A NOTE ON ASYMPTOTIC JOINT NORMALITY

By C. L. Mallows

Bell Telephone Laboratories, Incorporated

The concept of asymptotic normality takes on some new aspects when the dimensionality of the vector random variable under consideration is allowed to increase indefinitely. A necessary and sufficient condition for joint asymptotic normality in a new (strong) sense, in the case of independence, is given.

1. Introduction. When confronted with an intractable problem concerning the joint distribution of some set of k random variables, it is standard practice to embed this set in a sequence of sets of the same dimensionality, and to prove a theorem concerning the asymptotic distribution. Formally, one considers an array $\{X_{nj}, j = 1, \cdots, k, n = 1, 2, \cdots\}$ and proves that $\mathscr{L}(X_{n1}, \cdots, X_{nk}) \to F$ as $n \to \infty$ where F is some k-dimensional distribution, and we refer (throughout) to the usual weak-$*$ convergence, which is equivalent to convergence in the Lévy metric. Such a theorem is of practical value only if the limiting result provides a good approximation to the original problem. However the theoretician has unlimited flexibility in choosing the sequence in which to embed his problem; qualitative and quantitative differences may ensue from different choices.

Our purpose here is to suggest that in some cases it may be appropriate to consider a different kind of sequence, in which the dimensionality changes (increases) as the sequence evolves. Formally, now consider an array $\{X_{nj}, j = 1, \cdots, k_n; n = 1, 2, \cdots\}$ where $\{k_n\}$ is some increasing sequence of integers. One situation in which this framework seems appropriate is the following. Consider a 2^n factorial experiment, in which $k + 1 = 2^n$ observations are taken, one at each of the possible settings of n two-level design variables. Display the observations in a vector $\mathbf{Y} = (Y_0, Y_1, \cdots, Y_k)'$. Then the mean and (multiples of) the k contrasts (main effects and interactions) are given by elements of a vector $\mathbf{X} = (X_0, \cdots, X_k)'$ where $\mathbf{X} = 2^{-\frac{1}{2}n}\mathbf{HY}$, and where $\mathbf{H}$ is a $2^n \times 2^n$ Hadamard matrix (so that all its elements are ± 1, and $\mathbf{H'H} = 2^n\mathbf{I}_k$). If $Y_0, \cdots, Y_k$ are i.i.d. (independent and identically distributed) and Normal, then the contrasts $X_1, \cdots, X_k$ will also be i.i.d. and Normal; this provides an appropriate reference distribution for the technique of half-normal plotting [3], in which the ordered values of $|X_1|, \cdots, |X_k|$ are compared with their expected values (or approximations to these), Normality being assumed. However if $Y_0, \cdots, Y_k$ are i.i.d. but nonnormal, then the contrasts $X_1, \cdots, X_k$ will still have identical marginal distributions, but these will not be normal,

---

and the contrasts will not be independent. It is of interest to know whether in this situation the half-normal plotting technique retains validity. Here we are interested in a subtle aspect of the joint distribution of all the contrasts, and it seems natural to embed the problem in a sequence obtained by letting n increase indefinitely; so that the dimensionality of the problem increases also. We return to this problem briefly in the last section of this paper.

Another problem where this approach is natural is in the study of the approximate joint normality of the "pseudo-values" that arise in application of the jackknife technique (for which see e.g. [7], [1]). Let θ be a parameter to be estimated, and suppose we have a set of $N = kn$ independent, identically distributed observations divided into n groups of k observations each. If $\hat{\theta}_N$ is an estimate of θ based on all N observations, and $\hat{\theta}_{(i)}$ is the estimate obtained when the ith group is discarded, the k pseudo-values are defined as

$$\hat{\theta}_i = n\hat{\theta}_N - (n-1)\hat{\theta}_{(i)} \qquad\qquad i = 1, \cdots, k$$

and are clearly identically distributed, in fact interchangeable. However, they are sometimes asserted to be also approximately independent, and if n is large they can be expected to be approximately normal. One is thus tempted to examine the set of pseudo-values for conformity with the hypothesis of independent normality, in the hope that any departures may be indicative of some important phenomena in the data. For this procedure to be valid, one needs assurance that approximate normality will obtain when k and n are both large.

