The Annals of Statistics
1980, Vol. 8, No. 5, 1017–1022

TESTS BASED ON LINEAR COMBINATIONS OF THE ORTHOGONAL COMPONENTS OF THE CRAMÉR-VON MISES STATISTIC WHEN PARAMETERS ARE ESTIMATED[1]

BY DAVID SCHOENFELD

Harvard School of Public Health and Sidney Farber Cancer Inst.

In a previous work, the author showed how linear combinations of the orthogonal components of the Cramér-von Mises statistic could be used to test fit to a fully specified distribution function. In this paper, the results are extended to the case where r parameters are estimated from the data. It is shown that if the coefficient vector of the linear combination is orthogonal to a specified r dimensional subspace, then the asymptotic distribution of that combination is the same whether the parameters are estimated or known exactly.

1. Introduction. The orthogonal components of the Cramér-von Mises statistic were introduced by Durbin and Knott (1972) to test goodness of fit to a completely specified distribution function. Schoenfeld (1977) examined the asymptotic properties of linear combinations of a generalization of these components. He defined a contiguous family of alternative distributions and showed that for any member of this family an asymptotically most powerful test could be found based on a linear combination of the components. The asymptotic power and efficiency of these tests were shown to have simple expressions.

This paper extends Schoenfeld's results to the case where r parameters of the hypothetical distribution function are estimated from the data. The components can be computed using estimates of the parameters. If the coefficient vector of a linear combination of the components is orthogonal to a specified r dimensional subspace, then the asymptotic distribution of that combination is the same as if the true parameter values were used. This theory is applied to the case where location and scale parameters are unknown.

2. Definitions and earlier results. Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables which under the null hypothesis has a known distribution function $F(X, \boldsymbol{\theta})$ that depends on an r-dimensional parameter $\boldsymbol{\theta}$. If $\boldsymbol{\theta}$ has a known value $\boldsymbol{\theta}_0$, we let $U_i = F(X_i, \boldsymbol{\theta}_0)$. The null hypothesis is that $\{U_i\}_{i=1,n}$ have a uniform distribution. Let $\{1, d_1, d_2, \dots\}$ be a specified orthonormal basis for the space of square integrable functions on the unit interval. The jth generalized orthogonal component is defined by

$$V_{nj} = V_{nj}(U_1, U_2, \cdots, U_n) = n^{-\frac{1}{2}}\Sigma_{i=1}^{n} d_j(U_i). \tag{2.1}$$

Received October 1978; revised March 1979.

[1]Partial support was provided by grants NSF-GP32052X2, NCI-CA23415-02, and NCI-CA-23316-02.

AMS 1970 *subject classifications.* Primary 62G10; secondary 62G20.

Key words and phrases. Orthogonal components, Cramér-von Mises statistic, tests of composite hypothesis.

The orthogonal components as defined by Durbin and Knott have $d_j(u) = 2^{\frac{1}{2}} \cos j\pi u$.

Schoenfeld considered a sequence of alternative densities of the form

$$p_n(u) = 1 + h(u)/n^{\frac{1}{2}} + k_n(u)/n \tag{2.2}$$

where $h(u)$ is square integrable and $k_n(u)$ is dominated by a function that is square integrable. Schoenfeld showed that if $c_j = \int h(u)d_j(u)du$ and if $m \to \infty$ as $n \to \infty$, then the test which rejects when

$$\Sigma_{j=1}^{m} c_j V_{nj} > K$$

is asymptotically most powerful. The test based on $\Sigma_{j=1}^{m} b_j V_{nj}$, with $\Sigma_{j=1}^{m} a_j b_j > 0$, has relative asymptotic efficiency

$$(\Sigma_{j=1}^{m} c_j b_j)^2 / (\Sigma_{j=1}^{\infty} c_j^2)(\Sigma_{j=1}^{m} b_j^2).$$

If $\boldsymbol{\theta}$ is unknown and one has an estimator, $\hat{\boldsymbol{\theta}}$, based on $X_1, X_2, \cdots, X_n$ one can compute the components substituting $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$. Define

$$\hat{U}_i = F(X_i, \hat{\boldsymbol{\theta}})$$

and

$$\hat{V}_{nj} = n^{-\frac{1}{2}} \Sigma_{i=1}^{n} d_j(\hat{U}_i). \tag{2.3}$$

3. Results. The main theorem of this paper shows that when $\{b\}_{j=1}, {}_m$ are chosen orthogonal to a specified r dimensional subspace, then tests based on $\Sigma_{j=1}^{m} b_j \hat{V}_{nj}$ will be asymptotically equivalent to those based on $\Sigma_{j=1}^{m} b_j V_{nj}$. In this case all the results of the earlier paper can be applied.

