2 \log \lambda_n$ is asymptotically distributed as a linear combination of independent chi-square random variables*, i.e.,

$$-2 \log \lambda_n \to_{\mathscr{L}} c_1(\theta^*)\chi_1^2 + c_2(\theta^*)\chi_2^2 + \cdots + c_r(\theta^*)\chi_r^2$$

where $\chi_1^2, \chi_2^2, \cdots, \chi_r^2$ are independent chi-square random variables with one degree of freedom and $c_1(\theta^*) \geqq c_2(\theta^*) \geqq \cdots \geqq c_r(\theta^*) \geqq 0$ are the eigenvalues of the matrix $M(\theta^*)W(\theta^*)$.

(2) *Assume existence of a $\theta_0^* \in \Theta_0$ uniquely maximizing $Eg(X, \theta)$ over Θ_0, and assume $\{P_\theta, \Theta_0\}$ to be regular with respect to $Q_{\theta^*,\gamma}$. (By the convention of Remark 2.1, $Q_{\theta^*,\gamma}$ should be relabeled $Q_{\theta_0^*,\gamma}$ when referring to the regular model $\{P_\theta, \Theta_0\}$.) If $\theta^* \neq \theta_0^*$,* i.e. *if $\theta^* \notin \Theta_0$, then*

$$-\log \lambda_n/n \to Eg(X, \theta^*) - Eg(X, \theta_0^*) \tag{2.1}$$

almost surely as $n \to \infty$.

The proof is given in Foutz and Srivastava (1974). A sketch of the proof follows:

Let $\hat{\theta}_{r,n}$ represent the first r components of $\hat{\theta}_n$, and let θ_r^* be the first r components of θ^*. It may be argued step by step as in Roy (1957) that

$$-2 \log \lambda_n = Z'(W + \mathscr{E})Z, \tag{2.2}$$

where $Z = n^{\frac{1}{2}}(\hat{\theta}_{r,n} - \theta_r^*)$, and the $r \times r$ dimensional matrix $\mathscr{E}$ converges in probability to 0 when $Q_{\theta^*,\gamma}$ is the underlying distribution.

From (2.2) and A2 it follows that $-2 \log \lambda_n$ is asymptotically distributed as $Y'WY$ where Y is an r variate normal random vector with mean 0 and covariance matrix M. Part (1) of the theorem is now a consequence of a result on the distribution of quadratic forms in normal random vectors found in Johnson and Kotz (1970, page 150).

To prove part (2), let $\hat{\theta}_{0,n}$ be the maximum likelihood estimate of θ constructed from the regular model $\{P_\theta, \Theta_0\}$, and write

$$-\log \lambda_n/n = \frac{1}{n} \sum_{i=1}^n g(X_i, \hat{\theta}_n) - \frac{1}{n} \sum_{i=1}^n g(X_i, \hat{\theta}_{0,n}) .$$

Use the almost sure convergence of $\hat{\theta}_n$ to θ^* and of $\hat{\theta}_{0,n}$ to θ_0^* along with property A4 of regular models with respect to $Q_{\theta^*,\gamma}$ to conclude (2.1).

We now introduce the concepts needed in Section 4 to propose measures of the efficiency of the likelihood ratio test when the model is incorrect. The following definitions of a standard sequence, the level attained by a standard sequence, and the approximate slope of a standard sequence are modified versions of the concepts of Bahadur (1960):

DEFINITION 2.1. A sequence $\{T_n\} = \{T_n(x_1, x_2, \cdots, x_n)\}$ of measurable functions is called a standard sequence for testing $H\colon \Theta_0$ in the model $\{Q_{\theta,\gamma}, \Theta\}$, γ fixed, if the following conditions are satisfied:

(i) Let $\{Q^n_{\theta,\gamma}, \Theta\}$ be the family of joint distributions of $X_1, X_2, \cdots, X_n$ corresponding to the family $\{Q_{\theta,\gamma}, \Theta\}$ of distributions for X. For every $\theta \in \Theta_0$, there is a continuous distribution function, $G_{\theta,\gamma}(t)$ such that

$$\lim_{n\to\infty} Q^n_{\theta,\gamma}[(x_1, \cdots, x_n)\colon T_n(x_1, \cdots, x_n) \leqq t] = G_{\theta,\gamma}(t) \tag{2.3}$$

for every t.