2. Finite-dimensional asymptotic normality. With this as background, we propose to discuss the concept of asymptotic joint normality of a triangular array of variables $\mathscr{X} = \{X_{nj}, j = 1, \cdots, k_n; n = 1, 2, \cdots\}$ where $k_n \to \infty$. We denote the standard d-dimensional normal distribution by Φ_d.

DEFINITION. $\mathscr{X}$ is coordinatewise asympotically standard normal (c.a.s.n.) if for every sequence $\{m_n\}$ $(1 \leq m_n \leq k_n, n = 1, 2, \cdots)$ we have $\mathscr{L}(X_{n,m_n}) \to \Phi_1$.

DEFINITION. $\mathscr{X}$ is d-dimensionally coordinatewise asymptotically standard normal (d-c.a.s.n.) if for every sequence of d-vectors $\{(m_{n1}, \cdots, m_{nd}); 1 \leq m_{ni} \leq k_n; m_{ni} \neq m_{nj}$ for $i \neq j$ whenever $k_n \geq d; n = 1, 2, \cdots\}$ we have $\mathscr{L}(X_{n,m_{n1}}, \cdots, X_{n,m_{nd}}) \to \Phi_d$. We observe the trivial

THEOREM 1. *If $\mathscr{X}$ is row-wise independent (i.e. for each n, $\{X_{nj}, j = 1 \cdots k_n\}$ are mutually independent) then $\mathscr{X}$ is c.a.s.n. if and only if (iff) it is d-c.a.s.n. for every d.*

PROOF. For every d, $\mathscr{X}$ is d-c.a.s.n. iff the sequence $((X_{nj}, j = 1, \cdots, k_n), n = 1, 2, \cdots)$ is asymptotically normal in the usual sense.

In what follows, any sequence that is indexed by (n, j) without qualification is to be read in lexicographical order, as in the above proof.

3. Some examples. Our first example is of a triangular array $\mathscr{X}$ which is row-wise independent and d-c.a.s.n. for every d and yet for which the sequence of standardized row means is not asymptotically normal.

EXAMPLE 1. Let $k_n = n$ and let the elements $\{X_{n1}, \cdots, X_{nn}\}$ be i.i.d. with distribution F_n, where $F_n \to \Phi_1$ and yet $\mathscr{X}$ fails to satisfy the well-known necessary and sufficient conditions for the central limit theorem to hold. (See e.g. page 316 of [4].) For example, let $F_n = p_n B(a_n) + (1 - p_n)\Phi_1$ where $B(a)$ is the two-point binomial distribution, assigning probability $\frac{1}{2}$ to each of the points $\pm a$. To make $F_n \to \Phi_1$ we need only $p_n \to 0$; but if $n^{-1}a_n{}^2 \not\to 0$ and $np_n \not\to 0$ then $\mathscr{L}(n^{-\frac{1}{2}}(X_{n1} + \cdots + X_{nn})) \not\to \Phi_1$.

EXAMPLE 2. In the 2^n-factorial experiment situation outlined above, we have $k_n = 2^n - 1$, $\mathbf{X}_n' = (X_{n1}, \cdots, X_{nk_n})' = 2^{-\frac{1}{2}n}\mathbf{H}_0(Y_{n0}, Y_{n1}, \cdots, Y_{nk_n})'$ where $\mathbf{H}_0$ is a $k_n \times k_n + 1$ matrix obtained by deleting the row $(1, 1, \cdots, 1)$ from a suitable Hadamard matrix. If $Y_{n0}, \cdots, Y_{nk_n}$ are i.i.d. with $\mathscr{L}(Y_{n0}) = F_n$, then $\mathscr{X}$ is d-c.a.s.n. for all d as soon as the $\mathscr{Y}$ array satisfies the Lindeberg conditions. However since

$$2^{-\frac{1}{2}n}\mathbf{H}_0'\mathbf{X}_n = (Y_{n0} - \bar{Y}_n, \cdots, Y_{nk_n} - \bar{Y}_n) \qquad \text{where} \qquad \bar{Y}_n = 2^{-n} \sum_j Y_{nj},$$

clearly we should hesitate to call $\mathscr{X}$ approximately multinormal unless also $F_n \to \Phi_1$.

