TRANSFORMATION-PARAMETER/STRUCTURAL MODELS:
ASYMPTOTIC CONDITIONAL DISTRIBUTIONS

D. Brenner, D.A.S. Fraser, F.R.S.C., and P. McDunnough

This is a preliminary report on the large sample (asymptotic) properties of the conditional distributions that arise in the statistical analysis of transformation-parameter [4]/structural [2,3] models. The almost-sure convergence of the standardized structural distribution to a unit normal is derived from a corresponding convergence of the likelihood function. The statistical literature contains a variety of results on the limiting nature of the likelihood function but these have focussed on specific conditional point-estimates, posterior distributions from proper priors, or the central (local) shape of the likelihood function. In this report we obtain the almost-sure convergence of the structural distribution in the context of the location model. The techniques employed are being extended to handle the general transformation-parameter/structural model.

TRANSFORMATION-PARAMETER/STRUCTURAL MODELS. The general structural model has the form: \( y = \theta z \) with \( \theta \in G \), \( z \in (S, A, P) \), where \( G \) is a group of measurable transformations acting exactly on \( S \); foundational support may be found in [1]. With an observed \( y_0 \), the distribution \( P \) is replaced by the conditional distribution \( P^{Gy_0} \) given the observed orbit \( Gy_0 \). The corresponding transformation parameter model has the form \( y = (S, A, P; \theta^{-1}) \), \( \theta \in G \) yields the conditional model \( G \)-orbit \( Gy_0 \).

For a sample \( y_n = (y_1, \ldots, y_n) \) the n-fold independent product is \( y_n = \theta z \) with \( \theta \) in \( G \), \( z \in (S^n, A^n, P^n) \), where the transformations act coordinate-wise. Under regularity conditions, it is the conditional measure \( P^{Gy_n} \), suitably normalized, that converges almost-surely to the unit normal \( \phi \).
THE STRUCTURAL DISTRIBUTION. For what follows we specify some general
smoothness assumptions about the underlying structural model. S1: The
sample space $S$ and the group $G$ are presented as open sets of $R^N$ and $R^L$
respectively. S2: The functions $g=g_1g_2$ and $y=g_1g_2z$ are continuously
differentiable in the arguments $g_1$ and $g_2$ in $G$ and $z$ in $S$. S3: There
is a global cross-section $Q$ and the projection $D: S \times Q$ defines an equiva-
 mantener map $[z]$ by $z=[z]D(z)$ so that $[z]$ and $D(z)$ are continuously
differentiable.

Since the asymptotic properties being considered here are
essentially "local" ones, these assumptions can be relaxed so that $S$ and
$G$ can be taken to be differentiable manifolds and $Q$ can define local
cross-sections; the assumptions as stated cover most statistical applications.

Let $J_N(g,x)=\partial g_x/\partial x$ and define $J_N(z)=J_N([z],D(z))$; let $J_L(g,h)=\partial g/h/\partial h$
and define $J_L(g)=J_L(g,0)$: this gives (locally) the invariant measures
$M(dx)=|J_N(z)|^{-1}dz$ and $\mu(dg)=|J_L(g)|^{-1}dg$ standardized with respect to
Euclidean volume at $Q$ and $g$ respectively, where $|A|$ is the absolute value
of the determinant of $A$. We make a final assumption S4: The measure $P$
is given by a density $p$ with respect to $M$, and $p$ is twice continuously
differentiable.

The marginal density for the orbit $Gx$, using coordinates on the
cross-section $Q$ is then $k(D)=\int_G p(gD)J(D)\mu(dg)$ relative to Euclidean
volume at $D$ orthogonal to $GD$ and $J(D)=|A'A|^{-1}$ where $A=\partial g/\partial g$. The con-
ditional distribution then has density $k(D)^{-1}p(gD)J(D)$ relative to the
(left) haar measure $\mu$ on $G$. The structural distribution for $\theta$ given
$D$ is then $k(D)^{-1}p(\theta y)J(D)\Delta([y])\mu(d\theta^{-1})$ where $\Delta(g)=|\partial \mu(g)/\partial \mu(g^{-1})|$ is the
modular function of the group. The likelihood function is the equivalence
class $L(\theta;y)=\{ cp(\theta^{-1}y): c \in \mathbb{R}^+ \}$. We consider the limiting properties of the
normalized likelihood.
ASYMPTOTIC STRUCTURAL DISTRIBUTION. For a sample of size \( n \) from the structural model satisfying assumptions S1-S4, the compound model satisfies the same assumptions. Let \( \hat{\theta}_n = \hat{\theta}_n(y_n) \) be a value maximizing the likelihood function; it is called the maximum likelihood estimate of \( \theta \) (MLE(\( \theta \))).

