The Annals of Mathematical Statistics
1972, Vol. 43, No. 2, 447-454

FIXED ALTERNATIVES AND WALD'S FORMULATION OF THE NONCENTRAL ASYMPTOTIC BEHAVIOR OF THE LIKELIHOOD RATIO STATISTIC[1]

BY T. W. F. STROUD

Queen's University at Kingston

1. Introduction and summary. Let X be a random vector, taking values in p-dimensional Euclidean space $\mathscr{E}^p$ with density $f(x; \theta)$. The parameter θ belongs to a subset Θ of a Euclidean space $\mathscr{E}^q$ and is unkown. Let g be a function over the parameter space having continuous first partial derivatives and taking values in $\mathscr{E}^r$ $(r \leqq q)$. To test the hypothesis $g(\theta) = 0$ against the alternative $g(\theta) \neq 0$ using a sample of n independent observations of X, one frequently uses the Neyman-Pearson generalized likelihood ratio test statistic λ_n. The limiting distribution of $-2 \ln \lambda_n$ under the null hypothesis, as $n \to \infty$, was shown by Wilks (1938) to be chi-square with r degrees of freedom (assuming regularity conditions). If $\{\theta_n\}$ is a *sequence* of alternatives converging to a point of the null hypothesis *at the rate* $n^{-\frac{1}{2}}$, the limiting distribution is noncentral chi-square with noncentrality parameter equal to the limit of $n[g(\theta_n)]' \sum_g^{-1} (\theta_n)[g(\theta_n)]$, where $\sum_g (\theta)$ is the asymptotic covariance matrix of the quantity $n^{\frac{1}{2}}[g(\hat{\theta}) - g(\theta)]$ as $n \to \infty$ with θ fixed ($\hat{\theta}$ denoting the maximum-likelihood estimator of θ based on sample size n).

This noncentral convergence was first proved by Wald (1943), along with a number of other results, on the basis of some rather severe uniformity conditions. Davidson and Lever (1970) have proved the result using more intuitive assumptions. Feder (1968) has obtained asymptotic noncentral chi-square for the case where both the hypothesis and alternative regions are cones in Θ; this is essentially a generalization of $g(\theta) = 0$ versus $g(\theta) \neq 0$, since the hypothesis $g(\theta) = 0$ is locally equivalent to a hyperplane and $g(\theta) \neq 0$ to its complement. Despite the generality, Feder's assumptions are quite mild compared with Wald's.

The result appears in Wald's paper as a special case of a more general statement entitled "Theorem IX." This theorem states that for $\theta \in \Theta$ and $-\infty < t < \infty$ the relationship

$$(1.1) \qquad P_\theta[-2 \ln \lambda_n < t] - P_\theta[K_n < t] \to 0$$

holds *uniformly in t and θ*, where K_n has a noncentral chi-square distribution with r degrees of freedom and noncentrality parameter equal to $n[g(\theta)]' \sum_g^{-1} (\theta)[g(\theta)]$.

Received November 1969.

[1] Partially supported by National Research Council of Canada, Grant A7220. A portion of the theorem was first obtained in the author's doctoral dissertation at Stanford University, which research was supported in part by the National Science Foundation, Grant GP-6681.

This formulation of Wald is too strong. It will be shown by counterexample that, if θ is held *fixed* while $n \to \infty$, relationship (1.1) fails to hold uniformly in t. The counterexample is that of testing the value of the mean of a normal distribution with unknown mean and variance.

Wald's proof of Theorem IX treats two cases separately, case (i) where θ_n approaches the null hypothesis set at the rate $n^{-\frac{1}{2}}$ or faster, and case (ii) where it does not. The proof of (1.1) in case (i) requires convergence of θ_n at the rate $n^{-\frac{1}{2}}$ in order that the Taylor series expansion of the logarithm behave nicely. In case (ii) there is no reason at all to believe the distribution of K_n to be a good approximation to that of $-2\ln\lambda_n$. From Wald's paper (page 480, line following (212)) one gets the impression that Wald felt that the statement of uniform convergence of (1.1) in case (ii) was trivial, since *pointwise* convergence *is* trivial (because both terms tend to zero for fixed t). But, since K_n does not converge in distribution to a random variable in case (ii), there is really no reason why pointwise convergence should imply uniform convergence.

