# INVARIANCE UNDER DEPENDENCE BY MIXING

By D.R. Jensen

*Virginia Polytechnic Institute and State University*

This paper is concerned with arrays of conditionally independent random elements that become dependent by mixing. The principal focus is the preservation of properties known to hold under independence. Findings are reported in the context of limit theory, including laws of large numbers and central limit theory, and topics in statistical inference. Several standard results, ranging from Berry-Esseén bounds in central limit theory to the use of Friedman's (1937) test in the analysis of two-way data, are seen to remain valid under certain models for dependence. The class of limit laws for standardized sums is expanded to include dependent cases, as are bounds on rates of convergence to these limits.

**1. Introduction.** Independence and conditional independence are central to probability theory and its applications, supporting the theory of Markov chains, Bayesian analysis, limit theory, and the foundations of statistical inference. See Dawid (1979), for example. A systematic study of conditionally independent events dates back at least to de Finetti (1937), who postulated these as models for conditional independence in a random environment. More recently, conditionally independent events have been studied as models for weak dependence (cf. Dykstra et al. (1973), Shaked (1977), and Tong (1980), for example), but there the primary focus centers on inequalities relating unconditional joint probabilities to products of marginal probabilities.

The assumption of independence pervades much of mathematical statistics, including essential portions of parametric and nonparametric statistical inference. That independence is a fundamental mathematical concept is not in doubt. Much less clear is the extent to which it mimics reality. Indeed, apart from highly specialized models, there appear to be no omnibus empirical tests for genuine independence. On physical grounds it even may be argued that observable phenomena at best can be only conditionally independent, owing to the common background energy attributed to the big bang.

*AMS 1980 subject classifications.* Primary 60E05; secondary 60F05, 62H05.

*Key words and phrases.* Conditional independence, models for dependence, limit theory, unconditional inference.

In view of these uncertainties, it is essential to regard independence as but one of many model assumptions subject to misspecification. It then is pertinent to examine questions of robustness and even invariance of critical properties to the assumption of independence. That is the focus of this paper, an outline of which follows.

Supporting developments and models for dependence are given in Sections 2 and 3. In Section 4 we consider topics in limit theory. These topics include laws of large numbers, central limit theory, a study of the types of limit laws achieved by conditionally independent sequences, and versions of Berry-Esseén bounds appropriate for these. Section 5 studies topics in statistical inference under dependence. These include the relative sensitivities of experiments, the use of Friedman's (1937) test in the analysis of two-way data under dependence, and the use of Anderson's (1984) classification statistic under dependence by scaling. We infer that numerous properties of these procedures continue to hold exactly under certain models for dependence.

**2. Preliminaries.** To fix notation $\Re^k$ and $\Re^k_+$ are Euclidean $k$-space and its positive orthant; $F_{n\times k}$ is the space of real $(n \times k)$ matrices; $S_k$ and $S_k^+$ consist of symmetric $(k \times k)$ matrices and their positive semidefinite varieties; $\mathbf{X}' = [X_1, \ldots, X_k]$ is the transpose of $\mathbf{X} \in \Re^k$; and $\| \bullet \|$ is the Euclidean norm on $\Re^k$. Special arrays are the $(k \times k)$ identity $\mathbf{I}_k$, the unit vector $\mathbf{1}_k = [1, \ldots, 1]' \in \Re^k$, the Kronecker product $\mathbf{A} \times \mathbf{B} = [a_{ij}\mathbf{B}]$, and the block diagonal array Diag $(\mathbf{A}_1, \ldots, \mathbf{A}_r)$. Cumulative distribution, probability density, and characteristic functions are abbreviated as *cdf*, *pdf*, and *chf*, respectively, with $\mathcal{L}(\mathbf{X})$ as the distribution of $\mathbf{X}$. Let $(\Omega, B(\Omega), Q)$ be a probability space. We are concerned with the probability measure $P$ on $(\Re^k, B(\Re^k))$, the *cdf* $F(\mathbf{x})$, and the *chf* $\phi_x(\mathbf{t})$ generated by $\mathbf{X}(\omega) \in \Re^k$, as well as sequences $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ on $(\Re^k)^n$ which, on occasion, are independent and identically distributed (*iid*). Moment arrays for $\mathbf{X} = [X_1, \ldots, X_k]' \in \Re^k$, when defined, are the expected vector value $E(\mathbf{X}) = \theta \in \Re^k$, the dispersion matrix $D(\mathbf{X}) = E(\mathbf{X} - \theta)(\mathbf{X} - \theta)' \in S_k^+$, and the absolute central moments $\{\beta_{\delta j} = E \mid X_j - \theta_j \mid^\delta; 1 \leq j \leq k\}$ of order $\delta > 0$. Some special distributions on $\Re^k$ are the Gaussian law $N_k(\theta, \Sigma)$ having the mean $\theta \in \Re^k$, the dispersion matrix $\Sigma \in S_k^+$, and the *pdf* $f_N(\mathbf{x}; \theta, \Sigma)$, and mixtures of these. In particular, $H_k(\theta, G) = \int_{S_k^+} N_k(\theta, \mathbf{S})dG(\mathbf{S})$ is a Gaussian mixture over $S_k^+$ with respect to $G(\bullet)$ having the *pdf*

