On asymptotic normality of likelihood and conditional analysis

D. BRENNER, D. A. S. FRASER, and Philip McDUNNOUGH

University of Toronto

Key words and phrases: Asymptotic normality, likelihood analysis, normalized likelihood, convergence of likelihood functions, conditional inference.

AMS 1980 subject classifications: Primary 62E20; secondary 62A05.

ABSTRACT

The likelihood function from a large sample is commonly assumed to be approximately a normal density function. The literature supports, under mild conditions, an approximate normal shape about the maximum; but typically a stronger result is needed: that the normalized likelihood itself is approximately a normal density. In a transformation-parameter context, we consider the likelihood normalized relative to right-invariant measure, and in the location case under moderate conditions show that the standardized version converges almost surely to the standard normal. Also in a transformation-parameter context, we show that almost sure convergence of the normalized and standardized likelihood to a standard normal implies that the standardized distribution for conditional inference converges almost surely to a corresponding standard normal. This latter result is of immediate use for a range of estimating, testing, and confidence procedures on a conditional-inference basis.

1. INTRODUCTION

Almost all statistical methods based on large-sample likelihood make some reference to normality. For point estimates based on likelihood there is a wide literature concerning the limiting normality of their distributions; this makes possible the numerical solution of many practical inference problems that would otherwise be quite inaccessible. For some other statistical procedures references are made to a normal shape for the large-sample likelihood itself. The literature does support under mild conditions an approximate normal shape about the maximum; but in most cases a stronger result is needed: that the normalized likelihood itself converge to a normal density function. In this paper we consider this convergence of the normalized likelihood function and prove the convergence subject to moderate conditions in the location-parameter case. Also we give a theorem for transformation-parameter models that establishes the transfer of limiting normality for the normalized likelihood to limiting normality for the conditional distribution used for conditional inference.

2. BACKGROUND

An “asymptotic normal form” for the likelihood function from a large sample has had some moderate acknowledgement in the literature. Most attention however has focused on the convergence of point estimates derived from the likelihood, in particular on limiting normal distribution theory for these estimates. This began in an extensive way with the investigation of maximum-likelihood estimation by Fisher (1934) and came to some crystallization in the strong pointwise result due to Wald...
The likelihood function, itself in its logarithmic form, was examined as a function of the random sample (Fraser 1968) and shown to converge almost surely in the neighbourhood of the true value to a quadratic function with maximum at the maximum-likelihood estimate. Walker (1968) in the case of a real parameter showed that the posterior distribution from a proper prior converges weakly to a normal density. Hájek (1971) proposed a definition of asymptotically normal likelihood with the purpose of deriving properties of certain estimators. Heyde and Johnstone (1979) extended the Walker results to stochastic processes, again with respect to a proper prior. A likelihood function normalized with respect to a right-invariant prior is covered by the preceding only if the right-invariant prior is proper, a very special case found with compact groups but not (for instance) with the location, location-scale, and regression models. The need for coverage in these latter cases is the central concern of this paper.

Transformation-parameter models and associated conditional-inference methods have a history with basic roots in the examination by Fisher (1934) of location and location-scale models. Pitman (1939) examined invariant location-scale confidence intervals, and Peisakoff (1951) gave a decision-theoretic framework. The case of the regression model was examined in Fraser (1961) and Verhagen (1961), and the general logic and analysis of the transformation-parameter model was developed in a series of works (Fraser 1961, 1968, 1979). Conditional analysis was shown to be necessary in the group-model case in Fraser (1979) and shown to be consistent only in the group case (Brenner and Fraser, 1979).

In two recent papers, Hinkley (1978) and Efron and Hinkley (1978), the suggestion is given to use a normal approximation for the conditional distribution obtained by conditional-inference arguments; the particular location and variance are taken to be the maximum-likelihood estimate and the inverse of the observed information. This is supported by a basically heuristic argument which assumes normal form for the normalized likelihood function. Numerical support, however, for the approximate normal form has been found in the extensive simulations for the location-scale case using the conditional-inference computer program in Fraser (1976, 1979).

