The Annals of Statistics
1987, Vol. 15, No. 1, 398–412

ON THE EFFECT OF SUBSTITUTING PARAMETER ESTIMATORS IN LIMITING χ^2 U AND V STATISTICS

BY TERTIUS DE WET AND RONALD H. RANDLES

Institute for Maritime Technology and University of Florida

Consider statistics $T_n(\lambda)$ that take the form of limiting chi-square (degenerate) U or V statistics. Here the phrase "limiting chi-square" means they have the same asymptotic distribution as a weighted sum of (possibly infinitely many) independent χ_1^2 random variables. This paper examines the limiting distribution of $T_n(\hat{\lambda})$ and compares it to that of $T_n(\lambda)$, where $\hat{\lambda}$ denotes a consistent estimator of λ based on the same data. Whether or not $T_n(\hat{\lambda})$ and $T_n(\lambda)$ have the same limiting distribution is primarily a question of whether or not a certain mean function has a zero derivative. Some statistics that are appropriate for testing hypotheses are used to illustrate the theory.

1. Introduction. In many hypothesis testing and estimation problems, we need to know the effects of substituting an estimator of a parameter into a statistic. Let $X_1, \dots, X_n$ denote a random sample and consider statistics of the form

$$T_n(\hat{\lambda}) = T_n(X_1, \dots, X_n; \hat{\lambda}).$$

The statistic $\hat{\lambda}$ is a consistent estimator of the parameter λ. This paper investigates the limiting distribution of $T_n(\hat{\lambda})$ and compares it to the limiting distribution of $T_n(\lambda)$, thus exploring the large sample effect (if any) of substituting an estimator into $T_n(\lambda)$. We shall see how to determine whether $T_n(\hat{\lambda})$ has the same limiting distribution as $T_n(\lambda)$ and, if not, what change results.

Investigations of these questions for particular statistics are common in the literature. General theorems are available for certain classes of statistics with asymptotic normal distributions. Sukhatme (1958) and Randles (1982) investigated cases in which $T_n(\lambda)$ is a U statistic. Pierce (1982) provides a general limiting normal distribution theorem for settings in which $\hat{\lambda}$ is an efficient estimator of λ. Fligner and Hettmansperger (1979) give a result applicable when one can obtain the weak convergence of the conditional distribution of $T_n(\hat{\lambda})$ given $\hat{\lambda}$. For cases in which $T_n(\lambda)$ is an L statistic, Parr (1982) and Randles (1982) have derived limiting normal distribution results. Iverson and Randles (1983) establish weak and strong consistency and a LIL when $T_n(\lambda)$ is either a U statistic or an L statistic.

The present paper describes the effects of auxiliary estimators on the limiting distribution for cases in which $T_n(\lambda)$ is a U or V statistic with a limiting χ^2-type distribution. Gregory (1977) proved a basic limit theorem for limiting χ^2-type (degenerate) U statistics. Serfling (1980) gives an independent proof. Neuhaus (1977) proved the two-sample U-statistic analogue to this result. Each of these

Received September 1984; revised March 1986.
AMS 1980 *subject classifications.* Primary 62E20; secondary 62G10.
Key words and phrases. Limiting χ^2 distributions, U statistics, V statistics, asymptotic distribution.

authors discussed the corresponding result for V statistics as well. Examples of useful limiting χ^2-type U or V statistics are given by de Wet and Venter (1972, 1973, 1976) and de Wet, Venter and Van Wyk (1979).

Several examples of limiting χ^2-type U and V statistics with kernels that contain an estimator have appeared in the literature. de Wet and Venter (1972) study a statistic used to test for normality when the mean and variance are nuisance parameters that must be estimated. Gregory (1977) discusses a Cramér–von Mises test statistic computed on data that have been standardized by location and scale estimators. Pearson χ^2 statistics for goodness-of-fit in which the cell boundaries are functions of estimated parameters have been described by a number of authors. See the general results of Moore and Spruill (1975), as well as their references.

The present paper examines $T_n(\hat{\lambda})$ in situations in which $T_n(\lambda)$ is a limiting χ^2-type V or U statistic. With kernels of a particular form, general results are obtained that characterize the distribution of $T_n(\hat{\lambda})$ and relate it to that of $T_n(\lambda)$. Conditions are specified under which these two quantities have the same limiting χ^2-type distribution (referred to as Case I) and have different χ^2-type limiting distributions (Case II).

Section 2 gives the motivation for our approach and statements of the main theorems. Section 3 contains examples that illustrate applications of the results. It includes settings in which the estimator changes the limiting distributions and others in which it does not. Proofs of the theorems are given in Section 4.

2. Motivation and statement of results. Let $X_1, X_2, \dots, X_n$, possibly vector valued, be i.i.d. with d.f. $F(x)$. We consider statistics of the form

$$V_n(\hat{\lambda}) = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} h\big(X_i, X_j; \hat{\lambda}\big), \tag{2.1}$$

where the estimator $\hat{\lambda}$ is also a function of $X_1, \dots, X_n$ and consistently estimates the p-vector parameter λ. The kernel $h(\cdot)$ is assumed to satisfy

$$h(x, y; \gamma) = h(y, x; \gamma) \tag{2.2}$$

and

$$\int h(x, t; \lambda)\, dF(t) \equiv 0 \tag{2.3}$$

for every x, y and $\gamma \in R^p$. When γ is known, $V_n(\gamma)$ is a degree-2 V statistic. Equation (2.3) shows that $V_n(\lambda)$ is a member of the (degenerate) class of degree-2 V statistics with χ^2-type limiting distributions. (See, e.g., Theorem 6.4.1B in Serfling (1980).)