THEOREM 1. *Let V_{nj} and $\hat{V}_{nj}$ be defined as in* (2.1) *and* (2.3). *Let $F_0(x) = F(x, \boldsymbol{\theta}_0)$ and $f(x, \boldsymbol{\theta}) = d/dx\, F(x, \boldsymbol{\theta})$ and $F_0^{-1}(u) = \inf\{x: F_0(x) = u\}$. Define*

$$a_{ji} = \int_0^1 d_j(u) \frac{\partial}{\partial \theta_i} \log f(F_0^{-1}(u), \boldsymbol{\theta})\,|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} du.$$

Let $\{b_j\}_{j=1}, m$ be m real numbers. If, for $i = 1, 2, \cdots, r$,

$$\Sigma_{j=1}^{m} b_j a_{ji} = 0, \tag{3.1}$$

then

$$\Sigma_{j=1}^{m} b_j \hat{V}_{nj} \to_p \Sigma_{j=1}^{m} b_j V_{nj}.$$

The following regularity conditions are necessary:

(i) *The distribution function $F(x, \boldsymbol{\theta})$ is continuous in x for all θ.*

(ii) *Let $F_0(x) = F(x, \boldsymbol{\theta}_0)$. Then the functions*

$$g_i(u, \boldsymbol{\theta}_0) = \frac{\partial}{\partial \theta_i} F(F_0^{-1}(u), \boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

are continuous in $u \in [0, 1]$ with

$$d_j(1) g_i(1, \boldsymbol{\theta}_0) = d_j(0) g_i(0, \boldsymbol{\theta}_0). \tag{3.2}$$

(iii) *The following formula holds*:

$$\frac{\partial^2}{\partial u \partial \theta_i} F(F_0^{-1}(u), \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{\partial^2}{\partial \theta_i \partial u} F(F_0^{-1}(u), \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

(iv) *The estimator* $\hat{\boldsymbol{\theta}}_n$ *obeys the following condition, satisfied by first order efficient estimators* (*Rao*, 1973, *page* 348):

$$n^{\frac{1}{2}}(\hat{\theta}_{ni} - \theta_{0i}) \to_p n^{-\frac{1}{2}} \Sigma_{j=1}^n l_i(X_j, \boldsymbol{\theta}_0)$$

where $E[l_i(X, \boldsymbol{\theta}_0)] = 0$, $E[l_i(X, \boldsymbol{\theta}_0) l_j(X, \boldsymbol{\theta}_0)] < \infty$ *and this matrix is nonnegative definite.*

(v) *The functions* $d_j(u)$ *are bounded.*

Equation (3.2) holds whenever $g_i(1, \boldsymbol{\theta}_0) = g_i(0, \boldsymbol{\theta}_0) = 0$ which is a requirement that $F(x, \boldsymbol{\theta})$ is well behaved near $\pm\infty$. Condition (iii) holds whenever integration and differentiation can be interchanged. The condition that $d_j(u)$ be bounded is satisfied for the components suggested by Durbin and Knott, as well as for many other commonly used orthonormal bases. Other regularity conditions on $d_j(u)$ and $g_i(u, \boldsymbol{\theta}_0)$ might be substituted for the boundedness at $d_j(u)$ if one used a more elementary proof based on the Taylor expansion of $\hat{U}_{nj}$ about $\boldsymbol{\theta}$. These conditions, however, might be harder to verify. See Neymann (1959).

PROOF. Let $F_n(u)$ be the empirical distribution function of $\{U_1, U_2, \cdots, U_n\}$ and let $\hat{F}_n(u)$ be defined similarly using $\{\hat{U}_1, \hat{U}_2, \cdots, \hat{U}_n\}$. The condition (i), (iv) and the fact that $g_i(u, \theta_0)$ is continuous for all $u \in [0, 1]$ imply that the results of Durbin (1973) hold. In particular, using Lemma 2 with $\gamma = 0$, one has that

$$n^{\frac{1}{2}}[F_n(u) - u] = n^{\frac{1}{2}}[\hat{F}_n(u) - u] + \Sigma_{i=1}^r W_{ni} g_i(u, \boldsymbol{\theta}_0) + E_n \tag{3.3}$$

where $W_{n1} \ldots W_{nr}$ are asymptotically multivariate normal and $E_n \to_p 0$ under the null hypothesis that $X_1, X_2, \ldots X_n$ have cdf $F(X, \boldsymbol{\theta}_0)$.

