Austral. J. Statist., 18 (1, 2), 1976, 53–61

THE RELATIONSHIP BETWEEN A CLASS OF ASYMPTOTICALLY NORMAL ESTIMATORS AND GOODNESS OF FIT TESTS[1]

G. M. TALLIS AND P. L. CHESSON[2]
Department of Statistics, University of Adelaide

Summary

This paper considers a class of asymptotically normal estimators defined implicitly in terms of asymptotically normal statistics. The estimators are used as a basis for a general goodness of fit test where asymptotic null and non-null distributions are obtained. Certain standard tests are obtained as special cases of these results.

1. Introduction

Let X_n be a sequence of random vectors with components $X_i^{(n)}$, $i=1,2,\ldots,k$. It is assumed that $n^{\frac{1}{2}}(X_n-\mu(\theta))\xrightarrow{L}N(\gamma, V(\theta))$ where $\theta\varepsilon\Omega$ is a q-dimensional vector of unknown parameters, γ is in the range of V and $q<\text{rank } V\equiv\rho(V)=l\leqslant k$.

Define the general class of estimators of θ, θ_n, implicitly by

$$\Phi(X_n, \theta_n)=0,\ \Phi'=(\Phi_1,\ldots,\Phi_q)$$

where Φ satisfies the Implicit Function Theorem at $(\mu(\theta), \theta)$ i.e. $\Phi(\mu(\theta), \theta)=0$ and the continuous derivatives satisfy

$$\det\ (\partial\Phi_i(\mu(\theta),\theta)/\partial\theta_j)\neq 0.$$

Then θ_n is consistent for θ and $n^{\frac{1}{2}}(\theta_n-\theta)$ is asymptotically normally distributed.

In the paper the use of θ_n in certain goodness of fit tests is examined. Specifically, we investigate situations where goodness of fit is measured by closeness of X_n to $\mu(\theta_n)$. The aim of the work is to establish general procedures which include some standard tests as special cases. The results are later generalized to allow the usual norm, $n^{\frac{1}{2}}$, to be replaced by a sequence of random variables I_n. The structure is also used to discuss minimum quadratic form estimation first introduced, in a limited context, by Gurland and Dahiya (1972).

2. Main Result

To obtain the main result of this paper we make use of the following important lemma. The proof may be found in Rao and Mitra (1971, p. 173).

[1] Manuscript received June 3, 1975; final revision March 8, 1976.

[2] The work of this author was supported by a C.S.I.R.O. Postgraduate Studentship.

Lemma 1. *Let Y be distributed as $N(\mu,\Sigma)$ and Σ^g be any generalized inverse (g-inverse) of Σ i.e. $\Sigma\Sigma^g\Sigma=\Sigma$. If μ is in the range of Σ, then $Y'\Sigma^gY$ has a $\chi^2(d,\lambda)$ distribution where $d=\rho(\Sigma)$ and $\lambda=\mu'\Sigma^g\mu$.*

Throughout this paper it is assumed that $\mu(\theta)$ has continuous first order partial derivations with respect to $\theta\varepsilon\Omega$ and $\Phi(X_n,\theta)$ has continuous first order partial derivatives with respect to X_n and $\theta\varepsilon\Omega$.

We now define for i, $r=1,2,\ldots,q$ and $j=1,2,\ldots,k$, $M=[\partial\Phi_i/\partial X_j]$; $N=[-\partial\Phi_i/\partial\theta_r]$; $Q=[\partial\mu_j(\theta)/\partial\theta_r]$ and $\Sigma=(I-A')\;V(I-A)$ where $A'=QN^{-1}M$. We make the further assumptions that the range of $Q(\theta)$ is contained in the range of $V(\theta)$ and that N^{-1} exists for $\theta\varepsilon\Omega$.

When $\rho(V)=l<k$ we introduce a $(k-l)\times k$ matrix H such that $VH'=0$ and $\rho\begin{pmatrix}V\\H\end{pmatrix}=k$. Such a matrix always exists.