$$f(\mathbf{x}; \theta, G) = \int_{S_k^+} f_N(\mathbf{x}; \theta, \mathbf{S})dG(\mathbf{S}); \tag{1}$$

and $\mathcal{H}_k = \{H_k(\theta, G); \theta \in \Re^k, G \in M(S_k^+)\}$ is the class of all such mixtures, where $M(S_k^+)$ is the class of all probability measures on $S_k^+$. Specifically, $\mathcal{H}_{k0} = \{H_{k0}(\mathbf{0}, G); G \in M(S_k^+)\} \subset \mathcal{H}_k$ consists of dispersion mixtures of zero-mean Gaussian laws on $\Re^k$. The special case for which $\mathbf{S} = s\Sigma$, with $\Sigma$ fixed and $s$ random on $\Re^1_+$, yields scale mixtures of Gaussian laws within the class of ellipsoidal distributions on $\Re^k$. The Gaussian law on $F_{n\times k}$ representing an *iid* sample

$\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_n]'$ from $N_k(\theta, \Sigma)$ is denoted by $N_{n\times k}(\mathbf{1}_n \times \theta', \mathbf{I}_n \times \Sigma)$.

**3. Models for Dependence.** We develop models for dependence through mixing conditionally independent arrays. These consist of sequences of random elements arising in limit theory, as well as two-way arrays with applications in inference. It is natural to use *chf*'s, for which bounded convergence applies directly.

*3.1. Conditionally Independent Sequences.* We consider random sequences in $(\Re^k)^n$, on occasion specializing to the case $k = 1$. Let $(\Gamma, B(\Gamma), \mu)$ be a probability space, and consider a collection $\{\phi_i(\mathbf{t}_i; \gamma); 1 \le i \le n\}$ of $B(\Gamma)$-measurable functions such that each $\phi_i(\mathbf{t}_i; \gamma)$ is a *chf* on $\Re^k$ for each $\gamma \in \Gamma$. Let $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ be a sequence on $(\Re^k)^n$ whose joint *chf* is given by

$$\phi_{\mathbf{X}}(\mathbf{t}_1, \ldots, \mathbf{t}_n) = \int_\Gamma \Pi_{i=1}^n \phi_i(\mathbf{t}_i; \gamma) d\mu(\gamma) \tag{2}$$

with $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_n] \in F_{k\times n}$. Clearly $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ are conditionally independent with mixing measure $\mu$. The special case for which $\{\phi_i(\mathbf{t}_i; \gamma) = \phi(\mathbf{t}_i; \gamma); 1 \le i \le n\}$ follows when $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ are conditionally *iid*. For the latter case with $k = 1, \{\{X_1, \ldots, X_n\}; n = 1, 2, \ldots\}$ is a de Finetti sequence on $\Re^\infty$.

Observe that $\gamma$ may be a scalar, a vector, a matrix, or an arbitrary random element to be denoted by $\boldsymbol{\gamma}$. The concept of dependence by scaling is made precise in the following definition, where $\gamma$ and the typical element $\mathbf{Z}_i$ conform for multiplication.

Definition 1. Random elements $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ are said to be *dependent by scaling* if there are independent random elements $\{\mathbf{Z}_1, \ldots, \mathbf{Z}_n\}$ and a random element $\boldsymbol{\gamma}$, independent of $\{\mathbf{Z}_1, \ldots, \mathbf{Z}_n\}$, such that $\mathcal{L}(\mathbf{X}_1, \ldots, \mathbf{X}_n) = \mathcal{L}(\gamma\mathbf{Z}_1, \ldots, \gamma\mathbf{Z}_n)$. In particular, random vectors $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ on $(\Re^k)^n$ are called *dependent by coordinate scaling* if $\mathcal{L}(\mathbf{X}_1, \ldots, \mathbf{X}_n) = \mathcal{L}(\boldsymbol{\gamma}\mathbf{Z}_1, \ldots, \boldsymbol{\gamma}\mathbf{Z}_n)$ such that $\boldsymbol{\gamma} = \mathrm{Diag}(\gamma_1, \ldots, \gamma_k)$.

Standard limit theorems depend heavily on moments. It is useful to distinguish orders of dependence under mixing, at issue being the manner in which the conditional moments depend on the mixing variable.

