stochastic processes and functional analysis

in celebration of M.M. Rao's 65th birthday

edited by

Jerome A. Goldstein
University of Memphis
Memphis, Tennessee

Neil E. Gretsky
University of California-Riverside
Riverside, California

J. J. Uhl, Jr.
University of Illinois

Marcel Dekker, Inc. **New York•Basel•Hong Kong**

Library of Congress Cataloging-in-Publication Data

Stochastic processes and functional analysis : in celebration of M. M. Rao's 65th birthday / edited by Jerome A. Goldstein, Neil E. Gretsky, John Jerry Uhl, Jr.
p. cm. — (Lecture notes in pure and applied mathematics : v. 186)
Held at the Univ. of Calif. —Riverside, Nov. 18-20, 1994.
"Published writings of M. M. Rao": p.
ISBN 0-8247-9801-5 (pbk. : alk. paper)
1. Stochastic processes—Congresses. 2. Functional analysis--Congresses. I. Rao, M. M. (Malempati Madhusudana). II. Goldstein, Jerome A. III. Gretsky, Neil E. IV. Uhl, J. J. (J. Jerry) V. Series.
QA274.A1S7665 1997
515'.7—dc21

96-48137
CIP

The publisher offers discounts on this book when ordered in bulk quantities. For more information, write to Special Sales/Professional Marketing at the address below.

This book is printed on acid-free paper.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Current printing (last digit):
10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

Preface

A conference in Modern Analysis and Probability on the occasion of his 65th birthday was held in honor of M. M. Rao at the University of California, Riverside. Over a hundred friends, colleagues, students, and other mathematicians attended during the three-day meeting. The Mathematics Department provided copious amounts of coffee and doughnuts. An on-campus dinner was held on a Friday night and an off-campus dinner was held Saturday night following a gala reception at the home of M. M. and his wife, Durgamba. Support for the conference was supplied by the College of Natural and Agricultural Sciences at the University of California–Riverside, the Mathematics Department at the University of California–Riverside, and the National Science Foundation.

This festschrift volume contains most of the talks given at the conference as well as several that were contributed later. The beginning portions of the book include a biography of M. M. Rao, a bibliography of his published writings, an ancestral mathematical family tree, and a list of Ph.D. theses written under Rao and his students.

The talks at the conference included four keynote addresses by Rao, Jean Bourgain, S. R. S. Varadhan, and Michael Crandall. All but Crandall's talk are contained here; Crandall's talk will appear in a paper that will be published elsewhere. Rao's paper is an account of that portion of his work which originated in problems arising in applications. It is organized by area and features the work of his students as much as his own. The breadth and depth of Rao's mathematical work and its impact on analysis, probability, and stochastic processes can be seen not only by what is included in this paper but also by the portion of his bibliography which is *not* in this paper. The editors enjoyed this paper immensely. Even as good as the paper is, it does not capture the charm and the emotion with which the talk was given.

Jean Bourgain's paper is a long, densely written survey (an "exposé" in his terminology) of persistency of quasi-periodic solutions of linear or integrable partial differential equations after Hamiltonian perturbation. Much of the original work is due to Bourgain and is not in print elsewhere. The talk given by Varadhan reported on joint work with H. T. Yau concerning scaling limits for lattice gas models. This provides a way to give a simplified description of the state of a large system of interacting particles which is evolving in time. The results typify recent deep research involving hydrodynamic limits, which establish that nonlinear partial differential equations govern many large particle systems in the limit.

The remaining eighteen papers are original contributions in probability and statistics, stochastic processes, Banach space theory, measure theory, and differential equations—both deterministic and stochastic.

Many other people attended the conference who did not give talks for one reason or another. Although we cannot list all of them (our sincere apologies) we would like to mention two esteemed intellectual colleagues of M. M. Rao, Mannie Parzen and Howard Tucker, as well as two former students, William Kraynek and Marc Mehlman.

Jerome A. Goldstein
Neil E. Gretsky
J. J. Uhl, Jr.

Multivariate Distributions with Gaussian Conditional Structure

BARRY C. ARNOLD Department of Statistics, University of California, Riverside, California

JACEK WESOLOWSKI Mathematical Institute, Warsaw University of Technology, Warsaw, Poland

Key words: *quasi-Gaussian distributions, classical normal distribution, normal conditionals distribution, elliptical contours, linear regression, mixtures, Kagan class.*

ABSTRACT

Multivariate distributions exhibiting some features of the conditional structure associated with the classical normal model are investigated. Features considered include conditional distributions of subvectors and conditional moments. Our understanding of the classical normal model is enhanced by the study of such quasi-Gaussian distributions together with investigation of additional assumptions required to characterize the classical normal model. Special attention is paid to the class of distributions exhibiting Gaussian conditional structure of the second order, i.e. those in which the conditional moments of orders one and two match the Gaussian model.