Theorem 1. *If $W_n=n(X_n-\mu(\theta_n))'\;(\Sigma+M'M+H'H)^{-1}\;(X_n-\mu(\theta_n))$ then W_n is asymptotically distributed as $\chi^2(l-q,\lambda)$, where*

$$\lambda=\gamma'(I-A)\;[\Sigma+M'M+H'H]^{-1}\;(I-A')\gamma.$$

Note (i) M and N are evaluated at $(\mu(\theta),\theta)$ and Q, V and H at θ. However, if V is a continuous function of θ, θ can be replaced by a consistent estimator in $(\Sigma+M'M+H'H)^{-1}$ without altering the asymptotic result. For instance, θ_n is a suitable candidate.

(ii) Implicit differentiation of $\Phi(\mu(\theta),\theta)$ with respect to θ shows that $MQ=N$ and $\rho(Q)=q$.

(iii) Since $\begin{pmatrix}M\\H\end{pmatrix}(QH')=\begin{bmatrix}N & MH'\\O & HH'\end{bmatrix}$, $\rho(M'\vdots H')=q+k-l$

(iv) The assumption that N^{-1} exists is a restatement of the Implicit Function Theorem property of Φ.

Proof of Theorem 1. The essence of the proof is that since $n^{\frac{1}{2}}(X_n-\mu(\theta))\overset{L}{\rightarrow}N(\gamma,V)$, $X_n\overset{P}{\rightarrow}\mu(\theta)$. The Implicit Function Theorem guarantees the existence of a suitably regular function f such that $\theta_n=f(X_n)$. This is enough to ensure that

$$n^{\frac{1}{2}}(X_n-\mu(\theta_n))\overset{L}{\rightarrow}N((I-A')\gamma,\Sigma)$$

and Lemma 1 then gives the required result. We now provide details.

From the Implicit Function Theorem, it follows that there exists a neighbourhood of $(\mu(\theta),\theta)$, K, and a vector function f such that $\theta_n=f(X_n)$ whenever $(X_n,\theta_n)\varepsilon\mathrm{K}$. The components of f have continuous first order partial derivatives and $f(\mu(\theta))=\theta$. Moreover, if Df is the matrix $[\partial f_i(x)/\partial x_j]$, then $Df=N^{-1}M$.

Thus the vector function $h(x)=x-\mu(f(x))$ is continuously differentiable and $Dh=I-QN^{-1}M=I-A'$. Hence

$$n^{\frac{1}{2}}\,h(X_n)=n^{\frac{1}{2}}(X_n-\mu(\theta_n))$$

has the same asymptotic distribution as $n^{\frac{1}{2}}\;(I-A')(X_n-\mu(\theta))\overset{L}{\rightarrow}N((I-A')\gamma,\Sigma)$ where

$$\Sigma=(I-A')\;V(I-A).$$

Now let Σ^g be a g-inverse of Σ and notice that since γ is in the range of V, $(I-A')\gamma$ is in the range of Σ. It follows immediately from Lemma 1 that

$$n(X_n-\mu(\theta_n))'\,\Sigma^g\,(X_n-\mu(\theta_n))\xrightarrow{L}\chi^2(d,\lambda)$$

where $d=\rho(\Sigma)$ and $\lambda=\gamma'(I-A')\,\Sigma^g\,(I-A)\gamma$.

Now $V=P'I^{(l)}I^{(l)}P$ where P is non-singular and $I^{(l)}=\text{diag}(1,\ldots,1,0,\ldots,0)$, there being l 1's, and $\Sigma=(I-A')P'I^{(l)}I^{(l)}P(I-A)=E'E$, say.

Clearly $\rho(\Sigma)=\rho(E)$, $EM'=0$, $EH'=0$ and hence the null space of E has dimension greater than or equal to $\rho(M'\vdots H')=q+k-l$. Thus $\rho(E)\leqslant k-q-k+l=l-q$.

