The Annals of Statistics
1980, Vol. 8, No. 2, 285–298

THE EMPIRICAL DISTRIBUTION FUNCTION OF RESIDUALS FROM GENERALISED REGRESSION

BY R. M. LOYNES

University of Sheffield

Generalised residuals, as defined by Cox and Snell, may be thought of as residuals from generalised regression. Under regularity conditions on the regression and on the estimator of the unknown parameters, the asymptotic behaviour of the empirical distribution of these residuals is determined. The addition of a random adjustment to the maximum likelihood estimator leads to the familiar Brownian bridge as the limit.

1. Introduction. The use of (estimated) residuals for the general linear regression model

$$\mathbf{X} = B\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{1.1}$$

is well known: defined as $\mathbf{e} = \mathbf{X} - B\tilde{\boldsymbol{\theta}}$, where $\tilde{\boldsymbol{\theta}}$ is an estimator of $\boldsymbol{\theta}$ (usually the least squares estimator), they are used to test the adequacy of the model, and, in case it proves to be not adequate, to suggest modifications.

There are many situations in which the model (1.1) is inappropriate, however, and a more general model is called for. Cox and Snell (1968) defined 'generalised residuals' for the model

$$X_i = g_i(\varepsilon_i, \boldsymbol{\theta}), \qquad 1 \leqslant i \leqslant n \tag{1.2}$$

by writing

$$e_i = h_i(X_i, \tilde{\boldsymbol{\theta}}), \tag{1.3}$$

where h_i is supposed uniquely defined by the requirement that $X_i = g_i(\varepsilon_i, \boldsymbol{\theta})$ if and only if $\varepsilon_i = h_i(X_i, \boldsymbol{\theta})$, and, taking $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}$ the maximum likelihood estimator (MLE), found approximations valid for large n to the distribution of the individual e_i; see also Loynes (1969). In (1.2) the ε_i are assumed independent and identically distributed and for this reason we regard the model as defining a generalised regression. There is still a lack of precision in the description however, for we may think of the distribution of ε_i as containing further unknown parameters, such as the variance in model (1.1) with the usual assumptions, or as being completely specified. We shall assume the latter, taking any unknown parameter into the $\boldsymbol{\theta}$, and consider elsewhere the other possibility. With this understanding the model

Received September 1977; revised November 1978.

AMS 1970 *subject classifications.* Primary 62E20, 62G10; secondary 60F05.

Key words and phrases. Generalised residuals, empirical distribution function, general nonlinear regression, random parameter adjustment.

(1.2) is essentially equivalent to the assumption

(1.4) X_i independent with distribution functions $F_i(x, \boldsymbol{\theta})$,

for suitable F_i: that (1.2) implies (1.4) is obvious, and the converse is a consequence of the fact that under weak assumptions $X_i = F_i^{-1}(U_i, \boldsymbol{\theta})$, where $U_i = F_i(X_i, \boldsymbol{\theta})$ are independent variables, uniform on [0, 1]. Again using the probability integral transform and its inverse, if the distribution of ε_i in (1.2) is assumed completely known, it may as well be assumed uniform on [0, 1], and then the definition (1.3) leads to

$$e_i = F_i(X_i, \tilde{\boldsymbol{\theta}}) \tag{1.5}$$

where $F_i(x, \boldsymbol{\theta})$ is the distribution function of X_i, as in (1.4).

In this paper our interest is in the use of the e_i in a test of goodness-of-fit of a model, and in particular in tests based on the empirical distribution function of the e_i. Plainly explicit, exact, results are not to be expected in any such general context, at least for finite n, and we shall in fact prove a limit theorem valid as $n \to \infty$; for this the results of Cox and Snell are not sufficient, since we need some knowledge of the joint distribution of the e_i. In fact we change the details slightly in order to allow consideration of certain alternative hypotheses. We follow Durbin (1973) by assuming that $\boldsymbol{\theta} = [\boldsymbol{\theta}_1', \boldsymbol{\theta}_2']'$, where $\boldsymbol{\theta}_1$ is a vector of p_1 parameters, $\boldsymbol{\theta}_2$ a vector of p_2 parameters; $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$ is the null hypothesis, and $\boldsymbol{\theta}_2$ consists of nuisance parameters, that are estimated from the data, and then $\tilde{\boldsymbol{\theta}}_n = [\boldsymbol{\theta}_{10}', \tilde{\boldsymbol{\theta}}_{2n}']'$. Moreover we consider, for a given γ, the sequence of hypotheses $H_n(\gamma)$: $\boldsymbol{\theta} = \boldsymbol{\theta}_n = [\boldsymbol{\theta}_{1n}', \boldsymbol{\theta}_{20}']'$, where $\boldsymbol{\theta}_{1n} = \boldsymbol{\theta}_{10} + \gamma n^{-\frac{1}{2}}$ and $\boldsymbol{\theta}_{20}$ is fixed; the null hypothesis corresponds to $\gamma = 0$. With this feature of varying n we need to replace (1.4) by

(1.6) for $1 \leqslant i \leqslant n$, $X_{i,n}$ are independent with distribution functions $F_{in}(x, \boldsymbol{\theta}_n)$,

and (1.5) by

$$e_{in} = F_{in}(X_{in}, \tilde{\boldsymbol{\theta}}_n), \tag{1.7}$$

but we shall drop the additional suffix n whenever possible. In contrast to the identically distributed case, it is here essential to have a 'triangular array' in which the functions F_{in} indeed depend on n. The empirical distribution of the e_i is defined in the standard way as

$$\hat{F}_n(t) = n^{-1}\Sigma_i I(e_i \leqslant t) \tag{1.8}$$

where I is the indicator function, and the corresponding empirical process by

$$\hat{y}_n(t) = n^{\frac{1}{2}}(\hat{F}_n(t) - t). \tag{1.9}$$

Then the main result of this paper (Theorem 1) is that under $H_n(\gamma)$ and with regularity conditions

$$\hat{y}_n \Rightarrow y; \tag{1.10}$$

i.e., $\hat{y}_n$ converges weakly (in distribution) in $D[0, 1]$ to a process y, which will turn out to be Gaussian. From this result for $\gamma = 0$, in principle, asymptotically valid tests, such as Kolmogorov-Smirnov tests, can be constructed (cf. Hajék and Šidák (1967)). For $\gamma \neq 0$ we can obtain the asymptotic power against $H_n(\gamma)$. In general, unfortunately, in this case of estimated parameter values the distribution of the limiting process y, under H_0 depends on the F_i and even on the particular value of $\boldsymbol{\theta}_0 = [\boldsymbol{\theta}_{10}', \boldsymbol{\theta}_{20}']'$, which makes such tests not nonparametric and thus largely unusable for lack of suitable tables of critical values; we show (in Theorem 2, Corollary 2) that by choosing $\tilde{\boldsymbol{\theta}}_{2n}$ to be the MLE adjusted by a small random quantity, the limit process y can be made the familiar Brownian bridge. If this is done the tests using critical values from the standard tables again become asymptotically nonparametric: the disquiet one feels at using randomised tests has to be balanced against the fact that if one wishes to use this kind of test, the alternative is to construct special tables for every model of interest.