(ii) For every $\theta \in \Theta_0$, there is a constant, $a(\theta, \gamma)$, $0 < a < \infty$, such that

$$\log\{1 - G_{\theta,\gamma}(t)\} = -at\{1 + o_{\theta,\gamma}(1)\}/2 \tag{2.4}$$

where, as $t \to \infty$, $o_{\theta,\gamma}(1) \to 0$ uniformly for $\theta \in \Theta_0$.

(iii) There exists a function $b(\theta, \gamma)$ on $\Theta - \Theta_0$ with $0 < b < \infty$ such that for each $\theta \in \Theta - \Theta_0$, $T_n/n \to b(\theta, \gamma)$ almost surely when the common probability distribution of the X_i's is $Q_{\theta,\gamma}$.

DEFINITION 2.2. Let $\{T_n\}$ be a standard sequence for testing $H\colon \Theta_0$ in the model $\{Q_{\theta,\gamma}, \Theta\}$, and let $G_{\theta,\gamma}(t)$ be defined by (2.3). For any given data $x_1, x_2, \cdots, x_n$ the approximate level attained by $\{T_n\}$ for testing H in the model $\{Q_{\theta,\gamma}, \Theta\}$ is defined by

$$L_{\gamma,n}(x_1, x_2, \cdots, x_n) = \sup\{1 - G_{\theta,\gamma}(T_n(x_1, x_2, \cdots, x_n)); \theta \in \Theta_0\} . \tag{2.5}$$

Write $L_{\gamma,n} = L_{\gamma,n}(X_1, X_2, \cdots, X_n)$.

A measure of the rate at which $L_{\gamma,n} \to 0$ when the underlying distribution is $Q^*_{\theta,\gamma}$, $\theta^* \in \Theta - \Theta_0$, is given by

THEOREM 2.2. *Let $L_{\gamma,n}$ be the approximate level attained by the standard sequence $\{T_n\}$ for testing $H\colon \Theta_0$ in the model $\{Q_{\theta,\gamma}, \Theta\}$. Define*

$$S(\theta^*, \gamma) = \inf\{a(\theta, \gamma); \theta \in \Theta_0\} b(\theta^*, \gamma) . \tag{2.6}$$

When the underlying distribution is $Q^*_{\theta,\gamma}$ with $\theta^* \in \Theta - \Theta_0$, $-2\log L_{\gamma,n}/n$ converges almost surely to $S(\theta^*, \gamma)$ as $n \to \infty$.

PROOF. For every t, let $f_{\theta,\gamma}(t)$ be the $o_{\theta,\gamma}(1)$ term on the right of (2.4). By (2.5),

$$(-2\log L_{\gamma,n})/n = [-2\log\sup\{1 - G_{\theta,\gamma}(T_n); \theta\in\Theta_0\}]/n$$
$$= \inf[-2\log\{1 - G_{\theta,\gamma}(T_n)\}; \theta\in\Theta_0]/n .$$

Use (2.4) to write the above as

(2.7) $$(-2\log L_{\gamma,n})/n = \inf[a(\theta,\gamma)T_n\{1 + f_{\theta,\gamma}(T_n)\}; \theta\in\Theta_0]/n .$$

Let θ' be any fixed point in Θ_0, then expression (2.7) implies the inequalities

(2.8) $$(-2\log L_{\gamma,n})/n \leqq a(\theta',\gamma)\{1 + f_{\theta',\gamma}(T_n)\}T_n/n$$

and

(2.9) $$\inf\{a(\theta,\gamma); \theta\in\Theta_0\}T_n/n + \inf\{a(\theta,\gamma); \theta\in\Theta_0\}\inf\{f_{\theta,\gamma}(T_n); \theta\in\Theta_0\}T_n/n \leqq (-2\log L_{\gamma,n})/n .$$