But for $k\times k$ matrices E_1 and E_2, $\rho(E_1)=r$ and $\rho(E_2)=s$, $\rho(E_1E_2)\geqslant r+s-k$. Put $E_1=I^{(l)}P$, $E_2=(I-A)$, then $\rho(I^{(l)}P)=l$ and $\rho(I-A)=tr(I-A)=k-q$ and $\rho(E)\geqslant l+k-q-k=l-q$. Finally, $\rho(\Sigma)=l-q$, $\rho(E'\vdots M'\vdots H')=k$ and a g-inverse of Σ is $(\Sigma+M'M+H'H)^{-1}$, Rao (1973, p. 34).

We may note, from the conditions of Theorem 1, that there is essentially a unique solution of the equation $\Phi(X_n,\theta_n)=0$ which is a consistent estimator of θ. This is because, if θ_n satisfies the equation and is a consistent estimator, (X_n,θ_n) belongs to **K** with probability tending to 1 and so $\theta_n=f(X_n)$ with probability tending to 1. However, in the presence of multiple solutions, we are still left with the problem of deciding which is the consistent estimator.

Some of these solutions can be eliminated, at least in theory, on the grounds that they do not correspond to functions determined implicitly at points $(\mu(\theta),\theta)$ for $\theta\varepsilon\Omega$. To discriminate between the remaining solutions we need some condition on $\mu(\theta)$. One possible condition is, for fixed θ and every $\delta>0$ there exists an $\varepsilon>0$ such that $||\theta-\theta_1||\geqslant\delta$ implies $||\mu(\theta)-\mu(\theta_1)||\geqslant\varepsilon$. This condition embodies the same idea as Rao's assumption 1.1, Rao (1973, p. 359).

If f is the function implicitly determined by Φ at $(\mu(\theta),\theta)$ there are spherical neighbourhoods **M** and **N** of $\mu(\theta)$ and θ respectively such that $\{(x,f(x))|x\varepsilon\mathbf{M}\}$ is the set of zeros of Φ in $\mathbf{M}\times\mathbf{N}$. Suppose g is another function determined implicitly by Φ. Either g is part of an extension of f to a continuously differentiable function f^* satisfying $\Phi(x,f^*(x))=0$; or for every $x\varepsilon\mathbf{M}$ $g(x)\notin\mathbf{N}$. Only the latter case is of concern.

Let δ be the radius of **N**, then for $x\in\mathbf{M},||\theta-g(x)||\geqslant\delta$ and so $||\mu(\theta)-\mu(g(x))||\geqslant\varepsilon$ which implies that if $||x-\mu(\theta)||\leqslant\varepsilon/2$, then $||x-\mu(g(x))||\geqslant\varepsilon/2$. We see from this that $X_n\xrightarrow{P}\mu(\theta)$ implies that $||X_n-\mu(g(X_n))||\geqslant\varepsilon/2$ with probability tending to one where ε is independent of g.

On the other hand $||X_n-\mu(f(X_n))||\xrightarrow{P}0$. Hence an asymptotically acceptable means of discriminating between the remaining solutions is to choose θ_n as the solution of $\Phi(X_n,\theta_n)=0$ corresponding to implicit functions determined at points $(\mu(\theta),\theta),\theta\varepsilon\Omega$, for which $||X_n-\mu(\theta_n)||$ is minimal. Under this procedure Theorem 1 remains valid.

It seems more natural to use W_n for the purpose of choosing the appropriate solution of $\Phi(X_n,\theta_n)=0$. This also leads to a valid asymptotic procedure if there is only a finite number of implicit functions. In general, however, different conditions on $\mu(\theta)$, or alternatively conditions on Φ and V, are necessary if W_n is to be used to choose θ_n.

3. Random Norms and Numerical Solution of $\Phi(X_n,\theta_n)=0$.

In some practical situations the usual norm $n^{\frac{1}{2}}$ is replaced by a sequence of random variables I_n. That is we assume a limit law of the form

$$I_n(X_n-\mu(\theta))\xrightarrow{L}N(\gamma,V)$$

where $I_n\xrightarrow{P}\infty$.

