The Annals of Statistics
1982, Vol 10, No. 2, 475–478

THE ASYMPTOTIC EFFECT OF SUBSTITUTING ESTIMATORS FOR PARAMETERS IN CERTAIN TYPES OF STATISTICS[1]

BY DONALD A. PIERCE

Oregon State University

In a variety of statistical problems, one is interested in the limiting distribution of statistics $\hat{T}_n = T_n(y_1, y_2, \cdots, y_n; \hat{\lambda}_n)$, where $\hat{\lambda}_n$ is an estimator of a parameter in the distribution of the y_i and where the limiting distribution of $T_n = T_n(y_1, y_2, \cdots, y_n; \lambda)$ is relatively easy to find. For cases in which the limiting distribution of T_n is normal with mean independent of λ, a useful method is given for finding the limiting distribution of $\hat{T}_n$. A simple application to testing normality in regression models is given.

1. Introduction and main result. The result here is of interest primarily in hypothesis testing, using statistics which involve substituting estimates for nuisance parameters. Special cases have been derived before; some of these are mentioned later. The aim is to present the result in a quite general setting, using tools which avoid ordinarily tedious calculations. There are close ties to results of Randles (1982), which will be pointed out.

Let $y_1, y_2, \cdots$ be a sequence of random variables whose joint distribution depends on a parameter λ, possibly vector-valued. It is not necessary that these observations be independent or identically distributed. Let $\hat{\lambda}_n = \hat{\lambda}_n(y_1, \cdots, y_n)$ be an asymptotically normal and efficient sequence of estimators. It is desired to find the limiting distribution of a statistic $\hat{T}_n = T_n(y_1, \cdots, y_n; \hat{\lambda}_n)$, where at the true λ the corresponding sequence $T_n = T_n(y_1, \cdots, y_n; \lambda)$ has a limiting normal distribution *with an asymptotic mean which is constant in* λ. Without loss of generality, this asymptotic mean will be taken as zero.

A generalization is possible to statistics $S_n = S_n(y_1, \cdots, y_n; \lambda)$ with asymptotic mean $\mu_n(\lambda)$ which is not constant, by taking $T_n = S_n - \mu_n(\lambda)$. But then the result will be for the limiting distribution of $\hat{S}_n - \mu_n(\hat{\lambda}_n)$, rather than $\hat{S}_n - \mu_n(\lambda)$. The emphasis of Randles (1982) is on the more conventional latter case. The former case may be useful, however, when λ is a nuisance parameter.

It is apparent that under regularity conditions $\hat{T}_n$ is asymptotically normal with mean zero. The primary point here involves calculation of its asymptotic variance. Three regularity conditions, and then the main result, are now given.

The first assumption is that for every λ there is joint convergence in law to normality:

$$(1.1)\qquad \begin{bmatrix} \sqrt{b}\, T_n \\ \sqrt{n}(\hat{\lambda}_n - \bar{\lambda}) \end{bmatrix} \to \begin{bmatrix} L \\ \hat{\delta} \end{bmatrix} \sim N\left[0, \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}\right].$$

The dispersion matrices may depend continuously on λ. It is assumed for convenience that V_{22} is nonsingular.

The second condition is that there is a matrix B, possibly depending continuously on λ, such that

$$(1.2)\qquad \sqrt{n}\hat{T}_n = \sqrt{n}T_n + B\sqrt{n}(\hat{\lambda}_n - \lambda) + o_p(1).$$

When T_n is differentiable in λ, this ordinarily follows from a first-order expansion, and $B = \lim E(\partial T_n/\partial\lambda)$. More generally, however, (1.2) often holds when T_n is only asymp-

Received April 1981; revised August 1981.

[1] Research supported by PHS Grant CA27532, National Cancer Institute, DHHS.

AMS 1970 *subject classification.* Primary 62E20, secondary 62F05.

Key words and phrases. Asymptotic distributions, goodness-of-fit tests, nuisance parameters, residuals.

totically smooth in λ. For example, this occurs in applications mentioned later where T_n is a functional of the empirical distribution function of probability integral transformations $\{F_i(y_i; \lambda), i = 1, \cdots, n\}$. Section 4 of Pierce and Kopecky (1979) illustrates an approach. Randles (1982) in effect provides a more general treatment of the validity of (1.2). Note that his function $\mu(\cdot)$ is not the asymptotic mean of T_n, but is rather a function of λ and another variable.

The third condition is that $\hat{\lambda}_n$ is asymptotically efficient. This is defined as follows in order to dispense with superefficient estimators, which quite generally can improve on regular efficient estimators only on sets of Lebesgue measure zero; see, for example, LeCam (1953), Zacks (1971, Chapter 4). There should be no estimator $\hat{\lambda}_n^*$, asymptotically normal as above, with dispersion matrix V_{22}^* such that (i) $V_{22} - V_{22}^*$ is nonnegative definite for all λ, and (ii) for some x and all λ in some open interval, $x' V_{22}^* x < x' V_{22} x$.