EXAMPLE 3. For comparison we also exhibit an array that is d-c.a.s.n. for every d (but not row-wise independent) with $\mathscr{L}(n^{-\frac{1}{2}}(X_{n1} + \cdots + X_{nn})) \to G$ where G is an arbitrary distribution. Take $k_n = n$, and let $y = \{Y_{nj}\}$ be an array of the same shape as $\mathscr{X}$, all its members being i.i.d. standard normal. Let Z be distributed according to G, and set

$$X_{nj} = Y_{nj} - n^{-1}(Y_{n1} + \cdots + Y_{nn}) + n^{-\frac{1}{2}}Z .$$

Then $\mathscr{X}$ is d-c.a.s.n. for every d, and yet $n^{-\frac{1}{2}}(X_{n1} + \cdots + X_{nn}) = Z$.

Confidence in appropriateness of a multinormal approximation to the distribution of some k_0-dimensional vector variable X will be shaken if some natural array $\mathscr{X}$ in which X is embedded exhibits behavior of the type just described, even if it does manage to be d-c.a.s.n., either for $d = k_0$ only or even for all d.

EXAMPLE 4. A final example is presented to show one way in which intuition may be misleading. Let $(X_{n1}, \cdots, X_{nk_n})$ be i.i.d. with distribution $F_n = (1 - p_n)\Phi_1 + p_n G$ where $G'(x) \sim |x|^{-\alpha}$ for $|x|$ large, where $\alpha > 1$. Then if $k_n = n$, we have $\mathscr{L}(n^{-\frac{1}{2}}(X_{n1} + \cdots + X_{nn})) \to \Phi_1$ provided $n^{\frac{1}{2}(3-\alpha)}p_n \to 0$, for example if $\alpha = \frac{5}{2}$, $p_n = n^{-\frac{1}{3}}$. However in this case if $k_n = n^2$ then

$$\mathscr{L}(n^{-1}(X_{n1} + \cdots + X_{nn^2})) \not\to \Phi_1$$

since $n^{(3-\alpha)}p_n \not\to 0$. Thus it is possible that the n-fold convolution of a df F_n

can approach normality while a further stage of n-fold convolving destroys this limiting behavior.

4. Joint asymptotic normality. One way of defining a concept of joint asymptotic normality, stronger than that of "d-c.a.s.n. for all d," is as follows.

DEFINITION. The array $\mathscr{X}$ is "jointly asymptotically normal" (j.a.n.) if for every constant array $\mathscr{A}$, (of the same shape as $\mathscr{X}$) satisfying $\sum_{j=1}^{k_n} a_{nj}^2 = 1$, we have $\mathscr{L}(\sum_{j=1}^{k_n} a_{nj} X_{nj}) \to \Phi_1$.

It is clear that j.a.n. implies d-c.a.s.n. for all d; each of the first three examples given above demonstrates that the converse is false. In Example 2 above, $\mathscr{X}$ is j.a.n. iff $\mathscr{Y}$ is also. Notice that it is not more restrictive to require that $\mathscr{L}(\Sigma a_{njm} X_{nj}, m = 1, \cdots, d) \to \Phi_d$ where for each n, (a_{njm}) is an $k_n \times d$ matrix with standardized orthogonal columns.

5. A sufficient condition for joint asymptotic normality. We now study conditions that ensure j.a.n. in the case of row-wise independence. Our first theorem is due to P. J. Bickel, who give an entirely different proof [2].

THEOREM 2. *If $\mathscr{X}$ is row-wise independent and c.a.s.n., with $E(X_{nj}) = 0$, $E(X_{nj}^2) \to 1$, then $\mathscr{X}$ is* j.a.n.