The type of result contained in Theorem 1, for the case of identically distributed X_i, goes back at least to Darling (1955). A modern treatment, again for this special case, which in many ways sets the pattern we shall follow, is due to Durbin (1973); recently newer treatments have become available, e.g., Csörgö et al. (1977). In the case treated here, in which parameters are estimated and the X_i are *not* identically distributed, there seems to be no previous work, not even for the linear model (1.1).

The possibility of randomly adjusting the parameter estimate can also be traced back, and references will be given later.

2. The main results.

The basic description of the model and the basic definitions are in (1.6) to (1.10) and thereabouts: the material up to that point is merely motivational. For clarity the main results are in the present section (at the end), while the rather long proof of Theorem 1 is deferred to Section 3. First we list a number of assumptions and definitions.

A1. There is a neighbourhood, $\mathfrak{N}$, of $\boldsymbol{\theta}_0$, to which attention is confined: if necessary by choosing n sufficiently large that $\boldsymbol{\theta}_n \in \mathfrak{N}$, and by discarding an event, whose probability, by A8, is small for large n, in the complement of which $\tilde{\boldsymbol{\theta}}_n \in \mathfrak{N}$.

A2. For fixed i and n, there exist (possibly infinite) a, b independent of $\boldsymbol{\theta}$ such that, for $\boldsymbol{\theta} \in \mathfrak{N}$, $F_{in}\ (a, \boldsymbol{\theta}) = 0$, $F_{in}(b, \boldsymbol{\theta}) = 1$, and $F_{in}(x, \boldsymbol{\theta})$ is continuous and strictly increasing for $a < x < b$.

A3. If $x_{in}(t, \xi, \eta) = F_{in}(F_{in}^{-1}(t, \xi), \eta)$, then there exists a continuous (vector) function $\psi = (\psi_1', \psi_2')'$ such that,

$$\sup\nolimits_{0 \leqslant t \leqslant 1,\, |\xi - \boldsymbol{\theta}_0| \leqslant Ln^{-\frac{1}{2}}} \Big| n^{-\frac{1}{2}} \Sigma (x_{in}(t, \xi, \boldsymbol{\theta}_n) - t) \\ - n^{\frac{1}{2}} (\xi - \boldsymbol{\theta}_n)' \psi(t) \Big| \rightarrow 0,$$

as $n \rightarrow \infty$, for every $L < \infty$.

The distance $|\xi - \boldsymbol{\theta}_0|$ in A3 may in principle denote any of the usual, equivalent, distances indifferently; some of the constructions below, however, are more easily described if it is taken as the maximum of the absolute values of the various components rather than, say, the Pythagorean distance.

A4. If

$$K_n(\varepsilon) = \sup_t n^{-\frac{1}{2}}\Sigma_i \sup_{|\xi_1-\xi_2|\leqslant \varepsilon n^{-\frac{1}{2}},\, |\xi_1-\theta_0|<Ln^{-\frac{1}{2}},\, |\xi_2-\theta_0|<Ln^{-\frac{1}{2}}}$$
$$|x_{in}(t, \xi_1, \theta_n) - x_{in}(t, \xi_2, \theta_n)|$$

then $K_n(\varepsilon) \to 0$ as $\varepsilon \to 0$, uniformly in n, for every $L < \infty$.

A5. If, for $\lambda > 0$,

$$L_n(\lambda) = \sup_t n^{-\frac{1}{2}}\Sigma_i \sup_{|\xi-\theta_0|\leqslant Ln^{-\frac{1}{2}}}\left(x_{in}\left(t + \lambda n^{-\frac{1}{2}}, \xi, \theta_n\right) - x_{in}(t, \xi, \theta_n)\right)$$

then $L_n(\lambda) \to 0$ as $\lambda \to 0$, uniformly in n for every $L < \infty$. Notice that, since

$$\frac{\partial x_{in}(t, \xi, \eta)}{\partial t} = \frac{f_{in}\left(F_{in}^{-1}(t, \xi), \eta\right)}{f_{in}\left(F_{in}^{-1}(t, \xi), \xi\right)},$$

the term whose supremum is required has a derivative in t of constant sign if the density f_{in} has monotone likelihood ratio. In that case attention may be confined to the two end points $t = 0$ and $t = 1 - \lambda n^{-\frac{1}{2}}$.

A6. $n^{\frac{1}{2}}(\tilde{\theta}_{2n} - \theta_{20}) = n^{-\frac{1}{2}}\Sigma_i \mathbf{l}_{in}(X_i, \theta_n) + A\gamma + \mathbf{Z}_n + \mathbf{l}_n$, where

(i) $E[\mathbf{l}_{in}(X_i, \theta_n)|\theta = \theta_n] = 0$, for all i and n;

(ii) $n^{-1}\Sigma_i E[\mathbf{l}_{in}(X_i, \theta_n)\mathbf{l}_{in}(X_i, \theta_n)'|\theta = \theta_n] \to J$; a fixed matrix;

(iii) A is a fixed finite matrix (of order $p_2 \times p_1$);

(iv) $\mathbf{l}_n \to \mathbf{0}$, in probability;

(v) $\mathbf{Z}_n$ is $N(\mathbf{0}, \Sigma)$, and is independent of the X_i, and Σ is constant;

(vi) $n^{-1}\Sigma_i \mathbf{h}_{in}(t, \theta_n) \to \mathbf{h}(t)$ for each t, where

$$\mathbf{h}_{in}(t, \theta) = \int_{-\infty}^{F_{in}^{-1}(t, \theta)} \mathbf{l}_{in}(x, \theta)dF_{in}(x, \theta),$$

and $\mathbf{h}$ is arbitrary;

(vii) under $H_n(\gamma)$, $n^{-\frac{1}{2}}\Sigma_i \mathbf{l}_{in}(X_i, \theta_n)$ and any finite set of

$$n^{-\frac{1}{2}}\Sigma_i\{I(F_{in}(X_i, \theta_n) \leqslant t) - t\}$$

with varying t have asymptotically a joint multivariate normal distribution, with parameters equal to the limits of the appropriate first and second moments.