This kind of behaviour is observed as a result of sequential estimation procedures (see Anscombe (1952)) or in connection with point estimation of parameters from realizations of stochastic processes (see Heyde (1973)). On the other hand, more mundane situations covered by these extensions arise, for example, when the sample size N is beyond the control of the experimenter and is determined by a latent random process with parameter α. In this case N may $\xrightarrow{P}\infty$ as $\alpha\to\infty$, say, α being unknown. One may wish to use the random norm $N^{\frac{1}{2}}$.

In the present case there is no additional difficulty in defining a consistent estimator, θ_n, implicitly by the equation

$$\Phi(X_n,\theta_n)=0$$

since the assumptions of this section guarantee the essential requirement $X_n\xrightarrow{P}\mu(\theta)$. Following similar reasoning to that given by Rao (*loc. cit.*, p. 387), for the special case $I_n=n^{\frac{1}{2}}$, it is easy to see that $I_n(X_n-\mu(\theta_n))$ has the same asymptotic distribution as

$$(I-A')I_n(X_n-\mu(\theta)).$$

This yields a corollary to Theorem 1.

Corollary. *Let the conditions on μ, V and Φ imposed in Section 2 hold, then*

$$W_n=I_n^2(X_n-\mu(\theta_n))'\,(\Sigma+M'M+H'H)^{-1}\,(X_n-\mu(\theta_n))$$

is asymptotically distributed as $\chi^2(l-q,\lambda)$ with λ as in Theorem 1.

In some cases I_n may be a function of θ, $I_n(\theta)$ say. This is certainly likely when I_n is the information about θ as defined by Heyde (*loc. cit.*) for stochastic processes. If $\log I_n(\theta)$ is continuous at θ uniformly in n, almost surely, then $\frac{I_n(\theta)}{I_n(\theta_n^*)}\xrightarrow{P}1$ where θ_n^* is any consistent estimator of θ e.g. θ_n. In this case the corollary holds with $I_n(\theta_n^*)$ replacing $I_n(\theta)$.

In applications the equations $\Phi(X_n,\theta_n)=0$ may not be soluble in closed form. Numerical procedures will have to be sought and the appropriate methods will probably be dictated by the form of Φ and μ.

However, a general iteration procedure is obtained by using

$$\theta_n^{(j)} = \theta_n^{(j-1)} + N^{-1}\Phi(X_n, \theta_n^{(j-1)})$$

where N is evaluated at $\theta_n^{(j-1)}$. An initial guess, $\theta_n^{(0)}$, is required and this will have to be set on rational grounds. The usual care must be taken in particular cases but, for large $n, \theta_n^{(j)}$ will converge to θ_n with high probability.

4. Specializations

(a) *Standard Goodness of Fit Tests*

Let Y be a random variable with distribution function $F(y,\theta)$, $\theta \varepsilon \Omega$ a q-vector of unknown parameters. For $y_0 = -\infty < y_1 < y_2 < \ldots < y_k = \infty$ put

$$p_i(\theta) = F(y_i, \theta) - F(y_{i-1}, \theta),\ i = 1, 2, \ldots, k.$$

A random sample of size n is taken from $F(y,\theta)$ and N_i is the number in this sample for which $y_{i-1} < Y \leqslant y_i$. The theory of Section 2 will be used to establish certain classical results with regard to the distribution of the test statistic $W_n = \sum_{i=1}^{k} (N_i - np_i(\theta_n))^2 / n\, p_i(\theta_n)$, where θ_n represents certain estimates of θ.