In the same paper, Wald (1943) also described a test procedure based only on the unrestricted maximum-likelihood estimator $\hat{\theta}_n$. This procedure rejects for large values of the statistic

$$Q_n = n[g(\hat{\theta}_n)]' \textstyle\sum_g^{-1}(\hat{\theta}_n)[g(\hat{\theta}_n)] .$$

Wald claimed in his paper that (1.1) again holds uniformly in t and θ if $-2\ln\lambda_n$ is replaced by Q_n. This claim too is false, in the stated generality, as the same counterexample will demonstrate.

Keeping θ as a fixed alternative while $n \to \infty$ has the disadvantage that the limiting behavior of each of the quantities $-2\ln\lambda_n$, Q_n and K_n is degenerate in the sense that the probability mass moves out to infinity with increasing n. However, statement (1.1), uniform in t for fixed θ, has meaning here since both $-2\ln\lambda_n$ (or Q_n) and K_n may be related to quantities with genuine limiting normal distributions which must be identical or at least very similar in order for (1.1) to be uniform in t. The precise result is embodied in a theorem presented in Section 2 of this paper.

In Sections 3 and 4 we consider the case of X normally distributed with mean μ and variance σ^2, where $-\infty < \mu < \infty$, $0 < \sigma_1 < \sigma < \sigma_2$, and the hypothesis to be tested is $\mu = 0$. It is shown in Sections 3 and 4, respectively, that for this problem the relationships

$$P_\theta[Q_n < t] - P_\theta[K_n < t] \to 0$$

and

$$P_\theta[-2\ln\lambda_n < t] - P_\theta[K_n < t] \to 0$$

fail to be uniform in t when $\theta = (\mu, \sigma)$ is fixed and satisfies $\mu \neq 0$, $\sigma_1^2 < \sigma^2 < \sigma_2^2 - \mu^2$. The space of values of σ has been truncated in order to satisfy Wald's regularity conditions.

In the following section boldface letters denote vectors and matrices. The law of the random vector $\mathbf{x}$ is denoted throughout by $\mathscr{L}(\mathbf{x})$. In particular, $\mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ refers to a normal law with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. By $\mathscr{L}(\mathbf{x}_n) \to \mathscr{L}(\mathbf{y})$ or $\mathscr{L}(\mathbf{x}_n) \to \mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is meant, respectively, that the law of $\mathbf{x}_n$ converges to the law of $\mathbf{y}$ or to the stated normal law, as $n \to \infty$. The definitions of the Mann-Wald symbols O_p and o_p may be found in Chernoff ((1956), Section 2), as may the statements of some basic results of large-sample theory which are used freely in the proof of the theorem.

2. Wald's uniform convergence relation (for fixed θ) and the asymptotic normality of related quantities. Consider the statement (1.1), with θ fixed, where $-2\,ln\,\lambda_n$ is replaced by an arbitrary sample statistic (or other sequence of random variables) J_n. The statement now reads

$$(2.1) \qquad P_\theta[J_n < t] - P_\theta[K_n < t] \to 0 \quad \text{uniformly in} \quad t\,.$$

If $g(\theta) \neq 0$, the noncentrality parameter for K_n becomes infinite as $n \to \infty$, so that K_n does not possess a limiting distribution, and neither does J_n if (2.1) holds. However, K_n is distributed as $n\mathbf{y}'\mathbf{y}$, where $n^{\frac{1}{2}}(\mathbf{y} - \boldsymbol{\eta})$ has the r-variate normal distribution $\mathscr{N}(\mathbf{0}, \mathbf{I})$ and $\boldsymbol{\eta} = \sum_g^{-\frac{1}{2}}(\theta)[g(\theta)]$. If J_n has a representation $J_n \sim n\mathbf{h}_n{}'\mathbf{h}_n$, where for some fixed $\boldsymbol{\xi}$ the random vector $n^{\frac{1}{2}}(\mathbf{h}_n - \boldsymbol{\xi})$ has a limiting nonsingular r-variate normal distribution $\mathscr{N}(\mathbf{0}, \boldsymbol{\Gamma})$, one may ask the following questions, where $|\cdot|$ denotes the usual vector norm:

(1) Does (2.1) hold if $\boldsymbol{\Gamma} = \mathbf{I}$ and $|\boldsymbol{\xi}| = |\boldsymbol{\eta}|$?
(2) Can (2.1) hold if $\boldsymbol{\Gamma}$ and $\boldsymbol{\xi}$ are otherwise?