The main result is that under these conditions

$$\sqrt{n}\hat{T}_n \to \hat{L} \sim N(0, V_{11} - B V_{22} B'). \tag{1.3}$$

The essence of the argument, given later, is as follows. From (1.1) and (1.2) it follows that $\sqrt{n}\hat{T}_n \to \hat{L}$, which is normal with the distribution of $L + B\hat{\delta}$. The result (1.3) then follows from the fact that $\hat{L}$ and $\hat{\delta}$ are independent. That this is true is an asymptotic version of the fundamental result that minimum variance unbiased estimators ($\hat{\delta}$) are uncorrelated with statistics having constant expectation ($\hat{L}$); see, for example, Rao (1973, Section 5a.2).

Observe that $\mathrm{Var}(\hat{L}) = \mathrm{Var}(L + B\hat{\delta})$ would ordinarily be calculated as $V_{11} + BV_{22}B' + V_{12}B' + BV_{21}$. Computation of V_{12} is often difficult, involving representation of $\hat{\delta}$ in terms of efficient scores. The two expressions for $\mathrm{Var}(\hat{L})$ suggest that $B = -V_{12}V_{22}^{-1}$, which is discussed later.

2. An illustrative application. Consider the standard simple linear regression model for independent observations $y_i \sim N(\alpha + \beta x_i, \sigma^2)$, where the x_i are fixed and $\lambda = (\alpha, \beta, \sigma)$. Using maximum likelihood estimators, consider the limiting distribution of the skewness statistic

$$\sqrt{n}\hat{T}_n = n^{-1/2} \textstyle\sum_{i=1}^{n} \{(y_i - \hat{\alpha}_n - \hat{\beta}_n x_i)/\hat{\sigma}_n\}^3.$$

It is easily calculated that the limiting distribution of $\sqrt{n}T_n$, involving true parameters and hence true errors, is $N(0, 15)$. The result (1.3) will be used to show that the limiting distribution of $\sqrt{n}\hat{T}_n$ is $N(0, 6)$, assuming ordinary regularity of the x_i so that $\sqrt{n}\hat{\lambda}_n$ is asymptotically normal.

Before giving this calculation it is interesting to note that similar calculations yield precisely the same limiting distribution when either (i) β and x_i are replaced by vectors of fixed dimension, or (ii) the terms βx_i are omitted so that the y_i are identically distributed. More general results of this nature are given in Pierce and Kopecky (1979). It is also noted that Anscombe (1961) gave formulas for the exact variance of $\hat{T}_n$, but the asymptotic result is less than obvious from that approach.

It will be assumed that $\Sigma x_i/n \to \bar{x}$ and $\Sigma(x_i - \bar{x})^2/n \to s^2$. The matrix B can be found by a first-order approximation to be

$$B = (-3/\sigma)[1, \bar{x}, 0].$$

Further, the upper left 2×2 part of V_{22} is

$$(\sigma^2/s^2)\begin{bmatrix} s^2 + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix},$$

and as previously noted, $V_{11} = 15$. Direct calculation using (1.3) shows that $\sqrt{n}\hat{T}_n \to \hat{L} \sim N(0, 6)$.

3. Discussion of applications. In many applications $\hat{T}_n$ is a statistic for testing validity of the model for the y_i, where T_n is a statistic that could be rather easily used if λ were known. Of particular interest in this context are cases where T_n is a function of the probability integral transformations $F_i(y_i; \lambda)$. The results here are then closely related to those of Durbin (1973), Pierce and Kopecky (1979), and Loynes (1980). Somewhat more generally, the results here are useful in the analysis of generalized residuals, as developed by Cox and Snell (1968). In fact, (1.3) yields results equivalent to, but much simpler than, equations (19) and (20) of Cox and Snell (1971), where the simplification discussed at the end of Section 1 above was understandably overlooked.

Another general area of application, although overlapping somewhat with the above, is where $\hat{T}_n$ relates to inference about parameters other than λ. For example, $n\hat{T}_n$ may be the derivative of the log likelihood function with respect to the additional parameters, evaluated at an hypothesized value for them and at $\lambda = \hat{\lambda}_n$. In this sense the results here are closely related to standard results on use of efficient scores for testing composite hypotheses; see, for example, Cox and Hinkley (1974, page 323). It should be emphasized that when λ does not completely specify the distribution of the y_i, then (1.3) is only valid where $\hat{\lambda}_n$ is asymptotically efficient when treating the remaining parameters as known.

Another type of situation where these results may be useful is in developing distribution theory for tests of separate hypotheses, as suggested by Cox (1962).

An interesting application of the result is given by Habib and Thomas (1981), where asymptotic distribution theory for goodness-of-fit tests for censored data, with estimated parameters, is developed.

Finally, the results of Pierce and Kopecky (1979) and Loynes (1980) relate to asymptotic distributions of statistics which are permutationally symmetric functions of generalized residuals $\{F_i(y_i; \hat{\lambda}); i = 1, 2, \cdots, n\}$. This symmetry may be an extremely limiting feature for statistics intended to test adequacy of a generalized regression model; departures of interest very often will involve relationships between the generalized residuals and potentially explanatory regression variables. The results of this paper apply without the restriction of this type of symmetry.