From (iii) of Definition 2.1 we know that $T_n/n \to b(\theta^*,\gamma) > 0$ almost surely as $n\to\infty$. Thus, $T_n\to\infty$ and $f_{\theta'\gamma}(T_n)\to 0$ almost surely as $n\to\infty$. It can now be seen that the right of (2.8) converges almost surely to $a(\theta',\gamma)b(\theta^*,\gamma)$ as $n\to\infty$, and since θ' was chosen arbitrarily in Θ_0, we may conclude

$$\lim_{n\to\infty} -2\log L_{\gamma,n}/n \leqq \inf\{a(\theta,\gamma); \theta\in\Theta_0\}b(\theta^*,\gamma)$$

almost surely. In addition, since $f_{\theta,\gamma}(x)\to 0$ uniformly for $\theta\in\Theta_0$ as $x\to\infty$, it follows that $\inf\{f_{\theta,\gamma}(T_n); \theta\in\Theta_0\}\to 0$ almost surely as T_n converges almost surely to ∞. The left of (2.9) is now seen to converge almost surely to $\inf\{a(\theta,\gamma); \theta\in\Theta_0\}b(\theta^*\gamma)$, giving

$$\lim_{n\to\infty} -2\log L_{\gamma,n}/n \geqq \inf\{a(\theta,\gamma); \theta\in\Theta_0\}b(\theta^*,\gamma)$$

almost surely. This proves Theorem 2.3.

The function $S(\theta,\gamma)$ in (2.6) is defined to be the approximate slope of $\{T_n\}$ in the model $\{Q_{\theta,\gamma}, \Theta\}$.

3. Approximate slope of the likelihood ratio. As in Theorem 2.1, let λ_n be the likelihood ratio statistics constructed from the incorrect model $\{P_\theta, \Theta\}$ for testing $H\colon \Theta_0$. Conditions for the sequence of test statistics $\{-2\log\lambda_n\}$ to be a standard sequence for testing $H\colon\Theta_0$ in the correct model $\{Q_{\theta,\gamma}\Theta\}$ are given in

THEOREM 3.1. *Assume $\{P_\theta, \Theta\}$ and $\{P_\theta', \Theta_0\}$ to be regular with respect to $\{Q_{\theta,\gamma}, \Theta\}$. For each $\theta^*\in\Theta$, let $M(\theta^*)W(\theta)^*$ be the matrix specified by Theorem 2.1, and let $c_1(\theta^*,\gamma)$ be the largest eigenvalue of $M(\theta^*)W(\theta^*)$. If*

(3.1) $$0 < \inf\{c_1(\theta,\gamma); \theta\in\Theta_0\} \leqq \sup\{c_1(\theta,\gamma); \theta\in\Theta_0\} < \infty ,$$

then $\{-2\log\lambda_n\}$ is a standard sequence for testing $H\colon\Theta_0$ in the model $\{Q_{\theta,\gamma}, \Theta\}$.

In order to prove the theorem, the following preliminary results will be needed:

LEMMA 3.1. *Let $\Phi(t)$ be the standard normal distribution. Then*

$$\log\{1 - \Phi(t)\} = t^2\{1 + o(1)\}/2 ,$$

where

$$t^{-2}\{2\log t + \log(2\pi)\} < o(1) < t^{-2}\{\log(2\pi) - 2\log(t^{-1} + t^{-3})\}$$

for every $t > 1$.

The proof is given in Bahadur (1960).

The proof of the following lemma is given incorrectly in Bahadur (1960):

LEMMA 3.2. *Let* $F_k(t)$ *be the chi-square distribution function with* k *degrees of freedom. Then*

$$\log\{1 - F_k(t)\} = -t\{1 + o_k(1)\}/2 ,$$

where

$$-2k\log(t/2)/x \leqq o_k(1) \leqq \{\log(2\pi) - 2\log(t^{-\frac{1}{2}} - t^{-\frac{3}{2}})\}/t .$$

PROOF. Let m be a positive integer such that $k + 1 \geqq 2m \geqq k$. For any fixed $t > 0$, let Z be a Poisson random variable with mean $t/2$. It follows that

$$(3.2)\qquad 2\{1 - \Phi(t^{\frac{1}{2}})\} = 1 - F_1(t) \leqq 1 - F_k(t) \leqq 1 - F_{2m}(t) = P[Z \leqq m - 1] .$$

Also,

$$(3.3)\qquad P[Z \leqq m - 1] = e^{-t/2}\textstyle\sum_{z=0}^{m}(t/2)^z/z! .$$

Expressions (3.2) and (3.3) give

$$1 - \Phi(t^{\frac{1}{2}}) \leqq 1 - F_k(t) \leqq e^{-t/2}(t/2)^k .$$

Take logarithms above and use Lemma 3.1 to obtain

$$-t[1 + 2\{\log(2\pi) - \log(t^{-\frac{1}{2}} - t^{-\frac{3}{2}})\}/t]/2 \leqq \log\{1 - F_k(t)\}$$
$$\leqq -t\{1 - 2k\log(t/2)/t\}/2 .$$

This proves Lemma 3.2.