1 THE CLASSICAL MULTIVARIATE NORMAL DISTRIBUTION

A random vector $\underline{X} = (X_1, X_2, \ldots, X_k)$ is said to have a classical multivariate normal distribution if it admits a representation of the form

$$\underline{X} = \underline{\mu} + \Sigma^{1/2}\underline{Z}$$

where $Z_1, Z_2, \ldots, Z_k$ are i.i.d. standard univariate normal random variables. In such a case we write $\underline{X} \sim N^{(k)}(\underline{\mu}, \Sigma)$. Here $\underline{\mu}\epsilon\mathbf{R}^k$ and Σ is a non-negative definite $k \times k$ matrix. Such random variables have remarkable properties. For example:

1. All one dimensional marginals are normal.
2. All ℓ dimensional marginals, $\ell < k$, are ℓ-variate normal.
3. All linear combinations are normal. In fact for any $\ell \times k$ matrix B we have
$$B\underline{X} \sim N^{(\ell)}(B\underline{\mu}, B\Sigma B')\ .$$
4. All conditionals are normal. Thus if we partition $\underline{X} = (\dot{\underline{X}}, \ddot{\underline{X}})$ then the conditional distribution of $\dot{\underline{X}}$ given $\ddot{\underline{X}} = \ddot{\underline{x}}$ is multivariate normal.
5. All regressions are linear. Thus for any i and any $j_1, j_2, \ldots, j_\ell (\neq i)$ $E(X_i|X_{j_1}, \ldots, X_{j_\ell})$ is a linear function of $X_{j_1}, X_{j_2}, \ldots, X_{j_\ell}$.
6. All conditional variances are constant. Thus $var(X_i|X_{j_1}, \ldots, X_{j_\ell})$ is nonrandom for any i, and any $j_1, j_2, \ldots, j_\ell (\neq i)$.
7. If Σ is positive definite, the joint density of $\underline{X}$ is elliptically contoured.
8. $\underline{X}$ has linear structure, i.e. $\underline{X}$ admits a representation of the form
$$\underline{X} = \underline{a}_0 + A\underline{Z}$$
where the Z_i's are independent random variables.

Most of these properties, taken individually, fail to characterize the classical multivariate normal distribution. Combinations of these properties can be used to characterize the classical model . Condition 3 does characterize the classical model. Condition 4 also will characterize the classical model provided $k > 2$. None of the others alone will do it. Conditions 7 and 8, together, will characterize the classical distribution.

The present paper will focus mainly on two issues: the possibility of weakening the assumption of property 4 and still preserving a k-variate normal characterization (Section 2), and a discussion of models which, though not classical normal, mimic the conditional moment structure of the classical models (Section 3 and 4). Additional conditions for such structures leading to multivariate normality are outlined in Section 5.

Some useful notational conventions follow. Suppose $\underline{X}$ denotes a k-dimensional random vector and $\underline{x}\epsilon\mathbf{R}^k$. A partition of $\underline{X}$ into two subvectors of dimension $\dot{k}$ and $\ddot{k}$ with $\dot{k}+\ddot{k}=k$ will be denoted by $(\dot{\underline{X}},\ddot{\underline{X}})$. The corresponding partition of $\underline{x}$ will be denoted $(\dot{\underline{x}},\ddot{\underline{x}})$. X_i will denote the ith coordinate of $\underline{X}$. $\underline{X}_{(i)}$ is the $k-1$ dimensional vector obtained from $\underline{X}$ by deleting X_i. $\underline{X}_{(i,j)}$ is obtained from $\underline{X}$ by deleting X_i and X_j. Analogously real vectors $\underline{x}_{(i)}$ and $\underline{x}_{(i,j)}$ are defined.

2 CONDITIONAL CHARACTERIZATIONS OF THE CLASSICAL NORMAL MODEL

Suppose that for each i and for each $\underline{x}_{(i)}\epsilon\mathbf{R}^{k-1}$ the conditional distribution of X_i given $\underline{X}_{(i)}=\underline{x}_{(i)}$ is normal with a mean and variance that may depend on $\underline{x}_{(i)}$, i.e.

$$X_i|\underline{X}_{(i)}=\underline{x}_{(i)}\sim N(\mu_i(\underline{x}_{(i)}),\sigma_i^2(\underline{x}_{(i)}))\ . \tag{1}$$

In this case, generalizing the early results of Bhattacharya (1943) and solving an appropriate set of functional equations, one may verify that $\underline{X}$ must have what we may call a k-variate normal conditionals distribution with density of the form:

$$f_{\underline{X}}(\underline{x})=\exp[-\frac{1}{2}G(\underline{x})] \tag{2}$$

where

$$G(\underline{x})=\sum_{i_1=0}^{2}\sum_{i_2=0}^{2}\cdots\sum_{i_k=0}^{2}\gamma_{i_1i_2,\ldots,i_k}(\prod_{j=1}^{k}x_j^{i_j})\ . \tag{3}$$

There are necessary restrictions on the ranges of the γ's in (2.3) in order to ensure integrability and to ensure that all expressions for conditional variances are uniformly positive. Of course $\gamma_{00\ldots0}$ is not really a parameter, it is a normalizing factor that is a function of the remaining γ's chosen to ensure that the integral of the joint density is 1. If $\underline{X}$, of dimension k, has a density of the form (2.2) we will write

$$\underline{X}\sim NC^{(k)}(\underline{\gamma})\ .$$

See Arnold, Castillo and Sarabia (1992) for a more detailed introduction to the normal conditionals model.

