TESTING GOODNESS-OF-FIT FOR THE DISTRIBUTION
OF ERRORS IN REGRESSION MODELS

by

DONALD A. PIERCE
OREGON STATE UNIVERSITY

and

STANFORD UNIVERSITY

and

KENNETH J. KOPECKY
FRED HUTCHINSON CANCER
RESEARCH CENTER

NOTICE
This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

TECHNICAL REPORT NO. 16

June 8, 1978

STUDY ON STATISTICS AND ENVIRONMENTAL
FACTORS IN HEALTH

PREPARED UNDER SUPPORT TO SIMS FROM
ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION (ERDA)
ROCKEFELLER FOUNDATION
SLOAN FOUNDATION
ENVIRONMENTAL PROTECTION AGENCY (EPA)
NATIONAL SCIENCE FOUNDATION (NSF)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Testing Goodness-of-fit for the Distribution
of Errors in Regression Models

Donald A. Pierce
Oregon State University¹

Kenneth J. Kopecky
Fred Hutchinson Cancer
Research Center

SUMMARY

Asymptotic results are given for the problem of testing goodness-of-fit for any specified distribution of errors in multiple regression models. In particular, the results apply to the case of testing for normality in standard regression and experimental design models. For a very wide class of goodness-of fit statistics, in particular those which depend only on the empirical distribution of the residuals, it is shown that the limiting distributions under the null hypothesis are precisely the same in regression models as in ordinary location-scale models.

Some Key Words: Asymptotic distribution; chi-square tests; goodness-of-fit tests; regression models; residuals; tests of normality.

1. INTRODUCTION

Let $y_1, y_2, \dots, y_n$ be statistically independent observations following the model

$$y_i = x_i' \beta + \sigma e_i, \quad i = 1, 2, \dots, n,$$

where the x_i are given vectors and (β, σ) are unknown parameters. The

¹This research was carried out during a visit to Stanford University.

problem considered here is testing whether the e_i are from a specified distribution, such as the standard normal. It is assumed, and is crucial here, that the regression model includes a constant term, that is, it can be written such that the first element of each x_i vector is 1.

Define residuals by $\hat{e}_i = (y_i - x_i' \hat{\beta}) / \hat{\sigma}$, where $(\hat{\beta}, \hat{\sigma})$ are maximum likelihood estimators. The results here pertain to the asymptotic distribution under the null hypotheses of any goodness-of-fit statistic which depends only on the empirical distribution of the $\hat{e}_i$; that is, those which are permutationally symmetric functions of the residuals. This is a wide and natural class of goodness-of-fit statistics, although others are worth consideration since the $\hat{e}_i$ are not identically distributed.

Considerable attention has been given to this problem for the very special case that β is a scalar and $x_i' \beta = \beta$ for all i , the case of identically distributed observations from a location-scale family. The basic result here, which seems surprising, is that regardless of the (fixed) dimension of β , the limiting distribution under the null hypothesis of any goodness-of-fit statistic of the type described above is precisely the same for the regression model as it is for the ordinary location-scale model.

Many asymptotic results are available for the location-scale problems. For testing normality, with both parameters unknown, the limiting distributions of most of the common general types of statistics are now available. A convenient reference is Pearson and Hartley (1972). As is well-known the ordinary chi-square statistic does not have precisely a limiting chi-square distribution when $(\bar{x}, s)$ are used for estimators; tables are given by Dahiya and Gurland (1972). A modification which does have a limiting chi-square distribution

was developed by Rao and Robson (1974); see also Moore (1977). For results on the Neyman smooth test see Thomas and Pierce (1978). For the case of extreme value, or Weibull, distributions see Stephens (1977).

The actual result here is given in terms of the empirical distribution of $H(\hat{e}_i)$, where $H(\cdot)$ is the hypothesized, and true, distribution function of the true residuals e_i . Write $\hat{u}_i = H(\hat{e}_i)$ and $\hat{F}_n(\cdot)$ for the empirical distribution function of the $\hat{u}_i$. That is, for $0 \leq t \leq 1$, $\hat{F}_n(t)$ is the proportion of $\hat{u}_1, \hat{u}_2, \dots, \hat{u}_n$ less than or equal to t . Consider the stochastic process $\hat{y}_n(t) = \sqrt{n}\{\hat{F}_n(t) - t\}$. It will be shown that under ordinary regularity conditions $\hat{y}_n$ converges to a stochastic process $\hat{y}$, the distribution of which depends on $H(\cdot)$ but is the same for the regression model as for the ordinary location-scale model. The asymptotic distribution of any goodness-of-fit statistic which is a suitably continuous function of $\hat{y}_n$ is then determined by the distribution of $\hat{y}$. For discussion of the necessary continuity see Durbin (1973b), Sections 3.2, 4.4, and 4.5.