LEMMA 3.3. *Let* $G(t; c_1, c_2, \cdots, c_r) = P[c_1\chi_1^2 + c_2\chi_2^2 + \cdots + c_r\chi_r^2 \leqq t]$ *where* $c_1 \geqq c_2 \geqq \cdots \geqq c_r \geqq 0$ *and* $\chi_1^2, \chi_2^2, \cdots, \chi_r^2$ *are independent chi-square random variables with* 1 *degree of freedom. Then*

$$\log\{1 - G(t; c_1, c_2, \cdots, c_r)\} = -t\{1 + o(1)\}/(2c_1) ,$$

where

$$(3.4)\qquad -2c_1 r\log(t/2c_1)/t \leqq o(1) \leqq c_1[\log(2\pi) - 2\log\{(c_1/t)^{\frac{1}{2}} - (c_1/t)^{\frac{3}{2}}\}]/t$$

for $t > c_1$.

PROOF. Let $F_k(t)$ be the chi-square distribution function with k degrees of freedom, then

$$1 - F_1(t/c_1) = 1 - P[c_1\chi_1^2 \leqq t] \leqq 1 - G(t; c_1, c_2, \cdots, c_r) ,$$

and

$$1 - G(t; c_1, c_2, \cdots, c_r) \leqq 1 - P[\textstyle\sum_{i=1}^{r} c_1\chi_i^2 \leqq t] = 1 - F_r(t/c_1) .$$

Take logarithms in these inequalities to obtain

$$(3.5)\qquad \log\{1 - F_1(t/c_1)\} \leqq \log\{1 - G(t; c_1, c_2, \cdots, c_r)\} \leqq \log\{1 - F_r(t/c_1)\} .$$

Lemma 3.2 can be applied to $\log\{1 - F_1(t/c_1)\}$ and $\log\{1 - F_r(t/c_1)\}$ in (3.5) to obtain

$$-t\{1 + c_1[\log(2\pi) - \log\{(c_1/t)^{\frac{1}{2}} - (c_1/t)^{\frac{3}{2}}\}]/t\}/(2c_1)$$
$$\leqq \log\{1 - G(t; c_1, c_2, \cdots, c_r)\} \leqq -t\{1 - 2rc_1 \log(t/2c_1)/t\}/(2c_1)\,.$$

This proves Lemma 3.3.

PROOF OF THEOREM 3.1. The proof involves verifying conditions (i), (ii), and (iii) of Definition 2.1.

By Theorem 2.1, (i) is satisfied with

$$G_{\theta,\gamma}(x) = P[\textstyle\sum_{i=1}^{r} c_i(\theta, \gamma)\chi_i^2 \leqq x] \tag{3.6}$$

where $c_1(\theta, \gamma) \geqq c_2(\theta, \gamma) \geqq \cdots \geqq c_r(\theta, \gamma) \geqq 0$ and $\chi_1^2, \chi_2^2, \cdots, \chi_r^2$ are independent chi-square random variables with 1 degree of freedom.

From Lemma 3.3, the distribution function in (3.6) satisfies (ii) with $a(\theta, \gamma) = 1/c_1(\theta, \gamma)$ and $o_{\theta,\gamma}(1)$ satisfying expression (3.4) for $t > c_1(\theta, \gamma)$. It follows from (3.1) and (3.4) that $o_{\theta,\gamma}(1) \to 0$ uniformly for $\theta \in \Theta_0$ as $t \to \infty$.

The validity of (iii) is a direct consequence of part 2 of Theorem 2.1.