Definition 2. Conditionally independent random variables are said to be *dependent of order r* if their conditional moments of order $r$ depend on the mixing parameter $\gamma \in \Gamma$.

*3.2. Conditionally Independent Ensembles.* We assemble a collection of random elements in a two-way array having $r$ rows. In particular, let $\{\mathbf{X}_i \in F_{n_i\times k_i}; 1 \le i \le r\}$ be random having the *chf*'s $\{\{\phi_i(\mathbf{T}_i; \gamma_i); \gamma_i \in \Gamma_i\}; 1 \le i \le r\}$, and let $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_\mathbf{r}]$. Our model for an ensemble of conditionally independent random components is given by the joint *chf*

$$\phi_{\mathbf{X}}(\mathbf{T}_1, \ldots, \mathbf{T}_r) = \int_\Gamma \Pi_{i=1}^r \phi_i(\mathbf{T}_i; \gamma_i) d\mu(\gamma_1, \ldots, \gamma_r) \tag{3}$$

where $\mu(\bullet, \ldots, \bullet)$ is a mixing measure on $\Gamma = \Gamma_1 \times \cdots \times \Gamma_r$. If $\mu(\bullet)$ is concentrated along the equiangular line $\{\gamma_1 = \cdots = \gamma_r\}$ in $\Gamma$, then this model reduces to that of Section 3.1. Different versions of (3) are developed in Section 5.

**4. Dependent Limit Theorems.** We consider various modes of stochastic convergence, giving unconditional laws of large numbers and results in central limit theory for conditionally independent sequences. The classical limit theorems depend heavily on moments. Here we require conditional moments of given order, whereas unconditional moments need not be defined. Our basic approach uses *chf*s. Together with the Dominated Convergence Theorem, this justifies limits of mixtures as mixtures of limits, so that unconditional versions of the classical limit theorems follow on mixing.

*4.1. Laws of Large Numbers.* Let $\{\mathbf{X}_n; n = 1,2,\ldots\}$ be a sequence of conditionally independent random elements on $\Re^k$, and let $\{\theta_n; n = 1,2,\ldots\}$ be a sequence of parameters. We are concerned with the unconditional convergence of $\{(\mathbf{X}_n - \theta_n); n = 1,2,\ldots\}$ in various stochastic modes. Under the modes of Section 3.1 our basic tools are the representation $P(A) = \int_\Gamma P_\gamma(A)d\mu(\gamma)$, the variance formula $\mathrm{Var}(X_n) = E_\gamma[\mathrm{Var}(X_n \mid \gamma)] + \mathrm{Var}_\gamma[E(X_n \mid \gamma)]$ on $\Re^1$, and the corresponding dispersion formula $D(\mathbf{X}_n) = E_\gamma[D(\mathbf{X}_n \mid \gamma)] + D[E(\mathbf{X}_n \mid \gamma)]$ on $\Re^k$.

Suppose (i) that $\{\mathbf{X}_n; n = 1,2,\ldots\}$ are conditionally independent with mixing parameter $\gamma$, (ii) that $\{\theta_n; n = 1,2,\ldots\}$ is a sequence of constant elements not depending on $\gamma \in \Gamma$, and (iii) that $(\mathbf{X}_n - \theta_n) \to \mathbf{0}$ in some mode for each $\gamma \in \Gamma$. Assumption (ii) assures unconditional convergence to a constant rather than to a random variable. Basic relationships among various conditional and unconditional modes of convergence are summarized in Table 1, arranged in pairs as Cases 1–4 in which conditional convergence in the first mode of each pair implies unconditional convergence in the second mode of that pair.

TABLE 1. Unconditional modes of convergence implied by conditional convergence for each $\gamma \in \Gamma$.

| Case | Conditional Mode | Unconditional Mode |
|---|---|---|
| 1 | Almost Sure | Almost Sure |
| 2 | Mean Square | Mean Square* |
| 3 | Mean Square | In Probability |
| 4 | In Probability | In Probability |

*Assuming unconditional second moments.

The claims in Table 1 are easily verified using standard arguments. For Case 1, to show that almost sure conditional convergence implies almost sure convergence unconditionally, let $A = \{\omega : (\mathbf{X}_n(\omega) - \theta_n) \to \mathbf{0}\}$. Then $\{P_\gamma(A) = 1; \gamma \in \Gamma\}$ by hypothesis, so that $P(A) = \int_\Gamma P_\gamma(A)d\mu(\gamma) = 1$ unconditionally. For Case 2, with $E(X_n \mid \gamma) = \mu_n(\gamma)$ on $\Re^1$, it follows that $E[(X_n - \theta_n)^2] = E_\gamma[\text{Var}(X_n \mid \gamma) + (\mu_n(\gamma) - \theta_n)^2]$, where the expression $[\text{Var}(X_n \mid \gamma) + (\mu_n(\gamma) - \theta_n)^2] \to 0$ for each $\gamma \in \Gamma$ by hypothesis. Parallel arguments apply in the vector case. Case 3 follows from suitable versions of Chebychev inequalities on $\Re^1$ and $\Re^k$ applied conditionally. Case 4 follows on using bounded convergence. Observe that Case 2 applies automatically whenever the conditional moments to second order do not depend on $\gamma$.

