ASYMPTOTIC STUDENTIZATION IN TESTING OF HYPOTHESES

By Herman Chernoff[1]

Cowles Commission for Research in Economics

1. Summary. A method suggested by Wald for finding critical regions of almost constant size and various modifications are considered. Under reasonable conditions the sth step of this method gives a critical region of size $\alpha + R_s(\theta)$ where θ is the unknown value of the nuisance parameter, $R_s(\theta) = O(N^{-s/2})$ and N is the sample size. The first step of this method gives the region which is obtained by assuming that an estimate $\hat{\theta}$ of the nuisance parameter is actually equal to θ.

2. Introduction. The problem of nuisance parameters often arises in the testing of hypotheses in the following form: It is desired to construct a test of a hypothesis H so that the probability of rejecting H if it is true is equal to α. However the probability distribution of the data is not uniquely determined by H. Indeed, if the hypothesis is true then the observations have a distribution depending on a nuisance parameter θ whose value is not known. Generally a critical region will have a size which depends on the value of θ. Neyman has done considerable work on the problem of finding similar regions, i.e., regions whose size is independent of θ.

Wald has suggested the following method of finding critical regions whose size is almost independent of θ. Suppose that t is a statistic such that if θ were known then the critical region $t \leq c_1(\theta)$ would be a good critical region for testing the hypothesis H. Suppose also that $\hat{\theta}$ is an estimate of θ and that $g(t, \hat{\theta} \mid \theta)$ represents the joint distribution of t, $\hat{\theta}$ under H when θ is the value of the nuisance parameter. Then consider the regions

$$
\begin{aligned}
t &\leq c_1(\hat{\theta}) && \text{where} \quad \Pr\{t \leq c_1(\theta)\} = \alpha \quad \text{independent of } \theta; \\
t &\leq c_1(\hat{\theta}) + c_2(\hat{\theta}) && \text{“} \quad \Pr\{t - c_1(\hat{\theta}) \leq c_2(\theta)\} = \alpha \quad \text{independent of } \theta; \\
t &\leq c_1(\hat{\theta}) + \cdots + c_s(\hat{\theta}) && \text{“} \quad \Pr\{t - c_1(\hat{\theta}) \cdots - c_{s-1}(\hat{\theta}) \leq c_s(\theta)\} = \alpha \\
&&& \qquad \text{independent of } \theta.
\end{aligned}
$$

Under the assumption that $\hat{\theta}$ is close to θ it is reasonable to expect that $\Pr\{t \leq c_1(\hat{\theta})\}$ would be close to α. It might also be expected that $\Pr\{t \leq c_1(\hat{\theta}) + c_2(\hat{\theta})\}$ would be even closer to α.

This method has been shown to have good properties when considered from the asymptotic point of view. Suppose that t, $\hat{\theta}$ are two sequences of statistics

[1] This paper is based on a dissertation written under the supervision of Professor Abraham Wald and submitted as partial fulfilment of the requirements for Ph.D. in the Graduate Division of Applied Mathematics of Brown University.

(depending on N, the size of the sample or an analogous variable) with distribution represented by $g(t, \hat{\theta} \mid \theta)$ where N is understood to be present. Then it has been shown that under reasonable conditions, with modifications for the sake of calculation,

$$| \Pr\{t \leq c_1(\hat{\theta}) + \cdots + c_s(\hat{\theta})\} - \alpha | = O(N^{-s/2}).$$

The statement of the theorem presenting this result will be given in section 4. It has also been shown that if roughly speaking $\hat{\theta}$ is distributed almost symmetrically about θ, the above result may be obtained in half the steps, i.e.,

$$| Pr\{t \leq c_1(\hat{\theta}) + \cdots + c_s(\hat{\theta})\} - \alpha | = O(N^{-s}).$$

It is true that under relatively weak conditions and for fixed N it is possible for any $\epsilon > 0$ to obtain a function $h(\hat{\theta})$ such that $| \Pr\{t \leq h(\hat{\theta})\} - \alpha | < \epsilon$. However such a critical region can have very poor properties from the point of view of the alternative hypotheses especially if $h(\hat{\theta})$ is a very wildly oscillating function. On the other hand this objection does not apply to Wald's method for large N because

$$\begin{aligned} &| c_1^{(r)}(\theta) | \leq M && r = 0, 1, \cdots, s; \\ &| c_2^{(r)}(\theta) | \leq MN^{-1/2} && r = 0, 1, \cdots, s - 1; \\ &| c_s^{(r)}(\theta) | \leq MN^{-(s-1)/2} && r = 0, 1, \end{aligned}$$

and hence $c_1(\hat{\theta}) + \cdots + c_2(\hat{\theta})$ is almost constant over "that small range in which $\hat{\theta}$ will probably fall."

In the above it has been implied that θ is a one dimensional variable. However the results are easily extended to the case where θ is a k-dimensional variable.