Let

$$s(u) = \Sigma_{j=1}^m b_j d_j(u),$$

and let $\int_0^1 ds$ be the signed Riemann-Stieltjes integral induced by s. This is a linear functional on the space of integrable functions on [0, 1]. When this functional is applied to (3.3), that is, when we integrate (3.3) with respect to ds, the first term is

$$-\Sigma_{j=1}^m b_j V_{nj},$$

the second term is

$$-\Sigma_{j=1}^m b_j \hat{V}_{nj},$$

the third term is

$$-\Sigma_{i=1}^r W_{ni}(\Sigma_{j=1}^m b_j a_{ji}),$$

and the fourth term goes to zero in probability.

Using integration by parts and noting that $d_j(u)$ have zero integrals one can show that

$$\int_0^1 u\,ds = s(1).$$

The integral of the first term in (3.3) is

$$n^{-\frac{1}{2}}\sum_{i=1}^{n}\big(i\,s(U_{(i+1)}) - is(U_{(i)})\big) - n^{\frac{1}{2}}s(1)$$

where $0 = U_0 < U_{(1)} < U_{(2)}, \cdots, < U_{(n)} < U_{(n+1)} = 1$ and $\{U_{(i)}\}_{i=1,n}$ are the order statistics of $\{U_i\}_{i=1,n}$. This expression collapses to

$$-\,n^{-\frac{1}{2}}\sum_{i=1}^{n}s(U_i).$$

However

$$-\,n^{-\frac{1}{2}}\sum_{i=1}^{n}s(U_i) = -n^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{j=1}^{m}b_j d_j(U_i) = -\sum_{j=1}^{m}b_jV_{nj}.$$

The argument for the second term in (3.3) is exactly the same.

Now, use integration by parts and (3.2) to show that

$$\int_0^1 g_i(u,\boldsymbol{\theta}_0)\,ds(u) = -\int_0^1 s(u)\frac{\partial}{\partial u}g_i(u,\boldsymbol{\theta}_0)\,du.$$

However, (iii) implies that

$$\frac{\partial}{\partial u}\frac{\partial}{\partial \theta_i}F(F_0^{-1}(u),\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{\partial}{\partial \theta_i}f(F_0^{-1}(u),\boldsymbol{\theta})\big[f(F_0^{-1}(u),\boldsymbol{\theta}_0)\big]^{-1}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

$$= \frac{\partial}{\partial \theta_i}\log f(F_0^{-1}(u),\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

Therefore

$$\int_0^1 g_i(u,\boldsymbol{\theta}_0)\,ds(u) = -\int_0^1\Big(\sum_{j=1}^{m}b_jd_j(u)\Big)\frac{\partial}{\partial \theta_i}\log f(F_0^{-1}(u),\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}du$$

$$= -\sum_{j=1}^{m}b_ja_{ji}.$$

Since $\int_0^1 ds$ is a bounded linear functional it is continuous in the uniform metric and hence continuous at zero in the Skorohod metric (Billingsley 1968). Therefore $\int_0^1 E_n\,ds \to_p 0$.

Therefore the condition that

$$\sum_{j=1}^{m}b_ja_{ji} = 0$$

implies that under the null hypothesis

$$\sum_{j=1}^{m}b_j\hat{V}_{nj} \to_p \sum_{j=1}^{m}b_jV_{nj}.$$

COROLLARY 1. *Suppose that $F_n(X,\boldsymbol{\theta})$ is a sequence of alternative distributions to $F(X,\theta)$. Suppose further that $\frac{\partial}{\partial u}F_n(F_0^{-1}(u),\boldsymbol{\theta}_0)$ can be expressed as a sequence $1 + h(u)/n^{\frac{1}{2}} + k_n(u)/n$ with $h(u)$ square integrable and $k_n(u)$ dominated by a square integrable function. Then, if the conditions of the main theorem hold,*

$$\sum_{j=1}^{m}b_j\hat{V}_{nj} \to_p \sum_{j=1}^{m}b_jV_{nj}$$

when $X_1, X_2, \cdots, X_n$ have the distribution $F_n(X,\theta_0)$.

PROOF. This follows by the fact that the sequence of distributions induced by this sequence of densities is contiguous to the sequence of uniform distributions on the unit n-cube. See Schoenfeld (1977).

4. Example: Location and scale parameters. Let $G((x - \theta_1)/\theta_2, \gamma)$ be a three parameter family of distributions with θ_1 and θ_2 the location and scale parameter. The parameter γ is a shape parameter which distinguishes the null hypothesis, $\gamma = 0$, from the alternative hypotheses. For instance, if the null hypothesis is the condition that the sample has the normal distribution function Φ, then

$$G((X - \theta_1)/\theta_2, \gamma) = \Phi((X - \theta_1)/\theta_2 + \gamma)/2 + \Phi((X - \theta_1)/\theta_2 - \gamma)/2$$

represents a bimodel alternative.