Various laws of large numbers follow from the foregoing developments unconditionally without difficulty. Details are supplied for the following version of Khintchine's Theorem on $\Re^k$, where moments of the unconditional distribution are not required in order to validate the result.

THEOREM 4.1. *(Khintchine). Let* $\{\mathbf{X}_n; n = 1, 2, \ldots\}$ *be conditionally iid on* $\Re^k$ *having the conditional mean* $E(\mathbf{X}_n \mid \gamma) = \theta$. *Then the sample mean* $\bar{\mathbf{X}}_n = (\mathbf{X}_1 + \cdots + \mathbf{X}_n)/n$ *is weakly consistent for* $\theta \in \Re^k$.

PROOF. Use *chf*s and dominated convergence to write

$$\lim_{n\to\infty} \phi_{\bar{\mathbf{X}}_n}(\mathbf{t}) = \int_\Gamma \lim_{n\to\infty} [\phi(\mathbf{t}/n; \gamma)]^n d\mu(\gamma). \tag{4}$$

Expanding under the integral gives

$$\lim_{n\to\infty} [1 + i\mathbf{t}'\theta/n + o(\| \mathbf{t} \| /n)^n = e^{i\mathbf{t}'\theta} \tag{5}$$

so that $\lim_{n\to\infty} \phi_{\bar{\mathbf{X}}_n}(\mathbf{t}) = e^{i\mathbf{t}'\theta}$ because $\theta$ does not depend on $\gamma$. This is the *chf* of the distribution on $\Re^k$ degenerate at $\theta$, thus completing our proof.

It deserves emphasis that moments of the unconditional distribution are not required to exist, yet a weak law of large numbers nonetheless applies unconditionally on $\Re^k$.

*4.2. Central Limit Theory.* We consider scalar and vector sequences of the types described in Section 3.1. Not only are the limit laws characterized, but dependent versions of local and global Berry-Esseén bounds are developed. In particular, $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ is a conditionally independent sequence on $(\Re^k)^n$ having the typical *cdf* $G_i(\bullet)$ and moments $E(\mathbf{X}_i \mid \gamma) = \theta$ and $D(\mathbf{X}_i \mid \gamma) = \Sigma_i(\gamma)$. Consider $\mathbf{Z}_n = n^{1/2}(\bar{\mathbf{X}}_n - \theta)$; and let $\phi_n(\mathbf{t})$, $F_n(\mathbf{x})$, and $P_n(\bullet)$ respectively be the *chf*, *cdf*, and probability measure induced by $\mathbf{Z}_n$, with $P(\bullet)$ as the weak limit $P(\bullet) = \lim_{n\to\infty} P_n(\bullet)$. Define the Lindeberg function on $\Re^k$ as

$$L_n(z; \gamma) = n^{-1}\Sigma_{i=1}^n \int_{\|\mathbf{x}\| > z} \| \mathbf{x} \|^2 \, dG_i(\mathbf{x}). \tag{6}$$

A principal result is the following.

THEOREM 4.2. *Let* $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ *be conditionally independent on* $(\Re^k)^n$ *having the typical conditional moments* $E(\mathbf{X}_i \mid \gamma) = \mathbf{0}$ *and* $D(\mathbf{X}_i \mid \gamma) = \Sigma_i(\gamma)$. *Suppose that as* $n \to \infty$, $n^{-1} \sum_{i=1}^n \Sigma_i(\gamma) \to \Sigma(\gamma) \neq \mathbf{0}$ *and* $L_n(n^{1/2}\varepsilon; \gamma) \to 0$ *for each* $\varepsilon > 0$

*and $\gamma \in \Gamma$. Then the limit distribution of $\mathbf{Z}_n = n^{1/2}\bar{\mathbf{X}}_n$ exists in the class $\mathcal{H}_{k0}$, and its chf is given by*

$$\lim_{n\to\infty} \phi_n(\mathbf{t}) = \int_\Gamma e^{-\mathbf{t}'\boldsymbol{\Sigma}(\gamma)\mathbf{t}/2} d\mu(\gamma). \tag{7}$$

PROOF. Write the *chf* $\phi_n(\mathbf{t})$ in its mixture representation and use dominated convergence to get

$$\lim_{n\to\infty} \phi_n(\mathbf{t}) = \int_\Gamma \lim_{n\to\infty} [\Pi_{i=1}^n \phi_i(\mathbf{t}/n^{1/2}; \gamma)] d\mu(\gamma) \tag{8}$$

where

$$\lim_{n\to\infty} [\Pi_{i=1}^n \phi_i(\mathbf{t}/n^{1/2}; \gamma)] = e^{-\mathbf{t}'\boldsymbol{\Sigma}(\gamma)\mathbf{t}/2} \tag{9}$$

using standard arguments. This completes our proof.