Before giving our proof of this theorem, it is convenient to introduce a metric on the class $\mathscr{F}_0$ of distributions with zero mean and finite variance $\mathscr{F}_0 = \{F: \int x \, dF = 0, \int x^2 \, dF < \infty\}$. Define $\rho: \mathscr{F}_0 \times \mathscr{F}_0 \to R_+$ by

$$\rho^2(F, G) = \int_0^1 (f(w) - g(w))^2 \, dw$$

where $f(w)$ is the essentially unique) monotone function satisfying $f(F(x)) = x$ a.e. (F). To verify that ρ is a metric we observe that $\rho \geq 0$ with equality iff $F \equiv G$, and that (using Schwarz) ρ satisfies the triangle inequality.

LEMMA 1.

$$\rho(F_n, G) \to 0 \text{ iff } \{F_n \to G \text{ (in the weak-* topology) } and \int x^2 \, dF_n \to \int x^2 \, dG\}.$$

PROOF. Suppose $\rho \to 0$. Then for each interval (p, q) we have

$$\int_p^q (f_n(w) - g(w))^2 \, dw \to 0,$$

so $F_n \to G$. Also

$$\rho^2(F_n, G) \geq \{(\int f^2 \, dw)^{\frac{1}{2}} - (\int g^2 \, dw)^{\frac{1}{2}}\}^2$$
$$= ((\int x^2 \, dF_n)^{\frac{1}{2}} - (\int x^2 \, dG)^{\frac{1}{2}})^2$$

so that $\int x^2 \, dF_n \to \int x^2 \, dG$. Conversely, take $\varepsilon > 0$, and choose p so that

$$\int_0^p + \int_{1-p}^1 g^2 \, dw \leq \varepsilon.$$

Then for n sufficiently large we shall have (since $F_n \to G$)

$$\int_p^{1-p} (f - g)^2 \, dw < \varepsilon$$

and
$$\left|\int_p^{1-p} (f^2 - g^2)\, dw\right| < \varepsilon \, ,$$
and (since $\int x^2\, dF_n \to \int x^2\, dG$)
$$\left|\int_0^1 (f^2 - g^2)\, dw\right| < \varepsilon \, .$$
Thus
$$\left|\int_0^p + \int_{1-p}^1 f^2\, dw\right| < 3\varepsilon \, ,$$
and
$$\int_0^1 (f - g)^2\, dw < \varepsilon + (\varepsilon^{\frac{1}{2}} + (3\varepsilon)^{\frac{1}{2}})^2 < 9\varepsilon \, .$$

Another representation of ρ is very convenient.

LEMMA 2.
$$\rho^2(F_1, F_2) = \min_{\lambda \in \Lambda(F_1, F_2)} \int (x - y)^2\, d\lambda(x, y)$$
where $\Lambda(F_1, F_2)$ is the class of bivariate distribution functions λ on $R \times R$ such that $\lambda(x, \infty) = F_1(x)$, $\lambda(\infty, y) = F_2(y)$.

PROOF. First suppose that F_1, F_2 have bounded support, so that $F_1(\alpha_1) = 0$, $F_1(\beta_1) = 1$, $F_2(\alpha_2) = 0$, $F_2(\beta_2) = 1$ with α_1, β_1, α_2, β_2 finite. We have to show that $\max \int xy\, d\lambda(x, y) = C_{12}$ where $C_{12} = \int f_1(p) f_2(p)\, dp$.

Integrating by parts, we have
$$\int xy\, d\lambda(x, y) = \beta_1 \int y\, dF_2 + \beta_2 \int x\, dF_1 - \beta_1 \beta_2 + \int\int \lambda(x, y)\, dx\, dy \, .$$

However $\lambda(x, y) \leqq \lambda_0(x, y)$ where $\lambda_0(x, y) = \min(F_1(x), F_2(y))$, and since $\lambda_0 \in \Lambda(F_1, F_2)$ we have immediately that
$$\max \int xy\, d\lambda(x, y) = \int xy\, d\lambda_0(x, y) = C_{12} \, .$$