To provide a general solution for a useful class of statistics without requiring the kernel h to be differentiable in $\hat{\lambda}$, we consider kernels of the form

$$h(x_1, x_2; \gamma) = \int_{-\infty}^{\infty} g(x_1, t; \gamma) g(x_2, t; \gamma)\, dM(t) \tag{2.4}$$

for some function g and M a finite positive measure, possibly defined over a vector valued space. In Section 3 a number of examples are discussed for which this representation holds. Now, using (2.4), (2.1) can be written as

$$V_n(\hat{\lambda}) = \int_{-\infty}^{\infty} \left[n^{-1} \sum_{i=1}^{n} g(X_i, t; \hat{\lambda}) \right]^2 dM(t). \tag{2.5}$$

Since the term in square brackets is (for each fixed t) a degree-1 V statistic with an estimated parameter, we motivate an asymptotic approximation for $V_n(\hat{\lambda})$ by first examining a degree-1 V statistic that has an estimator inserted in its kernel,

$$V_{1n}(\hat{\lambda}) = n^{-1} \sum_{i=1}^{n} h(X_i; \hat{\lambda}).$$

Taking $\hat{\lambda} = \lambda(F_n)$, a functional of the empirical d.f. F_n, $V_{1n}(\hat{\lambda})$ can be written functionally as

$$T_1(F) = \int h(x; \lambda(F))\, dF(x). \tag{2.6}$$

The first Gâteaux differential of $T_1(\cdot)$ at F in the direction of F_n produces

$$\begin{aligned} d_1T_1(F; F_n - F) = {} & \int h(x; \lambda)\, d(F_n - F) \\ & + \mathbf{d}_1\theta_1(\lambda)'\, d_1\lambda(F; F_n - F). \end{aligned} \tag{2.7}$$

Here $\mathbf{d}_1\theta_1(\lambda)$ denotes the vector of partial derivatives of $\theta_1(\gamma) = E_\lambda[h(X_1; \gamma)]$ with respect to γ, where the expectation assumes λ is the actual parameter value. Also, $d_1\lambda(F; F_n - F)$ denotes the vector of Gâteaux differentials of the components of $\lambda(\cdot)$ at F in the direction of F_n. The differential in (2.7) motivates the theorems in Sukhatme (1958) and Randles (1982), which asymptotically approximate $V_{1n}(\hat{\lambda}) - \theta_1(\lambda)$ with

$$V_{1n}(\lambda) - \theta_1(\lambda) + \mathbf{d}_1\theta_1(\lambda)'(\hat{\lambda} - \lambda).$$

This is a useful approximation, because often $h(x; \gamma)$ is not differentiable in γ at $\gamma = \lambda$, but yet $\theta_1(\gamma)$ is. Note that if $\theta_1(\gamma)$ has a zero derivative at $\gamma = \lambda$, then the limiting distribution of $V_{1n}(\lambda)$ is not affected if λ has to be estimated.

We now apply these ideas to $V_n(\hat{\lambda})$ as given by (2.5). With $\hat{\lambda} = \lambda(F_n)$, write (2.5) functionally as

$$T(F) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(x, t; \lambda)\, dF(x) \right]^2 dM(t).$$

Taking the first differential of the interior of this square bracket in the fashion of (2.7), we are motivated to approximate $V_n(\hat{\lambda})$ with

$$\int_{-\infty}^{\infty} \left[n^{-1} \sum_{i=1}^{n} g(X_i, t; \lambda) + \mathbf{d}_1\mu(t; \lambda)'(\hat{\lambda} - \lambda) \right]^2 dM(t), \tag{2.8}$$

where $\mathbf{d}_1\mu(t; \lambda)$ represents the vector of partial derivatives of $\mu(t; \gamma) = E_\lambda[g(X_1, t; \gamma)]$ at $\gamma = \lambda$.

REMARK. Note again the fact that the presence of an estimated parameter in $V_n(\hat{\lambda})$ has an effect on the limiting distribution if $\mu(t;\gamma)$ has a nonzero vector of derivatives at $\gamma = \lambda$. Also, if $\hat{\lambda}$ is asymptotically linear in the observations (see Condition 2.10), (2.8) leads to the integral of the square of a linear statistic, thus having a χ^2-type limiting distribution. The approximation (2.8) has the advantage that $g(x, t;\gamma)$ need not be differentiable in γ.

We now state the conditions and the main result showing the validity of the approximation in (2.8). Condition 2.10 is the usual asymptotic linearity applied to $\hat{\lambda}$. Conditions 2.9 and 2.11 are basically smoothness conditions on the function $g(\ )$ and hence indirectly on the kernel $h(\ )$.

CONDITION 2.9. Suppose

$$\mu(t;\gamma) = E_{\lambda}[g(X_1, t;\gamma)]$$

exists and satisfies

$$\mu(t;\lambda) \equiv 0$$

for every t and every γ in a neighborhood of $\gamma = \lambda$. In addition, assume $\mu(t, \gamma)$ has an $L_2(R, M)$ differential at $\gamma = \lambda$ with partial derivative vector $\mathbf{d}_1\mu(t;\lambda)$, satisfying

$$\int_{-\infty}^{\infty} [\mathbf{d}_1\mu(t;\lambda)_r]^2\, dM(t) < \infty$$

for $r = 1, \dots, p$, where $\mathbf{d}_1\mu(\ )_r$ is the rth component of the vector $\mathbf{d}_1\mu(\ )$. The existence of an $L_2(R, M)$ differential means for any $\varepsilon > 0$ there is a bounded sphere $\mathfrak{C}$ in R^p centered at λ, such that $\gamma \in \mathfrak{C}$ implies

$$\|\gamma - \lambda\|^{-2} \int_{-\infty}^{\infty} [\mu(t;\gamma) - \mathbf{d}_1\mu(t;\lambda)'(\gamma - \lambda)]^2\, dM(t) < \varepsilon.$$

CONDITION 2.10. Suppose

$$\hat{\lambda} = \lambda + n^{-1} \sum_{i=1}^{n} \alpha(X_i) + \mathbf{o}_p(n^{-1/2}),$$

where $E[\alpha(X_i)_r] = 0$ and $E[\alpha(X_i)_r \alpha(X_i)_{r'}]$ is finite for all $1 \le r \le r' \le p$.