Our central concern lies with the normality of the normalized likelihood function and with the transfer of this normality from the likelihood to the conditional distribution for conditional inference.

3. LIMITING FORM OF LIKELIHOOD AND THE CONDITIONAL DENSITY

Let \((y_1, \ldots, y_n)\) be a sample from a distribution on \(\mathbb{R}^n\) with density function \(f(y; \theta)\) with respect to Euclidean measure. The density function for the sample is \(f_n(y; \theta) = \prod_{i=1}^{n} f(y_i; \theta)\). The likelihood function from the sample \(y\) is the function of \(\theta\) given by \(L_n(\theta; y) = c f_n(y; \theta)\), where \(c > 0\) is arbitrary. Also of interest is the distribution of the random function \(L_n(\cdot; y)\) induced by the \(\theta\)-distribution for \(y\).

For the location-model case we have \(f(y; \theta) = p(y - \theta)\). Conditional analysis for this model was based on an implicit ancillarity principle in Fisher (1934); in fact it becomes automatic in the context of objective error (Fraser 1968, 1979; Brenner and Fraser 1981). Let \(\hat{\theta}(y)\) be a location estimate (say maximum likelihood) and \(d = y - \hat{\theta}(y)1\) be the residual vector. The likelihood function (of \(\theta\)) is

\[
L_n(\theta; y) = c \prod_{i=1}^{n} p(y_i - \theta) = c \prod_{i=1}^{n} p(\hat{\theta} - \theta + d_i). \tag{3.1}
\]
The conditional-inference distribution of \( \hat{\theta} \) given \( d \) is

\[
k_n(d) \prod_{i=1}^{n} (\hat{\theta} - \theta + d_i).
\] (3.2)

The clear close connection between these was noted by Fisher (1934): “The frequency distribution is the mirror image of the likelihood function”.

We now give notation for the general transformation-parameter model and demonstrate a close parallel connection between the likelihood function and the conditional inference distribution.

Let \( z \) have density function \( p(z) \) on \( \mathbb{R}^p \), and let \( \theta \) in \( \Omega \) be an invertible transformation \( \mathbb{R}^p \to \mathbb{R}^p \). We are concerned with the distribution of \((y_1, \ldots, y_n)\) where \( y_i = \theta z_i \) and \((z_1, \ldots, z_n)\) is a sample from \( p(z) \). As an illustration consider the regression model \( y = X\beta + \sigma z \) where \( z \) is \( \mathcal{N}_d(0, \mathbf{I}) \) and \( \theta = (\beta, \sigma) \).

Towards presenting the likelihood function and the conditional-inference distribution we introduce some assumptions and notation.

**Assumption S1.** The class \( \Omega \) of transformations is a group.

**Assumption S2.** The class \( \Omega \) is presented as an open subset of \( \mathbb{R}^L \) (or as an \( L \)-dimensional submanifold of a Euclidean space).

**Assumption S3.** The functions \( \theta = \theta_1 \theta_2 \) and \( y = \theta_1 \theta_2 z \) are continuously differentiable in \( \theta_1, \theta_2, z \).

**Assumption S4.** There exists a continuously differentiable function \( \hat{\theta}_n(y) = \hat{\theta}_n(y_1, \ldots, y_n) \) from \( \mathbb{R}^{np} \) to \( \Omega \) such that \( \hat{\theta}_n(gy) = g \hat{\theta}_n(y) \) for any \( g \) in \( \Omega \).

Typically the maximum-likelihood estimator (for instance when it is uniquely determined) satisfies this last assumption.

Now let \( D_n(y) = \hat{\theta}_n^{-1}(y)y \), so that

\[
y = \hat{\theta}_n(y)D_n(y), \quad z = \hat{\theta}_n(z)D_n(z).
\] (3.3)

The equation \( y = \theta z \) then gives

\[
\hat{\theta}_n(y) = \theta \hat{\theta}_n(z), \quad D_n(y) = D_n(z),
\]

from which we note that \( D_n(y) \) is invariant.