The answer to (1) will be seen to be yes. The answer to (2) is that $|\boldsymbol{\xi}| = |\boldsymbol{\eta}|$ is always necessary, but that $\boldsymbol{\Gamma} = \mathbf{I}$ may be relaxed slightly, assuming we are dealing with the noncentral case ($\boldsymbol{\eta} \neq \mathbf{0}$). (In the central case the answer to (2) is no; this will be treated briefly at the end of the section.) The condition weaker than $\boldsymbol{\Gamma} = \mathbf{I}$ that may cause (2.1) to hold is that the norm of $\boldsymbol{\xi}$ under the inner product given by $\boldsymbol{\Gamma}$ be the same as under the inner product given by $\mathbf{I}$. This can only occur (for $\boldsymbol{\Gamma} \neq \mathbf{I}$) if $r \geqq 2$, and cannot occur if either $\boldsymbol{\Gamma} < \mathbf{I}$ or $\boldsymbol{\Gamma} > \mathbf{I}$. The following theorem treats both (1) and (2) in the noncentral case.

THEOREM. *Let $\{\mathbf{h}_n\}$ be a sequence of random column-vectors in $\mathscr{E}^r$ such that $\mathscr{L}[n^{\frac{1}{2}}(\mathbf{h}_n - \boldsymbol{\xi})] \to \mathscr{N}(\mathbf{0}, \boldsymbol{\Gamma})$, where $\boldsymbol{\xi}$ and $\boldsymbol{\Gamma}$ are fixed, $\boldsymbol{\xi} \in \mathscr{E}^r$ and $\boldsymbol{\Gamma}$ is nonsingular. For each n, let K_n be a real random variable with the noncentral chi-square distribution with r degrees of freedom and noncentrality parameter $n\boldsymbol{\eta}'\boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is a fixed r-dimensional vector, $\boldsymbol{\eta} \neq \mathbf{0}$. Then, as $n \to \infty$,*

$$(2.2) \qquad P[n\mathbf{h}_n{}'\mathbf{h}_n \leqq t] - P[K_n \leqq t] \to 0 \quad \textit{uniformly for all real} \quad t$$

if and only if $\boldsymbol{\eta}'\boldsymbol{\eta} = \boldsymbol{\xi}'\boldsymbol{\xi} = \boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi}$.

PROOF. First some results are developed which concern $n\mathbf{h}_n'\mathbf{h}_n$ and K_n separately. From these results the condition $\boldsymbol{\eta}'\boldsymbol{\eta} = \boldsymbol{\xi}'\boldsymbol{\xi} = \boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi}$ is then proved sufficient and necessary.

Define $J_n = n\mathbf{h}_n'\mathbf{h}_n$ and $\mathbf{w}_n = n^{\frac{1}{2}}(\mathbf{h}_n - \boldsymbol{\xi})$. From the given asymptotic normality of $\mathbf{w}_n$ a statement concerning the cumulative distribution function (cdf) of J_n will be derived. From the definitions,

$$J_n = n(\boldsymbol{\xi} + n^{-\frac{1}{2}}\mathbf{w}_n)'(\boldsymbol{\xi} + n^{-\frac{1}{2}}\mathbf{w}_n)$$

which implies

$$n^{-\frac{1}{2}}(J_n - n\boldsymbol{\xi}'\boldsymbol{\xi}) = 2\boldsymbol{\xi}'\mathbf{w}_n + n^{-\frac{1}{2}}\mathbf{w}_n'\mathbf{w}_n \,. \tag{2.3}$$

Since $\boldsymbol{\eta} \neq \mathbf{0}$ is assumed, $\boldsymbol{\xi} \neq \mathbf{0}$ is clearly necessary for (2.2) to hold; otherwise $J_n = \mathbf{w}_n'\mathbf{w}_n$ has a non-degenerate limiting distribution while K_n does not. Hence we assume $\boldsymbol{\xi} \neq \mathbf{0}$ throughout.