Having established that $\{-2 \log \lambda_n\}$ is a standard sequence for testing H: Θ_0 in $\{Q_{\theta,\gamma}, \Theta\}$, we may now apply Theorem 2.2 to obtain an expression for the approximate slope of $\{-2 \log \lambda_n\}$ in

THEOREM 3.2. *Under the conditions of Theorem* 3.1, *the approximate slope of* $\{-2 \log \lambda_n\}$ *for testing* H: Θ_0 *in* $\{Q_{\theta,\gamma}, \Theta\}$ *is given by*

$$S(\theta^*, \gamma) = \inf\{1/c_1(\theta, \gamma);\, \theta \in \Theta_0\} b(\theta^*, \gamma) \tag{3.7}$$

for $\theta^* \in \Theta - \Theta_0$. *The constants* $\{c_1(\theta, \gamma), \theta \in \Theta_0\}$ *are specified in Theorem* 3.1, *and* $b(\theta^*, \gamma)$ *is the almost sure limit of* $-2 \log \lambda_n/n$ *when the distribution of* X *is* $Q_{\theta^*,\gamma}$ *for* $\theta^* \in \Theta - \Theta_0$.

Note that by part (2) of Theorem 2.1, the limit $b(\theta^*, \gamma)$ in (3.7) is $2Eg(X, \theta^*) - 2Eg(X, \theta_0^*)$ where the expectations are taken under $Q_{\theta^*,\gamma}$ and θ_0^* maximizes $Eg(X, \theta)$ over Θ_0.

PROOF. By Theorem 3.1 $\{-2 \log \lambda_n\}$ is a standard sequence for testing H in the model $\{Q_{\theta,\gamma}, \Theta\}$. In the proof of Theorem 3.1 it is shown that for $\{-2 \log \lambda_n\}$, the constants $a(\theta, \gamma)$ and $b(\theta, \gamma)$ of Definition 2.1 are given by $a(\theta, \gamma) = 1/c_1(\theta, \gamma)$, for $\theta \in \Theta_0$, and $b(\theta, \gamma)$ equal to the almost sure limit of $(-2 \log \lambda_n)/n$ under the distribution $Q_{\theta,\gamma}$, $\theta \in \Theta - \Theta_0$. The application of Theorem 2.2 to the standard sequence $\{-2 \log \lambda_n\}$ now shows its approximate slope to be

$$S(\theta, \gamma) = \inf\{a(\theta, \gamma);\, \theta \in \Theta_0\} b(\theta, \gamma) = \inf\{c_1(\theta, \gamma)^{-1};\, \theta \in \Theta_0\} b(\theta, \gamma)$$

for $\theta \in \Theta - \Theta_0$. This proves Theorem 3.2.

4. Measures of efficiency. The concept of Bahadur efficiency is applied in this section to obtain measures of the efficiency of the likelihood ratio test when the model is incorrect. Bahadur's measure of efficiency is based on the following concepts of the exact level and the exact slope of a test:

Let $\{T_n\}$ be a sequence of test statistics for testing $H: \Theta_0$ in the model $\{Q_{\theta,\gamma}, \Theta\}$, γ fixed. Let $\{Q_{\theta,\gamma}^{(n)}, \Theta\}$ be the family of joint distributions of $X_1, X_2, \cdots, X_n$ corresponding to the family $\{Q_{\theta,\gamma}, \Theta\}$ of distributions for X. For any $\theta \in \Theta_0$ and for any t, the distribution function of T_n is

$$G_{\theta,\gamma}^{(n)}(t) = Q_{\theta,\gamma}^{(n)}[(x_1, x_2, \cdots, x_n): T_n(x_1, x_2, \cdots, x_n) \leqq t] .$$

Given the data $x_1, x_2, \cdots, x_n$, the exact level of the test for H based on $\{T_n\}$ is defined to be

$$L_n^*(x_1, x_2, \cdots, x_n) = \sup\{1 - G_{\theta,\gamma}^{(n)}(T_n(x_1, x_2, \cdots, x_n)): \theta \in \Theta_0\} .$$

Write $L_n^* = L_n^*(X_1, X_2, \cdots, X_n)$. Suppose that there is a function $S^*(\theta)$ defined on $\Theta - \Theta_0$ such that $0 < S^*(\theta) < \infty$ and such that $-2 \log L_n^*/n \to S^*(\theta)$ almost surely under the distribution $Q_{\theta,\gamma}$, for $\theta \in \Theta - \Theta_0$. The limit, $S^*(\theta)$, is defined to be the exact slope of the sequence $\{T_n\}$ for testing H.