The direct application of the method is often quite difficult because of the calculations involved. Modifications can be applied which simplify the calculations. Such modification usually consist of changing the $c_r(\theta)$ by a small amount provided the remainder is simple and "well behaved." A case where considerable simplifications can be made is that where $g_1(t \mid \hat{\theta}, \theta)$, the conditional distribution of t, can be expanded in a Taylor Expansion,

$$\begin{aligned} g_1(t \mid \hat{\theta},\theta) = g_1(c_1(\theta) \mid \theta, \theta) + (t - c_1(\theta)) \frac{\partial g_1}{\partial t} + (\hat{\theta} - \theta) \frac{\partial g_1}{\partial \theta} \\ + \cdots + \frac{1}{s!} \sum_{j=0}^{s} (t - c_1(\theta))^j (\hat{\theta} - \theta)^{s-j} \frac{\partial^s}{\partial t^j \partial \hat{\theta}^{s-j}} g_1(t' \mid \hat{\theta}', \theta), \end{aligned}$$

where the partial derivatives "behave." This case will be described in detail in section 3, and an example previously treated by Welch (see [1]) will be discussed in section 4.

Another case where simplifications often arise is the asymptotic case, that is the case where $g(t, \hat{\theta} \mid \theta)$ has an asymptotic expansion. The asymptotic case may also be regarded as an extension of the following partition principle which is very useful. If $g(t, \hat{\theta} \mid \theta) = g_0(t, \hat{\theta} \mid \theta) + h(t, \hat{\theta} \mid \theta)$ and $\iint |h| \, dt \, d\hat{\theta} \leq MN^{-s/2}$ and if $\varphi(\hat{\theta})$ is such that

$$\left| \int_{-\infty}^{\infty} d\hat{\theta} \int_{-\infty}^{\varphi(\theta)} dt \, g_0(t, \hat{\theta} \mid \theta) - \alpha \right| \leq MN^{-s/2},$$

then $|\Pr\{t \leq \varphi(\hat{\theta})\} - \alpha| \leq MN^{-s/2}$. Thus our theorems apply to $g(t, \hat{\theta} \mid \theta)$ if $g = g_0 + h$ where g_0 has sufficient differentiability properties.

3. The Taylor expansion treatment. Let $g(t, \hat{\theta} \mid \theta) = g_1(t \mid \hat{\theta}, \theta) g_2(\hat{\theta} \mid \theta)$ where g_1 is the conditional density of t given $\hat{\theta}$ and $g_2(\hat{\theta} \mid \theta)$ is the marginal density of $\hat{\theta}$. $g_3(t \mid \theta) = \int_{-\infty}^{\infty} d\hat{\theta} g(t, \hat{\theta} \mid \theta)$ is the marginal density of t. In what follows we shall use M as a generic bound. Thus the statement $f(t, \theta) < M(\theta_1, \theta_2)$, $\theta_1 \leq \theta \leq \theta_2$, means that there is a constant M depending on (θ_1, θ_2) and independent of t, θ, N so that $f(t, \theta) < M(\theta_1, \theta_2)$ $\theta_1 \leq \theta \leq \theta_2$.

First we obtain $c_1(\theta)$ so that $\Pr\{t \leq c_1(\theta)\} = \alpha$.
Then we have

THEOREM 1. *If for every finite interval* (θ_1, θ_2),

(i) $$\left| \frac{\partial^p g_3}{\partial \theta^p} (t \mid \theta + \Delta) \right| < G_1(t, \theta) < G_2(t), \quad |\Delta| \leq \Delta'(\theta_1, \theta_2, N), \quad p = 0, 1, \cdots s,$$
$$\theta_1 \leq \theta, \theta + \Delta \leq \theta_2,$$

where $\int_{-\infty}^{\infty} G_2(t) \, dt < M(\theta_1, \theta_2)$, G_1 *and* G_2 *may depend on* N, θ_1, *and* θ_2

(ii) $$\frac{\partial^{p+q} g_3(t \mid \theta)}{\partial \theta^p \partial t^q} \text{ is continuous in } t, \theta \text{ and}$$

bounded in absolute value by $M(C_1, C_2, \theta_1, \theta_2)$ *for* $p + q < s$, $\theta_1 \leq \theta \leq \theta_2$, $C_1 \leq t \leq C_2$;

(iii) $0 < \dfrac{1}{M(C_1, C_2, \theta_1, \theta_2)} < g_3(t \mid \theta)$ *for* $\theta_1 \leq \theta \leq \theta_2$, $C_1 \leq t \leq C_2$;

(iv) $0 < \alpha < 1$,

then $\Pr\{t \leq c_1(\theta)\} = \alpha$ *defines* $c_1(\theta)$ *uniquely and so that* $|c_1^{(p)}(\theta)| \leq M(\theta_1, \theta_2)$ *for* $p = 0, 1, \cdots, s$ $\theta_1 \leq \theta \leq \theta_2$.