It is known for instance that when $\theta_n = \hat{\theta}_n$, the maximum likelihood estimator, then $W_n \xrightarrow{L} \chi^2(k-q-1)$. This result holds whenever $\hat{\theta}_n$ is replaced by an asymptotically efficient estimator θ_n^* since then

$$n^{\frac{1}{2}}(\theta_n^* - \hat{\theta}_n) \xrightarrow{P} 0.$$

In the notation of Section 2, put $\mu(\theta) = p(\theta)$, $X_i^{(n)} = N_i/n$, $V = \Lambda^{-1}(I - \varphi\varphi')\Lambda^{-1}$, $\Lambda^{-1} = \text{diag}\,(p_1^{\frac{1}{2}}(\theta), \ldots, p_k^{\frac{1}{2}}(\theta))$, $\varphi = \Lambda^{-1}1$, and let $H = 1' = (1, \ldots, 1)$. Given an appropriate Φ Theorem 1 states that

$$n(X_n - p(\theta_n))'\ (\Sigma + M'M + 11')^{-1}\ (X_n - p(\theta_n)) \xrightarrow{L} \chi^2(k-q-1).$$

For the non-null case consider a sequence of alternatives by replacing $F(y,\theta)$ with $F_n(y,\theta) = F(y,\theta) + n^{-\frac{1}{2}}G(y)$. Under the alternative nX_n is a multinomial with mean $np(\theta) + n^{\frac{1}{2}}\gamma$ where $\gamma'1 = 0$. It can be shown that $n^{\frac{1}{2}}(X_n - p(\theta)) \xrightarrow{L} N(\gamma, V)$. Notice that γ belongs to the range of V since $\rho(V) = k-1$ and $V1 = 0$. Hence Theorem 1 applies giving the asymptotic distribution of

$$n(X_n - p(\theta_n))'\ (\Sigma + M'M + 11')^{-1}\ (X_n - p(\theta_n)) \text{ as } \chi^2(k-q-1, \lambda)$$

with λ specified by Theorem 1 and the circumstances of the problem.

(i) Maximum Likelihood, $\theta_n = \hat{\theta}_n$.

When θ is estimated by maximum likelihood Φ_i takes the form

$$\Phi_i(X_n, \theta_n) = \sum_{j=1}^{k} X_j^{(n)}\ \partial \ln p_j(\hat{\theta}_n)/\partial \theta_i = 0.$$

The condition $\Phi(p(\theta), \theta) = 0$ of Theorem 1 is satisfied and in this case $M = R'\Lambda$ for $R = \Lambda Q$; $N = R'R = \mathrm{J}$, the information matrix with assumed inverse J^{-1} for $\theta \varepsilon \Omega$; and $\rho(M' \vdots 1) = q+1$.

Let $W_n = n(X_n - p(\hat{\theta}_n))'\Lambda^2 \ (X_n - p(\hat{\theta}_n))$, then using the special forms of M, Q and N

$$\Sigma = \Lambda^{-1}[I - RJ^{-1}R'] \, [I - \varphi\varphi'] \, [I - RJ^{-1}R']\Lambda^{-1}.$$

Straightforward algebra shows that Λ^2 is a g-inverse of Σ and $\rho(\Sigma) = \rho(\Lambda\Sigma\Lambda) = \text{tr}[I - RJ^{-1}R' - \varphi\varphi'] = k - q - 1$ thus establishing the classical result $W_n \xrightarrow{L} \chi^2(k-q-1)$. In the non-null case $W_n \xrightarrow{L} \chi^2(k-q-1, \lambda)$ where $\lambda = \gamma'\Lambda(I - RJ^{-1}R')\Lambda\gamma$.

(ii) A Moment Estimator, $\theta_n = \bar{\theta}_n$.

Let $T = [t_{ij}]$, $t_{ij} = \bar{y}_j^i$, $j = 1, 2, \ldots, k$, $i = 1, 2, \ldots, q$, $\bar{y}_j \varepsilon (y_{j-1}, y_j]$ then a moment estimator of θ is defined by

$$\Phi(X_n, \bar{\theta}_n) = T(X_n - p(\bar{\theta}_n)) = 0.$$

In this case again $\Phi(p(\theta), \theta) = 0$, $M = T$, $N = TQ$ and N^{-1} is assumed to exist. Moreover, $\rho(T' \vdots 1) = q + 1$ and the general results of Theorem 1 can be applied using these special forms of M and N.