For the density function \( y \) we define a Jacobian determinant

\[
J_n(\theta, z) = \left| \frac{\partial \theta z}{\partial z} \right|
\] (3.4)

and let

\[
J_n(z) = J_n(\hat{\theta}_n(z), D_n(z)).
\] (3.5)

The chain rule for derivatives with Assumptions S1, S2 gives

\[
J_n(\theta, z) = J_n(\theta z)J_n^{-1}(z).
\] (3.6)

The density function for \( y \) is obtained from the density \( p_n(z) = \prod p(z_i) \) for \( z \):

\[
f_n(y : \theta) = p_n(\theta^{-1} y) J_n(\theta^{-1} y) J_n^{-1}(y).
\] (3.7)
The likelihood function is available immediately from (3.7):

\[ L_n(\theta : y) = cp_n(\theta^{-1}y)J_n(\theta^{-1}y) \]
\[ = cp_n(\theta^{-1}\hat{\theta}_n D_n)J_n(\theta^{-1}\hat{\theta}_n D_n) \]
\[ = ch_n(\theta^{-1}\hat{\theta}_n : D_n), \]  

(3.8)
say, where we write \( \hat{\theta}_n = \hat{\theta}_n(y) \) and \( D_n = D_n(y) \). The conditional-inference distribution for \( \hat{\theta}_n(y) \) given \( D_n(y) = D_n \) is obtained in Fraser (1968, 1979):

\[ k_n(D_n)p_n(\theta^{-1}\hat{\theta}_n D_n)J_n(\theta^{-1}\hat{\theta}_n D_n), \]  

(3.9)
a density with respect to left-invariant measure on \( \Omega \). The close connection between (3.8) and (3.9) can be emphasized by recording the density for \( g = \theta^{-1}\hat{\theta}_n \),

\[ h_n(g : D_n) = k_n(D_n)p_n(gD_n)J_n(gD_n), \]  

(3.10)
and noting that \( g = \theta^{-1}\hat{\theta}_n \) as a function of \( \theta \) gives (3.8) and as a function of \( \hat{\theta}_n \) gives (3.9).

Tests, estimates, and confidence regions for \( \theta \) all follow from the distribution (3.10) and the “pivotal” equation \( g = \theta^{-1}\hat{\theta}_n \).

Our focal concern in this section lies with the transfer of the limiting normality of the likelihood to limiting normality of the conditional-inference distribution. For this we need additional differentiability assumptions.

**Assumption S5.** The maximum likelihood estimator satisfies Assumption S4.

**Assumption S6.** The density \( p(z) \) is twice continuously differentiable.

The observed information matrix is

\[ I(\hat{\theta}_n(y), D_n) = -\frac{\partial^2}{\partial \theta^2} \ln h_n(\theta^{-1}\hat{\theta}_n : D_n) \bigg|_{\hat{\theta}_n}. \]  

(3.11)

If the distribution of \( \hat{\theta}_n \) (given \( D_n \)) is location-normal, then \( I(\hat{\theta}_n(y), D_n) \) is the inverse variance matrix. On the other hand, if the likelihood (3.8) is treated as a distribution that is location normal in form, then \( I(\hat{\theta}_n(y), D_n) \) is the inverse variance matrix of that distribution.

**Theorem 3.1.** For a transformation-parameter model satisfying Assumptions S1-S6, if the likelihood function standardized with respect to location \( \hat{\theta}_n(y) \) and scale variance matrix \( I^{-1}(\hat{\theta}_n, D_n) \) converges in probability (or almost surely) to the standard normal, and if the maximum diagonal element of the scale matrix goes to zero, then the conditional-inference distribution for \( \hat{\theta}_n \) standardized with respect to location \( \theta \) and scale variance matrix \( I^{-1}(\hat{\theta}_n, D_n) \) converges in probability (or almost surely) to the standard normal.