Now we remove the boundedness condition. First, notice that in all cases $\lambda_0 \in \Lambda(F_1, F_2)$ so that C_{12} is a possible value of $\int xy\, d\lambda$. We show that for all $\lambda \in \Lambda$ and $\varepsilon > 0$, $\int xy\, d\lambda < C_{12} + \varepsilon$. Choose ε. Since $\int x\, dF_i = 0$ and $\int x^2\, dF_i < \infty$ for $i = 1, 2$, we can choose δ so small that
$$\delta < \varepsilon \, , \qquad \int_0^\delta + \int_{1-\delta}^1 f_i^2(p)\, dp < \varepsilon \, , \qquad f_i(\delta) = \alpha_i < 0 \, , \qquad f_i(1 - \delta) = \beta_i > 0 \, ,$$
$i = 1, 2$. Replace f_i by $f_{i0}(p) = \alpha_i$ for $0 < p < \delta$, $= f_i(p)$ for $\delta < p < 1 - \delta$, $= \beta_i$ for $1 - \delta < p < 1$, and replace F_i by the induced df's. F_{i0}, $i = 1, 2$. Then $F_{i0}(x) = F_i(x)$ for $\alpha_i < x < \beta_i$. Using Schwarz, we find $|\int f_1 f_2\, dp - \int f_{10} f_{20}\, dp| < 2\varepsilon$. Also, given any $\lambda \in \Lambda(F_1, F_2)$ we can construct $\lambda^0 \in \Lambda(F_{10}, F_{20})$ by setting $\lambda^0(x, y) = 0$ for $x < \alpha_1$ and for $y < \alpha_2$, $= \lambda(x, y)$ for $\alpha_1 \leqq x < \beta_1$, $\alpha_2 \leqq y < \beta_2$, $= F_1(x)$ for $\alpha_1 \leqq x < \beta_1$, $\beta_2 \leqq y$, $= F_2(y)$ for $\beta_1 \leqq x$, $\alpha_2 \leqq y < \beta_2$, $= 1$ for $\beta_1 \leqq x$, $\beta_2 \leqq y$. Then (again using Schwarz)
$$\left|\int xy\, d\lambda - \int xy\, d\lambda^0\right| \leqq 2\left(\int_{-\infty}^{\alpha_1} + \int_{\beta_1}^\infty\right) \int_{-\infty}^\infty + 2 \int_{-\infty}^\infty \left(\int_{-\infty}^{\alpha_2} + \int_{\beta_2}^\infty\right) |xy|\, d\lambda(x, y) \leqq A\varepsilon$$
where
$$A = 2\left(\int y^2\, dF_2(y)\right)^{\frac{1}{2}} + 2\left(\int x^2\, dF_1(x)\right)^{\frac{1}{2}} \, .$$

However we have already shown that $\int xy\,d\lambda^0 \leqq \int f_{10}f_{20}\,dp$, so the result follows.

The metric ρ is introduced because it has one very useful property, as follows.

LEMMA 3. *If*

$$H = \mathscr{L}(\sum_{i=1}^{k} a_i X_i)$$

where $X_1, X_2, \cdots$ *are independent with* $\mathscr{L}(X_i) = F_i \in \mathscr{F}_0$, *and* $\Sigma a_i^2 = 1$, *then* $H \in \mathscr{F}_0$ *and*

$$\rho^2(H, \Phi_1) \leqq \sum_{i=1}^{k} a_i^2 \rho^2(F_i, \Phi_1) .$$

PROOF. Using the representation of Lemma 2 we have (for $k = 2$)

$$
\begin{aligned}
\rho^2(H, \Phi_1) &= \min_{\lambda \in \Lambda(H,\Phi_1)} \int (x - u)^2 \, d\lambda(x, u) \\
&\leqq \inf_{(\lambda,\mu) \in \Lambda(F_1,\Phi_1) \times \Lambda(F_2,\Phi_1)} \int (a_1 y + a_2 z - a_1 v - a_2 w)^2 \, d\lambda(y, v) \, d\mu(z, w) \\
&= \inf_{\lambda \in \Lambda(F_1,\Phi_1)} \inf_{\mu \in \Lambda(F_2,\Phi_1)} \int a_1^2 (y - v)^2 \, d\lambda(y, v) + \int a_2^2(z - w)^2 \, d\mu(z,w) \\
&= a_1^2 \rho^2(F_1, \Phi_1) + a_2^2 \rho^2(F_2, \Phi_1)
\end{aligned}
$$