This condition A6 largely parallels assumptions of Durbin, of course, though the introduction of $\mathbf{Z}_n$ is new. Something similar to (vi) seems to be needed in Durbin's work, though it is not explicitly there; condition (vii) is given in this form for generality and simplicity of statement—Lindeberg-type conditions would presumably often be used in particular applications.

A7: (i) For all i, n, $F_{in}(x, \theta)$ has a density $f_{in}(x, \theta)$ such that

$$\partial \log f_{in}(x, \theta_n)/\partial\theta_2 \text{ exists, and } E\left[\frac{\partial \log f_{in}(X_i, \theta_n)}{\partial\theta_2}\,\middle|\, \theta = \theta_n\right] = 0.$$

(ii) For all n,

$$\mathcal{I}(\boldsymbol{\theta}_n) = n^{-1}\Sigma_i E\left[\frac{\partial \log f_{in}(X_i, \boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}_2}\frac{\partial \log f_{in}(X_i, \boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}_2'}\middle| \boldsymbol{\theta} = \boldsymbol{\theta}_n\right]$$

exists, and converges to the finite positive-definite matrix $\mathcal{I}$ as $n \rightarrow \infty$.

(iii)

$$\mathcal{I}_{21}^{(n)} = n^{-1}\Sigma_i E\left[\frac{\partial \log f_{in}(X_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}\frac{\partial \log f_{in}(X_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1'}\middle| \boldsymbol{\theta} = \boldsymbol{\theta}_0\right]$$

exists, and converges to a finite limit $\mathcal{I}_{21}$ as $n \rightarrow \infty$.

(iv)

$$\frac{\partial F_{in}(x, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} = \int_{-\infty}^{x} \frac{\partial f_{in}(y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} dy$$

for all x when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

(v) Assumption A3.

(vi) $-n^{-1}\partial/(\partial\boldsymbol{\theta}_2)\Sigma_i x_{in}(t, \boldsymbol{\theta}_n, \boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n}$ converges to the same limit $\psi_2(t)$, for each t, which appears in A3.

(vii) Under $H_n(\gamma)$, $n^{-\frac{1}{2}}\Sigma_i\partial/(\partial\boldsymbol{\theta}_2)\log f_{in}(X_i, \boldsymbol{\theta}_n)$ and any finite set of $n^{-\frac{1}{2}}\Sigma\{I(F_{in}(X_i, \boldsymbol{\theta}_n) < t) - t\}$ with varying t have asymptotically a joint multivariate normal distribution, with parameters equal to the limits of the appropriate first and second moments.

Definition. If A7 holds, $\tilde{\boldsymbol{\theta}}_{2n}$ is a randomly adjusted efficient estimator of $\boldsymbol{\theta}_{20}$ relative to the sequence of alternatives $\{H_n(\gamma)\}$ when

$$n^{\frac{1}{2}}(\tilde{\boldsymbol{\theta}}_{2n} - \boldsymbol{\theta}_{20}) = n^{-\frac{1}{2}}\mathcal{I}^{-1}\Sigma_i\frac{\partial \log f_{in}(X_i, \boldsymbol{\theta}_n)}{\partial\, \boldsymbol{\theta}_2} + \mathcal{I}^{-1}\mathcal{I}_{21}\boldsymbol{\gamma} + \mathbf{Z}_n + \boldsymbol{\varepsilon}_n,$$

where $\mathbf{Z}_n$ is $N(0, \Sigma)$ independent of the X_i, and $\boldsymbol{\varepsilon}_n \rightarrow \mathbf{0}$ in probability. If this condition holds with $\mathbf{Z}_n = \mathbf{0}$, then $\tilde{\boldsymbol{\theta}}_{2n}$ is efficient.

Proposition 1. *If* A7 *holds, a (randomly adjusted) efficient estimator satisfies* A6.

A8. $n^{\frac{1}{2}}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n) \Rightarrow_{\mathcal{D}} T$, T being some random variable.
The notation $\Rightarrow_{\mathcal{D}}$ means convergence in distribution; in the present case this is in Euclidean space, but elsewhere may be in $D = D[0, 1]$.

Proposition 2. A6 $\Rightarrow$ A8.

A9 (a). For any choice of $\boldsymbol{\eta}_n$ in the neighborhood $|\boldsymbol{\xi} - \boldsymbol{\theta}_0| < \mathrm{L}n^{-\frac{1}{2}}$,

$$n^{-1}\Sigma_i\frac{\partial}{\partial\boldsymbol{\eta}}x_{in}(t, \boldsymbol{\eta}, \boldsymbol{\theta}_n)\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}_n}$$

converges to the continuous function ψ uniformly in t.

A9 (b). For any choices of ξ_n, η_n in the neighbourhood $|\xi - \theta_0| < Ln^{-\frac{1}{2}}$,

$$-n^{-1}\Sigma_i \frac{\partial}{\partial\eta} x_{in}(t, \xi, \eta)\Big|_{\xi=\xi_n,\, \eta=\eta_n}$$

converges to the continuous function ψ uniformly in t.

Proposition 3. A9 (a) *or* A9 (b) $\Rightarrow$ A3.

It is superficially attractive to deal with A3 by introducing derivatives as in A9; unfortunately A9 is in general too restrictive, though there are cases of interest which it covers. The proof of Proposition 3 follows immediately from the mean value theorem, since

$$\Sigma_i x_{in}(t, \xi, \theta_n) = \Sigma x_{in}(t, \theta_n, \theta_n) + (\xi - \theta_n)' \frac{\partial}{\partial\eta} \Sigma x_{in}(t, \eta, \theta_n)$$

where η lies between ξ and θ_n; similarly

$$\Sigma x_{in}(t, \xi, \theta_n) = \Sigma x_{in}(t, \xi, \xi) + (\theta_n - \xi)' \frac{\partial}{\partial\eta} x_{in}(t, \xi, \eta),$$

where η is between θ_n and ξ.