PROOF. Since $g_3(t \mid \theta)$ is positive, $c_1(\theta)$ is uniquely defined by condition (i). From this and conditions (i) and (ii) it follows that $c_1'(\theta)$ exists and is given by

$$\int_{-\infty}^{c_1(\theta)} dt \frac{\partial g_3}{\partial \theta} (t \mid \theta) + c_1'(\theta) g_3(c_1(\theta) \mid \theta). \tag{1}$$

We may continue in this fashion differentiating formally $p \leq s$ times to get

$$(2) \quad \int_{-\infty}^{c_1(\theta)} dt \frac{\partial^p g_3(t \mid \theta)}{\partial \theta^p} + \sum [c_1^{(j_1)}(\theta)]^{i_1} [c_1^{(j_2)}(\theta)]^{i_2} \cdots [c_1^{(j_k)}(\theta)]^{i_k} \frac{\partial^{i+j} g_3(c_1(\theta) \mid \theta)}{\partial \theta^i \partial t^j}$$

$$+ c_1^{(p)}(\theta) g_3(c_1(\theta) \mid \theta) = 0, \qquad j_1, j_2, \cdots, j_k, i_1, \cdots, i_k, i + j < p.$$

From the continuity and positiveness it follows that $c^{(p)}(\theta)$ is continuous. Since $\int_{-\infty}^{\infty} G_2(t)\, dt < M(\theta_1, \theta_2)$ it follows that there is a constant $M(\theta_1, \theta_2)$ so that

$$\int_{-\infty}^{-M(\theta_1,\theta_2)} G_2(t)\, dt < \alpha, \qquad \int_{M(\theta_1,\theta_2)}^{\infty} G_2(t)\, dt < 1 - \alpha.$$

Thus

$$| c_1(\theta) | \leq M(\theta_1, \theta_2).$$

From (1) and condition (i) it follows easily that $| c_1'(\theta) | \leq M(\theta_1, \theta_2)$. Similarly we obtain $| c_1^{(p)}(\theta) | \leq M(\theta_1, \theta_2)$ for $\theta_1 \leq \theta \leq \theta_2$.

While the conditions (i) to (iv) suffice to insure the results of the theorem they are not necessary. It is often possible to obtain these properties of $c_1(\theta)$ in particular examples where $g_3(t, \theta)$ does vanish at points so long as $g_3(c_1(\theta), \theta)$ behaves well.

DEFINITION 1. *$\varphi_m(\hat{\theta})$ is an admissible function of order $m (m \leq s$, s fixed in advance) if $\varphi_m(\hat{\theta}) = c_1(\hat{\theta}) + \cdots + c_m(\hat{\theta})$ where* $\Pr\{t \leq c_1(\theta)\} = \alpha$ *and*

$$(3) \quad | c_i^{(p)}(\theta) | \leq M(\theta_1, \theta_2) N^{-(i-1)/2}, \qquad p = 0, 1, \cdots, s + 1 - i, \theta_1 \leq \theta \leq \theta_2.$$

Now let

$$(4) \qquad H_k(\theta) = N^{k/2} E(\hat{\theta} - \theta)^k = N^{k/2} \int_{-\infty}^{\infty} (\hat{\theta} - \theta)^k g_2(\hat{\theta} \mid \theta)\, d\hat{\theta} \quad \text{and}$$

$$(5) \qquad G_{pq}(\theta) = \frac{\partial^{p+q}}{\partial \hat{\theta}^p \partial t^q} g_1(t \mid \hat{\theta}, \theta) \big|_{t=c_1(\theta), \hat{\theta}=\theta}.$$

We have

THEOREM 2. *If*

(i) $\Pr\{t \leq c_1(\theta)\} = \alpha, \quad 0 < \alpha < 1,$ *and* $| c_1^{(p)}(\theta) | \leq M(\theta_1, \theta_2),$
$\theta_1 \leq \theta \leq \theta_2, p = 0, 1, \cdots, s;$

(ii) $\delta = \delta(N) = O(1)$ *is a function of N such that*

$$\int_{|\hat{\theta}-\theta| \geq \delta} d\hat{\theta}\, | \hat{\theta} - \theta |^k g_2(\hat{\theta} \mid \theta) \leq M(\theta_1, \theta_2) N^{-s/2}, \theta_1 \leq \theta \leq \theta_2, k = 0, 1, \cdots, s;$$

(iii) $\left| \frac{\partial^{p+q}}{\partial t^p \partial \hat{\theta}^q} g_1(t \mid \hat{\theta}, \theta) \right| \leq M(\theta_1, \theta_2), \qquad p + q = s,$
$| t - c_1(\theta) | < \rho, | \hat{\theta} - \theta | \leq \delta,$

where

$$\rho = \underset{|\hat{\theta}-\theta|\leq\delta}{\text{Max.}} \; | c_1(\hat{\theta}) - c_1(\theta) | + N^{-(1/2)+\eta}, \quad \eta > 0, \quad \theta_1 \leq \theta \leq \theta_2 ;$$

(iv) $| H_k^{(p)}(\theta) | \leq M(\theta_1, \theta_2)$ *for* $p = 0, 1, \cdots, s - k, k = 1, \cdots, s,$
$\theta_1 \leq \theta \leq \theta_2$;

(v) $| G_{pq}^{(l)}(\theta) | \leq M(\theta_1, \theta_2)$ *for* $l = 0, 1, \cdots, s - p - q,$
$p + q \leq s - 1$;

(vi) $\varphi_m(\hat{\theta})$ *is an admissible function of order* $m \leq s$,
then

$$\Pr\{t \leq \varphi_m(\hat{\theta})\} = \alpha + r_{m1}(\theta)N^{-1/2} + \cdots + r_{ms}(\theta)N^{-s/2} \tag{6}$$

where

$| r_{mj}^{(p)}(\theta) | \leq M(\theta_1, \theta_2)$ *for* $p = 0, 1, \cdots, s - j, \; j \leq s, \; \theta_1 \leq \theta \leq \theta_2$.