**Proof.** The assumptions assure that the transformations from \( \theta \) to \( g \) by means of \( g = \theta^{-1}\hat{\theta}_n \), and from \( g \) to \( \hat{\theta}_n \) by means of \( g = \theta^{-1}\hat{\theta}_n \), are twice continuously differentiable. As a result the full transformation from \( \theta \) to \( \hat{\theta}_n \) in any neighborhood can be approximated by its derivative, which is linear. The mapping of the multivariate normal distributions discussed before the theorem then determines the mapping of the limiting normal distributions. Q.E.D.
4. LIMITING NORMALITY

In the introduction we expressed our central interest in the asymptotic normality of the conditional inference distribution. The preceding section showed that such asymptotic normality follows from a strong type of asymptotic normality of the likelihood function.

In this section we consider this strong asymptotic normality of the likelihood function for the case of transformation-parameter models and give the detailed theorem for the location model.

Consider a statistical model for \( y = (y_1, \ldots, y_n)' \) with location parameter \( \theta \). For this we assume a density and initially allow the generality of statistical dependence: \( y = \theta 1 + z \), where \( z \) has density \( p_n(z) \) which in the independent case is \( \prod_i^n p(z_i) \).

The likelihood function from \( y \) is given by

\[
L_n(\theta) = c p_n(y - \theta 1)
\]

(4.1)

with log-likelihood

\[
l_n(\theta) = a + \ln p_n(y - \theta 1)
\]

where \( a \) is arbitrary. Our main result is concerned with the normalized likelihood function

\[
\pi_n(\theta) = \frac{p_n(y - \theta 1)}{\int p_n(y - u 1) \, du}
\]

(4.2)

In particular we are concerned with the distribution for the "variable" \( \theta \) standardized with respect to location \( \hat{\theta}_n(y) \) and to scale \( \sigma_n \), where

\[
\sigma_n^2(y) = -l_n''(\hat{\theta}_n) = -\frac{\partial^2}{\partial \theta^2} \ln L_n(\theta) \bigg|_{\hat{\theta}_n}
\]

is the information at the maximum-likelihood estimate \( \hat{\theta}_n \). The corresponding standardized density in terms of \( \tau = (\theta - \hat{\theta}_n)/\sigma_n \) is

\[
q_n(\tau) = \sigma_n \pi_n(\hat{\theta}_n + \tau \sigma_n).
\]

Theorem 4.3, recorded later in this section, shows that

\[
q_n(\tau) \overset{a.s.}{\to} \frac{1}{\sqrt{2\pi}} e^{-\tau^2/2}
\]

(4.3)

subject to suitable regularly conditions. However, by a theorem of Scheffé (1947) the pointwise convergence of the density implies the pointwise convergence of the distribution function. Thus

\[
Q_n((a, b)) = \int_{\hat{\theta}_n + a \sigma_n}^{\hat{\theta}_n + b \sigma_n} \pi_n(\theta) \, d\theta \overset{a.s.}{\to} \int_{a}^{b} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz.
\]

(4.4)

Theorem 3.1 then establishes a corresponding asymptotic normal result for the conditional-inference distribution.

The local normal shape of the likelihood function near its maximum is relatively
easy to establish (Fraser 1968, p. 302 ff). For this we let

$$b_n(\tau) = \frac{q_n(\tau)}{\max_i q_n(t)} = \frac{\pi_n(\hat{\theta} + \tau \sigma_n)}{\max_i \pi_n(\hat{\theta} + i\tau \sigma_n)} = \frac{L_n(\hat{\theta} + \tau \sigma_n)}{\max_i L_n(\hat{\theta} + i\tau \sigma_n)},$$

(4.5)

and show in Theorem 4.1 that $b_n(\tau) \rightarrow \exp\{-\tau^2/2\}$ under moderate conditions. We then address the stronger result implicit in (4.3).

Towards Theorem 4.1 we record some preliminary assumptions.

**Assumption M1.** $\hat{\theta}_n(y) \xrightarrow{a.s.} \theta$.

**Assumption M2.** $l''_n(\theta)/l''_n(\hat{\theta}_n) \xrightarrow{a.s.} 1$.

**Assumption M3.** For each $\tau_0$,

$$\sup_{-\tau_0 < \tau < \tau_0} \left| \frac{l''(\hat{\theta}_n + \tau \sigma_n) - l''(\hat{\theta}_n)}{l''(\hat{\theta}_n)} \right| \xrightarrow{a.s.} 0.$$

The preceding assumptions M2 and M3 are implied by the following assumptions M4 and M5.