where in the second step we have used the fact that to every pair (λ, μ) in $\Lambda(F_1, \Phi_1) \times \Lambda(F_2, \Phi_1)$ there corresponds a unique $\lambda \in \Lambda(H, \Phi_1)$ induced by the projection $X = a_1 Y + a_2 Z$, $U = a_1 V + a_2 W$, and where in the third step we have used the facts that $\int x\,dF_1 = \int x\,dF_2 = \int x\,d\Phi_1 = 0$. The general result follows by induction on k.

PROOF OF THEOREM 2. Put $\rho_{nj} = \rho(F_{nj}, \Phi_1)$, $\rho_n = \max_j \rho_{nj}$. Then since $\mathscr{X}$ is c.a.s.n. and $EX_{nj}^2 \to 1$, $\rho_n \to 0$. From Lemma 3 it follows that for any array of constants $\{a_{nj}\}$ (with $\sum_j a_{nj}^2 = 1$) we have $\rho(\mathscr{L}(\sum_j a_{nj} X_{nj}), \Phi_1) \leqq \rho_n$, so that $\mathscr{L}(\sum_j a_{nj} X_{nj}) \to \Phi_1$; i.e. $\mathscr{X}$ is j.a.n.

REMARK 1. Clearly the conditions of Theorem 2 are not necessary; $\mathscr{X}$ can be j.a.n. even if a bounded number of elements of each row of $\mathscr{X}$ fail to satisfy $E(X_{nj}^2) \to 1$.

REMARK 2. It is impossible to define a metric δ on the class of all distribution functions with zero mean such that δ induces the weak-* topology and such that $a^2 + b^2 = 1$ implies

$$\delta(\mathscr{L}(aX + bY), \Phi_1) \leqq \max(\delta(\mathscr{L}(X), \Phi_1), \delta(\mathscr{L}(Y), \Phi_1)) .$$

Such a metric, if it existed, could be used to prove that independence + c.a.s.n. $\Rightarrow$ j.a.n., which is false by Example 1.

6. Necessary and sufficient conditions. We now present an n and s condition for j.a.n. in the case row-wise independence. It turns out that there is a sequence of "worst case" linear combinations; once these are under control all the others follow. We need a technical result.

LEMMA 4. *Suppose* $v_1, v_2, \cdots, v_k$ *are non-increasing right-continuous functions*

of y for $y \geqq 0$, where $v(0) = \infty$ is allowed, and for $z > 0$ set $v(z) = \inf \max_j v_j(y_j)$ where the infimum is taken over nonnegative $y_1, \cdots, y_k$ summing to z. Then for each z the infimum is attained at some point $(\eta_1, \cdots, \eta_k)$, $v(z)$ is non-increasing and right-continuous, and for all z, j either (i) $\eta_j = 0$ and $v_j(0) \leqq v(z)$, *or* (ii) $v_j(\eta_j) = v(z)$ *or* (iii) $v_j(\eta_j) < v(z)$, $v_j(\eta_j - 0) > v(z)$.

The proof by induction on k is straightforward.

For each n, we apply this result to the functions

$$V_{nj}(a_{nj}) = \int x^2 I_{\{|a_{nj}x| < 1\}} \, dF_{nj}(x) \qquad\qquad j = 1, \cdots, k_n$$