Since by hypothesis $\mathscr{L}(\mathbf{w}_n) \to \mathscr{N}(\mathbf{0}, \boldsymbol{\Gamma})$, it is immediate that $\mathbf{w}_n = O_p(1)$ and hence $n^{-\frac{1}{2}}\mathbf{w}_n'\mathbf{w}_n = o_p(1)$. Therefore, from (2.3),

$$\mathscr{L}[n^{-\frac{1}{2}}(J_n - n\boldsymbol{\xi}'\boldsymbol{\xi})] \to \lim \mathscr{L}(2\boldsymbol{\xi}'\mathbf{w}_n) = \mathscr{N}(0, 4\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi}) \,.$$

Using the theorem of Pólya (1920) that a sequence of cdf's converging pointwise to a continuous cdf must converge uniformly, we may write for any given $\varepsilon > 0$

$$\left|P[n^{-\frac{1}{2}}(J_n - n\boldsymbol{\xi}'\boldsymbol{\xi}) \leqq a] - \Phi\left(\frac{a}{2(\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi})^{\frac{1}{2}}}\right)\right| < \varepsilon/2 \tag{2.4}$$

for all real a, if n is sufficiently large, where the notation $\Phi(\cdot)$ refers to the standard normal cdf. We may now write (2.4), in terms of the cdf of J_n, as

$$\left|P[J_n \leqq n^{\frac{1}{2}}a + n\boldsymbol{\xi}'\boldsymbol{\xi}] - \Phi\left(\frac{a}{2(\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi})^{\frac{1}{2}}}\right)\right| < \varepsilon/2 \tag{2.5}$$

for all real a, if n is sufficiently large.

By definition, K_n is distributed as $(\mathbf{z} + n^{\frac{1}{2}}\boldsymbol{\eta})'(\mathbf{z} + n^{\frac{1}{2}}\boldsymbol{\eta})$ where $\mathbf{z}$ has the r-variate normal distribution $\mathscr{N}(\mathbf{0}, \mathbf{I})$. From this it follows that

$$n^{-\frac{1}{2}}(K_n - n\boldsymbol{\eta}'\boldsymbol{\eta}) = 2\boldsymbol{\eta}'\mathbf{z} + n^{-\frac{1}{2}}\mathbf{z}'\mathbf{z}$$

which, since $n^{-\frac{1}{2}}\mathbf{z}'\mathbf{z} = o_p(1)$, implies

$$\mathscr{L}[n^{-\frac{1}{2}}(K_n - n\boldsymbol{\eta}'\boldsymbol{\eta})] \to \mathscr{L}(2\boldsymbol{\eta}'\mathbf{z}) = \mathscr{N}(0, 4\boldsymbol{\eta}'\boldsymbol{\eta}) \,.$$

Using Pólya's theorem as before, a result resembling (2.4) is obtained which may be rewritten in parallel form to (2.5) as follows:

$$\left|P[K_n \leqq n^{\frac{1}{2}}b + n\boldsymbol{\eta}'\boldsymbol{\eta}] - \Phi\left(\frac{b}{2(\boldsymbol{\eta}'\boldsymbol{\eta})^{\frac{1}{2}}}\right)\right| < \varepsilon/2 \tag{2.6}$$

for all real b, if n is sufficiently large.

Sufficiency of the condition stated in the theorem will now be proved. Suppose $\boldsymbol{\eta}'\boldsymbol{\eta} = \boldsymbol{\xi}'\boldsymbol{\xi} = \boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi}$. For $a = b$ the left sides of (2.5) and (2.6) are identical, except for the quantities J_n and K_n. Combining (2.5) and (2.6) yields

$$|P[J_n \leqq n^{\frac{1}{2}}b + n\boldsymbol{\eta}'\boldsymbol{\eta}] - P[K_n \leqq n^{\frac{1}{2}}b + n\boldsymbol{\eta}'\boldsymbol{\eta}]| < \varepsilon$$

for all real b, if n is sufficiently large. But this is equivalent to (2.2), so that sufficiency is proved.

To prove necessity, we develop some further consequences of (2.5) and (2.6), and then show that the failure of either $\boldsymbol{\eta}'\boldsymbol{\eta} = \boldsymbol{\xi}'\boldsymbol{\xi}$ or $\boldsymbol{\eta}'\boldsymbol{\eta} = \boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi}$ leads to the failure of uniformity or (2.2).