CONDITION 2.11. Suppose there is a number $M^* > 0$ and a neighborhood $K(\lambda)$ of λ such that

(a) if $\gamma \in K(\lambda)$ and $D(\gamma, d)$ is a sphere centered at γ with radius d such that $D(\gamma, d) \subset K(\lambda)$, then

$$(2.12)\qquad \int_{-\infty}^{\infty} \left\{ E\left[\sup_{\gamma' \in D(\gamma, d)} |g(X_i, t;\gamma') - g(X_i, t;\gamma)| \right] \right\}^2 dM(t) \le M^* d^2,$$

and

(b) for any $\varepsilon > 0$ there is a $d^* > 0$ such that $0 < d < d^*$, $\gamma \in K(\lambda)$ and $D(\gamma, d) \subset K(\lambda)$ imply

$$(2.13)\qquad \int_{-\infty}^{\infty} E\left[\sup_{\gamma' \in D(\gamma, d)} |g(X_i, t;\gamma') - g(X_i, t;\gamma)|^4 \right] dM(t) < \varepsilon.$$

We now state our first result. To this end, let

$$V_n = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} h_*(X_i, X_j) \tag{2.14}$$

with

$$\begin{aligned} h_*(x, y) = \int_{-\infty}^{\infty} &[g(x, t; \lambda) + \mathbf{d}_1 \mu(t; \lambda)' \alpha(x)] \\ &\times [g(y, t; \lambda) + \mathbf{d}_1 \mu(t; \lambda)' \alpha(y)] \, dM(t). \end{aligned} \tag{2.15}$$

THEOREM 2.16. *Let $X_1, \dots, X_n$ be i.i.d. with d.f. $F(x)$. Suppose Conditions* 2.9, 2.10 *and* 2.11 *hold, and*

$$E\left[h_*^2(X_1, X_2)\right] < \infty \quad \text{and} \quad E\left[h_*(X_1, X_1)\right] < \infty. \tag{2.17}$$

Let $\{\delta_k\}$ denote the eigenvalues of the operator A defined by

$$Aq(x) = \int_{-\infty}^{\infty} h_*(x, y) q(y) \, dF(y).$$

Then

$$n\left(V_n(\hat{\lambda}) - V_n\right) \to_p 0$$

and

$$nV_n(\hat{\lambda}) \to_d \sum_{k=1}^{\infty} \delta_k \chi_{1k}^2,$$

where χ_{1k}^2, $k = 1, 2, \dots$, are independent χ_1^2 variates.

REMARK. Note that $nV_n(\hat{\lambda})$ has the same limiting distribution as nV_n, the latter being a degree-2 V statistic with a degenerate kernel h_*. Note also that the effect of the estimator $\hat{\lambda}$ is captured in $\mathbf{d}_1\mu(t; \lambda)$. For Case I situations

$$\mathbf{d}_1 \mu(t; \lambda) \equiv \mathbf{0}, \tag{2.18}$$

for all t and hence $nV_n(\hat{\lambda})$ and $nV_n(\lambda)$ have the same limiting distribution. Thus, the limiting distribution of $nV_n(\hat{\lambda})$ is the same as when λ is known. For Case II situations $\mathbf{d}_1\mu(t; \lambda) \neq \mathbf{0}$ and the limiting distribution is affected by substitution of the estimator $\hat{\lambda}$. Examples of both types will be presented in Section 3.

Next we state a result for U statistics with estimators in their kernel which corresponds to Theorem 2.16. Define

$$U_n(\hat{\lambda}) = \binom{n}{2}^{-1} \sum_{i<j} \int_{-\infty}^{\infty} g\left(X_i, t; \hat{\lambda}\right) g\left(X_j, t; \hat{\lambda}\right) dM(t)$$

and

$$U_n = \binom{n}{2}^{-1} \sum_{i<j} \int_{-\infty}^{\infty} [g(X_i, t;\lambda) + \mathbf{d}_1\mu(t;\lambda)'\alpha(X_i)] \times [g(X_j, t;\lambda) + \mathbf{d}_1\mu(t;\lambda)'\alpha(X_j)]\, dM(t).$$

The following theorem establishes the limiting distribution of $nU_n(\hat{\lambda})$.

THEOREM 2.19. *Suppose the assumptions of Theorem* 2.16 *hold and, in addition,*

$$E\left[\int_{-\infty}^{\infty} g^2(X_1, t;\lambda)\, dM(t)\right] < \infty.$$

Then

$$nU_n(\hat{\lambda}) \to_d \sum_{k=1}^{\infty} (\delta_k \chi_{1k}^2 - \delta_k'), \tag{2.20}$$

where $\{\delta_k\}$ *are defined in Theorem* 2.16, $\{\delta_k'\}$ *are the eigenvalues of the operator* A *defined by*

$$Aq(x) = \int_{-\infty}^{\infty} h(x, y;\lambda) q(y)\, dF(y)$$

and χ_{1k}^2, $k = 1, 2, \dots,$ *are independent* χ_1^2 *variates.*

REMARK. This result shows the peculiar effect that the estimator $\hat{\lambda}$ produces in degenerate U statistics where the limiting distribution depends on both the eigenvalues when λ is known (δ_k') and when it is estimated (δ_k). Of course these sets of eigenvalues are the same when (2.18) is satisfied. Note also that the right side of (2.20) is just

$$\sum_{k=1}^{\infty} \delta_k(\chi_{1k}^2 - 1) + \sum_{k=1}^{\infty} (\delta_k - \delta_k').$$

The quantity $\Sigma(\delta_k - \delta_k')$ was labelled γ in Theorem 4.3 of Gregory (1977) when he proved a corresponding result for the particular case of a Cramér–von Mises statistic.