PROOF. Expand $g_1(t \mid \hat{\theta}, \theta)$ in a Taylor Expansion about $t = c_1(\theta)$, $\hat{\theta} = \theta$, with remainder terms of order s in $t - c_1(\theta)$, $\hat{\theta} - \theta$, and expand $c_i(\theta)$ about $\hat{\theta} = \theta$ where the remainder term is of order $s + 1 - i$. Then for $| \hat{\theta} - \theta | \leq \delta$, we have

$$\int_{c_1(\theta)}^{\varphi_m(\hat{\theta})} g_1(t \mid \hat{\theta}, \theta)\, dt = P\{(\hat{\theta} - \theta)^i, c_i^{(p)}(\theta), G_{pq}\} + RN^{-s/2}, \tag{7}$$

where P is a polynomial and $| R | \leq M(\theta_1, \theta_2)\sum_{i=0}^{s} (\hat{\theta} - \theta)^{s-i}N^{-i/2}$ for $| \hat{\theta} - \theta | \leq \delta$. Integrating over $| \hat{\theta} - \theta | \leq \delta$, we use conditions (ii), (iv) and (v) and the theorem follows. By a similar argument we have

THEOREM 3. *If*

(i) *the conditions of Theorem* 2 *hold for each* (θ_1, θ_2) *so that*

$$-\infty \leq \beta_1 < \theta_1 < \theta_2 < \beta_2 \leq \infty$$

and

(ii) $g_1(c_1(\theta) \mid \theta, \theta) > (1/M(\theta_1, \theta_2)) > 0, \quad \theta_1 \leq \theta \leq \theta_2$,

then the sequence

$$\begin{aligned} \varphi_1(\hat{\theta}) &= c_1(\hat{\theta}); \\ \varphi_2(\hat{\theta}) &= c_1(\hat{\theta}) - r_{1,1}(\hat{\theta})N^{-1/2}; \\ \varphi_m(\hat{\theta}) &= \varphi_{m-1}(\hat{\theta}) - \frac{r_{m-1,m-1}(\hat{\theta})N^{-(m-1)/2}}{g_1(c_1(\hat{\theta}) \mid \hat{\theta}, \hat{\theta})}, \qquad m \leq s, \end{aligned} \tag{8}$$

is a sequence of admissible functions such that

(a) $$\Pr\{t \leq \varphi_m(\hat{\theta})\} = \alpha + R(\theta)N^{-m/2}, \qquad m \leq s,$$

where $|R(\theta)| \leq M(\theta_1, \theta_2)$ *for* $\beta_1 < \theta_1 \leq \theta \leq \theta_2 < \beta_2$.

These theorems permit us to obtain and to calculate critical regions whose size is asymptotically close to α.

In Theorem 2, condition (ii) was much stronger than necessary. It may be relaxed if we define

$$H_k(\theta) = \int_{|\hat\theta-\theta|\leq\delta} N^{k/2} g_2(\hat\theta \mid \theta)(\hat\theta - \theta)^k \, d\theta,$$

where

$$\Pr\{|\hat\theta - \theta| \geq \delta\} \leq M(\theta_1, \theta_2)N^{-s/2}, \qquad \delta = \delta(N) = O(1).$$

However this may complicate the calculations.

The symmetric case arises when the first moment almost vanishes, i.e.

$$(10) \quad |H_1^{(p)}(\theta)| \leq M(\theta_1, \theta_2)N^{-1/2}, \quad p = 0, 1, \cdots, s-1, \quad \theta_1 \leq \theta \leq \theta_2.$$

In this case we have instead of the sequence given in Theorem 3, the sequence

$$(11) \quad \begin{aligned} \varphi_1(\hat\theta) &= c_1(\hat\theta); \\ \varphi_2(\hat\theta) &= c_1(\hat\theta) - \frac{r_{1,2}(\hat\theta)N^{-1} + r_{1,3}(\hat\theta)N^{-3/2}}{g_1(c_1(\hat\theta) \mid \hat\theta, \hat\theta)}; \\ \varphi_m(\hat\theta) &= \varphi_{m-1}(\hat\theta) - \frac{r_{m-1,2m-2}(\hat\theta)N^{-(m-1)} + r_{m-1,2m-1}(\hat\theta)N^{-(2m-1/2)}}{g_1(c_1(\hat\theta) \mid \hat\theta, \hat\theta)}, \end{aligned}$$

which is a sequence of admissible functions such that

$$\Pr\{t \leq \varphi_m(\hat\theta)\} = \alpha + r_{m,2m}(\theta)N^{-m} + \cdots + r_{m,s}(\theta)N^{-s/2}$$

$$|r_{m,n}^{(p)}(\theta)| \leq M(\theta_1, \theta_2) \qquad \theta_1 \leq \theta \leq \theta_2 \qquad p = 0, 1, \cdots, s-n.$$