The ratio of the exact slopes of two sequences of test statistics provides a measure of the relative efficiencies of the corresponding tests for H: Let $\{T_n^1\}$ and $\{T_n^2\}$ be two sequences for testing $H: \Theta_0$ in the single model $\{Q_{\theta,\gamma}, \Theta\}$. Let $S_1^*(\theta)$ and $S_2^*(\theta)$ be their respective exact slopes. Bahadur's efficiency of $\{T_n^1\}$ relative to $\{T_n^2\}$ for testing H in the model is defined to be $A^*(\theta) = S_1^*(\theta)/S_2^*(\theta)$. For interpretations of $A^*(\theta)$ as a measure of relative efficiency, see Bahadur (1960), (1965), and (1967).

In contrast the study of this paper concerns only one sequence of test statistics, $\{-2 \log \lambda_n\}$, where λ_n is the likelihood ratio for testing $H: \Theta_0$ constructed from the assumed model $\{P_\theta, \Theta\}$. The interest is in examining the efficiency of $\{-2 \log \lambda_n\}$ for testing H when the underlying distribution of X is in a family $\{Q_{\theta,\gamma}, \Theta\}$, for some $\gamma \in \Gamma$.

Let $\{Q_{\theta,\gamma'}, \Theta\}$ be a second family of distributions of interest, and consider the problem of measuring the efficiency of $\{-2 \log \lambda_n\}$ for testing H when the underlying distribution of X is in $\{Q_{\theta,\gamma}, \Theta\}$ *relative* to its efficiency when the underlying distribution of X is in $\{Q_{\theta,\gamma'}, \Theta\}$. (The choice $\{Q_{\theta,\gamma'}, \Theta\} = \{P_\theta, \Theta\}$, for example, may provide a meaningful reference for examining the relative efficiency of $\{-2 \log \lambda_n\}$ when the underlying distribution of X is in $\{Q_{\theta,\gamma}, \Theta\}$.) For this problem let $S^*(\theta, \gamma)$ and $S^*(\theta, \gamma')$ be the exact slopes of $\{-2 \log \lambda_n\}$ for testing H in $\{Q_{\theta,\gamma}, \Theta\}$ and $\{Q_{\theta,\gamma'}, \Theta\}$ respectively. The analog of Bahadur's efficiency, $A^*(\theta)$, is now defined to be $B^*(\theta, \gamma, \gamma') = S^*(\theta, \gamma)/S^*(\theta, \gamma')$.

To interpret B^* as a measure of relative efficiency, take $\{a_n\}$ to be a sequence of possible levels attained with $a_n \to 0$ as $n \to \infty$. Then $B^*(\theta, \gamma, \gamma')$ equals the limiting ratio as $a_n \to 0$ of

(i) the sample size necessary to attain the exact level a_n when the underlying distribution of X is $Q_{\theta,\gamma'}$ in the model $\{Q_{\theta,\gamma'}, \Theta\}$
to

(ii) the sample size necessary at attain the exact level a_n when the underlying distribution of X is $Q_{\theta,\gamma}$, in the model $\{Q_{\theta,\gamma}, \Theta\}$.

(See Bahadur (1967), Proposition 8.) Thus, values of $B^*(\theta, \gamma, \gamma') > 1$ indicate that $\{-2 \log \lambda_n\}$ is more efficient for testing H in $\{Q_{\theta,\gamma}, \Theta\}$ than in $\{Q_{\theta,\gamma'}, \Theta\}$.

The general problem of determining whether a sequence of test statistics has an exact slope and of evaluating it is a nontrivial one. (See Bahadur (1967), page 309; (1960), page 282.) The difficulty results because the exact level of a test depends on the fixed sample null distribution of the test statistic.

In light of this difficulty, Bahadur (1960) and (1967) suggests an approximation to A^* based on approximate slopes.

The corresponding approximation to the efficiency, $B^*(\theta, \gamma, \gamma')$, of $\{-2 \log \lambda_n\}$ in the model $\{Q_{\theta,\gamma}, \Theta\}$ relative to the model $\{Q_{\theta,\gamma'}, \Theta\}$ can be formulated in terms of the approximate slopes $S(\theta, \gamma)$ and $S(\theta, \gamma')$ of $\{-2 \log \lambda_n\}$ in the respective models. Denote this approximation by $B(\theta, \gamma, \gamma') = S(\theta, \gamma)/S(\theta, \gamma')$.

Some aspects of the difficulties in approximating S^* by S are discussed in Bahadur (1967). In general, $S(\theta, \gamma)$ need not be close to $S^*(\theta, \gamma)$ for a particular $\theta \in \Theta - \Theta_0$. An important exception occurs for the likelihood ratio statistic when the model is correct. In this case, under general conditions, the exact and approximate slopes of the likelihood ratio statistic have been shown to be the same in Bahadur (1965).