3. Examples. In this section a number of examples to which the theory of Section 2 applies are discussed. These examples are given for illustrative purposes and we will not strive to give them in their most general form.

3.1. *Cramér–von Mises statistic for the regression model with random regressors.* Consider the linear regression model with random regressors

$$X_j = \mathbf{C}_j'\boldsymbol{\beta} + E_j, \qquad j = 1, 2, \dots, n, \tag{3.1}$$

with $\{\mathbf{C}_j\}$ i.i.d. p-dimensional random vectors with joint d.f. G, $\boldsymbol{\beta}$ the vector of unknown regression constants and $\{E_j\}$ i.i.d. with d.f. F. We will assume the $\mathbf{C}$'s to be independent of the E's. We wish to test the hypothesis

$$H_0\colon F = F_0$$

with F_0 completely specified. Let $\hat{\boldsymbol\beta}$ denote a consistent estimator of $\boldsymbol\beta$ satisfying Condition 2.10. Form the residuals

$$Y_j = X_j - \mathbf{C}_j'\hat{\boldsymbol\beta}, \qquad j = 1, 2, \dots, n.$$

Let $F_n(t)$ be the empirical distribution function of the $\{Y_j\}$, i.e.,

$$F_n(t) = n^{-1}\sum_{j=1}^{n} I(Y_j \le t).$$

The Cramér–von Mises statistic for testing H_0 computed on the residuals is then given by

$$T_n(\hat{\boldsymbol\beta}) \equiv n\int_{-\infty}^{\infty} (F_n(t) - F_0(t))^2\, dF_0(t). \tag{3.2}$$

Let $\mathbf{Z} \equiv (\mathbf{C}, E)$ be $(p+1)$ dimensional. We can then write

$$T_n(\hat{\boldsymbol\beta}) = n^{-1}\sum_{i,\,j=1}^{n} h(\mathbf{Z}_i, \mathbf{Z}_j; \hat{\boldsymbol\lambda})$$

with

$$\hat{\boldsymbol\lambda} = \hat{\boldsymbol\beta} - \boldsymbol\beta, \qquad \boldsymbol\lambda = \mathbf{0}$$

and

$$h(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol\gamma) = \int_{-\infty}^{\infty} g(\mathbf{z}_1, t; \boldsymbol\gamma) g(\mathbf{z}_2, t; \boldsymbol\gamma)\, dF_0(t), \tag{3.3}$$

where

$$g(\mathbf{z}, t; \boldsymbol\gamma) = I(e - \mathbf{c}'\boldsymbol\gamma \le t) - F_0(t),$$

using the notation $\mathbf{z} = (\mathbf{c}, e)$. Thus, $M(t) = F_0(t)$.

The mean function is

$$\begin{aligned}\mu(t;\boldsymbol\gamma) &= E[g(\mathbf{Z}, t; \boldsymbol\gamma)]\\ &= P(E - \mathbf{C}'\boldsymbol\gamma \le t) - F_0(t)\\ &= \int_{R^p} F_0(t + \mathbf{c}'\boldsymbol\gamma)\, dG(\mathbf{c}) - F_0(t)\\ &= \int_{R^p} (F_0(t + \mathbf{c}'\boldsymbol\gamma) - F_0(t))\, dG(\mathbf{c}).\end{aligned}$$

This clearly exists and $\mu(t; Q) \equiv 0$, for all t. Direct partial differentiation gives

$$\frac{\partial}{\partial \gamma_j}\mu(t;\boldsymbol\gamma) = \int_{R^p} c_j f_0(t + \mathbf{c}'\boldsymbol\gamma)\, dG(\mathbf{c}),$$

and therefore

$$\mathbf{d}_1\mu(t;\mathbf{0}) = f_0(t)(EC_1, EC_2, \dots, EC_p)',$$

provided that these moments are finite. Applying Theorem 2.16, we see that if $E[C_j] = 0$ for $j = 1, \dots, p$, then estimating the parameters has no influence on the null limiting distribution of the test statistic. Otherwise, we can always take $E[C_j] = 0$ for $j = 2, \dots, p$ and $C_1 \equiv 1$ so that the first term is the mean of the Y responses. From $\mathbf{d}\mu_1(\cdot)$ we see that the asymptotic distribution is only affected by the estimator of the mean and not by the other estimators in the $\hat{\boldsymbol{\beta}}$ vector. This is the same conclusion reached by Pierce and Kopecky (1979) for a very broad class of test statistics but with stronger conditions on the estimators $\hat{\boldsymbol{\beta}}$. The exact nature of the change is determined by comparing the eigenvalues of $h_*(x, y)$ in (2.15) to those of $h(x, y; \boldsymbol{\lambda})$ as in (2.4) and will be a function of the intercept estimator used.

To verify the conditions of Theorem 2.16, we need to make further assumptions about the underlying population. Details of the verification are available from the authors.