with the constraint $\Sigma_j a_{nj}^2 = 1$, obtaining a minimax V_n attained at $\alpha_{n1}, \cdots, \alpha_{nk_n}$. We can thus define an array $\mathcal{X}^T$ by setting $P\{X_{nj}^T = X_{nj} \text{ or } 0\} = 1$, with $P\{X_{nj}^T = X_{nj}\} = 1$ if $\alpha_{nj} = 0$ or if $\alpha_{nj} > 0$, $|\alpha_{nj} X_{nj}| < 1$; $= p_{nj}$ if $|\alpha_{nj} X_{nj}| = 1$; $= 0$ if $|\alpha_{nj} X_{nj}| > 1$, where p_{nj} is so chosen that $E(X_{nj}^T)^2 = V_n$ for all n, j for which $\alpha_{nj} > 0$. Now define $\mu_{nj}^T = E(X_{nj}^T)$, $\beta_{nj} = \alpha_{nj} \operatorname{sgn}(\mu_{nj}^T + 0)$, so that $|\beta_{nj}| = \alpha_{nj}$, $\beta_{nj} \mu_{nj}^T = \alpha_{nj} |\mu_{nj}^T|$.

THEOREM 3. *If $\mathcal{X}$ is row-wise independent, then $\mathcal{X}$ is j.a.n. iff* (i) $\mathcal{X}$ is (ii) $\mathscr{L}(\Sigma_j \beta_{nj} X_{nj}) \to \Phi_1$, *and* (iii) $\Sigma(\mu_{nj}^T)^2 \to 0$, *where $\{\beta_{nj}\}$ and $\{\mu_{nj}^T\}$ are as c.a.s.n., defined above.*

PROOF. Suppose (i) and (ii) hold, and define the array $\mathcal{X}^T$ as above. Since $\mathcal{X}$ is c.a.s.n. $V_{nj}(a) \to \int x^2 I_{\{|ax| < 1\}} \, d\Phi(x)$ uniformly in each a-interval not including the origin; it follows that $\max_j |\alpha_{nj}| \to 0$. Thus for all $\varepsilon > 0$ $\max_j P\{|\beta_{nj} X_{nj}| \geqq \varepsilon\} \to 0$, and since $\mathscr{L}(\Sigma_j \beta_{nj} X_{nj}) \to \Phi_1$ it follows from [4] page 316 that $\Sigma_j P\{|\beta_{nj} X_{nj}| \geqq 1\} \to 0$. Thus $\Sigma_j P\{X_{nj} = X_{nj}^T\} \to 1$, so that $P\{X_{nj} = X_{nj}^T, 1 \leqq j \leqq k_n\} \to 1$ and hence $\mathscr{L}(\Sigma_j \beta_{nj} X_{nj}^T) \to \Phi_1$. Appealing to [4] again we have $\Sigma_j \beta_{nj} \mu_{nj}^T \to 0$, $\Sigma_j \beta_{nj}^2 (V_n - (\mu_{nj}^T)^2) \to 1$. Thus $\Sigma_j \alpha_{nj} |\mu_{nj}^T| \to 0$, so that $\Sigma_j \beta_{nj}^2 (\mu_{nj}^T)^2 \to 0$, and hence $V_n = \Sigma_j \beta_{nj}^2 V_n \to 1$. However, since $\mathscr{L}(X_{nj}^T) \to \Phi_1$, $\liminf \operatorname{Var}(X_{nj}^T) \geqq 1$ and so $\operatorname{Var}(X_{nj}^T) \to 1$. Now define $\mathcal{X}^{TC}$, a centered version of $\mathcal{X}^T$, by $X_{nj}^{TC} = X_{nj}^T - \mu_{nj}^T$. Theorem 2 applies, and we deduce that $\mathcal{X}^{TC}$ is j.a.n.

Thus for any standard array $\{a_{nj}\}$ we have $\mathscr{L}(\Sigma_j a_{nj} X_{nj}^{TC}) \to \Phi_1$; thus

$$\mathscr{L}(\Sigma_j a_{nj} X_{nj} - \Sigma_j a_{nj} \mu_{nj}^T) \to \Phi_1 .$$

Assumption (iii) now ensures that $\mathscr{L}(\Sigma_j a_{nj} X_{nj}) \to \Phi_1$, i.e., $\mathcal{X}$ is j.a.n.