When $S(\theta, \gamma)$ does not equal $S^*(\theta, \gamma)$ for $\theta \in \Theta - \theta_0$, the approximation $S(\theta, \gamma) \approx S^*(\theta, \gamma)$ may be very good for nonnull θ in a neighborhood of $\theta_0 \in \Theta_0$. To be precise, let $\theta_0 \in \Theta_0$, and let $\{\theta_\alpha\}$ be a sequence in $\Theta - \Theta_0$ such that $\theta_\alpha \to \theta_0$. In many cases $S(\theta_\alpha, \gamma)/S^*(\theta_\alpha, \gamma) \to 1$ as $\theta_\alpha \to \theta_0$ in any direction. (See Bahadur (1960), (1967).) This result now leads to a natural local measure of asymptotic relative efficiency: for $\theta_0 \in \Theta_0$, define the local asymptotic efficiency of $\{-2 \log \lambda_n\}$ in the model $\{Q_{\theta,\gamma}, \Theta\}$ relative to the model $\{Q_{\theta,\gamma'}, \Theta\}$ as the limit of $S(\theta, \gamma)/S(\theta, \gamma')$ as $\theta \to \theta_0$ In cases where $S(\theta, \gamma)/S^*(\theta, \gamma) \to 1$ as $\theta \to \theta_0$, this local asymptotic efficiency precisely equals the corresponding local efficiency defined in terms of exact slopes, i.e.,

$$\lim_{\theta \to \theta_0} \frac{S(\theta, \gamma)}{S(\theta, \gamma')} = \lim_{\theta \to \theta_0} \frac{S^*(\theta, \gamma)}{S^*(\theta, \gamma')} ,$$

or

$$\lim_{\theta \to \theta_0} B(\theta, \gamma, \gamma') = \lim_{\theta \to \theta_0} B^*(\theta, \gamma, \gamma') .$$

An application of this measure of local asymptotic relative efficiency is illustrated next.

5. Example. Recall the example of Section 1. The test statistic λ_n of (1.2) is the likelihood ratio constructed from the normal model for testing $H: (\sigma^2, \mu) = (\sigma_0^2, \mu)$, for a fixed σ_0^2 and an unspecified μ. The problem is to examine the performance of λ_n for testing H when the distribution of X is in the family of contaminated normal distributions (1.3).

When the distribution of X is $Q_{\sigma_0^2,\mu,\gamma}$, the asymptotic distribution of $-2 \log \lambda_n$

may be evaluated by Theorem 2.1. The matrix $\Lambda^{-1}C(\Lambda')^{-1}$ of the theorem is calculated to be

$$\Lambda^{-1}C(\Lambda')^{-1} = \begin{pmatrix} -1/2\sigma_0^4 & 0 \\ 0 & -1/\sigma_0^2 \end{pmatrix}^{-1} \begin{pmatrix} 3[-1+(.9+.1\gamma^2)/(.9+.1\gamma)^2]/4\sigma_0^4 & 0 \\ 0 & 1/4\sigma_0^4 \end{pmatrix}$$
$$\times \begin{pmatrix} -1/2\sigma_0^4 & 0 \\ 0 & -1/\sigma_0^4 \end{pmatrix}^{-1}$$
$$= \begin{pmatrix} 3\sigma_0^4[-1 + (.9 + .1\gamma^2)/(.9 + .1\gamma)^2] & 0 \\ 0 & \frac{1}{4} \end{pmatrix}.$$

With $r = 1$, its upper $r \times r$ dimensional diagonal block is

$$M = 3\sigma_0^4[-1 + (.9 + .1\gamma^2)/(.9 + .1\gamma)^2] .$$

The element W of the theorem is $W = 1/2\sigma_0^4$, and the largest eigenvalue of the "matrix" MW *is*

$$c_1(\gamma) = 1.5\{(.9 + .1\gamma^2) \cdot (.9 + .1\gamma)^2\} - .5 .$$

It follows from Theorem 2.1 that $-2 \log \lambda_n$ is asymptotically distributed as $c_1(\gamma)\chi_1^2$, where χ_1^2 is a chi-square random variable with one degree of freedom.