The classical k-variate normal distribution is of course a special case of the normal conditionals model (2.2), since obviously it satisfies the required condition (2.1). It can be recognized by the fact that for such a distribution all coefficients $\gamma_{\underline{i}}$ for which $\sum_{j=1}^{k}i_j>2$ must be zero since, in order for (2.2) to represent a classical normal model, $G(\underline{x})$ must be a quadratic form.

Many characterization programs may be viewed as beginning with conditional normal requirements leading to the model (2.2), or some related submodel, and then imposing additional conditions to ensure vanishing of the "unwanted" coefficients (i.e. $\gamma_{\underline{i}}$'s with $\sum_{j=1}^{k}i_j>2$).

To begin with, we may recall that the classical normal distribution actually has far more conditional normal distributions associated with it than those alluded to in (2.1). In fact, if $\underline{X} \sim N^{(k)}(\underline{\mu}, \Sigma)$, then for any partition of $\underline{X}$ into subvectors $\dot{\underline{X}}$ and $\ddot{\underline{X}}$ of dimensions $\dot{k}$ and $\ddot{k}$ with $k = \dot{k} + \ddot{k}$ we have

$$\dot{\underline{X}}|\ddot{\underline{X}} = \ddot{\underline{x}} \sim N^{(\dot{k})}(\dot{\underline{\mu}}(\ddot{\underline{x}}), \dot{\Sigma}(\ddot{\underline{x}})) \ . \tag{4}$$

Since all subvectors of $\underline{X}$ are again classical normal, even more conditional distributions, analogous to those in (2.4) but now based on partitioning subvectors of $\underline{X}$, are again guaranteed to be normal.

Assumption (2.1) is not enough to guarantee the classical model. Assumption (2.4) is more than enough (provided $k > 2$, otherwise (2.1) and (2.4) coincide and fail to characterize the classical normal model). In fact one may prove (see Arnold, Castillo and Sarabia (1994)) that, for $k > 2$, a sufficient condition to guarantee a classical multivariate normal model is an assumption that for each i, j and each $\underline{x}_{(i,j)} \epsilon \mathbf{R}^{k-2}$

$$(X_i, X_j)|\underline{X}_{(i,j)} \sim N^{(2)}(\underline{\mu}_{(ij)}(\underline{x}_{(i,j)}), \Sigma_{ij}(\underline{x}_{(i,j)})) \ . \tag{5}$$

The key observation is that (2.5) implies that for each i, $X_i|\underline{X}_{(i,j)} = \underline{x}_{(i,j)}$ is normal, since the classical bivariate normal has normal marginals. Consequently (2.5) is enough to guarantee that

$$\underline{X} \sim NC^{(k)}(\underline{\gamma}) \tag{6}$$

and for each i,

$$\underline{X}_{(i)} \sim NC^{(k-1)}(\underline{\gamma}^{(i)}) \ . \tag{7}$$

However marginals of a normal conditionals distribution (2.2) can only be of the normal conditionals form if certain $\underline{\gamma}$'s are zero. In fact (2.7) guarantees that all the "unwanted" $\underline{\gamma}$'s are zero, and the fact that $\underline{X}$ must have a classical normal distribution is a consequence.

Of course the conditional mean functions and conditional variance functions which are encountered in the normal conditionals model (2.2) are not the familiar linear regressions and constant conditional variances associated with the classical model. If we are willing to assume, in addition to the assumption that each X_i given $\underline{X}_{(i)}$ is normal, that the conditional variances are constant, i.e. that

$$X_i|\underline{X}_{(i)} = \underline{x}_{(i)} \sim N(\mu_i(\underline{x}_{(i)}, \sigma_i^2) \ , \tag{8}$$

then the unwanted $\underline{\gamma}$'s in (2.2) are forced to be zero and we must have $\underline{X}, \sim N^{(k)}(\underline{\mu}, \Sigma)$, i.e. classical normal. An analogous alternative sufficient prescription is the requirement that in (2.1) each $\mu_i(\underline{x}_{(i)})$ be a linear function of $\underline{x}_{(i)}$.

It is indeed well known that for a classical k-variate normal random vector $\underline{X}$ we can explicitly write the parameters in the conditional distribution of $\dot{\underline{X}}$ given $\ddot{\underline{X}}$ in terms of the original parameters of the distribution of $\underline{X}$. Thus with $\underline{X} = (\dot{\underline{X}}, \ddot{\underline{X}})$ and $\underline{\mu} = (\dot{\underline{\mu}}, \ddot{\underline{\mu}})$ we have

$$\dot{\underline{X}}|\ddot{\underline{X}} = \ddot{\underline{x}} \sim N^{(\dot{k})}(\dot{\underline{\mu}} + \Sigma_{12}\Sigma_{22}^{-1}(\ddot{\underline{x}} - \ddot{\underline{\mu}}), \Sigma_{11.2}) \tag{9}$$

where

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

and

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \ .$$

The linear regressions and constant conditional variances are explicitly displayed in (2.9). Linear regressions are not that unusual in multivariate distributions. Constant conditional variances are unexpected. In some ways they are even counterintuitive. Taken together, the requirements of linear regressions and constant conditional variances seem potentially so restrictive as to possibly, alone, suffice to characterize the classical normal model. They don't. But verifying that they don't and asking what additional requirements will lead to characterizations is an interesting exercise that enriches our understanding of the real nature of the curious classical multivariate normal model. The topic will be addressed in the next section.