The approach is in two steps. In Section 2 we generalize a result of Durbin (1973) concerning the asymptotic distribution of $\hat{y}_n$, generalizing from identically distributed to nonidentically distributed observations. This result is given without proof, but is easily believed. A sketch of the proof for the normal case is given in Section 4; further details are available from the authors.¹ In Section 3 is given the complete argument for the more surprising part, that the regression case reduces to the location-scale one.

The regularity conditions needed here are essentially only those needed to establish the asymptotic normality of $(\hat{\beta}, \hat{\sigma})$, and also that the e_i have a continuous distribution. See Bradley and Gart (1962), and also the comments in Cox and Hinkley (1974, p. 297) regarding the former conditions.

¹See Appendix

2. GENERALIZATION OF DURBIN'S RESULTS

In this section we give some results, without proof, in a setting more general than discussed above. First consider Durbin's (1973) problem in which $y_1, y_2, \dots, y_n$ are independent and identically distributed with cumulative distribution function $F(y; \theta)$, where θ is a vector of parameters. Let $\hat{\theta}$ be the maximum likelihood estimator, I_{θ} be the Fisher information, per observation, for estimation of θ , and $\underline{g}(t, \theta) = \partial F(y; \theta) / \partial \theta$ evaluated at $y = F^{-1}(t; \theta)$. Write $\hat{u}_i = F(y_i; \hat{\theta})$ and $\hat{F}_n(t)$ for the empirical distribution function of $\hat{u}_1, \hat{u}_2, \dots, \hat{u}_n$. Define the stochastic process $\hat{y}_n$ by $\hat{y}_n(t) = \sqrt{n} \{ \hat{F}_n(t) - t \}$. Durbin showed that under ordinary regularity conditions $\hat{y}_n$ converges to $\hat{y}$, a normal process with covariance function

$$\text{Cov}\{\hat{y}(s), \hat{y}(t)\} = \min(s, t) - st - \underline{g}(s, \theta)' I_{\theta}^{-1} \underline{g}(t, \theta). \quad (2.1)$$

If the observations are not identically distributed, but the cumulative distribution function of y_i is $F_i(y; \theta)$, then the above result remains true, with $\hat{u}_i = F_i(y_i; \hat{\theta})$, provided that $\underline{g}(t, \theta)$ and I_{θ} in (2.1) are replaced by suitable averages as follows. All the limits below are as $n \rightarrow \infty$, and are assumed to exist. The information matrix I_{θ} is replaced by $\lim I_{n, \theta}$ where $I_{n, \theta}$ is n^{-1} times the total Fisher information in $y_1, y_2, \dots, y_n$. The vector $\underline{g}(t, \theta)$ is replaced by $\lim \underline{g}_n(t, \theta)$ where $\underline{g}_n(t, \theta) = n^{-1} \sum_{i=1}^n \partial F_i(y_i^*; \theta) / \partial \theta$ and $y_i^* = F_i^{-1}(t; \theta)$.

Regularity conditions which are sufficient for this generalization are that for θ in a neighborhood of the true value:

- i) For $i = 1, 2, \dots$ the F_i are continuous in y with densities f_i satisfying the conditions of Bradley and Gart (1962), which imply that the maximum likelihood estimator is asymptotically normal;
- ii) The limits I_{θ} and $g(t, \theta)$ exist, I_{θ} is positive definite and continuous in θ , and $g(t, \theta)$ is continuous in (t, θ) ;
- iii) There exists an $M < \infty$ such that

$$|\partial F_i(y; \theta) / \partial \theta| < M \text{ for all } y, \theta, i = 1, 2, \dots;$$

- iv) There exists an $M < \infty$ such that

$$|\text{Min}\{F_i(y; \theta), 1 - F_i(y; \theta)\} \partial \log f_i(y; \theta) / \partial \theta| < M \text{ for all } y, \theta, i = 1, 2, \dots$$

This last condition means in essence that the efficient scores $\partial \log f_i(y; \theta) / \partial \theta$ do not become large in an erratic way as $F_i(y; \theta) \rightarrow 0$ or 1 . If they were monotone in y , for example, then the expression in (iv) would necessarily approach zero at these extremes as a consequence of existence of the means of the efficient scores.