Conversely, if $\mathcal{X}$ is j.a.n. then (i) and (ii) are trivially necessary; the above argument then shows that $\mathcal{X}^{TC}$ is also j.a.n. Taking $a_{nj} = \mu_{nj}^T / (\Sigma_j (\mu_{nj}^T)^2)^{\frac{1}{2}}$ (or arbitrarily, if $\Sigma_j (\mu_{nj}^T)^2 = 0$) we deduce that (iii) must hold.

7. Two special structures. Clearly it would be desirable to have conditions for j.a.n. when row-wise independence was not present, but we have no results in this direction. We have examined to some extent two special structures. Let F be a distribution function that is nonnormal, but is absolutely continuous

with the zero mean, unit variance, and finite third moment. Then ([6]) in the situation of Example 2 above, if $F_n = F$ for all n, each sub-array of $\mathscr{X}$ of size $\{k_n\}$ is j.a.n. provided $2^{-n}k_n \to 0$.

Thus in this case, only a vanishingly small fraction of the contrasts can be j.a.n. Of course, it is possible for the order-statistics of the contrasts to exhibit approximate normal-theory behavior without full approximate joint normality obtaining; some results in this direction were obtained in [6]. For example, if $F_n = F$ for all n, it is easily shown that the moment-ratio statistics $g_1 = k_n^{\frac{1}{2}}\Sigma X^3/(\Sigma X^2)^{\frac{3}{2}}$ and $g_2 = k_n \Sigma X^4/(\Sigma X^2)^2$ computed from the contrasts will converge in mean square to their normal-theory values provided $\int x^8 \, dF < \infty$; under certain other assumptions of F it is possible to derive the asymptotic covariance function of the process $k_n^{\frac{1}{2}}(G_n(x) - F_n(x))$ where $F_n = \mathscr{L}(k_n^{-\frac{1}{2}}\Sigma Y_i)$, and G_n is the empiric distribution function of the contrasts. This covariance function differs from the normal-theory one if $\int x^4 \, dF \neq 3(\int x^2 \, dF)^2$.

For another example, consider a sequence of linear processes $Z_n = \{Z_{nt}, t = \cdots, -1, 0, 1, \cdots\}$, $n = 1, 2, \cdots$ where $Z_{nt} = \Sigma_u a_{nu} Y_{t-u}$ with $\cdots, Y_{-1}, Y_0, Y_1, \cdots$ being independent with $\mathscr{L}(Y_u) = F$ for all u, and where $\Sigma a_{nu}^2 = 1$, $\max_u |a_{nu}| \to 0$. Construct an array $\mathscr{W}$ by setting $W_{nj} = Z_{nt_{nj}}$, where the array $\{t_{nj}\}$ is so chosen that the eigenvalues of the covariance matrix Σ_n of $W_n = (W_{n1}, \cdots, W_{nk_n})$ are uniformly bounded away from zero, and finally construct $\mathscr{X}$ by setting $(X_{n1}, \cdots, X_{nk_n}) = X_n = W_n \Sigma_n^{\frac{1}{2}}$. Then $\mathscr{X}$ is j.a.n. (compare [5]).

Acknowledgment. Conversations with L. A. Jaeckel were very helpful.

REFERENCES

[1] ARVESEN, JAMES N. (1969). Jackknifing U-statistics. *Ann. Math. Statist.* **40** 2076–2100.
[2] BICKEL, P. J. (1970). Personal communication.
[3] DANIEL, C. (1959). Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics* **1** 311–342.
[4] LOÈVE, M. (1960). *Probability Theory* (3rd ed.). Van Nostrand, Princeton.
[5] MALLOWS, C. L. (1967). Linear processes are nearly Gaussian. *J. Appl. Prob.* **4** 313–329.
[6] MALLOWS, C. L. (1969). Joint normality induced by orthogonal transformations. Bell Telephone Laboratories Memorandum.
[7] MOSTELLER, F. and TUKEY, J. W. (1968). Data analysis, including statistics, in *Handbook of Social Psychology* 2nd ed. **2** (Lindzey and Aronson, eds.) Addison-Wesley, Reading, Chapter 10.