**Assumption M4.** $l''_n(\theta)$ is continuous in $\theta$.

**Assumption M5.** $\sigma_n^2 \xrightarrow{a.s.} 0$.

**Theorem 4.1.** Under Assumptions M3 and M4,

$$b_n(\tau) \xrightarrow{a.s.} \exp\{-\tau^2/2\}.$$

**Proof.** We expand $l_n(\theta)$ to the second order about $\hat{\theta}_n$:

$$\ln b_n(\tau) = l_n(\hat{\theta}_n + \tau \sigma_n) - l_n(\hat{\theta}_n) = \frac{\tau^2}{2} \sigma_n^2 l''_n(\hat{\theta}_n + \tau \sigma_n),$$

where $\tau^*$ is between $-\tau$ and $\tau$. By Assumption M3 this converges to $-\tau^2/2$, which completes the proof. Q.E.D.

The appropriate integrated value of the likelihood function is fairly easy to establish provided the integral is restricted to a neighbourhood of the true $\theta$. For this we introduce a wider-range version of Assumption M3:

**Assumption M6.** For each $\epsilon > 0$, there exists a $\delta$ such that

$$\lim_{n \to \infty} \sup_{|\theta - \hat{\theta}_n| < \delta} \left| \frac{l''_n(\theta') - l''_n(\hat{\theta}_n)}{l''_n(\hat{\theta}_n)} \right| \xrightarrow{a.s.} 0.$$

**Theorem 4.2.** Under Assumptions M1, M4, M6, for each $\epsilon > 0$ there exists a $\delta$ such that

$$\lim_{n \to \infty} \sup_{|\theta - \hat{\theta}_n| < \delta} \left| \int_{|t - \hat{\theta}_n| < \delta} \frac{p_n(y - tI)}{\sigma_n p_n(y - \hat{\theta}_n I)} dt - \sqrt{2\pi} \right| \xrightarrow{a.s.} 0 < \epsilon.$$

**Proof.** The integrand can be rewritten as

$$\frac{1}{\sigma_n} \exp\{l_n(t) - l_n(\hat{\theta}_n)\} = \frac{1}{\sigma_n} \exp\left\{ -\frac{(t - \hat{\theta}_n)^2}{2\sigma_n^2} (1 + r_n) \right\},$$
where
\[ r_n = \sigma_n^2 \left( l_n''(\hat{\theta}_n) - l_n''(\hat{\theta}_n) \right). \]

By Assumption M6, \( \lim |r_n| < \epsilon \) if \( \delta \) is small enough. The integral can be rewritten
\[ I_\delta = \int_{|t - \theta| \leq \delta} \frac{1}{\sigma_n} \exp \left\{ -\frac{(t - \hat{\theta}_n)^2}{2\sigma_n^2} \right\} (1 + r_n) \, dt; \]

this integral is bounded by the two values of
\[
\int_{|t - \theta| \leq \delta} \frac{1}{\sigma_n} \exp \left\{ -\frac{(t - \hat{\theta}_n)^2}{2\sigma_n^2} \right\} (1 + \epsilon) \, dt \\
= \sqrt{2\pi} \sqrt{1 + \epsilon} \ \Phi \left( \frac{\sqrt{1 + \epsilon} \theta + \delta - \hat{\theta}_n}{\sigma_n} \right) - \Phi \left( \frac{\sqrt{1 + \epsilon} \theta - \delta - \hat{\theta}_n}{\sigma_n} \right) \overset{\text{a.s.}}{\rightarrow} \sqrt{2\pi} \sqrt{1 + \epsilon}.
\]

This completes the proof. Q.E.D.