Let $\{a_n\}$ and $\{b_n\}$ be arbitrary sequences of real numbers. Then (2.5) and (2.6) hold for sufficiently large n, by uniformity, when a and b are replaced by a_n and b_n, respectively. Let the particular sequences $\{a_n\}$ and $\{b_n\}$ be defined in terms of a sequence $\{t_n\}$, whose definition is postponed, as follows:

$$a_n = n^{-\frac{1}{2}}t_n - n^{\frac{1}{2}}\boldsymbol{\xi}'\boldsymbol{\xi}$$
$$b_n = n^{-\frac{1}{2}}t_n - n^{\frac{1}{2}}\boldsymbol{\eta}'\boldsymbol{\eta} \,.$$

Inequalities (2.5) and (2.6) now appear as follows:

$$\left|P[J_n \leqq t_n] - \Phi\left(\frac{n^{-\frac{1}{2}}t_n - n^{\frac{1}{2}}\boldsymbol{\xi}'\boldsymbol{\xi}}{2(\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi})^{\frac{1}{2}}}\right)\right| < \varepsilon/2$$
$$\left|P[K_n \leqq t_n] - \Phi\left(\frac{n^{-\frac{1}{2}}t_n - n^{\frac{1}{2}}\boldsymbol{\eta}'\boldsymbol{\eta}}{2(\boldsymbol{\eta}'\boldsymbol{\eta})^{\frac{1}{2}}}\right)\right| < \varepsilon/2$$

for n sufficiently large. According to (2.2),

$$\Phi\left(\frac{n^{-\frac{1}{2}}t_n - n^{\frac{1}{2}}\boldsymbol{\xi}'\boldsymbol{\xi}}{2(\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi})^{\frac{1}{2}}}\right) - \Phi\left(\frac{n^{-\frac{1}{2}}t_n - n^{\frac{1}{2}}\boldsymbol{\eta}'\boldsymbol{\eta}}{2(\boldsymbol{\eta}'\boldsymbol{\eta})^{\frac{1}{2}}}\right) \to 0 \tag{2.7}$$

as $n \to \infty$. Now put

$$t_n = n\boldsymbol{\eta}'\boldsymbol{\eta} + (n\boldsymbol{\eta}'\boldsymbol{\eta})^{\frac{1}{2}} \,.$$

Then the second term of (2.7) becomes $\Phi(\frac{1}{2})$. Let us investigate the first term, which is

$$\Phi\left(\frac{n^{\frac{1}{2}}(\boldsymbol{\eta}'\boldsymbol{\eta} - \boldsymbol{\xi}'\boldsymbol{\xi}) + (\boldsymbol{\eta}'\boldsymbol{\eta})^{\frac{1}{2}}}{2(\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi})^{\frac{1}{2}}}\right) . \tag{2.8}$$

Clearly, if $\boldsymbol{\eta}'\boldsymbol{\eta} \neq \boldsymbol{\xi}'\boldsymbol{\xi}$, then the factor $n^{\frac{1}{2}}$ will dominate the expression and convergence to $\Phi(\frac{1}{2})$ becomes impossible. If $\boldsymbol{\eta}'\boldsymbol{\eta} = \boldsymbol{\xi}'\boldsymbol{\xi}$, (2.8) becomes $\Phi((\boldsymbol{\eta}'\boldsymbol{\eta})^{\frac{1}{2}}/2(\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi})^{\frac{1}{2}})$, which implies that for (2.7) to hold it is necessary that

$$(\boldsymbol{\eta}'\boldsymbol{\eta})^{\frac{1}{2}}/2(\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi})^{\frac{1}{2}} = \tfrac{1}{2} \,,$$

or, equivalently, $\boldsymbol{\eta}'\boldsymbol{\eta} = \boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi}$. The necessity of $\boldsymbol{\eta}'\boldsymbol{\eta} = \boldsymbol{\xi}'\boldsymbol{\xi} = \boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi}$ has been proved. □