Limit laws for conditionally independent sequences having conditional second moments are seen to be dispersion mixtures of Gaussian laws belonging to the class $\mathcal{H}_{k0}$. This was noted by Taylor et al. (1985) for exchangeable sequences on $\Re^1$. The class of limit distributions is thus larger than for the independent case. A special case of some interest is that $\boldsymbol{\Sigma}(\gamma) = \gamma\boldsymbol{\Sigma}$, with $\gamma$ a positive scalar, in which case the limit laws are scale mixtures of Gaussian laws. This class contains the ellipsoidal stable laws as a proper subclass.

COROLLARY 4.2.1. *Let $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ be dependent of order $r$ for $r > 2$ but not for $r \leq 2$. Then the limit distribution of $\mathbf{Z}_n$ is Gaussian, as when $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ are independent.*

*4.3. Berry-Esseén Bounds.* We first consider the scalar case, for which we seek global and local bounds as well as invariance properties of the usual Berry-Esseén bounds. Owing to space constraints, we consider only conditionally *iid* sequences on $\Re^k$; more general results follow along similar lines. The following result gives lower and upper bounds on the rate of convergence for conditionally independent sequences on $\Re^1$.

THEOREM 4.3. *Let $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ be conditionally independent random variables on $\Re^1$ having the typical cdf $G_i(\bullet; \gamma)$ with conditionally zero means, the variance $\sigma_i^2(\gamma)$, and finite third moments $\beta_{3i}(\gamma)$. Let $F_n(\bullet)$ be the cdf of $Z_n = n^{1/2}\bar{X}_n$, and let $F(\bullet) = \lim_{n\to\infty} F_n(\bullet)$ be its limit. Then there are absolute constants $c_1$ and $c_2$ such that the bounds*

$$\begin{aligned} c_1 \int_\Gamma L_n(S_n; \gamma) d\mu(\gamma) &\leq \sup_x \mid F_n(x) - F(x) \mid \\ &\leq c_2 \int_\Gamma \int_0^{S_n(\gamma)} (S_n(\gamma))^{-1} L_n(z; \gamma) dz d\mu(\gamma) \end{aligned} \tag{10}$$

hold whenever the integrals are defined, where $S_n^2(\gamma) = \sigma_1^2(\gamma) + \cdots + \sigma_n^2(\gamma)$ and $L_n(z;\gamma) = [S_n^2(\gamma)]^{-1}\Sigma_{i=1}^n \int_{||x||>z} x^2 dG_i(x;\gamma)$.

PROOF. Apply the bounds of Studnev and Ignat (1967) conditionally.

THEOREM 4.4. *Let the conditionally iid variables $\{X_1, \ldots, X_n\}$ on $\Re^1$ be dependent by scaling, having the conditional cdf $G(\bullet;\gamma)$ with moments as in Theorem 4.3. Then a global bound is given by*

$$\sup_x \mid F_n(x) - F(x) \mid \leq \frac{c\beta_3(1)}{n^{1/2}[\sigma^2(1)]^{3/2}} \tag{11}$$

where $c$ is an absolute constant. Moreover, if $G(\bullet;\gamma)$ either has an absolutely continuous component or is of the lattice type, then the local bound

$$\mid F_n(x) - F(x) \mid \leq \frac{c\beta_3(1)}{n^{1/2}[\sigma^2(1)]^{3/2}} \int_0^\infty (1 + x/\gamma)^{-1} d\mu(\gamma) \tag{12}$$

holds with $c$ an absolute constant whenever the integral is defined.

PROOF. The first conclusion follows on applying the bounds of Berry (1941) and Esseén (1945) conditionally and noting that the moment ratio is scale-invariant. The second follows on applying a result of Bikjalis (1966) conditionally.

We next turn to Berry-Esseén bounds on $\Re^k$. We consider only conditionally *iid* sequences on $(\Re^k)^n$, noting that more general results follow without difficulty along similar lines. Let $\mathcal{F}$ be the class of all measurable convex subsets of $\Re^k$.