For identically distributed observations from location-scale families the last term in (2.1) does not in fact depend on θ . In applying the generalization to the regression setting the terms $g_n(s, \theta)' I_{n, \theta}^{-1} g_n(t, \theta)$ will depend on neither n nor θ , so the limits of the individual factors need not be explicitly evaluated.

3. REGRESSION MODELS

We now apply the results of Section 2 to the case $F_i(y; \theta) = H\{(y - x_i' \beta) / \sigma\}$, so $\theta = (\beta, \sigma)$. We will evaluate $g_n(s, \theta)' I_{n, \theta}^{-1} g_n(t, \theta)$ and show that for

any n this is precisely the same as $\underline{g}(s, \underline{\theta})' \underline{I}_{\underline{\theta}}^{-1} \underline{g}(t, \underline{\theta})$ corresponding to the family $F(y; \underline{\theta}) = H\{(y - \beta)/\sigma\}$. Note that the $\hat{u}_i$ of Section 2 are in this setting the $\hat{u}_i = H(\hat{e}_i)$ of Section 1.

For the location-scale model the parameter β can be chosen as any quantile of the distribution F . It is convenient to choose this so that the information matrix for (β, σ) is diagonal; this can be achieved from any other parameterization through the triangular matrix which diagonalizes the information matrix. Further, the diagonal terms of the resulting information matrix can be expressed as $\sigma^{-2} i_{\beta\beta}$ and $\sigma^{-2} i_{\sigma\sigma}$, where $i_{\beta\beta}$ and $i_{\sigma\sigma}$ are intrinsic constants determined by $H(\cdot)$. Moreover, writing $h(\cdot)$ for the density of $H(\cdot)$ and y^* such that $H\{(y^* - \beta)/\sigma\} = t$, it can be seen that $\partial H\{(y^* - \beta)/\sigma\} / \partial \beta = -d(t)/\sigma$ and $\partial H\{(y^* - \beta)/\sigma\} / \partial \sigma = -g(t)d(t)/\sigma$, where $d(t) = h\{H^{-1}(t)\}$, $g(t) = H^{-1}(t)$. Thus for the location-scale case

$$\underline{g}(s, \underline{\theta})' \underline{I}_{\underline{\theta}}^{-1} \underline{g}(t, \underline{\theta}) = d(s)d(t)\{i_{\beta\beta}^{-1} + g(s)g(t)i_{\sigma\sigma}^{-1}\}. \quad (3.1)$$

In the regression case we have that

$$n \underline{g}_n(t, \underline{\theta}) = -d(t)/\sigma \begin{bmatrix} \underline{X}_n' & \underline{1}_n \\ n & g(t) \end{bmatrix}$$

where $\underline{1}_n$ is an n -vector of ones and $\underline{X}_n$ is a matrix with rows $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$.

Further

$$n \underline{I}_{n, \underline{\theta}} = \sigma^{-2} \begin{bmatrix} i_{\beta\beta} & \underline{X}_n' \underline{X}_n & 0 \\ 0 & n i_{\sigma\sigma} \end{bmatrix},$$

so

$$n \underline{g}_n(s, \underline{\theta})' \underline{I}_{n, \underline{\theta}}^{-1} \underline{g}_n(t, \underline{\theta}) = d(s)d(t)\{\underline{1}_n' \underline{X}_n (\underline{X}_n' \underline{X}_n)^{-1} \underline{X}_n' \underline{1}_n i_{\beta\beta}^{-1} + n g(s)g(t)i_{\sigma\sigma}^{-1}\}.$$

That $\tilde{g}_n(s, \theta)' I_{n\theta}^{-1} \tilde{g}_n(t, \theta)$ is equal to (3.1) follows from the fact that $X_n(X_n' X_n)^{-1} X_n' \tilde{1}_n = \tilde{1}_n$. This last fact is a consequence of the regression model containing a constant term; $\tilde{1}_n$ is thus in the span of the columns of X_n , and $X_n(X_n' X_n)^{-1} X_n'$ is the projection onto this span.