3.2. *Cramér–von Mises statistic for the location scale case.* In this case we have $X_1, X_2, \dots, X_n$ i.i.d. $F((x - \mu)/\sigma)$ and we wish to test

$$H_0\colon F(t) = F_0(t)$$

with F_0 completely specified. Let $\hat{\boldsymbol{\lambda}} = (\hat{\mu}, \hat{\sigma})$ be a consistent estimator for $\boldsymbol{\lambda} = (\mu, \sigma)$ satisfying Condition 2.10. Let $F_n(t)$ be the empirical distribution function of $X_1, X_2, \dots, X_n$. Then the Cramér–von Mises statistic is given by (see, e.g., Durbin (1973))

$$\begin{aligned} T_n(\hat{\boldsymbol{\lambda}}) &= n\int_{-\infty}^{\infty} (F_n(\hat{\mu} + \hat{\sigma}t) - F_0(t))^2 \, dF_0(t) \\ &= \frac{1}{n} \sum_{i,j=1}^{n} h\big(X_i, X_j; \hat{\boldsymbol{\lambda}}\big), \end{aligned} \tag{3.4}$$

where

$$\begin{aligned} h(x_1, x_2; \boldsymbol{\gamma}) = \int_{-\infty}^{\infty} &(I(x_1 \le \gamma_1 + \gamma_2 t) - F_0(t)) \\ &\times (I(x_2 \le \gamma_1 + \gamma_2 t) - F_0(t)) \, dF_0(t). \end{aligned}$$

In showing that the conditions of Theorem 2.16 hold, we use

$$g(x, t; \boldsymbol{\gamma}) = I(x \le \gamma_1 + \gamma_2 t) - F_0(t)$$

and

$$M(t) = F_0(t).$$

Thus,

$$\mu(t; \boldsymbol{\gamma}) = Eg(X, t; \boldsymbol{\gamma}) = F_0((\gamma_1 + \gamma_2 t - \mu)/\sigma) - F_0(t)$$

and

$$\mathbf{d}_1\mu(t; \boldsymbol{\lambda}) = \sigma^{-1} f_0(t)(1\ t)'$$

and the function h_* of (2.15) can be written easily (depending of course on which estimators $\hat{\mu}$, $\hat{\sigma}$ we use). See also Gregory (1977) for more general results along these lines.

3.3. *χ^2 goodness-of-fit test statistic.* As another example consider the following χ^2 statistic which is a special case of statistics considered by Moore and Spruill (1975). Let $X_1, X_2, \ldots, X_n$ again be i.i.d. $F((x-\mu)/\sigma)$ and suppose we wish to test

$$H_0\colon F = F_0,$$

with F_0 completely specified. Let $\hat{\lambda} = (\hat{\mu}, \hat{\sigma})$ be a consistent estimator for $\lambda = (\mu, \sigma)$ satisfying Condition 2.10 and let $F_n(t)$ be the empirical distribution function of $X_1, X_2, \ldots, X_n$. Put

$$p_j = F_0(b_j) - F_0(b_{j-1})$$

and

$$\hat{p}_j = F_n(\hat{\mu} + b_j\hat{\sigma}) - F_n(\hat{\mu} + b_{j-1}\hat{\sigma})$$

for $j = 1, \ldots, k$, where $-\infty = b_0 < b_1 < \cdots < b_k = \infty$. Form the χ^2 goodness-of-fit test statistic

$$T_n(\hat{\lambda}) = \sum_{j=1}^{k} p_j^{-1}(n\hat{p}_j - np_j)^2. \tag{3.5}$$

This is easily seen to be equal to

$$n^{-1} \sum_{i,j=1}^{n} h(X_i, X_j; \hat{\lambda})$$

with

$$\begin{aligned} h(x, y; \gamma) &= \sum_{j=1}^{k} p_j^{-1}\big(I(\gamma_1 + \gamma_2 b_{j-1} < x \le \gamma_1 + \gamma_2 b_j) - p_j\big) \\ &\quad \times \big(I(\gamma_1 + \gamma_2 b_{j-1} < y \le \gamma_1 + \gamma_2 b_j) - p_j\big) \\ &= \int g(x, \mathbf{t}; \gamma) g(y, \mathbf{t}; \gamma)\, dM(\mathbf{t}), \end{aligned}$$

with M a discrete measure placing mass p_j^{-1} on the point $\mathbf{t}_j = (b_{j-1}, b_j)$, $j = 1, 2, \ldots, k$, and

$$g(x, \mathbf{t}_j; \gamma) = I(\gamma_1 + \gamma_2 b_{j-1} < x \le \gamma_1 + \gamma_2 b_j) - p_j.$$

Take $\mu = 0$ and $\sigma = 1$ without loss of generality. The conditions of Theorem 2.16 can be shown to hold with

$$\mathbf{d}_1\mu(\mathbf{t}_j; \lambda) = \begin{bmatrix} f_0(b_j) - f_0(b_{j-1}) \\ b_j f_0(b_j) - b_{j-1} f_0(b_{j-1}) \end{bmatrix}, \qquad j = 2, \ldots, k-1,$$

$$\mathbf{d}_1\mu(\mathbf{t}_1; \lambda) = \begin{bmatrix} f_0(b_1) \\ b_1 f_0(b_1) \end{bmatrix}$$

and

$$\mathbf{d}_1\mu(\mathbf{t}_k;\boldsymbol{\lambda}) = \begin{bmatrix} -f_0(b_{k-1}) \\ -b_{k-1}f_0(b_{k-1}) \end{bmatrix}.$$

The function h_* for this case can easily be written as

$$h_*(x, y) = \sum_{j=1}^{k} p_j^{-1}\left[I(b_{j-1} < x \le b_j) - p_j + \mathbf{d}_1\mu(\mathbf{t}_j;\boldsymbol{\lambda})'\boldsymbol{\alpha}(x)\right]$$
$$\times\left[I(b_{j-1} < y \le b_j) - p_j + \mathbf{d}_1\mu(\mathbf{t}_j;\boldsymbol{\lambda})'\boldsymbol{\alpha}(y)\right],$$

which clearly satisfies the degeneracy condition. For given F_0 and $\boldsymbol{\alpha}$, the eigenvalues of h_* can be found and the limiting distribution of $T_n(\hat{\boldsymbol{\lambda}})$ written.