We must first transform the hypothesis to a sequence of alternatives to the uniform distribution in the form (2.2). Furthermore, we must compute a_{j1} and a_{j2} and then use them to find the appropriate test.

Use $G_0((X - \theta_1)/\theta_2)$ to denote G as a function of $(X - \theta_1)/\theta_2$ when $\gamma = 0$. Let $u = G_0((X - \theta_1)/\theta_2)$, then the distribution function of u when $\gamma \neq 0$ is

$$G(G_0^{-1}(u), \gamma)$$

which does not contain θ_1 or θ_2. Let $\gamma_n = \alpha/n^{\frac{1}{2}}$ be a sequence of alternative values of γ. The sequence of alternative distributions can be put in form (2.2) using Taylor's theorem to expand G about $\gamma = 0$. In this case, letting $g(z, \gamma) = \frac{\partial}{\partial z} G(z, \gamma)$,

$$h(u) = \alpha \frac{\partial}{\partial \gamma} \left(\log g(G_0^{-1}(u), \gamma)\right)\Big|_{\gamma = 0}.$$

It suffices to show that $h(u)$ is square integrable and the remainder term is dominated by a square integrable function.

Let $g'(z, \gamma) = \frac{\partial}{\partial z} g(z, \gamma)$ and let $J(u) = g'(G_0^{-1}(u), 0)/g(G_0^{-1}(u), 0)$. Then

$$a_{j1} = -\frac{1}{\theta_2} \int_0^1 J(u) d_j(u) du$$

$$a_{j2} = -\frac{1}{\theta_2} \int_0^1 [1 + uJ(u)] d_j(u) du.$$

Notice that $J(u)$ does not contain θ_1 or θ_2 so that a_{j1}, a_{j2} are parameter free except for a constant multiplier. We can evaluate a_{j1} and a_{j2} at $\theta_2 = 1$ without changing the set of b_j for which (3.1) holds.

The test based on $c_j = \int_0^1 d_j(u) h(u) du$ may not in general satisfy (3.1). However, if we let $\{c_j'\}_{j=1,m}$ be the projection of $\{c_j\}_{j=1,m}$ on the subspace orthogonal to $\{a_{j1}\}_{j=1,m}$ and $\{a_{j2}\}_{j=1,m}$, then the test based on $\{c_j'\}_{j=1,m}$ will be asymptotically parameter free and have the highest efficiency of those tests based on a linear combination of m components which is orthogonal to $\{a_{j1}\}_{j=1,m}$ and $\{a_{j2}\}_{j=1,m}$. The properties of this test are described in a previous paper (Schoenfeld 1977).

The vectors $\{a_{j1}\}_{j=1,m}$ and $\{a_{j2}\}_{j=1,m}$ are the same vectors that would be used for tests of location and scale shift alternatives to a simple null hypothesis. The results essentially show that to test for a compound hypothesis one must choose tests "perpendicular" to the tests which would distinguish different null hypothesis distributions.

5. Discussion. Durbin, Knott and Taylor (1975) also investigate the components when parameters are estimated from the data. With $d_j(u) = 2^{\frac{1}{2}} \cos j\pi u$ they show that asymptotically, the vector $\{\hat{V}_{nj}\}$ in the space of square summable sequences, is the projection of the vector $\{V_{nj}\}$ onto the subspace orthogonal to $\{a_{ji}\}_{j=1,\infty i=1,r}$. They then find new components which are asymptotically parameter free. Each of these new components is a function of all the $\{V_{nj}\}$. An alternate approach to extending the results of Schoenfeld (1977) would be to apply them to these redefined components.

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measure.* Wiley, New York.

DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1** 279–290.

DURBIN, J., AND KNOTT, [illegible]. (1972). Components of the Cramér-von Mises statistic I. *J. Roy. Statist. Soc.* Series B **34** 290–307.

DURBIN, J., KNOTT, M., AND TAYLOR, L. C. (1975). Components of the Cramér-von Mises statistic II. *J. Roy. Statist. Soc.* Series B **37** 216–237.

RAO, C.R. (1973). *Linear Statistical Inference and Its Applications.* Wiley, New York.

SCHOENFELD, D. (1977). Asymptotic properties of tests based on linear combinations of the orthogonal components of the Cramér-von Mises statistic. *Ann. Statist.* **5** 1017–1026.

DEPARTMENT OF BIOSTATISTICS
HARVARD SCHOOL OF PUBLIC HEALTH
677 HUNTINGTON AVENUE
BOSTON, MASSACHUSETTS 02115