4. SKETCH OF PROOF FOR NORMAL REGRESSION

Since the argument leading to the result of Section 2 is rather involved, it is helpful to consider the special case of normal regression with known scale parameter, where the main ideas can be more easily seen.

Let Φ denote the standard normal distribution function, and consider independent observations $y_1, y_2, \dots$ with distribution functions respectively $\Phi(y - \tilde{x}_i' \beta)$. Here $\tilde{x}_i' = (1, x_{i1}, x_{i2}, \dots, x_{ip})$ and by $p = 0$ we mean $x_i = 1$. Let $\hat{e}_{in} = y_i - \tilde{x}_i' \hat{\beta}_n$ for $i = 1, 2, \dots, n$; $n = 1, 2, \dots$; where $\hat{\beta}_n$ denotes the maximum likelihood estimator based on $y_1, y_2, \dots, y_n$. It will be convenient here to work directly with the empirical distribution of the $\hat{e}_{in}$, rather than $\Phi(\hat{e}_{in})$.

Let $\hat{F}_n$ be the empirical distribution function of $\hat{e}_{1n}, \hat{e}_{2n}, \dots, \hat{e}_{nn}$, and $\hat{w}_n(e) = \sqrt{n} [\hat{F}_n(e) - \Phi(e)]$. We will sketch an argument showing that $\hat{w}_n \rightarrow \hat{w}$, a stochastic process whose distribution is the same for any $p = 0, 1, \dots$.

We have

$$\begin{aligned}
 \hat{w}_n(e) &= \sqrt{n} [\hat{F}_n(e) - \Phi(e)] \\
 &= \sqrt{n} [n^{-1} \#\{i \leq n: y_i - x_i' \hat{\beta}_n \leq e\} - \Phi(e)] \\
 &= \sqrt{n} [n^{-1} \#\{i \leq n: y_i - x_i' \beta \leq e + x_i' (\hat{\beta}_n - \beta)\} - \Phi(e)] \\
 &= \sqrt{n} [n^{-1} \#\{i \leq n: y_i - x_i' \beta \leq e + x_i' (\hat{\beta}_n - \beta)\} - n^{-1} \sum_{i=1}^n \Phi\{e + x_i' (\hat{\beta}_n - \beta)\}] \\
 &\quad + n^{-1/2} \sum_{i=1}^n [\Phi\{e + x_i' (\hat{\beta}_n - \beta)\} - \Phi(e)] \\
 &= w_n(e) + z_n(e), \text{ say.}
 \end{aligned}$$

It can be shown that the terms $x_i' (\hat{\beta}_n - \beta)$ in $w_n(e)$ have asymptotically negligible effect, and hence that $w_n \rightarrow w$, the standard empirical process for independent, identically distributed observations, with

$$\text{Cov}\{w(e_1), w(e_2)\} = \min\{\Phi(e_1), \Phi(e_2)\} - \Phi(e_1) \Phi(e_2).$$

Further, by a Taylor's approximation with $\phi(e) = d\Phi(e)/de$

$$\begin{aligned}
 z_n(e) &\doteq n^{-1/2} \phi(e) \sum_{i=1}^n x_i' (\hat{\beta}_n - \beta) \\
 &= n^{-1/2} \phi(e) \sum_{i=1}^n (y_i - x_i' \beta)
 \end{aligned}$$

so $z_n \rightarrow z$ which is $\phi(e)$ times a standard normal variate.

Thus $w_n \rightarrow \hat{w} = w + z$. But $\hat{w}$ and z are independent since for each n $\hat{w}_n$ is a function of $\hat{e}_{1n}, \hat{e}_{2n}, \dots, \hat{e}_{nn}$ and z_n is a function of $\hat{\beta}_n - \beta$. Hence the covariance function of $\hat{w}$ plus the covariance function of z is equal to the covariance function of w given above. Thus

$$\text{Cov}\{\hat{w}(e_1), \hat{w}(e_2)\} = \min\{\Phi(e_1), \Phi(e_2)\} - \Phi(e_1) \Phi(e_2) - \phi(e_1) \phi(e_2)$$

for any $p = 0, 1, \dots$.