4. An example. The following example previously treated by Welch from a different point of view will furnish an illustration of the applicability of the theorems to the case where θ is a k dimensional parameter. It will also serve as an example of an extended type of symmetry. That is, it has the property that $|H_{2k+1}^{(p)}(\theta)| \leq M(\theta_1, \theta_2)N^{-1/2}$, and hence, in the sequence (11), the $r_{m,2m+1}(\theta)$ terms effectively vanish thereby simplifying the calculations considerably.

We suppose that t is a normally distributed variable with mean μ and variance $\sigma^2 = \lambda_1\sigma_1^2 + \cdots + \lambda_k\sigma_k^2$ where the λ_i are known positive constants, the σ_i^2 are unknown parameters each of which is independently estimated by s_i^2 where $N_i s_i^2/\sigma_i^2$ has the χ^2 distribution with N_i degrees of freedom. It is desired to test the hypothesis that $\mu = 0$ so that the probability of rejecting the hypothesis if it is true should equal α. Under the hypothesis the joint density distribution of $t, s_1^2, \cdots s_k^2$, is given by

$$(12) \qquad g(t, s_1^2, \cdots, s_k^2/\sigma_1^2, \cdots, \sigma_k^2) = \frac{e^{-t^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \prod_{i=1}^{k} g(s_i^2 \mid \sigma_i^2; N_i),$$

where the moments of $s_i^2 - \sigma_i^2 = \hat{\theta}_i - \theta_i$ are given by the coefficients of $u^k/k!$ in the expansion about $u = 0$ of $e^{-u\sigma_i^2}(1 - (2u\sigma_i^2/N_i))^{-N_i/2}$:

$$H_1(\sigma_i^2) = 0;$$
$$H_2(\sigma_i^2) = 2\sigma_i^4;$$
$$H_3(\sigma_i^2) = \tfrac{4}{3}\sigma_i^6 N_i^{-1/2};$$
$$H_4(\sigma_i^2) = (\tfrac{11}{6} + 2N_i^{-1})\sigma_i^8.$$

We define $c_1(\theta)$ by $\Pr\{t \le c_1(\theta)\} = \alpha$ where

$$\theta = (\sigma_1^2, \sigma_2^2, \cdots, \sigma_k^2) \qquad \text{and} \qquad \hat{\theta} = (s_1^2, s_2^2, \cdots, s_k^2),$$
$$c_1(\theta) = c_1\sigma.$$

Now $\alpha_2(\theta) - \alpha = \Pr\{c_1(\theta) \le t \le c_1(\hat{\theta})\}$ may be computed within terms of order N_i^{-2} by expanding

$$c_1(\hat{\theta}) \approx c_1\sigma + c_1 \sum \frac{1}{2\sigma}\lambda_i(s_i^2 - \sigma_i^2) - c_1 \sum \frac{1}{8\sigma^3}\lambda_i\lambda_j(s_i^2 - \sigma_i^2)(s_j^2 - \sigma_j^2)$$
$$-\frac{1}{\sqrt{2\pi\sigma^2}}e^{-t^2/2\sigma^2} \approx \frac{1}{\sqrt{2\pi\sigma^2}}e^{-c_1^2/2}\{1 + (t - c_1\sigma)(-c_1/\sigma)\},$$

whence

$$\alpha_2(\theta) - \alpha \approx \int_0^\infty \cdots \int_0^\infty ds_1^2 \cdots ds_k^2 \prod_{i=1}^k g(s_i^2 \mid \sigma_i^2; N_i) \left\{\frac{e^{-c_1^2/2}}{\sqrt{2\pi\sigma^2}}\right\}\left\{\frac{c_1}{2\sigma}\sum \lambda_i(s_i^2 - \sigma_i^2)\right.$$
$$\left. - \frac{c_1}{8\sigma^3}\sum \lambda_i\lambda_j(s_i^2 - \sigma_i^2)(s_j^2 - \sigma_j^2) - \frac{c_1}{2\sigma}\left(\frac{c_1^2}{4\sigma^2}\sum \lambda_i\lambda_j(s_i^2 - \sigma_i^2)(s_j^2 - \sigma_j^2)\right)\right\}$$
$$= -\frac{e^{-c_1^2/2}}{\sqrt{2\pi\sigma^2}}\left\{\frac{c_1 + c_1^3}{8\sigma^3}\right\}\{\sum 2\lambda_i^2\sigma_i^4 N_i^{-1}\} + O(\sum N_i^{-2}).$$

Thus

$$c_2(\theta) = \frac{c_1 + c_1^3}{4\sigma^3}\sum \lambda_i^2\sigma_i^4 N_i^{-1}$$

and

$$\alpha_3(\theta) = \Pr\{t \le c_1 s + \frac{c_1 + c_1^3}{4s^3}\sum \lambda_i^2 s_i^4 N_i^{-1}\} = \alpha + O(\sum N_i^{-2}),$$

where

$$s^2 = \sum \lambda_i s_i^2.$$

Further approximations become somewhat complex and should be carried out in a systematic fashion.