Before leaving the study of conditional normality assumptions to focus on conditional moment assumptions, it is worth returning to the list of 8 properties of the classical model listed in section 1. Which of these in addition to (2.1) (i.e. X_i given $\underline{X}_{(i)} = \underline{x}_{(i)}$ is normal $\forall i, \forall \underline{x}_{(i)}$) will guarantee classical multivariate normality. We have already considered properties 5 and 6. Property 1 has potential, since marginals of normal conditionals models are typically not of the normal conditionals form and a fortiori not (classical) normal. In fact, if all one dimensional marginals of $\underline{X}$ are normal and (2.1) holds then the unwanted $\underline{\gamma}$'s in (2.3) must disappear and the classical normal model is obtained. Turning to condition 2, it is a condition that subsumes 1 and consequently can be used to characterize the classical model. Actually far less is needed. For example, if in addition to (2.1), each $\underline{X}_{(i)}$ is classical $(k-1)$-variate normal, then $\underline{X}$ must be classical k-variate normal. Indeed any marginal normality statement sufficient to guarantee one dimensional normal marginals will obviously suffice. Turn next to condition 3. If k linearly independent linear combinations of the coordinates of $\underline{X}$ are normally distributed and if (2.1) holds then, by a suitable linear transformation, we have $\underline{Y} = B\underline{X}$ with normal conditionals (i.e. (2.1)) and normal one dimensional marginals. Then $\underline{Y}$ and consequently also $\underline{X}$ is a classical normal random vector. Next turn to condition 7, elliptical contours. This is easily dealt with. The contours of the normal conditionals density are determined by $G(\underline{x})$ in (2.3). Their form will be elliptical only if the unwanted $\underline{\gamma}$'s are all zero; i.e. only in the classical normal case.

Finally consider condition 8. The assumption of linear structure turns out to be particularly fruitful in conjunction with certain conditional moment assumptions, as we shall see in the next section. In conjunction with the normal conditionals assumption, i.e. (2.1), the role of linear structure is less evident. Assumption 8 does however imply the existence of a linear transformation of $\underline{X}$ (with density (2.2)) that has a density which can be factored. This does imply that the unwanted $\underline{\gamma}$'s in (2.3) must be zero and does indeed guarantee classical multivariate normality.

3 GAUSSIAN CONDITIONAL STRUCTURE

What can we say about collections of random variables which exhibit linear regression functions and constant conditional variances? We will, following Wesolowski (1991), call this property Gaussian conditional structure of the second order.

Formally we will say that a random element (or indexed collection of random variables) $X = \{X_\alpha : \alpha \epsilon A\}$ exhibits Gaussian conditional structure of the second order and write $X \epsilon GCS_2(A)$ if for any $n = 2, 3, \ldots$ and any $\alpha_1, \alpha_2, \ldots, \alpha_n \epsilon A$,

(i) $E(X_{\alpha_1}|X_{\alpha_2}, \ldots, X_{\alpha_n})$ is a linear function of $X_{\alpha_2}, X_{\alpha_3}, \ldots, X_{\alpha_n}$ and

(ii) $var(X_{\alpha_1}|X_{\alpha_2}, \ldots, X_{\alpha_n})$ is non-random.

At times it is convenient to use the term Gaussian conditional structure of the second order to refer to the distributions or probability measures associated with the random element rather than with the random element per se; we will do this at times without explanation and without fear of confusion. To avoid trivial examples we will usually implicitly assume that the X_α's are linearly independent and not uncorrelated. Collections of independent random variables could otherwise provide uninteresting examples of Gaussian conditional structure of the second order.

Observe that A could correspond to the natural numbers or the reals or positive reals. Consequently time series will be subsumed in the class of random elements under consideration. Spatial processes can be viewed as being random elements associated with a set A that is a subset of $\mathbf{R}^k$. Any normal process or, more generally, any Gaussian random element, will obviously exhibit Gaussian conditional structure of the second order. Our main focus will be however on random vectors of dimension k; i.e. on random elements where $A = \{1, 2, \ldots, k\}$. If $X = (X_1, \ldots, X_k)$ exhibits Gaussian conditional structure of the second order we will write $X \epsilon GCS_2(k)$.