Theorem 1. *Under assumptions* A1 *to* A6, *with the sequence of alternatives* $H_n(\gamma)$, $\hat{y}_n$ *converges weakly in* D *to a Gaussian process* y, *where* y *has mean function*

$$Ey(t) = (\psi_2'(t)A - \psi_1'(t))\gamma$$

and covariance function

$$C(y(t_1), y(t_2)) = \min(t_1, t_2) - t_1 t_2 + \psi_2'(t_1)(\Sigma + J)\psi_2(t_2)$$

$$+ \psi_2'(t_1)\mathbf{h}(t_2) + \psi_2'(t_2)\mathbf{h}(t_1).$$

Theorem 2. *Under assumptions* A1 *to* A5 *and* A7, *with the sequence of alternatives* H_n (γ), *if* $\tilde{\theta}_{2n}$ *is a randomly adjusted efficient estimator then* $\hat{y}_n$ *converges weakly in* D *to a Gaussian process* y, *where* y *has mean function*

$$Ey(t) = (\psi_2'(t)\mathcal{I}^{-1}\mathcal{I}_{21} - \psi_1'(t))\gamma$$

and covariance function

$$C(y(t_1), y(t_2)) = \min(t_1, t_2) - t_1 t_2 + \psi_2'(t_1)(\Sigma - \mathcal{I}^{-1})\psi_2(t_2).$$

Corollary 1. *If no random adjustment is made, the covariance function of* y *is*

$$C(y(t_1), y(t_2)) = \min(t_1, t_2) - t_1 t_2 - \psi_2'(t_1)\mathcal{I}^{-1}\psi_2(t_2).$$

Corollary 2. *If the covariance matrix of the random adjustment,* Σ, *is chosen equal to* $\mathcal{I}^{-1}$, *then the covariance function of* y *is*

$$C(y(t_1), y(t_2)) = \min(t_1, t_2) - t_1 t_2,$$

the same as for a Brownian bridge; *for the null hypothesis* $\gamma = 0$, y *is a Brownian bridge.*

Provided $\mathcal{I}^{-1}$ is continuous at $\boldsymbol{\theta}_0$, we may choose $\Sigma = \mathcal{I}^{-1}(\tilde{\boldsymbol{\theta}}_n)$ in Corollary 2, and the result remains valid.

The proofs of the corollaries are of course trivial, while Theorem 2 follows immediately from Theorem 1 if we note that, under the additional assumptions, the values $A = \mathcal{I}^{-1}\mathcal{I}_{21}$, $J = \mathcal{I}^{-1}$, and $\mathbf{h}(t) = -\mathcal{I}^{-1}\boldsymbol{\psi}_2(t)$, are readily determined.

3. Proof of Theorem 1. It will be plain that the following owes much to Durbin (1973). However the neat use of random time-change by Durbin cannot be imitated here, and a much longer proof, adapted from Rao and Sethuraman (1975), is necessary.

First note that

$$(3.1)\qquad \hat{F}_n(t) = n^{-1}\Sigma_{i=1}^n I(e_i \leqslant t) = n^{-1}\Sigma I\big(U_{in} \leqslant x_{in}(t, \tilde{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_n)\big)$$

where

$$(3.2)\qquad U_{in} = F_{in}(X_i, \boldsymbol{\theta}_n)$$

and U_{in}, $1 \leqslant i \leqslant n$, are independently and uniformly distributed on [0, 1] under $H_n(\boldsymbol{\gamma})$. Thus

$$(3.3)\qquad \hat{y}_n(t) = n^{-\frac{1}{2}}\Sigma_{i=1}^n\big(I(U_{in} \leqslant t) - t\big) + d_n(t) + R_n(t)$$

where

$$(3.4)\qquad d_n(t) = n^{-\frac{1}{2}}\Sigma_{i=1}^n\big(x_{in}(t, \tilde{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_n) - t\big)$$

and

$$(3.5)\quad R_n(t) = n^{-\frac{1}{2}}\Sigma_1^n\big(I\big(U_{in} \leqslant x_{in}(t, \tilde{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_n)\big) - I(U_{in} \leqslant t) - x_{in}(t, \tilde{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_n) + t\big).$$

Now under $H_n(\boldsymbol{\gamma})$ the first term in (3.4) converges weakly as usual to the Brownian bridge, and we shall show that R_n is asymptotically negligible, so that it will then be only the effect of d_n that will need consideration.

LEMMA 1. *Under assumptions* A1, A2, A4, A5 *and* A8, $R_n \to_p 0$.

It is of course easy to replace the conditions in Lemma 1 by others involving the suprema of the derivatives of x_{in}; it turns out, however, that the end-points, $t = 0$ and $t = 1$, give trouble with the derivatives even for otherwise well-behaved F_{in}, and it is often easier to avoid use of the derivatives.

PROOF. The method is that of Rao and Sethuraman, although the details are quite different.

Define

$$(3.6)\quad R_n(t, \boldsymbol{\xi}) = n^{-\frac{1}{2}}\Sigma_1^n\big\{I\big(U_{in} \leqslant x_{in}(t, \boldsymbol{\xi}, \boldsymbol{\theta}_n)\big) - I(U_{in} \leqslant t) - x_{in}(t, \boldsymbol{\xi}, \boldsymbol{\theta}_n) + t\big\},$$

so that $R_n(t) = R_n(t, \tilde{\boldsymbol{\theta}}_n)$. Since, given $\phi > 0$, we may find $L < \infty$ such that $P[|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| > L/n^{\frac{1}{2}}] < \phi$, it is sufficient to show that, for any $\omega > 0$ and any fixed L, $P[\sup_{0 \leqslant t \leqslant 1}\sup_{|\boldsymbol{\xi} - \boldsymbol{\theta}_0| \leqslant L/n^{\frac{1}{2}}}|R_n(t, \boldsymbol{\xi})| > \omega] \to 0$. In fact we shall merely deal with

$R_n(t, \xi)$, omitting the modulus sign, but it will be plain that there is no difficulty completing the argument.