The proof for the general result of Section 2 is basically similar to this, using asymptotic approximations to the maximum likelihood estimators in terms of efficient scores. The special properties of the normal case which made the above argument fairly simple hold in an asymptotic sense more generally. The only part of Durbin's (1973) proof which must be substantially changed is that related to the above assertion that $w_n \rightarrow w$.

APPENDIX

We sketch briefly here the main points of modification of Durbin's (1973) proof to obtain the result of Section 2. Define $\hat{y}_n(t)$ as in the present paper, and I_{θ} , $g(t, \theta)$ as the limits in Section 2. Write $|A|$ for the number of elements in a set A . We will write $\ell_i = \partial \log f_i / \partial \theta$, which is slightly different than Durbin's ℓ . Lemma and theorem numbers are Durbin's.

Lemma 2, equation 13, is modified to $\hat{y}_n(t) = U_n^{(t)} + V_n^{(t)}$, where

$$U_n^{(t)} = n^{-1/2} \left[|\{i \leq n : F_i(y_i; \theta) \leq \hat{t}_{ni}(t)\}| - \sum_{i=1}^n \hat{t}_{ni}(t) \right]$$

$$V_n^{(t)} = n^{-1/2} \sum_{i=1}^n \{\hat{t}_{ni}(t) - t\}$$

$$\hat{t}_{ni}(t) = F_i\{F_i^{-1}(t; \hat{\theta}_n); \theta\}.$$

It takes no essential changes in the first part of the proof of Lemma 2 to show that

$$\sup_t \left| V_n^{(t)} - I_{\theta}^{-1} n^{-1/2} \sum_{i=1}^n \ell_i(y_i; \theta) g(t, \theta) \right| \xrightarrow{P} 0.$$

Lemma 1 is modified to treat $U_n^{(t)}$ as follows. Define

$$y_n(t) = n^{-1/2} [|\{i \leq n : F_i(y_i; \theta) \leq t\}| - nt],$$

which has the same distribution as $y_n(t)$ of Lemma 1. The only really major change in the entire proof is in the necessary argument to show that

$\sup_t |U_n^{(t)} - y_n(t)| \xrightarrow{P} 0$. We assume this for now, and return to this proof below.

Thus we have that $\sup_t |\hat{y}_n(t) - z_n(t)| \xrightarrow{p} 0$, where

$$z_n(t) = y_n(t) + I_{\theta}^{-1} n^{-1/2} \sum_{i=1}^n \ell_i(y_i; \theta) g(t, \theta).$$

The arguments of Lemma 3, and those in the paragraph preceding Theorem 2, can be modified in a fairly straightforward manner to see that the finite dimensional distributions of z_n converge to normal with covariance function (2.1) of the present paper. Furthermore, the arguments of Lemma 4 apply without major modification since our z_n is not fundamentally different than Durbin's.

We now return to the modification of Lemma 1. The new proof is rather involved, and will only be briefly sketched here. It will first be shown that for any fixed t and $\epsilon > 0$

$$\Pr\{|U_n(t) - y_n(t)| > \epsilon\} = O_p(n^{-1/2}). \quad (A.1)$$

It is crucial to use the fact that for any fixed t , $\hat{t}_{ni}(t) - t = a_i(\hat{\theta}_n)$, a function of $\hat{\theta}_n$, and that from our assumptions there exists a finite M such that with probability arbitrarily near one

$$\max_{1 \leq i \leq n} |a_i(\hat{\theta}_n)| \leq M |\hat{\theta}_n - \theta| = O_p(n^{-1/2}). \quad (A.2)$$

By considering separately the cases that $\hat{t}_{ni}(t) \geq t$ and $\hat{t}_{ni}(t) < t$ the problem can be reduced to one of showing that if $Z_1, Z_2, \dots$ are independent uniform $[0,1]$ variates, with $\hat{\theta}_n$ depending on $Z_1, Z_2, \dots, Z_n$ and $\hat{\theta}_n - \theta = O_p(n^{-1/2})$, then for positive functions $a_i(\cdot)$ satisfying (A.2)

$$\Pr \left\{ h_n(\hat{\theta}_n) = n^{-\frac{1}{2}} \left[|\{i \leq n : Z_i \leq a_i(\hat{\theta}_n)\}| - \sum_{i=1}^n a_i(\hat{\theta}_n) \right] > \varepsilon \right\} = o_p(n^{-\frac{1}{2}}).$$

For any fixed sequence $\hat{\theta}_n = \theta + o_p(n^{-\frac{1}{2}})$ this follows from Chebychev's inequality, but in fact $\hat{\theta}_n$ depends on $Z_1, Z_2, \dots, Z_n$.