A remark is in order about the subscript 2 that appears in our definition of Gaussian conditional structure of the second order. One could obviously ask that the random element mimic the conditional moment structure of a Gaussian element with regard to more than the first two conditional moments. One could ask for the first j conditional moments to behave as they do for Gaussian elements. The class of random elements exhibiting such behavior would be denoted by $GCS_j(A)$ instead of $GCS_2(A)$. Our focus will be on $GCS_2(A)$. Only once will we briefly mention how we might construct non-Gaussian members of $GCS_j(A)$, for $j > 2$.

If $\underline{X} \epsilon GCS_2(k)$, it is natural to ask whether $\underline{X}$ must necessarily be Gaussian. The question is already meaningful and reasonably challenging when $k = 2$; i.e. in the case of bivariate distributions. Kagan, Linnik and Rao (1973) provide the following lemma indicating the nature of characteristic functions associated with $GCS_2(2)$ distributions.

Lemma 3.1: In order that the two-dimensional random vector (X, Y) satisfy (i) $E(Y|X) =$

$\alpha + \beta X$ and (ii) $var(Y|X) = \sigma^2$ (a constant), it is necessary and sufficient that the characteristic function of (X, Y) satisfies

$$\frac{\partial}{\partial t_2}\phi(t_1, t_2)|_{t_2=0} = i\alpha\phi(t_1, 0) + \beta\frac{d}{dt_1}\phi(t_1, 0) \tag{1}$$

and

$$\frac{\partial^2}{\partial t_2^2}\phi(t_1, t_2)|_{t_2=0} \equiv -(\sigma^2 + \alpha^2)\phi(t_1, 0) + 2i\alpha\beta\frac{d}{dt_1}\phi(t_1, 0) + \beta^2\frac{d^2}{dt_1^2}\phi(t_1, 0)\ . \tag{2}$$

If one, as do Kagan, Linnik, and Rao, then assumes that (X, Y) has linear structure (i.e., satisfies condition 8, of section 1), then we may verify that indeed (X, Y) must have a classical bivariate normal structure.

Examples of non-Gaussian characteristic functions satisfying the conditions of Lemma 3.1 are not that easy to visualize. It is in fact probably an inappropriate approach to the problem of verifying that there do exist random vectors with GCS_2 that are not classical normal random vectors. It is probably more fruitful to seek non-Gaussian density functions that will exhibit the required conditional properties (and a fortiori will have characteristic functions satisfying the conditions in the Lemma). The first example of this genre was provided by Kwapien sometime prior to 1985. It was first reported in Bryc and Plucinska (1985). It was in fact presented in terms of the joint characteristic function. He considers a random vector (X, Y) whose joint characteristic function is given by

$$\phi_{X,Y}(s, t) = p\ \cos(s + t) + (1 - p)cos(s - t)\ . \tag{3}$$

where $p \in (0, 1)$ and, to avoid independence, $p \neq 1/2$. It is obvious that (3.3) does not correspond to a Gaussian random vector and it is not hard to verify that conditions (3.1) and (3.2) hold, as do the parallel conditions corresponding to interchanging the roles of X and Y. Consequently (X, Y) with characteristic function (3.3) does exhibit Gaussian conditional structure of order 2, i.e. $(X, Y) \in GCS_2(2)$.

Where did (3.3) come from? And, why does it work? The picture is clearer if we look at the following joint discrete density of a random vector (X, Y)

$$f_{X,Y}(x, y) : \quad \begin{array}{c|cc|} & x & \\ y & -1 & 1 \\ -1 & \frac{1-p}{2} & \frac{p}{2} \\ 1 & \frac{p}{2} & \frac{1-p}{2} \end{array} \tag{4}$$

where $p \in (0, 1)$ and, to avoid independence, $p \neq 1/2$. It is readily verified that this is indeed Kwapien's example (the corresponding characteristic function is given by (3.3)). But the joint distribution (3.4) has marginals with only two possible values. This gives linear regression functions by default (any function with a two point domain is linear!). Constant conditional variances are a consequence of the fact that $p(1 - p) = (1 - p)p$.

The elegant simplicity of the Kwapien example would suggest ready extension to higher dimensions. However, only recently (Nguyen, Rempala and Wesolowski (1994)), have any other (other than relabeled versions of the Kwapien example) non-Gaussian examples been described in either two or more dimensions. Indeed there were some disturbing indications that Gaussian conditional structure of the second order might be more restrictive than one would initially imagine. Bryc and Plucinska (1985) showed that if we consider a random element where $A = \{1, 2, \ldots\}$ then under mild regularity conditions, Gaussian conditional structure of order 2 i.e. $GCS_2(1, 2, \ldots)$ is sufficient to guarantee that X must be a normal process. Earlier, in a series of papers (Plucinska (1983), Wesolowski (1984) and Bryc (1985)), an analogous result was obtained in case on which $A = \mathbf{R}^+$.