Subdivide the cube centred at $\boldsymbol{\theta}_0$ of side $2L/n^{\frac{1}{2}}$ into (approximately) $(2L/\varepsilon)^p$ cubes of side $\varepsilon/n^{\frac{1}{2}}$, where $\varepsilon > 0$, and, labelling them arbitrarily, let the kth such cube be L_k^*. Let $\xi_{kin}^1(t)$, $\xi_{kin}^2(t)$ be the values of $\xi \in L_k^*$ at which $x_{in}(t, \xi, \boldsymbol{\theta}_n)$ take its maximum and minimum values respectively. Then

$$x_{in}\big(t, \xi_{kin}^1(t), \boldsymbol{\theta}_n\big) - x_{in}\big(t, \xi_{kin}^2(t), \boldsymbol{\theta}_n\big) \leqslant q^*{}_{in}(t, \varepsilon),$$

where

(3.7)

$$q_{in}{}^*(t, \varepsilon) = \sup\nolimits_{|\xi_1 - \xi_2| \leqslant \varepsilon/n^{\frac{1}{2}},\, |\xi_1 - \boldsymbol{\theta}_0| < L/n^{\frac{1}{2}},\, |\xi_2 - \boldsymbol{\theta}_0| < L/n^{\frac{1}{2}}} |x_{in}(t, \xi_1, \boldsymbol{\theta}_n) - x_{in}(t, \xi_2, \boldsymbol{\theta}_n)|$$

and hence

$$\text{(3.8)} \qquad \sup\nolimits_{\xi \in L_k^*} R_n(t, \xi) \leqslant n^{-\frac{1}{2}} \Sigma\{I(U_{in} \leqslant x_{ink}(t)) - I(U_{in} \leqslant t) - x_{ink}(t) + t\}$$
$$+ n^{-\frac{1}{2}} \Sigma q_{in}^*(t, \varepsilon),$$

where we have also written $x_{ink}(t) = x_{in}(t, \xi_{kin}^1(t), \boldsymbol{\theta}_n)$.

The second term on the right-hand side is not greater than the $K_n(\varepsilon)$ of A4, and may thus be made arbitrarily small by suitable choice of ε, independently of n, t, and k. Now subdivide [0, 1] into (approximately) $n^{\frac{1}{2}}/\lambda$ intervals of length $\lambda/n^{\frac{1}{2}}$ by points t_s, where $t_s < t_{s+1}$, and write $T_s^* = \{t : t_s \leqslant t \leqslant t_{s+1}\}$. Then since $x_{in}(t, \xi, \boldsymbol{\theta})$ increases with t it follows that $x_{ink}(t)$ also does, and thus, for the first term on the right-hand side of (3.8), if $t \in T_s^*$

$$\text{(3.9)} \quad n^{-\frac{1}{2}} \Sigma\{I(U_{in} \leqslant x_{ink}(t)) - I(U_{in} \leqslant t) - x_{ink}(t) + t\}$$
$$\leqslant n^{-\frac{1}{2}} \Sigma\{I(U_{in} \leqslant x_{ink}(t_{s+1})) - I(U_{in} \leqslant t_s) - x_{ink}(t_{s+1}) + t_s\}$$
$$+ n^{-\frac{1}{2}} \Sigma\{t_{s+1} - t_s + x_{ink}(t_{s+1}) - x_{ink}(t_s)\}.$$

The second term here

$$= \lambda + n^{-\frac{1}{2}} \Sigma\{x_{in}\big(t_{s+1}, \xi_{kin}^1(t_{s+1}), \boldsymbol{\theta}_n\big) - x_{in}\big(t_{s+1}, \xi_{kin}^1(t_s), \boldsymbol{\theta}_n\big)$$
$$+ x_{in}\big(t_{s+1}, \xi_{kin}^1(t_s), \boldsymbol{\theta}_n\big) - x_{in}\big(t_s, \xi_{kin}^1(t_s), \boldsymbol{\theta}_n\big)\}$$
$$\leqslant \lambda + n^{-\frac{1}{2}} \Sigma q_{in}^*(t_{s+1}, \varepsilon) + n^{-\frac{1}{2}} \Sigma \zeta_{in}(\lambda, t_s)$$

where

$$\zeta_{in}(\lambda, t) = \sup\nolimits_{|\xi - \boldsymbol{\theta}_0| \leqslant \frac{L}{n^{\frac{1}{2}}}} \{x_{in}\big(t + \lambda/n^{\frac{1}{2}}, \xi, \boldsymbol{\theta}_n\big) - x_{in}(t, \xi, \boldsymbol{\theta}_n)\}$$

and by suitable choice of λ and ε this too can be made negligible for all n. The term that remains is

$$\text{(3.10)} \qquad n^{-\frac{1}{2}} \Sigma\{B_{inks} - p_{inks}\} \operatorname{sgn}(inks) = \Sigma W_{inks}, \text{ say}$$

where $p_{inks} = |x_{ink}(t_{s+1}) - t_s|$, B_{inks} are Bernoulli variables, with expectation p_{inks}, independent as i varies, and $\operatorname{sgn}(inks) = 1$ or -1 according to the sign of $x_{ink}(t_{s+1}) - t_s$. At this point the argument of Rao and Sethuraman needs a little adjustment, since the signs concerned are not independent of i. Nevertheless we can write

$$\Sigma W_{inks} = n^{-\frac{1}{2}}\{\Sigma_{i\in T_+}(B_{inks} - p_{inks}) - \Sigma_{i\in T_-}(B_{inks} - p_{inks})\}$$

where $T_+ = \{i{:}\operatorname{sgn}(inks) = +1\}$, $T_- = \{i{:}\operatorname{sgn}(inks) = -1\}$.

Thus

$$P[\Sigma W_{inks} > \omega/2] \leqslant P\left[\Sigma_{i\in T_+}(B_{inks} - p_{inks}) > \frac{n^{\frac{1}{2}}\omega}{4}\right]$$

$$+ P\left[-\Sigma_{i\in T_-}(B_{inks} - p_{inks}) > \frac{n^{\frac{1}{2}}\omega}{4}\right].$$

Rao and Sethuraman's argument (from (2.39) on) shows that the first term is not greater than

$$\exp\left(-\frac{tn^{\frac{1}{2}}\omega}{4}\right)\Pi_{T_+}E\exp(t(B_{inks} - p_{inks})),\ t > 0, \tag{3.11}$$

and in fact the product, over $i \in T_+$, may be replaced by the product over all i, each factor being not less than 1. Moreover

$$\begin{aligned}\Sigma_i p_{inks} &= \Sigma|x_{ink}(t_{s+1}) - t_s|\\ &\leqslant \Sigma|x_{in}(t_{s+1}, \xi^1_{kin}(t_{s+1}), \boldsymbol{\theta}_n) - x_{in}(t_{s+1}, \boldsymbol{\theta}_n, \boldsymbol{\theta}_n)| + \Sigma|t_{s+1} - t_s|\\ &\leqslant \frac{2L}{\varepsilon}\Sigma q^*_{in}(t_{s+1}, \varepsilon) + n^{\frac{1}{2}}\lambda\\ &\leqslant n^{\frac{1}{2}}(\lambda + 2K_n(\varepsilon)L/\varepsilon),\\ &\leqslant n^{\frac{1}{2}}(\lambda + 2KL\varepsilon^{-1}),\end{aligned}$$

where $K = K(\varepsilon) = \sup_n K_n(\varepsilon) < \infty$, and from this the argument may be completed as in Rao and Sethuraman: we may choose, for example, $t = \log(1 + \omega/(\omega + 4\lambda + 8KL\varepsilon^{-1}))$, and use the facts that $\delta \geqslant \log(1 + \delta) > \delta - \delta^2/2$ to show that the expression in (3.11) is not greater than $\exp - Cn^{\frac{1}{2}}$, where $C = \omega^2/\{8(\omega + 4\lambda + 8KL\varepsilon^{-1})\}$.