5. Remarks. The range of application in practical statistical problems of the theorems of section 2 may be somewhat more limited than that of the original method proposed by Wald. Concerning the original method, the following theorems have been established.

THEOREM 4. *If*

(i) $\Pr\{t \leq c_1(\theta)\} = \alpha, 0 < \alpha < 1,$ *where* $| c_1^{(p)}(\theta) | \leq M(\theta_1, \theta_2), \theta_1 \leq \theta \leq \theta_2,$
$p = 0, 1, \cdots, s;$

(ii) $\left| \frac{\partial^{i+j} g(t, \hat{\theta} + \Delta \mid \theta + \Delta)}{\partial t^i \partial \Delta^j} \right| \leq G(\hat{\theta}, \theta)$ *for* $i + j \leq s - 1, C_1 \leq t \leq C_2,$
$\theta_1 \leq \theta, \theta + \Delta \leq \theta_2, | \Delta | \leq \Delta',$ *where* $G(\hat{\theta}, \theta)$ *depends on* $C_1, C_2, \theta_1, \theta_2, N,$ *and is integrable in* $\hat{\theta}$ *over* $(-\infty, \infty)$;

(iii) $\left| \frac{\partial^{i+j} g(t, \hat{\theta} \mid \theta)}{\partial t^i \partial \Delta^j} \right| \leq L(\hat{\theta}, \theta), i + j \leq s - 1, C_1 \leq t \leq C_2, \theta_1 \leq \theta \leq \theta_2,$
where $\int_{-\infty}^{\infty} L(\hat{\theta}, \theta) | \hat{\theta} - \theta |^k \, d\hat{\theta} \leq M(\theta_1, \theta_2, C_1, C_2)N^{-k/2}, k = 0, 1,$

(iv) $0 < A(C_1, C_2, \theta_1, \theta_2) < A(t) \leq g_3(t \mid \theta) \leq B(t) < B(C_1, C_2, \theta_1, \theta_2) < \infty,$
$\theta_1 \leq \theta \leq \theta_2, C_1 \leq t \leq C_2,$
$\int_{-\infty}^{\infty} B(t) \, dt < M(\theta_1, \theta_2);$

(v) $g(t, \hat{\theta} \mid \theta) > 0,$

then a sequence $c_1^*(\hat{\theta}), c_2(\theta), c_2^*(\hat{\theta}), \cdots, c_s^*(\hat{\theta}),$ *exists where* $c_m(\theta)$ *is uniquely defined in* (θ_1, θ_2) *by* $\Pr\{t - c_1^*(\hat{\theta}) - \cdots - c_{m-1}^*(\hat{\theta}) \leq c_m(\theta)\} = \alpha,$ *and*

$$|c_m^{(p)}(\theta)| \leq M(\theta_1, \theta_2)N^{-(m-1)/2} \; p = 0, 1, \cdots, s - m + 1, \theta_1 \leq \theta \leq \theta_2$$

and $c_m^*(\theta)$ *is any function so that*

$$|c_m^{*(p)}(\theta) - c_m^{(p)}(\theta) | \leq MN^{-m/2} \text{ for } \theta_1 \leq \theta \leq \theta_2, p = 0, 1, \cdots, s - m,$$

and

$$| c_m^{*(p)}(\hat{\theta}) | \leq M(\theta_1, \theta_2)N^{-(m-1)/2} \; -\infty < \hat{\theta} < \infty, p = 0, 1, \cdots, s - m + 1.$$

Finally for $c_m^*(\theta)$ arbitrary within the above conditions,

$$| \Pr\{t - c_1^*(\hat{\theta}) - \cdots - c_s^*(\hat{\theta}) \leq 0\} - \alpha | \leq M(\theta_1, \theta_2)N^{-s/2} \text{ for } \theta_1 \leq \theta \leq \theta_2.$$

The conditions on the derivatives with respect to Δ are natural because the intuitive approach to the method seems to hinge on the assumption that $g(t, \hat{\theta} + \Delta \mid \theta + \Delta)$ changes gradually with respect to Δ "independent" of the value of N. This would not be true of $g(t, \hat{\theta} \mid \theta + \Delta)$ for large N.