REMARK. In the preceding examples the emphasis was on Case II statistics, the case most often encountered. The next two examples give Case I situations, i.e., where estimating the parameter does not change the limiting distribution.

3.4. *Test for independence in two-sample regression with random regressors.* Consider the model

$$X_{1i} = \mathbf{C}_{1i}'\boldsymbol{\beta} + E_{1i},$$
$$X_{2i} = \mathbf{C}_{2i}'\boldsymbol{\beta} + E_{2i}, \qquad i = 1, 2, \dots, n,$$

with $\{(\mathbf{C}_{1i}, \mathbf{C}_{2i})\}$ i.i.d. $2p$-dimensional random vectors with joint distribution G and $\boldsymbol{\beta}$, the vector of unknown regression constants. Assume (E_{1i}, E_{2i}) are i.i.d. with distribution $F(t_1, t_2)$ and marginals F_1, F_2. We assume the $\mathbf{C}$'s are independent of the E_1's and E_2's. Suppose we wish to test the independence of E_1 and E_2, i.e.,

$$H_0\colon F(t_1, t_2) = F_1(t_1)F_2(t_2), \quad \text{all } t_1, t_2.$$

Estimate $\boldsymbol{\beta}$ consistently by $\hat{\boldsymbol{\beta}}$, assumed to satisfy Condition 2.10, and form the residuals

$$Y_{1i} = X_{1i} - \mathbf{C}_{1i}'\hat{\boldsymbol{\beta}}, \qquad Y_{2i} = X_{2i} - \mathbf{C}_{2i}'\hat{\boldsymbol{\beta}}.$$

As a test statistic for H_0 we can use the Cramér–von Mises-type statistic

$$T_n(\hat{\boldsymbol{\beta}}) = n\int_{R^2}[F_n(t_1, t_2) - F_{1n}(t_1)F_{2n}(t_2)]^2\, dF_{1n}(t_1)\, dF_{2n}(t_2),$$

with F_n, F_{1n} and F_{2n}, respectively, the empirical distribution functions of $\{(Y_{1i}, Y_{2i})\}$, $\{Y_{1i}\}$ and $\{Y_{2i}\}$. (See, e.g., de Wet (1980) and Randles (1984), where statistics of this nature were previously studied.)

Assume throughout that H_0 is true. As in de Wet (1980) one can show that

$$T_n(\hat{\boldsymbol{\beta}}) - T_{1n}(\hat{\boldsymbol{\beta}}) \to_p 0 \quad \text{as } n \to \infty,$$

where

$$T_{1n}(\hat{\boldsymbol{\beta}}) = n\int_{R^2}[F_n(t_1, t_2) - F_1(t_1)F_{2n}(t_2) - F_{1n}(t_1)F_2(t_2)$$
$$+ F_1(t_1)F_2(t_2)]^2\, dF_1(t_1)\, dF_2(t_2).$$

Substituting for the empirical distribution functions and changing the order of integration and summation, we get

$$T_{1n}(\hat{\beta}) = n^{-1} \sum_{i,j=1}^{n} h(\mathbf{Z}_i, \mathbf{Z}_j; \hat{\beta}),$$

where

$$\mathbf{Z} = (\mathbf{C}_1, \mathbf{C}_2, E_1, E_2) \equiv (\mathbf{C}, \mathbf{E})$$

and

$$h(\mathbf{z}_1, \mathbf{z}_2; \gamma) = \int_{R^2} g(\mathbf{z}_1, \mathbf{t}; \gamma) g(\mathbf{z}_2, \mathbf{t}; \gamma)\, dF_1(t_1)\, dF_2(t_2),$$

with $\mathbf{t} = (t_1, t_2)'$ and

$$g(\mathbf{z}, \mathbf{t}; \gamma) = \prod_{k=1}^{2} \left[I(e_k - \mathbf{c}_k'(\gamma - \beta) \le t_k) - F_k(t_k) \right].$$

The conditions of Theorem 2.16 can be shown to hold in this case using

$$h_*(\mathbf{z}_1, \mathbf{z}_2) = \int_{R^2} [g(\mathbf{z}_1, \mathbf{t}, \beta) + \mathbf{d}_1 \mu(\mathbf{t}; \beta)' \alpha(\mathbf{z}_1)] \times [g(\mathbf{z}_2, \mathbf{t}, \beta) + \mathbf{d}_1 \mu(\mathbf{t}, \beta)' \alpha(\mathbf{z}_2)]\, dM(\mathbf{t})$$

and

$$\mu(\mathbf{t}; \gamma) = E_\beta[g(\mathbf{Z}, \mathbf{t}; \gamma)] = \int_{R^{2p}} \prod_{k=1}^{2} [F_k(t_k + \mathbf{c}_k'(\gamma - \beta)) - F_k(t_k)]\, dG(\mathbf{c}).$$

From this it follows directly that

$$\mathbf{d}_1 \mu(\mathbf{t}, \beta) \equiv \mathbf{0}$$

and thus $h_* = h$, giving the same limiting distribution whether the parameter is known or estimated.