(iii) Minimum Chi-square.

This case is covered by the next section.

(b) *Minimum Quadratic Form Estimators.*

Let

$$A(\theta_1, S(\theta_1)) = (X_n - \mu(\theta_1))' S(\theta_1) \ (X_n - \mu(\theta_1))$$

where $S(\theta_1)$ is positive definite for $\theta_1 \varepsilon \Omega$. Then the minimum quadratic form estimator for θ, θ_n^*, is defined by

$$A(\theta_n^*, S(\theta_n^*)) = \min_{\theta_1 \varepsilon \Omega} A(\theta_1, S(\theta_1)).$$

Thus $A(\theta_n^*, S(\theta_n^*))$ is a measure of agreement between X_n and $\mu(\theta)$.

The asymptotic distribution of a certain quadratic form in $X_n - \mu(\theta_n^*)$ will be derived with help of Theorem 1. In order to do this we need conditions which ensure that the minimum occurs in the interior of Ω so that $\partial A(\theta, S(\theta))/\partial\theta_i = 0$ at the minimum.

Suppose $\mu(\theta)$ satisfies the condition at θ that for every $\delta > 0$ there is an $\varepsilon > 0$ such that $||\theta - \theta_1|| \geqslant \delta$ implies $||\mu(\theta) - \mu(\theta_1)|| \geqslant \varepsilon$. Let $\lambda_1(\theta_1)$ and $\lambda_q(\theta_1)$ be maximum and minimum eigenvalues of $S(\theta_1)$ respectively. Suppose that $\lambda_q(\theta_1) > c$ for all $\theta_1 \varepsilon \Omega$, c being a positive constant. Finally, assume that θ lies in the interior of Ω.

Now choose $\delta > 0$ such that the spherical neighbourhood of θ with radius δ is contained in Ω. Let θ_1 satisfy $||\theta - \theta_1|| \geqslant \delta$ so that $||\mu(\theta) - \mu(\theta_1)|| \geqslant \varepsilon$, where ε is independent of θ_1. Suppose $||X_n - \mu(\theta)|| \leqslant \varepsilon/2$ then $||X_n - \mu(\theta_1)|| \geqslant \varepsilon/2$. From a well known inequality for quadratic forms $A(\theta_1, S(\theta_1)) \geqslant (\varepsilon/2)^2 \lambda_q(\theta_1) \geqslant (\varepsilon/2)^2 c^2$. But $A(\theta, S(\theta)) \leqslant \lambda_1(\theta)||X_n - \mu(\theta)||^2$ and it is therefore clear that with probability tending to one as $n \to \infty$, $A(\theta, S(\theta))$ is less than $A(\theta_1, S(\theta_1))$ for $||\theta - \theta_1|| \geqslant \delta$. Thus the global minimum satisfies $||\theta - \theta_n^*|| < \delta$ implying that θ_n^* belongs to the interior of Ω again with probability tending to one.

Note also that δ can be made arbitrarily small and so θ_n^* is consistent for θ.

Set

$$\begin{aligned}\Phi_i(X_n,\theta) &= \partial A(\theta,S(\theta))/\partial\theta_i \\ &= -(X_n-\mu(\theta))'S(\theta)\mu_i(\theta) \\ &\quad +2^{-1}(X_n-\mu(\theta))'S_i(\theta)(X_n-\mu(\theta))\end{aligned}$$

when $\mu_i(\theta)=\partial\mu(\theta)/\partial\theta_i$ and $S_i(\theta)=\partial S(\theta)/\partial\theta_i$. Note that $\Phi(\mu(\theta),\theta)=0$ and in the previous notation $N=Q'SQ$, $M=-Q'S$ and

$$\Sigma=(I-Q(Q'SQ)^{-1}Q'S)\ V(I-SQ(Q'SQ)^{-1}Q').$$

Hence Φ satisfies the Implicit Function Theorem at θ provided it is continuously differentiable and Q is of full rank.