LEMMA 2. *Under assumptions* A1, A2, A3 *and* A8,

$$d_n \to_{\mathscr{D}} T'\Psi.$$

It follows at once that $d_n - n^{\frac{1}{2}}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n)'\psi \to_p 0$, and the remainder of the proof is trivial.

The proof now is almost complete. By Lemma 1 we may ignore the last term in (3.3), and each of the first two terms converges in distribution, and their distributions are therefore tight; thus their joint distribution is tight. (Billingsley (1968) page 41). The joint distribution of these two terms therefore converges weakly provided their joint finite-dimensional distributions converge, as they do according to A6 (vii) and Lemma 2. Finally, the map $(g, h) \to g + h$ in D is continuous where $g + h$ are continuous, and the fact that $\hat{y}_n$ converges in distribution is now obvious. To evaluate the limiting distribution is also straight forward, and the proof of Theorem 1 is complete.

4. Random adjustment. The technique of making random adjustments has appeared several times in the literature, though it has usually been thought of as applying to the residuals, or observations, directly, rather than to the parameter estimate.

Durbin (1961) replaced the observed value of a sufficient statistic by an independent observation from an appropriate distribution in order to remove the effect of a nuisance parameter. Although this may well be related to the present work, since the MLE is asymptotically sufficient, it seems very difficult to apply his technique using the MLE. See also Durbin (1975).

Tiao and Guttman (1967) adjusted the residuals for a sample from $N(\mu, \sigma)$ in precisely the same way as the present approach, but without considering it as involving the use of a new parameter estimate.

Theil (1965, 1968) introduced a definition of BLUS residuals for the general linear model specifically in order that they should have a scalar covariance matrix. It turns out that these residuals are closely connected with our adjusted residuals. For simplicity, consider only the case of a sample from $N(\mu, \sigma^2)$, say X_i, $0 \leqslant i \leqslant n$. Although there are $n + 1$ observations, only n residuals can be defined, and these will be chosen to correspond to $i = 1, 2, \ldots, n$. Then we find

$$e_i = X_i - \bar{X} - \left(1 + (n+1)^{\frac{1}{2}}\right)^{-1}\left(X_0 - \bar{X}\right),$$

which to order $n^{-\frac{1}{2}}$ is equal to

$$e_i' = X_i - \left(\bar{X}_* + n^{-\frac{1}{2}}(X_0 - \mu)\right)$$

where $\bar{X}_*$ is the mean of $X_1, X_2, \cdots, X_n$, and since as far as the observations $X_i (i = 1, 2, \cdots, n)$ are concerned $X_0 - \mu$ is an independent normal random variable, this is exactly the type of random adjustment dealt with previously.

Rao (1972) exhibited for a simple random sample an adjustment to the process $\hat{y}_n$ which behaves as in the case with no nuisance parameters; Durbin (1975) noted that this is equivalent to the use of $\hat{\boldsymbol{\theta}}_2$, the MLE based on half the sample, rather than the full sample value $\hat{\boldsymbol{\theta}}$. This almost fits the present framework (which could be extended easily to cover it): if we write $\hat{\boldsymbol{\theta}}_2 = \hat{\boldsymbol{\theta}} + \sigma$, then σ is not independent of the observations, but it is (asymptotically) independent of $\hat{\boldsymbol{\theta}}$ and of the first term in (3.3), as is easily seen if the usual asymptotic expansion for the MLE is used, and this is sufficient. The absence in this case of *external* randomisation is counterbalanced by the need to choose a particular half-sample, as Durbin observed. In fact, one of the advantages of the formulation of the present paper is that it is made clear exactly how much randomisation is needed: a p_2-variate normal variable as in A6. (Strictly speaking this is the maximum amount of randomisation needed, although in the general case this will indeed be required: if some component of $\psi_2(t)$ vanishes for all t, then no randomisation is required for that component of $\tilde{\theta}_{2n}$. See Section 5 (iii) for an example.)

The use of random adjustments need not of course be confined to the situation in which empirical distribution functions are studied. One is, however, reluctant to use them on general philosophical grounds and also because one assumes that their use must in general lead to a loss of power. The only evidence available on this latter point is due to Stephens (1978), who shows that this is indeed so, although the extent of the loss varies from case to case. It is of course worth noting that in all the cases considered by Stephens, the alternatives belong to quite separate families from the null hypotheses, in contrast to the kind of alternative allowed in Theorem 1. But this question of power is in any case not, I think, very important, for if tables for the correct test were available no one would use the randomised test.

5. Examples.

(i) *Identically distributed observations.* The conditions imposed here are slightly stronger, but otherwise the results here are, except in allowing random adjustment, as in Durbin (1973); random adjustment can in fact be carried out under exactly his conditions.

(ii) *Linear hypothesis.* If no detailed distributional assumptions about $\boldsymbol{\varepsilon}$ are made, (1.1) does not provide a model in the sense used here, but it is possible to discuss random adjustment to a limited extent in terms of first and second moments. Supposing that B is of full rank for convenience, and as usual supposing $\boldsymbol{\varepsilon}$ to have uncorrelated elements with variance σ^2, the least squares estimator $\hat{\boldsymbol{\theta}} = (B'B)^{-1}B'\mathbf{X}$, and the corresponding residual vector $\mathbf{e} = \mathbf{X} - B\hat{\boldsymbol{\theta}} = (I - M)\boldsymbol{\varepsilon}$, where $M = B(B'B)^{-1}B'$; moreover the covariance matrix of $\mathbf{e}$, $V(\mathbf{e}) = (I - M)\sigma^2$. Suppose instead we use an estimator $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} + \boldsymbol{\delta}$, where δ has mean $\mathbf{0}$ and is uncorrelated with $\mathbf{X}$, and $\mathrm{V}(\boldsymbol{\delta}) = (B'B)^{-1}\sigma^2$; then the corresponding residuals $\mathbf{f} = \mathbf{X} - B\tilde{\boldsymbol{\theta}}$ have $V(\mathbf{f}) = \sigma^2 I$—i.e., are uncorrelated and have in fact the same structure as $\boldsymbol{\varepsilon}$. Such an adjustment is approximately equivalent to the use of Theil's BLUS residuals, as noted earlier.