Write $G_\gamma(t) = P[c_1(\gamma)\chi_1^2 \leqq t]$. The approximate level attained by $\{-2 \log \lambda_n\}$ for testing H in the contaminated normal model is $L_{\gamma,n} = 1 - G_\gamma(-2 \log \lambda_n)$.

The approximate slope of $\{-2 \log \lambda_n\}$ for testing H in the contaminated normal model may be evaluated by Theorem 3.2 to be $S(\sigma^2, \mu, \gamma) = S(\sigma^2, \gamma) = \{(\sigma^2/\sigma_0^2) - 1\}/c_1(\gamma)$.

A measure of the efficiency of $\{-2 \log \lambda_n\}$ for testing $H\colon \sigma^2 = \sigma_0^2$ in $\{Q_{\sigma^2,\mu,\gamma}, \sigma^2 > 0, -\infty < \mu < \infty\}$ relative to the second model $\{Q_{\sigma^2,\mu,\gamma'}, \sigma^2 > 0, -\infty < \mu < \infty\}$ is $B(\sigma^2, \gamma, \gamma') = S(\sigma^2, \gamma)/S(\sigma^2, \gamma')$. Note that in this example $B(\sigma^2, \gamma, \gamma')$ is independent of the alternative $\sigma^2 \neq \sigma_0^2$, thus

$$B(\sigma^2, \gamma, \gamma') = \lim_{\sigma^2 \to \sigma_0^2} S(\sigma^2, \gamma)/S(\sigma^2, \gamma') = \lim_{\sigma^2 \to \sigma_0^2} S^*(\sigma^2, \gamma)/S^*(\sigma^2, \gamma') .$$

This limit is the local measure of relative efficiency discussed in Section 4. The table gives values of $B(\sigma^2, \gamma, \gamma') = B(\gamma, \gamma')$ for $\gamma = .5$, 1, 2, 3, 5, and 10; and for $\gamma' = 1$. These values of $B(\gamma, \gamma')$ give the efficiency of $\{-2 \log \lambda_n\}$ for testing H in models of heavy tailed, contaminated normal distributions ($\gamma = .5, 2, 3, 5, 10$) relative to the normal model ($\gamma' = 1$).

TABLE 1
Local asymptotic relative efficiencies

γ	$B(\gamma, \gamma' = 1)$
.5	.96
1	1.00
2	.90
3	.72
5	.48
10	.25

From Table 1, it may be concluded that the likelihood ratio test for $H: \sigma^2 = \sigma_0^2$ constructed from the normal model becomes progressively less efficient as the true family of distributions for X becomes progressively heavier tailed.

REFERENCES

[1] BAHADUR, R. R. (1960). Stochastic comparison of tests. *Ann. Math. Statist.* **31** 276–295.

[2] BAHADUR, R. R. (1965). An optimal property of the likelihood ratio statistic. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 13–26.

[3] BAHADUR, R. R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Statist.* **38** 303–324.

[4] BERK, ROBERT H. (1972). Consistency and asymptotic normality of mle's for exponential models. *Ann. Math. Statist.* **43** 193–204.

[5] FOUTZ, ROBERT V. and SRIVASTAVA, R. C. (1974). The asymptotic distribution of the likelihood ratio when then the model is incorrect. Technical Report No. 110, Dept. of Statistics, Ohio State Univ.

[6] JOHNSON, N. L. and KOTZ, N. (1970). *Continuous Univariate Distributions* **2**. Houghton Mifflin, Boston.

[7] PERLMAN, M. (1970). On the strong consistency of approximate maximum likelihood estimators. *Proc. Sixth Berkeley Symp. Math Statist. Prob.* **1** 263–281.

[8] PFANZAGL, J. (1969). On the measurability and consistency of minimum contrast estimates. *Metrika* **14** 249–272.

[9] ROY, K. P. (1957). A note on the asymptotic distribution of likelihood ratio. *Bull. Calcutta Statist. Assoc.* **7** 73–77.

[10] WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math Statist.* **20** 595–601.

[11] WILKS, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **1** 60–62.

VIRGINIA POLYTECHNIC INSTITUTE
AND STATE UNIVERSITY
BLACKSBURG, VIRGINIA 24061

DEPARTMENT OF STATISTICS
OHIO STATE UNIVERSITY
COLUMBUS, OHIO 43210