The first step to finding non-Gaussian random vectors of dimension greater than 2 with Gaussian conditional structure of order 2, would focus on 3 dimensional examples and, following Kwapien's lead, would focus on simple discrete examples in which the conditional moment conditions will transform into relatively simple equations in the unknown cell probabilities. Thus for example we might seek a 3-dimensional discrete distribution whose second order conditional structure will match that of a 3 dimensional classical normal distribution of the form

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N^{(3)}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix} \right). \tag{5}$$

For such a distribution we will have $E(X_i) = 0$ and $var(X_i) = 1, i = 1, 2, 3$. The conditional moments will be given by

$$\begin{aligned} E(X_i|X_j, X_k) &= \frac{1}{3}(X_j + X_k) \\ var(X_i|X_j, X_k) &= \frac{2}{3} \\ E(X_i|X_j) &= X_j/2 \end{aligned} \tag{6}$$

and

$$var(X_i|X_j) = \frac{3}{4}$$

for all choices of $i \neq j \neq k$ (since the distribution is clearly symmetric). Following the lead of the Kwapien example we would seek a 3 dimensional random vector (Y_1, Y_2, Y_3) with a discrete distribution with possible values for each Y_i being $-m, m-1, \ldots, 1, 1, \ldots, m$ and with probabilities $p_{ijk} = P(Y_1 = i, Y_2 = j, Y_3 = k)$. The joint distribution should be exchangeable. The marginal means and variances should be 0's and 1's. The conditional means and variances of the Y_i's should agree with those in (3.6). Our task is then to solve for the unknown values of $\{p_{ijk} : -m \leq i \leq j \leq k \leq m\}$ subject to the given constraints. If $m \leq 3$, there are more constraints than variables. For $m = 4$, there are 120 variables (p_{ijk}'s) which must be non-negative and satisfy 91 linear constraints. A promising situation, although a solution is not guaranteed. Unfortunately efforts to solve such a system of equations have not proved successful. The search for a solution continues since a simple discrete example may shed additional light on the nature of the class of distributions with Gaussian conditional structure. However, the problem of constructing a non-Gaussian multivariate distributions (of finite dimension > 2) with Gaussian conditional structure was recently resolved by Nguyen, Rempala and Wesolowski (1994). The solution is ingenious but, retrospectively, obvious. Inspired by their examples, the following simple construction is possible.

Take $f_0(\underline{x})$ to be the joint density of a classical k-variate normal distribution with mean vector $\underline{\mu}$ and variance-covariance matrix Σ. We now construct a k-dimensional density which has the same conditional means and variances as does $f_0(\underline{x})$. Pick two distinct bounded densities g_1 and g_2 each supported on the interval $(-1, 1)$ and each having mean 0 and variance 1. There are of course a plethora of such densities. Now consider the new k-dimensional density defined by

$$f^*(\underline{x}) = f_0(\underline{x}) + c\prod_{i=1}^{k}[g_1(x_i) - g_2(x_i)] \tag{7}$$

where c is chosen small enough to guarantee that the expression in (3.7) is uniformly positive (possible since the g_i's are bounded densities). Obviously $f^*(\underline{x})$ is non-Gaussian but all of its marginals are Gaussian and it is readily verified that all of its first and second conditional moments match those of $f_0(\underline{x})$. The density $f^*(\underline{x})$ thus belongs to $GCS_2(k)$; and in fact $GCS_2(k)$ is not just non-empty but contains an enormous variety of distributions constructed in a fashion analogous to that used to define (3.7). It is indeed possible, by putting additional higher moment conditions on the g_i's (used in the construction of f^*), to find k-variate non-Gaussian distributions whose conditional moments up to the m'th order ($m > 2$) match those of a classical normal k-variate distribution.

4 THE STRUCTURE OF THE CLASS $GCS_2(k)$

From the discussion in Section 3, we are aware that the class $GCS_2(k)$ is quite extensive. Our goal in the present section is to identify characteristic properties of the class and to identify conditions sufficient to guarantee that a member of the class indeed is a classical Gaussian distribution. For notational simplicity, some of the discussion is restricted to the bivariate case (i.e. $k = 2$).

Suppose that $\underline{X}\epsilon GCS_2(k)$. Obviously any vectors of the form $\underline{Y} = (c_1X_1 + b_1, c_2X_2 + b_2, \ldots, c_kX_k + b_k)$ for $c_1, \ldots, c_k > 0$ and $\underline{b}\epsilon\mathbf{R}^k$ will again belong to $GCS_2(k)$. Consequently there is no loss in generality if we focus on standardized members of $GCS_2(k)$. These are random vectors $\underline{Z}\epsilon GCS_2(k)$ with the property that $E(Z_i) = 0$ and $var(Z_i) = 1$; $i = 1, 2, \ldots, k$. Throughout this section we will adopt the convention that if we use the notation $\underline{X}$, we are dealing with a general member of the class $GCS_2(k)$ while, if we use the notation $\underline{Z}$, we are referring to a standardized random vector in $GCS_2(k)$.