The preceding theorems have focused on the limiting normal shape of the likelihood and on an appropriate normal-type integrability in a neighbourhood of the true \( \theta \). Some related considerations of the likelihood may be found in the literature. Wald (1949) has shown, in the independent and identically distributed case, that
\[
\sup_{|t - \theta| \leq \delta} \frac{l_n(t) - l_n(\hat{\theta}_n)}{l_n''(\hat{\theta}_n)} \overset{\text{a.s.}}{\rightarrow} -\infty
\]

under essentially the following assumption.

**Assumption M7.** \( |l(\theta)| = |\ln f(y; \theta)| \leq M(y) \) with \( \delta(M(y); \theta) \) finite.

In addition, results due to Walker (1969) imply that, for a proper prior \( p_1(\theta) \),
\[
\int \frac{p_1(y - t1)}{\sigma_n p_1(y - \hat{\theta}_n1)} p_1(t) \, dt \overset{\text{p}}{\rightarrow} \sqrt{2\pi} p_1(\theta).
\]

As noted by Heyde and Johnstone (1979) this result continues to hold for arbitrary stochastic processes under conditions which do little more than ensure consistency of \( \hat{\theta} \) and certain continuity properties of \( l_n''(\hat{\theta}) \). For some related results see Basawa and Prakasa Rao (1980) and the references cited there.

For the remainder of the proof of the strong likelihood limit (4.3) we return to the initial assumption of a sample \((y_1, \ldots, y_n)\) from a distribution with density \( p(y - \theta) \) on \( \mathbb{R} \). Let \((z_1, \ldots, z_n)\) be the corresponding sample for \( z = y - \theta \) from the density \( p(z) \) on \( \mathbb{R} \).

We record a final assumption needed for the strong likelihood result.

**Assumption M8.** The density \( p(z) \) is strictly positive and satisfies
\begin{enumerate}
  \item[(a)] \( \sup_z |z|^{1+d} p(z) < \infty \) for some \( d > 0 \)
\end{enumerate}

and
\begin{enumerate}
  \item[(b)] There exists an \( x_0 > 0, \delta > 0, \) and an \( M \) such that for all \( x \leq x_0, P(p(z) \leq x) \leq Mx^\delta \).
\end{enumerate}
Assumption M8 essentially ensures that the tails of the likelihood distribution are not too long. In fact, if \( p(z) \) decreases for large enough \( |z| \) and is bounded away from 0 on finite intervals, then M8(a) implies M8(b). Most of the common statistical distributions, including the Student and stable law families, satisfy M8. We are now in a position to prove the main result.

**Theorem 4.3.** Under Assumptions M1, M2, M4, M5, M6, M7, and M8 the strong likelihood limit (4.3) obtains.

**Proof.** Using Theorems 4.1 and 4.2, it suffices to show that for each \( \delta > 0 \)

\[
\int_{|t-\theta|>\delta} \frac{p_n(t \mathbf{1})}{\sigma_n p_n(y - \hat{\theta}_n \mathbf{1})} dt \xrightarrow{\text{a.s.}} 0. \tag{4.6}
\]

We have with probability 1 that for any \( \delta > 0 \) there exists a \( c > 0 \) such that

\[
\sup_{|\theta^\prime - \theta| > \delta} \frac{l_n(\theta^\prime) - l_n(\hat{\theta})}{n} \leq -c
\]

for all sufficiently large \( n \); this follows from the Wald result mentioned earlier. Set \( b_n = e^{nc} \), where \( c' < c \). Then, writing the integrand in (4.6) as

\[
\frac{1}{\sigma_n} \exp[l_n(t) - l_n(\hat{\theta}_n)],
\]

we see that

\[
\int_{t > \theta} \frac{1}{\sigma_n} \frac{p_n(t \mathbf{1})}{p_n(y - \hat{\theta}_n \mathbf{1})} dt \leq b_n e^{-cn} \xrightarrow{\text{a.s.}} 0.
\]