The $c_i^*(\theta)$ where introduced in Theorem 4 because in practical examples it is usually found too difficult to compute $c_i(\theta)$ efficiently. On the other hand there are many alternative ways of obtaining functions with the properties of the $c_i^*(\theta)$. The $c_2(\theta)$, $c_3(\theta)$ etc. mentioned in Theorems 1, 2, 3 play the role of the $c_i^*(\theta)$ in Theorem 4 with the exception of the condition on $c_i^*(\theta)$ for outside (θ_1, θ_2). The exception is due to the fact that the Theorems 1, 2, 3 correspond to the "infinite case." Theorem 4 is applicable to those cases where one is willing to assume that θ lies in (θ_1, θ_2). It often happens that there is no such reason or that the conditions of the theorem hold only for every closed proper subinterval of (β_1, β_2) but not for $\beta_1 \leq \theta \leq \beta_2$ itself. In these cases we may apply

THEOREM 5. *If*

(i) *all of the conditions of Theorem 4 apply to every finite proper closed subinterval (θ_1, θ_2) of (β_1, β_2) where (β_1, β_2) may be an infinite interval;*

(ii) $\Pr\{|\hat{\theta} - \theta| \geq \delta(N)\} \leq M(\theta_1, \theta_2)N^{-s/2}$ *for $\beta_1 < \theta_1 \leq \theta \leq \theta_2 < \beta_2$, where $\delta(N) = 0(1)$ unless β_1 or β_2 is finite, in which case $\delta(N) = o(1)$, then a sequence $c_1^*(\hat{\theta})$, $c_2(\theta)$, $c_2^*(\hat{\theta})$, $\cdots$, $c_s^*(\hat{\theta})$, exists, where $c_m(\theta)$ is uniquely defined in (β_1, β_2) by $Pr\{t - c_1^*(\hat{\theta}) - c_2^*(\hat{\theta}) - \cdots - c_{m-1}^*(\hat{\theta}) \leq c_m(\theta)\} = \alpha$, so that for every (θ_1, θ_2),*

$$| c_m^{*(p)}(\theta) | \leq M(\theta_1, \theta_2)N^{-(m-1)/2} \text{ if } \beta_1 < \theta_1 \leq \theta \leq \theta_2 < \beta_2, \quad p = 0, 1, \cdots, s - m + 1'$$

and for $c_m^*(\theta)$ arbitrary within the above conditions

$$| \Pr\{t \leq c_1^*(\hat{\theta}) + \cdots + c_m^*(\hat{\theta})\} - \alpha | \leq M(\theta_1, \theta_2)N^{-m/2} \quad \text{if} \quad \beta_1 < \theta_1 \leq \theta \leq \theta_2 < \beta_2, m \leq s.$$

Essentially this theorem can be proved by reference to the proof of Theorem 4 applied to the function

$$\begin{aligned} g^*(t, \hat{\theta} \mid \theta) &= g(t, \hat{\theta} \mid \theta) \quad \text{for} \quad | \hat{\theta} - \theta | \leq \delta; \\ &= \quad 0 \qquad\qquad\quad | \hat{\theta} - \theta | > \delta. \end{aligned}$$

Some of the conditions in Theorems 4 and 5 are stronger than necessary. For example $g > 0$ may be replaced by a weaker condition where g is positive in a region about $t = c_1(\theta)$. On the other hand the condition $\Pr\{|\hat{\theta} - \theta| \geq \delta\} \leq MN^{-s/2}$ in Theorem 5 is necessary to the argument used in the proof. It is easy to construct trivial examples where the results of this theorem apply although this condition is not satisfied. However an example has also been constructed where all the conditions of Theorem 5 hold except for this condition and the method of Wald fails to give the results.

These theorems are very easily extended to the k-dimensional parameter case by replacing the conditions on the derivatives with respect to Δ by the same order mixed derivatives with respect to $\Delta_1, \Delta_2, \cdots, \Delta_k$ of

$$g(t, \hat{\theta}_1 + \Delta_1, \hat{\theta}_2 + \Delta_2, \cdots, \hat{\theta}_k + \Delta_k \mid \theta_1 + \Delta_1, \cdots, \theta_k + \Delta_k).$$

The symmetric case arises when the distribution of $\hat{\theta}$ is almost symmetric about θ. More exactly we have

THEOREM 6. *If*
(i) *All the conditions of Theorem* 4 *hold and* $L(\hat{\theta}, \theta)$ *has the additional property that*

$$\int_{-\infty}^{\infty} (\hat{\theta} - \theta)^2 L(\hat{\theta}, \theta)\, d\hat{\theta} < M(\theta_1, \theta_2)N^{-1}, \qquad \theta_1 \leq \theta \leq \theta_2,$$

and

$$\text{(ii)} \qquad \left| \frac{\partial g^{i+j}}{\partial \Delta^i \partial t^j}(t, \hat{\theta} \mid \theta) - \frac{\partial g^{i+j}}{\partial \Delta^i \partial t^j}(t, 2\theta - \hat{\theta} \mid \theta) \right| < L(\hat{\theta}, \theta) \mid \hat{\theta} - \theta \mid,$$

$$C_1 \leq t \leq C_2, \qquad \theta_1 \leq \theta \leq \theta_2, \qquad i + j \leq s - 1,$$

then it is possible to construct a sequence $c_1^*(\hat{\theta}), c_2(\theta), \cdots, c_r^*(\hat{\theta})$, *as in Theorem* 4 *so that*