THEOREM 4.5. *Let $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ be conditionally iid on $(\Re^k)^n$ having conditional moments $E(\mathbf{X} \mid \gamma) = \theta$, $D(\mathbf{X} \mid \gamma) = \Sigma(\gamma)$, and $\{\beta_{3j}(\gamma) = E(\mid X_j - \theta_j \mid^3 \mid \gamma); 1 \leq j \leq k\}$, with $X_j$ as the jth component of $\mathbf{X}$. Then global bounds on the rate of convergence of $P_n$ to $P$ are given by*

$$\sup_{A\in\mathcal{F}} \mid P_n(A) - P(A) \mid \leq \frac{ck^3}{n^{1/2}} \int_\Gamma \sum_{j=1}^k [\xi_{jj}(\gamma)]^{3/2} \beta_{3j}(\gamma) d\mu(\gamma) \tag{13}$$

whenever the integral is defined, where $\Xi(\gamma) = [\xi_{ij}(\gamma)] = [\Sigma(\gamma)]^{-1}$ and $c$ is an absolute constant not depending on $k$.

PROOF. The proof consists of modifying a result of Bergström (1969) in a form due to Jensen and Mayer (1975), and applying the result conditionally.

COROLLARY 4.5.1. *Let $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ be dependent by coordinate scaling with moments as in Theorem 4.5. Then*

$$\sup_{A\in\mathcal{F}} \mid P_n(A) - P(A) \mid \leq \frac{ck^3}{n^{1/2}} \sum_{j=1}^k [\xi_{jj}(1)]^{3/2} \beta_{3j}(1). \tag{14}$$

PROOF. The conclusion follows on noting that the expression under the integral on the right of (13) is invariant under scaling the coordinates on $\Re^k$.

COROLLARY 4.5.2. *Let $\{\mathbf{X}_1,\ldots,\mathbf{X}_n\}$ be dependent of order $r$ for some $r > 3$ but not for $r \leq 3$, with moments otherwise as in Theorem 4.5. Then*

$$\sup_{A\in\mathcal{F}} \mid P_n(A) - P(A) \mid \leq \frac{ck^3}{n^{1/2}} \sum_{j=1}^{k} (\xi_{jj})^{3/2}\beta_{3j}. \tag{15}$$

PROOF. This follows because the expression under the integral on the right of (13) does not depend on $\gamma$.

There is a rich literature on central limit theory and bounds of the Berry-Esseén type on $\Re^1$ and $\Re^k$. Many known results carry over to conditionally independent sequences along the lines illustrated here. For example, unconditional Edgeworth series expansions of order $s$ on $\Re^k$ will emerge as mixtures of usual Edgeworth series on $\Re^k$ as given in Chambers (1967). Moreover, on applying results of von Bahr (1967) conditionally, bounds on the errors of these Edgeworth mixtures can be obtained in a manner similar to the unconditional versions of Berry-Esseén bounds given here. If the expression under the integral on the right of (13) is uniformly bounded for all $\gamma \in \Gamma$, then that bound can be used as a nonintegral version of (13). Note, however, that the integral version depends on the particular mixture and thus gives a tighter bound.

In another direction, suppose that second moments are not defined conditionally, but that the conditional distributions are in the domain of attraction of a stable limit on $\Re^1$ or $\Re^k$ with index $\alpha$. Then limit distributions of standardized sums are conditionally stable with index $\alpha$, and unconditionally are mixtures of these. Again the class of limit laws is larger than for the independent case. In particular, the *chf*'s for these limits can be studied as mixtures of Lévy representations for stable *chf*'s on $\Re^1$ or $\Re^k$, as appropriate.

**5. Topics In Inference.** Many statistical procedures are based on the assumption of independence. A number of these remain valid despite dependencies of certain types. Three examples are given here using variations of the basic model of Section 3.2. In these cases the argument is the same: Conditional distributions of the statistics in question are seen to be free of the conditioning variables and thus are identical to their unconditional forms. In all such cases the classical assumption of independence may be replaced by the much weaker assumption of conditional independence.

EXAMPLE 5.1. *Sensitivities of Experiments.* The sensitivities of alternative experiments in the normal-theory analysis of variance may be studied as follows. Suppose that $\mathcal{L}(\mathbf{X}_1) = N_n(\theta_1, \sigma_1^2\mathbf{I}_n)$ and $\mathcal{L}(\mathbf{X}_2) = N_n(\theta_2, \sigma_2^2\mathbf{I}_n)$ are models for two independent experiments pertaining to a parameter $\theta$, and that a linear hypothesis $H : \mathbf{A}\theta = \mathbf{0}$ is to be tested in each experiment. Let $F_1$ and $F_2$ be the corresponding variance ratios having noncentrality parameters $\lambda_1$ and $\lambda_2$. Bradley and Schumann (1957b) studied the ratio $R = F_1/F_2$ both as a gauge of the relative sensitivities of the two experiments, and as a statistic for testing $H : \lambda_1 = \lambda_2$. The distribution of $R$ and various applications are treated in Bradley and Schumann (1957a,b), Schumann and Bradley (1958), Schumann and Bradley (1959); see also Shue and Bain (1982), Subrahmaniam (1979), and Zerbe and Goldgar (1980) for related work.