Thus we are concerned with random vectors $\underline{X}$ such that, with $Z_i = (X_i - E(X_i))/\sqrt{varX_i}$, $\underline{Z}$ satisfies: for any $i, j_1, \ldots, j_\ell$ $(\ell \leq k-1)$

$$\text{(i)} \qquad E(Z_i|Z_{j_1}, \ldots, Z_{j_\ell}) = \sum_{m=1}^{\ell} \delta_{\underline{j},i,m} Z_{j_m} \tag{1}$$

and

$$\text{(ii)} \qquad var(Z_i|Z_{j_1}, \ldots, Z_{j_k}) = \sigma^2_{\underline{j},i} \tag{2}$$

for constants $\delta_{\underline{j},i,m}\ \epsilon\ \mathbf{R}$ and $\sigma^2_{\underline{j},i}\ \epsilon\ \mathbf{R}^+$.

A random vector $\underline{Z}$ satisfying (4.1) and (4.2) will have a corresponding variance-covariance matrix $\Sigma = R$ (with unit entries in the diagonal). To avoid trivial cases we assume R is not a diagonal matrix. Clearly there are quite complicated inter-relationships that must hold among the coefficients appearing in (4.1) and (4.2) since they must be consistent with some diagonal variance-covariance matrix R. Of course, for a given R, there are many $GCS_2(k)$ distributions. It is convenient to introduce the notation $GCS_2(k, \Sigma)$ to denote all random vectors $\underline{X}$ with Gaussian conditional structure of the second order with a given k-dimensional variance-covariance matrix Σ. Analogously if we write $\underline{Z} \in GCS_2(k, R)$ we mean that $\underline{Z}$ is a standardized vector with Gaussian conditional structure of the second order and correlation matrix R.

If (4.1) and (4.2) hold, the joint characteristic function of $\underline{Z}$ is severely constrained. Conditions analogous to those displayed in equations (3.1) and (3.2) must hold for various first and second partial derivatives of the joint characteristic function. In the bivariate case, we have $\underline{Z} \in GCS_2(2)$ iff

(i) $$E(Z_1|Z_2) = \rho Z_2, E(Z_2|Z_1) = \rho Z_1 \tag{3}$$

and

(ii) $$var(Z_1|Z_2) = var(Z_2|Z_1) = 1 - \rho^2 \tag{4}$$

where

$$\rho = cov(Z_1, Z_2)(\in (-1, 1)) .$$

Conditions (3.1) and (3.2) may be rewritten for such standardized variables as follows.

<u>Lemma 4.1</u>: $\underline{Z} \in GCS_2(2)$ iff for some $\rho \in (-1, 1)$ its joint characteristic function $\phi(t_1, t_2)$ satisfies

$$\frac{\partial}{\partial t_1}\phi(t_1, t_2)|_{t_1=0} = \rho \frac{d}{dt_2}\phi(0, t_2) \tag{5}$$

$$\frac{\partial}{\partial t_2}\phi(t_1, t_2)|_{t_2=0} = \rho \frac{d}{dt_1}\phi(t_1, 0) \tag{6}$$

$$\frac{\partial^2}{\partial t_1^2}\phi(t_1, t_2)|_{t_1=0} = (\rho^2 - 1)\phi(0, t_2) + \rho^2 \frac{d^2}{dt_2^2}\phi(0, t_2) \tag{7}$$

and

$$\frac{\partial^2}{\partial t_2^2}\phi(t_1, t_2)|_{t_2=0} = (\rho^2 - 1)\phi(t_1, 0) + \rho^2 \frac{d^2}{dt_1^2}\phi(t_1, 0) . \tag{8}$$

It is not hard to verify that a classical bivariate normal random vector with unit variances and correlation coefficient ρ, has a joint characteristics function which satisfies (4.5) - (4.8). Similarly the joint characteristic function of the Kwapien distribution (3.3) clearly satisfies (4.5) - (4.8) with $\rho = 2p - 1$.

The class $GCS_2(k)$ contains Gaussian distributions, non-Gaussian densities as in (3.7) and, when $k = 2$, even discrete distributions. The common features of all the members can be expressed in terms of properties of conditional moments or of derivatives of the joint characteristic function. The class is however diverse. Some closure properties are however available for the class $GCS_2(k)$. For example each subclass $GCS_2(k,\Sigma)$, for fixed Z, is closed under mixtures.

Theorem 4.2: Suppose $\{\underline{X}_\alpha : \alpha \epsilon \Lambda\}$ is an indexed collection of random vectors with $X_\alpha \epsilon GCS_2(k,\Sigma)$ for every α. If we define Z to be a random vector with distribution function

$$F_{\underline{Z}}(\underline{z}) = \int_\Lambda F_{\underline{X}_\alpha}(\underline{z})dH(\alpha)$$

for any probability distribution H on Λ, then $\underline{Z} \epsilon GCS_2(k,\Sigma)$.

Proof: The bivariate case ($k=2$) with Λ of cardinality 2 was reported by Bryc (1985). The general result is straightforward if we write the joint characteristic function as a mixture

$$\phi_{\underline{Z}}(t) = \int_\Lambda \phi_{\underline{X}_\alpha}(t)dH(\alpha)$$

and observe that the conditions (4.5) - (4.8) (and their k-dimensional analogs) are preserved by mixtures since the covariance structures (and hence the coefficients in (4.5) - (4.8)) are the same for every α.