$$| c_m^{(p)}(\theta) | \leq M(\theta_1, \theta_2)N^{-(m-1)},$$
$$p = 0, 1, \cdots, s - 2m + 2, \theta_1 \leq \theta \leq \theta_2;$$
$$| c_m^{*(p)}(\theta) - c_m^{(p)}(\theta) | \leq M(\theta_1, \theta_2)N^{-m},$$
$$p = 0, 1, \cdots, s - 2m + 1, \theta_1 \leq \theta \leq \theta_2;$$
$$| c_m^{*(p)}(\hat{\theta}) | \leq M(\theta_1, \theta_2)N^{-(m-1)},$$
$$p = 0, 1, \cdots, s - 2m + 2, -\infty < \hat{\theta} < \infty;$$

and

$$| \Pr\{t \leq c_1^*(\hat{\theta}) + \cdots + c_r^*(\hat{\theta})\} - \alpha | \leq M(\theta_1, \theta_2)N^{-s/2},$$
$$\theta_1 \leq \theta \leq \theta_2, r = \left[\frac{s+1}{2}\right].$$

Theorem 5 can also be extended to the symmetric case.

It is often possible in the theory of statistics to obtain an asymptotic expansion of the distribution of t, $\hat{\theta}$. The treatment of such cases is often very simple because of the prominent role played by the normal distribution in such asymptotic expansions. Suppose that

$$g(t, \hat{\theta} \mid \theta) = \sqrt{N}\gamma(t, \psi \mid \theta),$$

where $\psi = \sqrt{N}(\hat{\theta} - \theta)$; γ = density distribution of (t, ψ);

$$\gamma(t, \psi \mid \theta) = \gamma_0(t, \psi \mid \theta) + N^{-1/2}\gamma_1(t, \psi \mid \theta) + \cdots + N^{-(s-1)/2}\gamma_{s-1}(t, \psi \mid \theta) + \rho(t, \psi \mid \theta)N^{-s/2},$$

$\gamma_0, \gamma_1, \cdots, \gamma_{s-1}$ are independent of N;

$$\iint | \rho | \, d\psi\, dt \leq M(\theta_1, \theta_2), \qquad \theta_1 \leq \theta \leq \theta_2;$$

$$\iint | \gamma_i | \, d\psi\, dt \leq M(\theta_1, \theta_2), \qquad \theta_1 \leq \theta \leq \theta_2.$$

Correspondingly we have

$$g(t, \hat{\theta} \mid \theta) = g_0(t, \hat{\theta} \mid \theta) + N^{-1/2}g_1(t, \hat{\theta} \mid \theta) + \cdots + N^{-(s-1)/2} g_{s-1}(t, \hat{\theta} \mid \theta) + r(t, \hat{\theta} \mid \theta)N^{-s/2},$$

where

$$g_i(t, \hat{\theta} \mid \theta) = \sqrt{N}\, \gamma_i(t, \psi \mid \theta), \qquad r(t, \hat{\theta} \mid \theta) = \sqrt{N}\rho(t, \psi \mid \theta).$$

Then if we define $c_1(\theta)$ by $\int_{-\infty}^{\infty} d\hat{\theta} \int_{-\infty}^{c_1(\theta)} dtg_0 = \alpha$,

$$c_m(\theta) \quad \text{by} \quad \int_{-\infty}^{\infty} d\hat{\theta} \int_{c_1(\hat{\theta})+\cdots+c_{m-1}(\hat{\theta})}^{c_1(\theta)+\cdots+c_{m-1}(\hat{\theta})+c_m(\theta)} dtg_0$$

$$= \alpha - \int_{-\infty}^{\infty} d\hat{\theta} \int_{-\infty}^{c_1(\theta)+\cdots+c_{m-1}(\theta)} dt[g_0 + g_1 N^{-1/2} + \cdots + g_{m-1}N^{-(m-1)/2}].$$

or by

$$c_m(\theta) \int_{-\infty}^{\infty} d\hat{\theta} g_0(c_1(\theta), \hat{\theta} \mid \theta)$$

$$= \alpha - \int_{-\infty}^{\infty} d\hat{\theta} \int_{-\infty}^{c_1(\hat{\theta})+\cdots+c_{m-1}(\hat{\theta})} dt[g_0 + g_1 N^{-1/2} + \cdots + g_{m-1}N^{-((m-1)/2)}],$$

we obtain

$$\mid \Pr\{t \leq c_1(\hat{\theta}) + \cdots + c_s(\hat{\theta})\} - \alpha \mid \leq M(\theta_1, \theta_2)N^{-s/2},$$

if g obeys the conditions of Theorem 4 except that we need only $s - i + 1$ derivatives for $g_i(t, \hat{\theta} \mid \theta)$. The above definitions of $c_m(\theta)$ correspond to the $c_m^*(\theta)$ in Theorem 4. Analogues of Theorems 5 and 6 also apply to the asymptotic case.

REFERENCE

[1] B. L. Welch, "On the studentization of several variances," *Annals of Math. Stat.*, Vol. 18 (1947), p. 118.