Now let $\mathbf{X}_1$ and $\mathbf{X}_2$ be conditionally independent as in Section 3.2 such that $\mathcal{L}(\gamma_1\mathbf{X}_1 \mid \gamma_1) = N_n(\theta_1, \sigma_1^2\mathbf{I}_n)$ and $\mathcal{L}(\gamma_2\mathbf{X}_2 \mid \gamma_2) = N_n(\theta_2, \sigma_2^2\mathbf{I}_n)$, where $(\gamma_1, \gamma_2)$ have some joint distribution on $\Re_+^2$. Because $F_1, F_2$ and $R$ are scale-invariant, their conditional and unconditional distributions are identical, so that all the standard properties continue to hold exactly under conditional independence of the type indicated.

EXAMPLE 5.2. *Friedman's Test.* Friedman's (1937) test is used widely in nonparametrics to compare the effectiveness of $k$ treatments using $n$ experimental subjects. Let $\{Y_{ij}; 1 \leq j \leq k,\ 1 \leq i \leq n\}$ be outcomes of an experiment such that $\mathbf{Y}_i' = [Y_{i1}, \ldots Y_{ik}]$ has an exchangeable distribution on $\Re^k$ for each $i = 1, \ldots, n$. If $R_{ij}$ denotes the rank of $Y_{ij}$ among $\{Y_{i1}, \ldots Y_{ik}\}$ and if $\{R_j = R_{1j} + \cdots + R_{nj}; 1 \leq j \leq k\}$, then Friedman's statistic is

$$X_r^2 = \frac{12}{nk(k+1)} \sum_{j=1}^{k} [R_j - n(k+1)/2]^2. \tag{16}$$

The standard assumption is that $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_n\}$ are mutually independent, in which case the exact small-sample null distribution is based on the $(k!)^n$ possible permutations, and the asymptotic distribution is $\chi^2(k-1)$.

Specializing the model of Section 3.2, let $\{\phi_i(t_1, \ldots, t_k); 1 \leq i \leq n\}$ be *chf*s of exchangeable distributions on $\Re^k$, and let $\Gamma_0$ be the class of all monotonic increasing functions $\gamma : \Re^1 \to \Re^1$. For a typical *chf* $\phi(t_1, \ldots, t_k)$ of $\mathbf{Y}' = [Y_1, \ldots Y_k]$, denote by $\phi(t_1, \ldots, t_k; \gamma)$ the joint *chf* of $[\gamma(Y_1), \ldots, \gamma(Y_k)]$. Now choose $[\gamma_1, \ldots, \gamma_n]$ from $\Gamma = \Gamma_0^n$ according to some probability measure $\mu(\bullet)$, and consider the joint distribution of $\{\gamma_i(Y_{ij}); 1 \leq j \leq k, 1 \leq i \leq n\}$. This has the form (3). It is well known that $\{[\gamma_i(Y_{i1}), \ldots \gamma_i(Y_{ik})]; 1 \leq i \leq n\}$ are again exchangeable vectors on $\Re^k$, but now they are dependent. Nonetheless, the conditional null distribution of $X_r^2$ does not depend on $[\gamma_1, \ldots, \gamma_n]$, and thus its exact small-sample and asymptotic distributions are precisely those occurring when responses from subject to subject are independent. For example, the $k$ responses within each subject may be randomly scaled, with a different scaling for different subjects.

EXAMPLE 5.3. *Classification Rules.* Given samples from two Gaussian populations, $N_k(\theta_1, \Sigma)$ and $N_k(\theta_2, \Sigma)$, and a random observation $\mathbf{X}$ from $N_k(\theta, \Sigma)$ having unknown origins, the problem of classification is to assign $\mathbf{X}$ to one of the two populations. In particular, suppose $\mathcal{L}(\mathbf{X}_1) = N_{n_1 \times k}(\mathbf{1}_{n_1} \times \theta_1', \mathbf{I}_{n_1} \times \Sigma)$, and let $(\bar{\mathbf{X}}_1, \mathbf{S}_1)$ be the corresponding sample mean vector and sample dispersion matrix.

Similarly, consider $\mathcal{L}(\mathbf{X}_2) = N_{n_2 \times k}(\mathbf{1}_{n_2} \times \theta_2', \mathbf{I}_{n_2} \times \Sigma)$ and $(\bar{\mathbf{X}}_2, \mathbf{S}_2)$ . The standard procedure uses the classification statistic

$$V = [\mathbf{X} - (1/2)(\bar{\mathbf{X}}_1 + \bar{\mathbf{X}}_2)]'\mathbf{S}^{-1}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2) \tag{17}$$

where $\mathbf{S}$ is the pooled sample estimator for $\Sigma$; see Anderson (1984), p. 210. Normal-theory properties of the usual classification rule using $V$ are based on the mutual independence of $\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}\}$.