If we specialise to the case in which the elements of $\boldsymbol{\varepsilon}$ are independent $N(0, \sigma^2)$, and $\boldsymbol{\delta}$ is normal, then $\mathbf{f}$ also has independent $N(0, \sigma^2)$ components; this simple exact result is only relevant when σ^2 is known. Notice that this is a much stronger result than could be obtained from Theorem 2, Corollary 2: the latter merely shows how to ensure that a particular function of the adjusted residuals behaves as though they were independent.

(iii) *Simple linear regression.* Suppose $X_{in} = \alpha + u_{in}\beta + \varepsilon_{in}$, where the ε_{in} are independent $N(0, 1)$ variables; the unknown variance case we leave for the moment. Assume for convenience $\Sigma_i u_{in} = 0$, that $\alpha_0 = \beta_0 = 0$, and that in the earlier notation $p_1 = 0$: there is no question of evaluating power against alternatives.

LEMMA 3. *If $n^{-1}\Sigma u_{in}^2 \to b$, then conditions* A4, A5, A7, A9(a) *are satisfied, the least squares estimator is efficient, and $\psi(t) = [\phi(\Phi^{-1}(t)), 0]'$, $\mathscr{I} =$* diag(1, b).

Clearly $F_{in}(x, \theta) = \Phi(x - \alpha - \beta u_{in})$, and so

$$x_{in}(t, \xi, \eta) = \Phi(\Phi^{-1}(t) + (\alpha_\xi - \alpha_\eta) + (\beta_\xi - \beta_\eta)u_{in}).$$

Now

$$\frac{\partial x_{in}}{\partial \xi}(t, \xi, \eta) = \begin{bmatrix} \phi(\Phi^{-1}(t) + (\alpha_\xi - \alpha_\eta) + (\beta_\xi - \beta_\eta)u_{in}) \\ u_{in}\phi(\Phi^{-1}(t) + (\alpha_\xi - \alpha_\eta) + (\beta_\xi - \beta_\eta)u_{in}) \end{bmatrix}$$

so that

$$K_n(\varepsilon) < \frac{\varepsilon}{n}\Sigma_i(1 + |u_{in}|)\sup \phi,$$

and A4 follows. Moreover A9(a) is easily proved, if the term $\phi(\Phi^{-1}(t) + \alpha_\eta + \beta_\eta u_{in})$ in the derivative of x_{in} is replaced by $\phi(\Phi^{-1}(t)) + (\alpha_\eta + \beta_\eta u_{in})\phi'(x_*)$, and the fact that ϕ' is bounded is recalled; thus A3 is satisfied by Proposition 3. Moving on, A7 (iii) is here vacuous, and A7 (i), (ii), (iv), (v) and (vi) are immediately verified. It is just slightly more difficult to deal with A7 (vii): it is easily seen that it is sufficient, by using the Cramér-Wold device to show that, for any k and m, $(l\Sigma u_{in}X_i + m\Sigma Y_i)n^{-\frac{1}{2}} \equiv T_n$ is asymptotically normal with the obvious parameters, where (X_i, Y_i) are independent for different i and identically distributed, X_i are $N(0, 1)$, and all moments of Y_i exist. Assume for convenience that, for each n, u_{in} have been reordered so that u_{in}^2 increases with i, let λ_n be a sequence decreasing to 0, and define I_n as the maximum value of i for which $n^{-1}u_{in}^2 < \lambda_n$. Then, for $i > I_n$, $u_{in}^2 > n\lambda_n$, and it follows that for large n $(n - I_n) < 2b/\lambda_n$. Now T_n can be expressed as the sum of 2 independent components, one involving $i < I_n$, and the other $i > I_n$; as far as the latter is concerned, the part involving X_i is already normal, while the part involving Y_i has variance proportional to $n^{-1}(n - I_n)$, and so is asymptotically negligible provided $n\lambda_n \to \infty$. Now we use Liaponov's theorem on the other component, and for this we have to show that $\Sigma E|lu_{in}X_i + mY_i|^3/[\Sigma E(lu_{in}X_i + mY_i)^2]^{\frac{3}{2}} \to 0$, (Loève Section 20.1.a(ii)). The numerator is not greater than a multiple of $\Sigma_{i<I_n}|u_{in}|^3 + I_n KE|Y_i|^3$, (Loève Section 9.3, '$c_r$-inequality'), where K is, constant, which, since $|u_{in}|^3 < u_{in}^2\lambda_n^{\frac{1}{2}}n^{\frac{1}{2}}$, is $o(n^{\frac{3}{2}})$. If the denominator is exactly of order $n^{\frac{3}{2}}$, as would usually be the case, the result follows: we can ensure this by everywhere replacing Y_i by $Y_i + Z_i$, where Z_i are independent $N(0, 1)$, and then proving that $T_n + n^{-\frac{1}{2}}\Sigma Z_i$ is asymptotically normal, from which the required result follows at once.

Finally, since the efficiency of the least squares estimator is easily checked, we deal with A5. The derivative of the term whose supremum is sought is easily seen to be of constant sign, and the supremum therefore occurs either at $t = 0$ or $t = 1$, and it is also easily seen that the worst case is either $t = 0$ and $\delta_{in} = \alpha_\xi + \beta_\xi u_{in} = L(1 + |u_{in}|)n^{-\frac{1}{2}}$ or $t = 1$ and δ_{in} the negative of this value; clearly it is sufficient to deal with the former, so we consider

$$n^{-\frac{1}{2}}\Sigma_i\Phi\big(\Phi^{-1}(\lambda n^{-\frac{1}{2}}) + \delta_{in}\big) \leqslant n^{-\frac{1}{2}}\Sigma_i\Big\{\Phi\big(\Phi^{-1}(\lambda n^{-\frac{1}{2}})\big) + \delta_{in}\phi\big(\Phi^{-1}(\lambda n^{-\frac{1}{2}}) + \delta_{in}\big)\Big\},$$

by the mean value theorem and the monotonicity of ϕ. The first term equals λ and is consequently well-behaved, while for the second we may note that, since for large n $|u_{in}| < (2b)^{\frac{1}{2}}n^{\frac{1}{2}}$ and thus δ_{in} is bounded for all i and n by a constant, say K, it is not greater than

$$n^{-\frac{1}{2}}\phi\big(\Phi^{-1}(\lambda n^{-\frac{1}{2}}) + K\big)\Sigma\delta_{in}.$$

Again, $\Sigma\delta_{in} < K'n^{\frac{1}{2}}$ for suitable K', so that we may consider $K'\phi(\Phi^{-1}(\lambda n^{-\frac{1}{2}}) + K) < K'\phi(\Phi^{-1}(\lambda) + K)$, from which the required result is obvious.