Let $\{K_n\}$ be a sequence of neighborhoods of the true θ of radius $O(n^{-\frac{1}{2}})$ such that except for negligible probability $\hat{\theta}_n \in K_n$ for all $n = 1, 2, \dots$. One can show that for any $\varepsilon > 0$ there is an m such that for each $n = 1, 2, \dots$ there are m points $\theta_{n1}, \theta_{n2}, \dots, \theta_{nm}$ in K_n such that

$$\sup_{\theta \in K_n} |h_n(\theta)| \leq \max_{1 \leq i \leq m} |h_n(\theta_{ni})| + \varepsilon.$$

Since m does not depend on n , Chebychev's inequality can be used to show that

$$\Pr \left\{ \max_{1 \leq i \leq m} |h_n(\theta_{ni})| < \varepsilon \right\} = o_p(n^{-\frac{1}{2}}).$$

Having established the convergence for fixed t , we now show that it is uniform in t . An argument similar to that yielding (A.1) shows that as $s \rightarrow t$

$$\max_{1 \leq i \leq n} \min\{F_i(t; \theta), 1 - F_i(t; \theta)\} |\hat{t}_{ni}(t) - t - \{\hat{t}_{ni}(s) - s\}| = o_p(|s - t| n^{-\frac{1}{2}}). \quad (A.3)$$

Writing $\varepsilon_n(t) = U_n(t) - y_n(t)$, an argument similar to the one sketched above, now using (A.3), shows that uniformly in t and $i = 1, 2, \dots$, as $\Delta \rightarrow 0$,

$$\min\{F_i(t, \theta), 1-F_i(t, \theta)\} \Pr\left\{\sup_{0 < s < \Delta} |\varepsilon_n(s) - \varepsilon_n(t)| > \varepsilon\right\} = O_p(\Delta n^{-\frac{1}{2}}). \quad (A.4)$$

Now consider a partition of the unit interval, $0 = t_0 < t_1 < \dots < t_k = 1$. Then

$$\Pr\left\{\sup_{0 < t < 1} |\varepsilon_n(t)| > \varepsilon\right\} \leq \sum_{i=1}^{k-1} \Pr\{|\varepsilon_n(t_i)| > \varepsilon\} + \sum_{i=1}^k \Pr\left\{\sup_{t_{i-1} < t < t_i} |\varepsilon_n(t) - \varepsilon_n(t_i)| > \varepsilon\right\}.$$

For a sequence of such partitions with $O(n^{\frac{1}{n}})$ elements, we can use (A.1) to bound the first sum and (A.4) to bound the second.

REFERENCES

- Bradley, R. A. and Gart, J. J. (1962). The asymptotic properties of maximum likelihood estimators when sampling from associated populations. Biometrika 49, 205-14.
- Cox, D. R. and Hinkley, D. V. (1974). Theoretical Statistics. London: Chapman and Hall.
- Dahiya, R. C. and Gurland, J. (1972). Pearson chi-square test of fit with random intervals, I. Null case. Biometrika 59, 147-53.
- Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. Annals of Statistics 1, 279-90.
- Durbin, J. (1973b). Distribution Theory for Tests Based on the Sample Distribution Function. Philadelphia: Society for Industrial and Applied Mathematics.
- Moore, D. S. (1977). Generalized inverses, Wald's method, and the construction of chi-squared tests of fit. J. Amer. Statist. Assn. 72, 131-37.
- Pearson, E. S. and Hartley, H. O. (1972). Biometrika Tables for Statisticians, Vol. 2. Cambridge: Cambridge University Press.
- Rao, K. C. and Robson, D. S. (1974). A chi-square statistic for goodness-of-fit tests within the exponential family. Communications in Statistics 3, 1139-53.
- Stephens, M. A. (1977). Goodness-of-fit for the extreme value distribution. Biometrika 64, 583-88.
- Thomas, D. R. and Pierce, D. A. (1978). Neyman's smooth goodness-of-fit test when the hypothesis is composite. To appear in J. Amer. Statist. Assn.