3.5. *A multivariate sign test.* Suppose we observe

$$\mathbf{X}_i = \mu + \mathbf{Z}_i,$$

where $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ are i.i.d. p vectors, each with an elliptically symmetric distribution with mean vector $\mathbf{0}$ and variance–covariance matrix Σ. Suppose also that $E[Z_{j1}^4] < \infty$ for $j = 1, \dots, p$ where Z_{j1} is the jth component of $\mathbf{Z}_1$. The p-vector μ is a location parameter. Without loss of generality we test

$$H_0: \mu = \mathbf{0} \quad \text{versus} \quad H_a: \mu \ne \mathbf{0}.$$

Let $\mathbf{S}$ denote the sample variance–covariance matrix and form

$$\mathbf{S}^{-1} = \mathbf{A}_n' \mathbf{A}_n,$$

where $\mathbf{A}_n$ is a lower triangular $p \times p$ matrix with positive diagonal elements.

Letting $\mathbf{A}_{tn}$ denote the tth row of $\mathbf{A}_n$, a possible sign test statistic is

$$(3.6)\quad \begin{aligned} V_n(\mathbf{A}_n) &= \sum_{t=1}^{p}\left[n^{-1}\sum_{i=1}^{n}\operatorname{sgn}(\mathbf{A}_{tn}\mathbf{X}_i)\right]^2 \\ &= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{p}\operatorname{sgn}(\mathbf{A}_{tn}\mathbf{X}_i)\operatorname{sgn}(\mathbf{A}_{tn}\mathbf{X}_j). \end{aligned}$$

Therefore,

$$g(\mathbf{x}, t;\gamma) = \operatorname{sgn}(\gamma_t\mathbf{x}), \qquad \mu(t;\gamma) = E[\operatorname{sgn}(\gamma_t\mathbf{X}_1)]$$

and $M(t)$ is the measure that places mass 1 on each of the integers $1, \ldots, p$. Here γ denotes a $p \times p$ matrix with tth row γ_t. For example, when $p = 3$, $V_n(\mathbf{A}_n)$ examines each pair of transformed observations $(\mathbf{A}_n\mathbf{X}_i, \mathbf{A}_n\mathbf{X}_j)$ and contributes $+6$ to the sum in $V_n(\mathbf{A}_n)$ if these two vectors are in the same octant, $+2$ if they are in adjacent octants (ones which share a common face), -6 if they are in opposite octants and -2 otherwise. We view $nV_n(\mathbf{A}_n)$ as a sign test statistic analogue to Hotelling's T^2, which may be written

$$T^2 = n\sum_{t=1}^{p}\left[n^{-1}\sum_{i=1}^{n}\mathbf{A}_{tn}\mathbf{X}_i\right]^2.$$

Theorem 2.16 will be used to establish the null hypothesis limiting distribution of $nV_n(\mathbf{A}_n)$. Under H_0, the mean function

$$\mu(t;\gamma) = E[\operatorname{sgn}(\gamma_t\mathbf{Z}_1)] = 0$$

for every γ and $t = 1, \ldots, p$, and hence

$$\mathbf{d}_1\mu(t;\mathbf{0}) = \mathbf{0}.$$

We also note that $\sqrt{n}(\mathbf{A}_n - \mathbf{A})$ is bounded in probability, where $\mathbf{A}$ is the lower triangular factor satisfying $\mathbf{A}'\mathbf{A} = \mathbf{\Sigma}^{-1}$. Theorem 2.16 shows that $nV_n(\mathbf{A}_n)$ has the same null limiting distribution as $nV_n(\mathbf{A})$. From this it readily follows that

$$nV_n(\mathbf{A}_n) \to_d \chi_p^2$$

under H_0. For other multivariate sign tests see Killeen and Hettmansperger (1972), Dietz (1982) and their references.

4. Proofs of the results. We now sketch the proofs of the results stated in Section 2. More detailed proofs are available from the authors. For this we need the following lemma, the proof of which is straightforward.

LEMMA 4.1. *Let $X_1, \ldots, X_n$ be i.i.d. and suppose $k_n(x, y) = k_n(y, x)$ for every x, y and n. In addition, assume for every x and n that*

$$E[k_n(x, X_2)] = 0$$

and

$$(4.2)\qquad E[k_n^2(X_1, X_i)] = o(n^2)$$

for $i = 1$ and 2, as $n \to \infty$. Then

$$W_n = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} k_n(X_i, X_j) \to_p 0.$$

PROOF OF THEOREM 2.16. First note that nV_n is a degenerate degree-2 V statistic and Theorem 6.4.1B in Serfling shows

$$nV_n \to_d \sum_{k=1}^{\infty} \delta_k \chi_{1k}^2. \tag{4.3}$$

Define

$$Y_n = \int_{-\infty}^{\infty} \left[n^{-1/2} \sum_{i=1}^{n} \{ g(X_i, t; \lambda) + \mu(t; \hat{\lambda}) \} \right]^2 dM(t).$$

By applying Holder's and Jensen's inequalities one can show that

$$Y_n - nV_n \to_p 0. \tag{4.4}$$

Now

$$\begin{aligned} nV_n(\hat{\lambda}) - Y_n &= \int_{-\infty}^{\infty} \left[n^{-1/2} \sum_{i=1}^{n} \{ g(X_i, t; \hat{\lambda}) - g(X_i, t; \lambda) - \mu(t, \hat{\lambda}) \} \right]^2 dM(t) \\ &\quad + 2 \int_{-\infty}^{\infty} \left[n^{-1/2} \sum_{i=1}^{n} \{ g(X_i, t; \lambda) + \mu(t; \hat{\lambda}) \} \right] \\ &\quad \times \left[n^{-1/2} \sum_{i=1}^{n} \{ g(X_i, t; \hat{\lambda}) - g(X_i, t; \lambda) - \mu(t; \hat{\lambda}) \} \right] dM(t) \\ &\equiv T_{3n} + 2T_{4n}. \end{aligned}$$