Now consider the central case $\boldsymbol{\eta} = \mathbf{0}$. Here K_n has the chi-square distribution χ_r^2 for every n. Clearly (2.2) fails if $\boldsymbol{\xi} \neq \mathbf{0}$, since then $n\mathbf{h}_n'\mathbf{h}_n$ fails to possess a limiting distribution. If $\boldsymbol{\xi} = \boldsymbol{\eta} = \mathbf{0}$, it is clear that $\boldsymbol{\Gamma} = \mathbf{I}$ is a necessary condition for (2.2), since $n\mathbf{h}_n'\mathbf{h}_n$ is asymptotically distributed as a weighted sum of r independent central χ_1^2 random variables, where the weights are the eigenvalues of $\boldsymbol{\Gamma}$. By considering characteristic functions one can easily see that each weight must be unity (and hence $\boldsymbol{\Gamma} = \mathbf{I}$) in order that the limiting distribution of $n\mathbf{h}_n'\mathbf{h}_n$ be χ_r^2. Thus, in the case $\boldsymbol{\eta} = \mathbf{0}$, the conditions $\boldsymbol{\xi} = \mathbf{0}$ and $\boldsymbol{\Gamma} = \mathbf{I}$ are necessary for (2.2). If $\boldsymbol{\xi} = \mathbf{0}$ and $\boldsymbol{\Gamma} = \mathbf{I}$ are *assumed* to hold, the limiting distribution of $n\mathbf{h}_n'\mathbf{h}_n$ is clearly χ_r^2. Uniformity follows from Pólya's theorem, making the conditions $\boldsymbol{\xi} = \mathbf{0}$ and $\boldsymbol{\Gamma} = \mathbf{I}$ sufficient as well as necessary in the central case.

3. Failure of Wald's uniform convergence relation for Q_n. Let $X_1, X_2, \cdots, X_n$ be real-valued random variables, independently normally distributed with unknown mean μ and variance σ^2, where $-\infty < \mu < \infty$ and $0 < \sigma_1 < \sigma < \sigma_2 < \infty$. We wish to test the hypothesis

$$H: \mu = 0$$

against all alternatives. Let $\boldsymbol{\theta}' = (\mu, \sigma)$; the likelihood equations are known to be satisfied at $\hat{\boldsymbol{\theta}} = (\bar{X}, \hat{\sigma})$, where $\bar{X} = \sum X_i/n$ and $\hat{\sigma}^2 = \sum(X_i - \bar{X})^2/n$. Wald's statistic Q_n is given by

$$Q_n = n\bar{X}^2/\hat{\sigma}^2 .$$

At issue is the statement (2.2) which appears in the theorem of the previous section with $\mathbf{h}_n = \bar{X}/\hat{\sigma}$ and the noncentrality parameter of K_n equal to $n\mu^2/\sigma^2$.

Although the assumptions I—VII of Wald's paper are far from intuitive, with a little work one can check that they are satisfied for the present example. Assumption I can be seen to hold if we interpret the region referred to by Wald as D_n to be the set of all sample points for which the likelihood equations have a solution, which appears to be the way this assumption was used by Wald. Assumptions II, IIIc and VII are guaranteed by the assumption made here that $\sigma < \sigma_2 < \infty$, and Assumptions IIIa, IIIb and V by $\sigma > \sigma_1 > 0$. Assumption IV is well known to hold for problems concerning the normal distribution, and Assumption VI is trivial since we are testing a co-ordinate of θ.

It is shown below that for fixed (μ, σ) the limiting distribution of $n^{\frac{1}{2}}(\mathbf{h}_n - \boldsymbol{\xi}) = n^{\frac{1}{2}}(\bar{X}/\hat{\sigma} - \mu/\sigma)$ is $\mathcal{N}(0, 1 + \mu^2/(2\sigma^2))$. When $\mu \neq 0$, the variance differs from unity, and hence, by the theorem, relation (2.2) fails to be uniform in t, contrary to the claim of Wald ((1943), page 480, relation (210)) that this relation *is* uniform in t (and in $\boldsymbol{\theta}$ as well).

To derive the limiting distribution of $n^{\frac{1}{2}}(\bar{X}/\hat{\sigma} - \mu/\sigma)$, consider first the limiting distribution of $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, known to be normal with zero mean and covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2/2 \end{bmatrix}.$$

Let $\psi(\boldsymbol{\theta}) = \mu/\sigma$; then the distribution of $n^{\frac{1}{2}}[\psi(\hat{\boldsymbol{\theta}}_n) - \psi(\boldsymbol{\theta})]$ is asymptotically normal (see Rao (1965), page 322) with mean zero and variance equal to $\boldsymbol{\alpha}'\mathbf{\Sigma}\boldsymbol{\alpha}$ where

$$\boldsymbol{\alpha}' = (\partial\psi/\partial\mu \quad \partial\psi/\partial\sigma) = (1/\sigma \quad -\mu/\sigma^2).$$

Thus the asymptotic variance of $n^{\frac{1}{2}}\bar{X}/\hat{\sigma} = n^{\frac{1}{2}}\psi(\hat{\boldsymbol{\theta}}_n)$ is $\boldsymbol{\alpha}'\mathbf{\Sigma}\boldsymbol{\alpha} = 1 + \mu^2/(2\sigma^2)$.