If θ_n^* is in the interior of Ω, an event with probability tending to one, it satisfies

$$\Phi(X_n,\theta_n^*)=0$$

and since θ_n^* is consistent for θ it must correspond to the implicit function determined by Φ at $(\mu(\theta),\theta)$. Thus under the conditions above, there are two asymptotically equivalent ways of defining θ_n^* and Theorem 1 can be applied with $\theta_n=\theta_n^*$.

On the other hand if $\lambda_q(\theta)$ is not bounded away from zero, the first definition of θ_n^* may not yield a useful estimator since θ_n^* may not belong to the interior of Ω or may not even exist. The second definition is still available, however, and leads to a useful estimator since Theorem 1 continues to apply in a particularly satisfying way. To see this note that M and N do not involve derivatives of $S(\theta)$ and any fixed positive definite matrix does satisfy the eigenvalue condition.

Hence defining ${}^*\theta_n$ by

$$A({}^*\theta_n,S(\theta))=\min_{\theta_1\varepsilon\Omega} A(\theta_1,S(\theta))$$

it is clear that $n^{\frac{1}{2}}(X_n-\mu(\theta_n))$ has the same asymptotic distribution whether $\theta_n=\theta_n^*$ or $\theta_n={}^*\theta_n$. Therefore this is also true of W_n.

If V is of full rank a natural choice for S is V^{-1}. It can be verified that V^{-1} is a g-inverse of Σ and $\rho(\Sigma)=k-q$. In the null case

$$W_n=n(X_n-\mu(\theta_n^*))'V(\theta_n^*)^{-1}(X_n-\mu(\theta_n^*))\xrightarrow{L}\chi^2(k-q).$$

In the non-null case γ belongs to the range of V since $\rho(V)=k$, and a non-central χ^2 results with $\lambda=\gamma'(V-Q(Q'V^{-1}Q)^{-1}Q')\gamma$.

The above theory trivially extends to the situation considered in the Corollary.

Examples

(1) Consider a continuous and strictly monotone distribution function which is specified up to a q dimensional vector of unknown parameters, $F(y,\theta)$, $\theta\varepsilon\Omega$. A random sample $Y_1, Y_2, \ldots, Y_n$ is drawn and let X_{p_j}, $j=1,2,\ldots,k$, be the p_jth quantile statistic defined in the usual way, $p_i<p_j$, $i<j$.

It is well known that under mild conditions the vector $X_n' = (X_{p_1^{(n)}}, \ldots, X_{p_k^{(n)}})$ is asymptotically normally distributed with mean vector $y' = (y_{p_1}, \ldots, y_{p_k})$ and covariance matrix $n^{-1}V$, where for $i<j$

$$v_{ij} = p_i q_i [f(y_{p_i}) f(y_{p_j})]^{-1}$$

with $p_i = F(y_{p_i})$, $q_i = 1-p_i$ and $f=F'$, Cramér (1946, p. 367). Since F depends on θ it follows that, in general, $y_{p_i} = F^{-1}(p_i)$ and V also depend on θ, $y_{p_i}(\theta)$ and $V(\theta)$ say. The previous results now apply with $\mu(\theta) = y(\theta)$ and $S(\theta) = V(\theta)$.

For the non-null case consider a sequence of alternatives of the form

$$F_n(y) = F(y,\theta) + n^{-\frac{1}{2}} G(y).$$

with suitable conditions on G it can be shown that

$$n^{\frac{1}{2}}(X_n - y(\theta)) \xrightarrow{L} N(\gamma, V)$$

where $\gamma_i = -G(y_{p_i})/f(y_{p_i})$. Since V is of full rank γ belongs to the range of V. Theorem 1 applies and W_n is asymptotically distributed as

$$\chi^2(k-q,\lambda),\ \lambda = \gamma'(V - Q(Q'V^{-1}Q)^{-1}Q')\gamma.$$

In the case where F is normal, $\theta' = (\mu,\sigma)$, $y_{p_i}(\theta) = \mu + \sigma a_i$ and $V(\theta) = \sigma^2 V_1$ with