4. Proof and related calculations. It is clear that (1.1) and (1.2) imply $\sqrt{n}\hat{T}_n \to \hat{L}$, which is normal with the distribution of $L + B\hat{\delta}$. The key fact that $E(\hat{L})$ being free of λ implies that $\mathrm{Cov}(\hat{L}, \hat{\delta}) = 0$ can be established as follows.

Write $\hat{V}_{11} = \mathrm{Var}(\hat{L})$ and $\hat{V}_{21} = \mathrm{Cov}(\hat{\delta}, \hat{L})$. For simplicity, assume that $\hat{V}_{11}$ is nonsingular. Consider the estimator suggested by the regression of $\hat{\lambda}_n$ on $\hat{T}_n$,

$$\hat{\lambda}_n^* = \hat{\lambda}_n - \hat{V}_{21}\hat{V}_{11}^{-1}\hat{T}_n,$$

where $\hat{V}_{21}$ and $\hat{V}_{11}$ are evaluated at $\hat{\lambda}_n$ if necessary, with asymptotically negligible effect. Then for all λ, $\sqrt{n}(\hat{\lambda}_n^* - \lambda) \to N(0, V_{22}^*)$, with $V_{22}^* = V_{22} - \hat{V}_{21}\hat{V}_{11}^{-1}\hat{V}_{12}$. Thus $V_{22} - V_{22}^*$ is nonnegative definite for all λ, and if $\hat{V}_{21}$ were nonzero for some λ, simple continuity arguments show that then $x'V_{22}^*x < x'V_{22}x$ for some x and all λ in an open interval. Thus the asymptotic efficiency of $\hat{\lambda}_n$ implies that $\hat{V}_{12} = 0$, which completes the argument.

The above argument seems useful for application to moderately complicated models, since the essence of the argument is separated from rather arbitrary regularity conditions which would imply the hypotheses used here. The following formal calculations along more conventional lines may be helpful, however, in understanding why the result is true.

Suppose that T_n is differentiable in λ and that (1.2) results from a first-order approximation, so that (i) $B = \lim E(\partial T_n/\partial\lambda)$. Assume also that the usual asymptotic approximation (ii) $U_n(\lambda)/\sqrt{n} \doteq V_{22}^{-1}\sqrt{n}(\hat{\lambda}_n - \lambda)$ holds, where $U_n(\lambda)$ is the derivative of the log likelihood $\ell_n(\lambda)$. If $E(T_n)$ is constant in λ, then

$$0 = \partial E(T_n)/\partial\lambda = \partial\left[\int T_n \exp\{\ell_n(\lambda)\}\, d(y_1, \cdots, y_n)\right]/\partial\lambda$$
$$= \int (\partial T_n/\partial\lambda)\exp\{\ell_n(\lambda)\}\, d(y_1, \cdots, y_n)$$
$$+ \int T_n U_n(\lambda)\exp\{\ell_n(\lambda)\}\, d(y_1, \cdots, y_n),$$

so that (iii) $E(\partial T_n/\partial\lambda) = -\mathrm{Cov}(\sqrt{n}T_n, U_n/\sqrt{n})$.

Assuming that these covariances converge to that of the limiting distribution, then (i), (ii), and (iii) together yield that

$$B = \lim E(\partial T_n/\partial\lambda) = -\lim \mathrm{Cov}(\sqrt{n}T_n, U_n/\sqrt{n})$$
$$= -\lim \mathrm{Cov}\{\sqrt{n}T_n, V_{22}^{-1}\sqrt{n}(\hat{\lambda}n - \lambda)\} = -\mathrm{Cov}(L, V_{22}^{-1}\hat{\delta}) = -V_{12}V_{22}^{-1}.$$

The calculations suggested at the end of Section 1 then yield the result.

Acknowledgement. The author expresses appreciation to the editor for suggesting improvements in presentation.

REFERENCES

ANSCOMBE, F. J. (1961). Examination of residuals. *Fourth Berkeley Symposium on Statistics and Probability*, Vol. I, 1–36. Univ. of California Press, Berkeley.

COX, D. R. (1962). Further results on tests of separate families of hypotheses. *J. Roy. Statist. Soc. B* **24** 406–424.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics.* Chapman and Hall, London.

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

COX, D. R. and SNELL, E. J. (1971). On test statistics calculated from residuals. *Biometrika* **58** 589–594.

DURBIN, J. (1973). Weak convergence of the sample distribution function where parameters are estimated. *Ann. Statist.* **1** 279-90.

HABIB, G. and THOMAS, D. R. (1981). Chi square goodness-of-fit tests for censored data. Unpublished.

LECAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. of California Publ. in Statist.* **1** 277–330.

LOYNES, R. M. (1980). The empirical distribution function of residuals from generalized regression. *Ann. Statist.* **8** 285–298.

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.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd Ed. Wiley, New York.

ZACKS, S. (1971). *The Theory of Statistical Inference.* Wiley, New York.

DEPARTMENT OF STATISTICS
OREGON STATE UNIVERSITY
CORVALLIS, OREGON 97331