Similarly the integral over \( -b_n \leq t - \theta \leq -\delta \) also converges to zero. Now, if \( z_{(j):n} \) denotes the \( j \)th order statistic of \( z_1, \ldots, z_n \), we have for any \( \varepsilon > 0 \) and some \( d > 0 \)

\[
P(z_{(j):k} < \varepsilon b_k, k \geq n) = P(z_{(n):n} < \varepsilon b_n) \prod_{k=n+1}^{\infty} P(z_k < \varepsilon b_k)
\]

\[
\geq \left(1 - \frac{\varepsilon(n)}{(eb_n)^d} \right)^{n} \prod_{k=n+1}^{\infty} \left[1 - \frac{\varepsilon^d}{(eb_k)^d} \right] \rightarrow 1 \quad \text{as} \quad n \rightarrow \infty.
\]

Consequently \( z_{(n)}/b_n \xrightarrow{\text{a.s.}} 0 \), so that for large enough \( A \)

\[
\int_{t > \theta} \frac{p_n(t \mathbf{1})}{\sigma_n p_n(y - \hat{\theta}_n \mathbf{1})} dt \leq \frac{1}{\sigma_n p_n(y - \hat{\theta}_n \mathbf{1})} \int_{t > \theta} A^n \frac{A^n t^{-1}}{(t - y(n))^{n(1+d)}} dt
\]

which is asymptotically equivalent to

\[
\frac{A^n}{n\sigma_n p_n(y - \hat{\theta}_n \mathbf{1}) b_n^{n(1+d)+1}} \leq \frac{A^n}{n\sigma_n p_n(y - \bar{\theta} \mathbf{1}) b_n^{n(1+d)+1}}.
\]

We now show using Assumption M8(b) that \( p_n(y - \bar{\theta} \mathbf{1}) e^{cn^2} \xrightarrow{\text{a.s.}} \infty \). For this we have

\[
P\left( \prod_{i=1}^{m} p(z_i) \geq e^{-cn^2} : \text{all} \ m \geq n \right) \geq \prod_{i=n}^{n} P(p(z_i) \geq e^{-cn}) \prod_{m=n+1}^{\infty} P(p(z_{m}) \geq e^{-cn})
\]

\[
\geq (1 - Me^{-\delta cn})^{n} \prod_{m=n+1}^{\infty} (1 - Me^{-\delta cn}) \rightarrow 1.
\]
Thus \( P\left( \prod_{i=1}^{n} p(z_i)e^{\alpha z_i} \geq 1 : \text{all } m \geq n \right) \rightarrow 1 \) for any \( c > 0 \). It follows that \( \prod_{i=1}^{n} p(z_i)e^{\alpha z_i} \). \\

\( \rightarrow \infty \), and then that the integral over \( t - \theta > b_n \) converges to zero. The case with \\
\( t - \theta < -b_n \), follows from symmetry. Q.E.D.

RéSUMÉ

Il est coutumier de supposer que la fonction de vraisemblance d’un grand échantillon a approximativement la forme d’une cloche de Gauss; et en effet plusieurs auteurs ont proposé des conditions qui garantissent que cette hypothèse est approximativement vérifiée autour du maximum. Cependant, on a généralement besoin de faire appel à un résultat plus puissant, à savoir: que la fonction de vraisemblance normalisée elle-même suit approximativement une loi normale. Dans le contexte des modèles de transformations paramétriques, nous étudions la fonction de vraisemblance telle que normalisée d’après une mesure invariante à droite, et nous démontrons, dans le cas du problème de “location,” que sous certaines conditions cette fonction standardisée tend presque-partout vers une loi normale centrée réduite. Toujours dans le contexte des modèles de transformations paramétriques, nous prouvons que la convergence presque-partout de la fonction de vraisemblance normalisée et standardisée vers une loi normale centrée réduite entraîne la convergence presque-partout de la distribution standardisée qui sert pour l’inference conditionnelle vers une autre loi normale centrée réduite. Ce dernier résultat a des applications immédiates dans les problèmes de tests d’hypothèses et d’estimation ponctuelle ou par intervalles basés sur l’inference conditionnelle.

REFERENCES


Received 20 March 1981
Revised 31 December 1981
Accepted 6 January 1982

Department of Statistics
University of Toronto
Toronto, Ontario M5S 1A1