But Holder's inequality shows that

$$T_{4n}^2 \le Y_n T_{3n}$$

and hence

$$nV_n(\hat{\lambda}) - Y_n \to_p 0, \tag{4.5}$$

provided

$$T_{3n} \to_p 0. \tag{4.6}$$

To consider T_{3n}, we define

$$\begin{aligned} Q_n(\mathbf{s}) = \int_{-\infty}^{\infty} \Bigg[n^{-1/2} \sum_{i=1}^{n} \{ & g(X_i, t; \lambda + n^{-1/2}\mathbf{s}) \\ & - g(X_i, t; \lambda) - \mu(t; \lambda + n^{-1/2}\mathbf{s}) \} \Bigg]^2 dM(t) \end{aligned}$$

and seek to show $Q_n(\sqrt{n}(\hat{\lambda} - \lambda)) \to_p 0$. This last step follows from Lemma 4.1 in much the same fashion as Theorem 3.1 of Sukhatme (1958). □

PROOF OF THEOREM 2.19. Write

$$n[U_n(\hat{\lambda}) - U_n] = n^2(n-1)^{-1}[V_n(\hat{\lambda}) - V_n] + (n-1)^{-1} \times \sum_{i=1}^{n} \int_{-\infty}^{\infty} \left[\{g(X_i, t; \lambda) + \mathbf{d}_1\mu(t;\lambda)'\alpha(X_i)\}^2 - g^2(X_i, t; \hat{\lambda})\right] dM(t)$$

$$\equiv n(n-1)^{-1}\{n[V_n(\hat{\lambda}) - V_n] + U_{1n} - U_{2n}(\hat{\lambda})\}.$$

Theorem 2.16 shows that $n[V_n(\hat{\lambda}) - V_n] \to_p 0$. Condition 2.17 and the SLLN show that $U_{1n} \to_p \Sigma_{k=1}^{\infty}\delta_k$. Theorem 2.11 in Iverson and Randles (1983) shows that $U_{2n}(\hat{\lambda}) \to_p \Sigma_{k=1}^{\infty}\delta_k'$. The degenerate U-statistic limit theorem in Gregory (1977) or Theorem 5.5.2 in Serfling (1980) shows that $nU_n \to_d \Sigma_{k=1}^{\infty}\delta_k(\chi_{1k}^2 - 1)$, and hence the conclusion follows. □

Acknowledgment. The authors would like to thank D. Boos for suggesting Example 3.4.

REFERENCES

DE WET, T. (1980). Cramér–von Mises tests for independence. *J. Multivariate Anal.* **10** 38–50.

DE WET, T. and VENTER, J. H. (1972). Asymptotic distributions of certain test criteria of normality. *South African Statist. J.* **6** 135–149.

DE WET, T. and VENTER, J. H. (1973). Asymptotic distributions for quadratic forms with applications to tests of fit. *Ann. Statist.* **1** 380–387.

DE WET, T. and VENTER, J. H. (1976). Asymptotic distributions of quadratic functionals of partial sums. *South African Statist. J.* **10** 97–105.

DE WET, T., VENTER, J. H. and VAN WYK, J. W. J. (1979). The null distributions of some test criteria of multivariate normality. *South African Statist. J.* **13** 153–176.

DIETZ, E. J. (1982). Bivariate nonparametric tests for the one-sample location problem. *J. Amer. Statist. Assoc.* **77** 163–169.

DURBIN, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function.* SIAM, Philadelphia.

FLIGNER, M. A. and HETTMANSPERGER, T. P. (1979). On the use of conditional asymptotic normality. *J. Roy. Statist. Soc. Ser. B* **41** 178–183.

GREGORY, G. G. (1977). Large sample theory for U-statistics and tests of fit. *Ann. Statist.* **5** 110–123.

IVERSON, H. K. and RANDLES, R. H. (1983). Convergence results for statistics with estimated parameters. Technical Report No. 213, Dept. of Statist., Univ. of Florida.

KILLEEN, T. J. and HETTMANSPERGER, T. P. (1972). Bivariate tests for location and their Bahadur efficiencies. *Ann. Math. Statist.* **43** 1507–1516.

MOORE, D. S. and SPRUILL, M. C. (1975). Unified large sample theory of general chi-squared statistics for tests of fit. *Ann. Statist.* **3** 599–616.

NEUHAUS, G. (1977). Functional limit theorems for U-statistics in the degenerate case. *J. Multivariate Anal.* **7** 424–439.

PARR, W. C. (1982). A note on adaptive L-statistics. *Comm. Statist. A—Theory Methods* **11** 1511–1518.

PIERCE, D. A. (1982). The asymptotic effects of substituting estimators for parameters in certain types of statistics. *Ann. Statist.* **10** 475–478.

PIERCE, D. A. and KOPECKY, K. J. (1979). Testing goodness-of-fit for the distribution of errors in regression models. *Biometrika* **66** 1–5.

RANDLES, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *Ann. Statist.* **10** 462–474.

RANDLES, R. H. (1984). On tests applied to residuals. *J. Amer. Statist. Assoc.* **79** 349–354.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

SUKHATME, B. V. (1958). Testing the hypothesis that two populations differ only in location. *Ann. Math. Statist.* **29** 60–78.

INSTITUTE FOR MARITIME TECHNOLOGY
P.O. BOX 181
SIMONSTOWN 7995
SOUTH AFRICA

DEPARTMENT OF STATISTICS
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611