4. Failure of Wald's uniform convergence relation for $-2\ln\lambda_n$. Let the random variables $X_1, \cdots, X_n$, the parameter $\boldsymbol{\theta}' = (\mu, \sigma)$ and the hypothesis H be as in the previous section. Denote the restricted maximum-likelihood estimator under H by $\hat{\hat{\boldsymbol{\theta}}}{}' = (0, \hat{\hat{\sigma}})$, where $\hat{\hat{\sigma}}^2 = \sum X_i^2/n$. The likelihood ratio statistic is given by $\lambda_n = (\hat{\sigma}/\hat{\hat{\sigma}})^n$, or equivalently

$$-2\ln\lambda_n = n\ln[1 + (\bar{X}/\hat{\sigma})^2].$$

We note that $\sum X_i^2/n$ converges to $\sigma^2 + \mu^2$ as $n \to \infty$, a.s. and in L_p for each $p \geqq 1$. Let us consider a fixed (μ, σ) satisfying $\mu \neq 0$, $\sigma_1^2 < \sigma^2 < \sigma_2^2 - \mu^2$. Then, as $n \to \infty$, $\hat{\sigma}^2$ and $\hat{\hat{\sigma}}^2$ converge to σ^2 and $\sigma^2 + \mu^2$, respectively, both of which lie in the open interval (σ_1^2, σ_2^2). We consider the theorem of Section 2 with $\mathbf{h}_n = \{\ln[1 + (\bar{X}/\hat{\sigma})^2]\}^{\frac{1}{2}}$. The stochastic limit of $\mathbf{h}_n$ as $n \to \infty$ (denoted by $\boldsymbol{\xi}$) is $\{\ln[1 + (\mu/\sigma)^2]\}^{\frac{1}{2}}$, so that

$$n^{\frac{1}{2}}[\{\ln[1 + (\bar{X}/\hat{\sigma})^2]\}^{\frac{1}{2}} - \{\ln[1 + (\mu/\sigma)^2]\}^{\frac{1}{2}}]$$

has a limiting normal distribution (see Rao, (1965), page 319). As in the previous section, the noncentral chi-square random variable referred to by Wald (our K_n) has noncentrality parameter equal to $n\mu^2/\sigma^2$, and the corresponding value of $\boldsymbol{\eta}$ is μ/σ. By the theorem of Section 2, the uniform convergence statement of Wald's Theorem IX for fixed $\boldsymbol{\theta}$ (our (2.2)) cannot hold unless $\boldsymbol{\xi}'\boldsymbol{\xi} = \boldsymbol{\eta}'\boldsymbol{\eta}$, i.e. unless $\ln[1 + (\mu/\sigma)^2] = \mu^2/\sigma^2$, which is clearly false when $\mu \neq 0$.

5. Acknowledgments. The author acknowledges the benefit of illuminating discussions with Professors Lucien LeCam, Paul I. Feder and Charles H. Kraft. The author is grateful for the assistance of Professor Herman Chernoff with part of the material, and for the encouragement and advice of Professor Ingram Olkin throughout this research. Special thanks are due to the associate editor whose criticism of an earlier version was most helpful, and to a referee for his comments.

REFERENCES

CHERNOFF, H. (1956). Large sample theory: parametric case. *Ann. Math. Statist.*, **27** 1-22.

DAVIDSON, R. R. and LEVER, W. E. (1970). The limiting distribution of the likelihood ratio statistic under a class of local alternatives. *Sankhyā Ser. A* **32** 209-224.

FEDER, P. I. (1968). On the distribution of the log likelihood ratio test statistic when the true parameter is "near" the boundaries of the hypothesis regions. *Ann. Math. Statist.* **39** 2044-2055.

PÓLYA, G. (1920). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Math. Z.* **8** 171-181.

RAO, C. R. (1965). *Linear Statistical Inference and Its Applications.* Wiley, New York.

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

WILKS, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **9** 60-62.