S5: \( \hat{\theta}_n(y_n) \) is uniquely determined and continuously differentiable. It follows that \( \hat{\theta}_n(z_n|z_n) \) and \( D_n(z_n) = \hat{\theta}_n^{-1}(z_n)z_n \) satisfy S3. We are concerned with the almost-sure limiting normality (non-degenerate) of the normalized likelihood/structural distribution \( k_n^2(D_n) = n^{-1} y_n^* J_n(D_n)^2 \Delta(y_n) \mu(\theta^{-1}). \)

LOCATION MODEL; LIMITING NORMALITY. The location model is obtained with \( S = \mathbb{R} \) and \( G \) given by the additive group with application to \( S \) by translation.

The structural density function with respect to Lebesgue measure \( \mu(\theta^{-1}) \) is \( p_n(\theta) = \prod_{i=1}^{n} p(y_i - \theta) \), and the log-likelihood is \( \ell_n(\theta) = \log(\prod_{i=1}^{n} p(y_i - \theta)) \).

For the asymptotic behavior of the log-likelihood we record some mild regularity conditions; derivatives are with respect to \( \theta \).

A1,2: \( \hat{\theta}_n(y_n) \xrightarrow{a.s.} \theta \); \( \ell_n''(\hat{\theta}_n(y_n)) \xrightarrow{a.s.} 1 \).

An information function at the maximum likelihood estimate defines the scale \( \sigma_n \) by \( \sigma_n^2 = -\ell_n''(\hat{\theta}_n) \).

A3: For each \( \tau_0 > 0 \), \( \sup_{|\tau| < \tau_0} \left| \frac{\ell_n(\hat{\theta}_n + \tau) - \ell_n(\hat{\theta}_n)}{\ell_n''(\hat{\theta}_n)} \right| \xrightarrow{a.s.} 0 \).

A4,5: \( \ell_n''(\theta) \) is continuous in \( \theta \); \( \frac{\sigma_n^2}{\ell_n''(\hat{\theta})} \xrightarrow{a.s.} 0 \).

A6: For each \( \varepsilon > 0 \), there exists a \( \delta \) such that

\[
\lim_{|\hat{\theta}_n - \theta| \to 0} \sup_{|\hat{\theta}_n - \theta| < \delta} \left| \frac{\ell_n''(\hat{\theta}_n) - \ell_n''(\hat{\theta}_n)}{\ell_n''(\hat{\theta}_n)} \right| < \varepsilon
\]
A7.8: The density \( p(z) \) is strictly positive and
\[
\sup |z|^{1+\epsilon} p(z) < \infty \text{ for some } \epsilon > 0.
\]

There exists an \( x_0 \geq 0 \) and an \( M > 0 \) such that for all \( x \leq x_0 \), \( P(p(z) < x) \leq Mx_0 \).

The structural distribution, standardized by location \( \hat{\theta}_n \) and scale \( \sigma_n \), has density \( q_n(\tau) = \sigma_n \pi_n(\hat{\theta}_n + \tau \sigma_n) \).

THEOREM: Under the assumptions A1 - A6, \( q_n(\tau) \) converges to \( (2\pi)^{-1/2} e^{-\tau^2/2} \) with probability one. From Scheffe's theorem [5] it follows that with probability one the measures \( Q_n \) defined by
\[
Q_n([\tau_1, \tau_2]) = \int_{\hat{\theta}_n + \tau_1 \sigma_n}^{\hat{\theta}_n + \tau_2 \sigma_n} \tau_n(\theta) d\theta
\]
converge weakly to the unit normal.

LIMITING LIKELIHOOD AND CONDITIONAL DISTRIBUTIONS. There is a close connection between the strong limiting form of the likelihood function and the limiting form of the conditional distribution. This connection is given in the following theorem; for general notation the observed information matrix is the negative Hessian
\[
I_n(\hat{\theta}_n, D_n) = - \frac{\partial^2}{\theta^2} \pi_n(\theta^{-1} \theta_n D_n) \bigg|_{\theta = \hat{\theta}_n}.
\]

THEOREM: For a sample from a transformation-parameter/structural model satisfying assumptions S1 - S4, if the structural density function standardized with respect to location \( \hat{\theta}_n(y_n) \) and scale matrix \( I_n(\hat{\theta}_n, D_n)^{-1} \) converges in probability (almost surely) to the unit normal then the conditional density for \( \hat{\theta}_n \) standardized with respect to location \( \theta \) and scale matrix \( I_n(\theta, D_n)^{-1} \) converges in probability (almost surely) to the unit normal.
This theorem with the results from the preceding section gives the almost-sure convergence of the conditional distribution for the location model.

REFERENCES


Received January 29, 1981