However, we now suppose that $\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}\}$ are conditionally independent, given a nonsingular random matrix $\gamma \in F_{k \times k}$, such that $\mathcal{L}(\mathbf{X}_1 \mid \gamma) = N_{n_1 \times k}(\mathbf{1}_{n_1} \times \theta_1'\gamma, \mathbf{I}_{n_1} \times \gamma'\Sigma\gamma), \mathcal{L}(\mathbf{X}_2 \mid \gamma) = N_{n_2 \times k}(\mathbf{1}_{n_2} \times \theta_2'\gamma, \mathbf{I}_{n_2} \times \gamma'\Sigma\gamma)$, and $\mathcal{L}(\mathbf{X} \mid \gamma) = N_k(\gamma'\theta, \gamma'\Sigma\gamma)$. This is seen to be a special case of model (3) where $r = 3$. Since the statistic $V$ is invariant under nonsingular linear transformations, its conditional distribution $\mathcal{L}(V \mid \gamma)$ is independent of $\gamma \in \Gamma$. It follows that all standard properties of the usual classification rules carry over to scale mixtures of the types indicated.

## REFERENCES


ANDERSON, T.W. (1984). *An Introduction to Multivariate Statistical Analysis, 2nd ed.* John Wiley and Sons, New York.

BAHR, B. VON (1967). Multi-dimensional integral limit theorems. *Arkiv för Matematik* **7** 71–88.

BERGSTRÖM, H. (1969). On the central limit theorem in $\Re^k$. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **14** 113–126.

BERRY, A.C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* **49** 122–136.

BIKJALIS, A. (1966). Estimates for the remainder term in the central limit theorem. (In Russian.) *Lit. Matem. Sb.* **6** 321–346.

BRADLEY, R.A. and SCHUMANN, D.E.W. (1957a). The comparison of the sensitivities of similar experiments: Applications. *Biometrics* **13** 496–510.

BRADLEY, R.A. and SCHUMANN, D.E.W. (1957b). The comparison of the sensitivities of similar experiments: Theory. *Ann. Math. Statist.* **28** 902–920.

CHAMBERS, J.M. (1967). On methods of asymptotic approximations for multivariate distributions. *Biometrika* **54** 367–383.

DAWID, A.P. (1979). Conditional independence in statistical theory. *J.R. Statist. Soc.* B **41** 1–31.

DYKSTRA, R.L., HEWETT, J.E. and THOMPSON, W.A. (1973). Events which are almost independent. *Ann. Statist.* **1** 674–681.

ESSEÉN, C.G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta Math.* **77** 1–125.

FINETTI, B. DE (1937). La prevision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré* **7** 1–68.

FRIEDMAN, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Statist. Assoc.* **32** 675–701.

JENSEN, D.R. and MAYER, L.S. (1975). Normal-theory approximations to tests for linear hypotheses. *Ann. Statist.* **3** 429–444.

SCHUMANN, D.E.W. and BRADLEY, R.A. (1958). The comparison of the sensitivities of similar experiments. *Virginia J. Sci.* **9** 303–314.

SCHUMANN, D.E.W. and BRADLEY, R.A. (1959). The comparison of the sensitivities of similar experiments: Model II of the analysis of variance. *Biometrics* **15** 405–416.

SHAKED, M. (1977). A concept of positive dependence for exchangeable random variables. *Ann. Statist.* **5** 505–515.

SHUE, W.K. and BAIN, L.J. (1982). Approximate percentiles for ratios of $F$-variates. *Comm. Statist. A* **11** 1831–1838.

STUDNEV, Y.P. and IGNAT, Y.I. (1967). Refinement of the central limit theorem and of its global version. *Theory Prob. Applic.* **12** 508–512.

SUBRAHMANIAM, K. (1979). Tests of significance for the odds ratio in a $2 \times 2$ table based on the ratio of two independent $F$ variates. *Comm. Statist. B* **8** 245–256.

TAYLOR, R.L., DAFFER, P.Z. and PATTERSON, R.F. (1985). *Limit Theorems for Sums of Exchangeable Random Variables.* Rowman and Allanheld, Totowa, NJ.

TONG, Y.L. (1980). *Probability Inequalities in Multivariate Distributions.* Academic Press, New York.

ZERBE, G.O. and GOLDGAR, D.E. (1980). Comparison of intraclass correlation coefficients with the ratio of two independent $F$ statistics. *Comm. Statist. A* **9** 1641–1645.

DEPARTMENT OF STATISTICS
VIRGINIA POLYTECHNIC INSTITUTE
BLACKSBURG, VA 24061