Thus Theorem 1 is valid for any estimator satisfying A6, and Theorem 2 is valid for the least-squares (ML) estimator, both under the conditions on u_{in} previously given. Notice the curious consequence of the fact that the β component of ψ vanishes: the limiting covariance function is unaffected by the particular estimate of β used, provided it satisfies A6, and thus if we wish to recover the Brownian bridge as a limit, we have randomly to adjust $\hat{\alpha}$, but $\hat{\beta}$ needs no adjustment. The comment at the end of (ii) above is relevant here. (It may also be checked that the random adjustment to $\hat{\alpha}$ suggested by Theorem 2 is the same as that in (ii).)

Now suppose the variance is unknown, and write $\Theta = (\alpha, \beta, \sigma)'$. Then choosing $\theta_0 = (0, 0, 1)'$ for convenience, the analysis goes through much as before, and Lemma 3 is valid in this case also, except that now $\psi(t) = [\phi(\Phi^{-1}(t)), 0, \Phi^{-1}(t)\phi(\Phi^{-1}(t))]'$, and $\mathcal{I} = \mathrm{diag}\,(1, b, 3)$.

(iv) *Feigl and Zelen model.* Cox and Snell (1968) and Loynes (1969) discussed a model proposed by Feigl and Zelen (1965), in which the observations have the structure

$$X_i = \alpha \exp\,(\beta d_i)\varepsilon_i$$

where ε_i are exponentially distributed, with unit mean. Again we shall ignore the possibility of alternative hypotheses.

LEMMA 4. *If* $n^{-1}\Sigma d_{in}^2 \to b$, *then conditions* A4, A5, A7, A9(a) *are satisfied, with* $\psi(t) = (-\alpha^{-1}(1 - t)\log(1 - t), 0)'$, $I = \mathrm{diag}(\alpha^{-2}, b)$, *and the* MLE *is efficient.*

The proof of this is straightforward, but rather long and tedious, and we do no more than sketch the bare outline. A4 follows without difficulty, by the use of the derivative of x_{in}, from the condition that $n^{-1}\Sigma|d_{in}|$ is bounded, if one uses the fact that ue^{ku} is bounded in $u < 0$ for $k > 0$. A9(a) follows fairly easily in a similar way using second derivatives of e^{ku}, except that one has to show that

$$n^{-1}\Sigma(e^{\pm d_{in}\beta_\xi} - 1) \to 0$$

provided $|\beta_\xi| < Ln^{-\frac{1}{2}}$. Now one has

$$e^{\pm d_{in}\beta_\xi} = 1 \pm \beta_\xi d_{in} + \tfrac{1}{2}\beta_\xi^2 d_{in}^2 e^{\theta\beta_\xi d_{in}},$$

for some θ with $|\theta| < 1$, and since from $n^{-1}\Sigma d_{in}^2 \to b$ one knows that $d_{in} = 0(n^{\frac{1}{2}})$, the rest follows. Also that there is no difficulty in dealing with A7, so that A5 alone remains: the maximum of each term again occurs either at $t = 0$ or at $t = 1$, and approximation in the above spirit suffices to show that the condition is satisfied. The efficiency of the MLE may also be dealt with.

Thus Theorems 1 and 2 apply under the given conditions.

Note added in proof. Since this work was completed two papers dealing with the empirical distribution function which are related to it have appeared. Mukantseva ((1977), *Theor. Prob. Appl.* **22** 591–602) deals only with normal linear regression, but needs essentially no conditions. The main result of Pierce and Kopecky ((1979), *Biometrika* **66** 1–6) is similar to that of the present work; no proofs are given.

REFERENCES

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *J. Roy. Statist. Soc. B* **30** 248–275.

Csörgö, M., Komlós, J., Major, P., Révész, P. and Tusnady, G. (1977). On the empirical process when parameters are estimated. *Proc. Seventh Prague Conference B* 87–97.

Darling, D. A. (1955). The Cramér-Smirnov test in the parametric case. *Ann. Math. Statist.* **26** 1–20.

Durbin, J. (1961). Some methods of constructing exact tests. *Biometrika* **48** 41–55.

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

Durbin, J. (1975). Tests of model specification based on residuals. In *A Survey of Statistical Design and Linear Models* (ed. J. N. Srivastava) North Holland, Amsterdam.

Hajék, J. and Šidák, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.

Loève, M. (1955). *Probability Theory* (1st edition). Van Nostrand, New York.

Loynes, R. M. (1969). On Cox and Snell's general definition of residuals. *J. Roy. Statist. Soc. B* **31** 103–106.

Rao, K. C. (1972). The Kolmogoroff, Cramér-von Mises, chi square statistics for goodness-of-fit in the parametric case. Abstract 133-6. *Bull. Inst. Math. Statist.* **1** 87.

Rao, J. S. and Sethuraman, J. (1975). Weak convergence of empirical distribution functions of random variables subject to perturbations and scale factors. *Ann. Statist.* **3** 299–313.

Stephens, M. A. (1978). On the half-sample method for goodness-of-fit. *J. Roy. Statist. Soc.* B **40** 64–70.

Theil, Henri (1965). The analysis of disturbances in regression analysis. *J. Amer. Statist. Assoc.* **60** 1067–1079.

Theil, Henri (1968). A simplication of the BLUS procedure for analysing regression disturbances. *J. Amer. Statist. Assoc.* **63** 242–251.

Tiao, G. C. and Guttman, Irwin (1967). Analysis of outliers with adjusted residuals. *Technometrics* **9** 541–559.

Department of Probability and Statistics
University of Sheffield
Sheffield S3 7RH
England