$$p_i = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{a_i} \exp\{-t^2/2\} dt$$

and V_1 is independent of θ. We are in a position to use $^*\theta_n$ since minimizing

$$n(X_n - y(\theta_1))' \, V_1^{-1}(X_n - y(\theta_1))$$

is equivalent to minimizing

$$n(X_n - y(\theta_1))' \, (\sigma^2 V_1)^{-1} \, (X_n - y(\theta_1))$$

with σ^2 fixed at the true value. Explicit expressions for $^*\mu$ and $^*\sigma$ are given by

$$^*\mu = \frac{1'V_1^{-1}X_n a'V_1^{-1}a - a'V_1^{-1}X_n 1'V_1^{-1}a}{1'V_1^{-1}1 a'V_1^{-1}a - (1'V_1^{-1}a)^2}$$

$$^*\sigma = \frac{-1'V_1^{-1}a 1'V_1^{-1}X_n + 1'V_1^{-1}1 a'V_1^{-1}X_n}{1'V_1^{-1}1 a'V_1^{-1}a - (1'V_1^{-1}a)^2}$$

where $1' = (1,1,1, \ldots, 1)$. These results continue to hold whenever θ_1 and θ_2 are location and scale parameters.

The idea of using order statistics to measure goodness of fit has been presented from a different point of view by Bofinger (1973).
(2) Suppose the sample $Y_1, \ldots, Y_n$ is obtained using Steins double sampling procedure for estimating the mean of a normal distribution with fixed precision, so that n is replaced by a random variable N_c [see e.g. Wetherill (1966, p. 189)].

A preliminary sample $Y_1, \ldots, Y_m$ is taken and the sample variance s_m^2 calculated. N_c is chosen as the least integer not less than

$$\max\ \{m, m^2 s_m^2/(m-2)^2 c^2\}.$$

We assume that m is kept fixed but let $c \to 0$. Then $N_c \xrightarrow{P} \infty$ and the quantile statistics $X_c' = (X_{p_1}^{(c)}, \ldots, X_{p_k}^{(c)})$ calculated from $Y_1, \ldots, Y_{N_c}$ satisfy the limit law

$$N_c^{\frac{1}{2}} (X_c - y) \xrightarrow{L} N(0, V).$$

This follows since it is true for the conditional distribution of N_c and X_c given $Y_1, \ldots, Y_m$. It therefore holds for the marginal distribution as well.

From the Corollary we conclude that the asymptotic quantile goodness of fit test in Example 1 is still valid. Clearly asymptotic goodness of fit tests based on proportions, as discussed in Section 2(a), also remain valid with this stopping rule.

Acknowledgements

We wish to thank a number of readers for helpful comments on an early draft of this paper; in particular we are grateful to Dr. T. J. O'Neill for assistance at an early stage of the work.

References

Anscombe, F. J. (1952). "Large sample theory of sequential estimation". *Proc. Camb. Phil. Soc.*, 48, 600–607.

Bofinger, Eve. (1973). "Goodness-of-fit test using sample quantiles". *J. Roy. Statist. Soc.*, B., 35, 277–284.

Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press.

Gurland, J., and Dahiya, R. C. (1972). "Tests of fit for continuous distributions based on generalized minimum chi-square". *Statistical Papers in Honor of George W. Snedecor*, 115–128. Iowa State University Press.

Heyde, C. C., and Feigin, P. D. "On efficiency and exponential families in stochastic process estimation". *A Modern Course on Statistical Distributions in Scientific Work*, Vol. I, 227–240. Reidel Publishing Company, Boston.

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

Rao, C. R., and Mitra, S. K. (1971). *Generalized Inverse of Matrices and its Applications.* Wiley, New York.

Wetherill, C. B. (1966). *Sequential Methods in Statistics.* Methuen's Monographs on Applied Probability and Statistics, Methuen, London.