Linear combinations of independent random vectors in $GCS_2(k,\Sigma)$ will yield random vectors in $GCS_2(k)$ but with a different covariance matrix. Specifically we have

Theorem 4.3: Suppose that $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ are independent members (not necessarily identically distributed) of $GCS_2(k,\Sigma)$ then for $(a,b,) \neq (0,0)$, $a\underline{X}^{(1)}+b\underline{X}^{(2)} \epsilon GCS_2(k,(a^2+b^2)\Sigma)$. In particular if $a^2+b^2=1$, then $a\underline{X}^{(1)}+b\underline{X}^{(2)} \epsilon GCS_2(k,\Sigma)$.

Proof: We provide a proof in the bivariate case. More extensive equations analogous to (4.5) - (4.8) must be verified in higher dimensional cases.

For simplicity and without loss of generality we assume that $\underline{X}^{(1)}$ and $\underline{X}^2$ have been standardized and we will denote them by $\underline{Z}^{(1)}$ and $Z^{(2)}$. By assumption $\underline{Z}^{(1)}$ and $Z^{(2)}$ have common correlation ratio ρ and their joint characteristic functions satisfy (4.5) - (4.8). Denote the joint characteristic functions of $\underline{Z}^{(1)}, \underline{Z}^{(2)}$ and $a\underline{Z}^{(1)}+b\underline{Z}^{(2)}$ by f_1, f_2 and f_3. Because $\underline{Z}^{(1)}$ and $\underline{Z}^{(2)}$ are independent we have

$$f_3(t_1,t_2) = f_1(at_1,at_2)f_2(bt_1,bt_2)\ .$$

Consequently, using (4.5) for f_1 and f_2

$$\frac{\partial}{\partial t_1}f_3(t_1,t_2)|_{t_1=0} = a\frac{\partial}{\partial t_1}f_1(at_1,at_2)f_2(bt_1,bt_2)$$

$$+f_1(at_1,at_2)b\frac{\partial}{\partial t_1}f_2(bt_1,bt_2)|_{t_1=0}$$

$$= a\rho\frac{d}{dt_2}f_1(0,at_2)f_2(0,bt_2)$$

$$+b\rho f_1(0, at_2)\frac{d}{dt_2}f_2(0, bt_2)$$

$$= \rho\frac{d}{dt_2}f_3(0, t_2)\ .$$

Thus (4.5) holds for f_3. Similarly (4.6) may be verified. Differentiating twice and using (4.7) for f_1 and f_2 we find

$$\frac{\partial^2}{\partial t_1^2}f_3(t_1, t_2)|_{t_1=0} = (a^2+b^2)(\rho^2-1)f_3(0, t_2)$$

$$+\rho^2\frac{d^2}{dt_2^2}f_3(0, t_2)\ .$$

When $a^2 + b^2 = 1$ this implies that (4.7) continues to hold for f_3. In parallel fashion, since (4.8) holds for f_1 and f_2 then, when $a^2 + b^2 = 1$, it continues to hold for f_3. Since conditions (4.5) - (4.8) are sufficient for membership in $GSC_2(2)$ the conclusion of the theorem follows.

Naturally we can extend Theorem 4.3 to deal with sums of more than 2 independent members of $GCS_2(\Sigma)$. Indeed we can consider infinite convolutions since clearly the class $GCS_2(k,\Sigma)$ is closed under weak convergence (i.e. if $\underline{X}^{(n)} \in GCS_2(k,\Sigma), n = 1, 2, \ldots$ and $\underline{X}^{(n)} \xrightarrow{d} \underline{X}^{(\infty)}$ then $\underline{X}^{(\infty)} \in GCS_2(k,\Sigma)$). Thus we may state

Theorem 4.4: Suppose $\underline{X}^{(1)}\underline{X}^{(2)}\cdots$ are independent random vectors each being a member of $GCS_2(k,\Sigma)$ (the same Σ for every $\underline{X}^{(i)}$). Define $\underline{Y} = \sum_{i=1}^{\infty} a_i\underline{X}^{(i)}$ where $\sum_{i=1}^{\infty} a_i^2 = 1$. It follows that $\underline{Y} \in GCS_2(k,\Sigma)$.

Example: (Uniform and Cantor marginals) Suppose that $\underline{X}^{(i)}, i = 1, 2, \ldots$ are i.i.d. Kwapien random vectors (with characteristic function (3.3) and joint density (3.4)). Consider a random vector $\underline{Y} = \sum_{i=1}^{\infty} a_i\underline{X}^{(i)}$ where $\sum_{i=1}^{\infty} a_i^2 < \infty$. Since each $\underline{X}^{(i)} \in GCS_2(2)$ with correlation $2p-1$, it follows that $\underline{Y} \in GCS_2(2)$ with the same correlation, $2p-1$. Particular choices for the a_i's yield interesting examples. If we choose $a_i = 1/2^i, i = 1, 2, \ldots, \underline{Y}$ will have a continuous bivariate distribution with uniform $(-1, 1)$ marginals (and Gaussian conditional structure). We conjecture but are unable to prove that this joint distribution is singular (unless $p = 1/2$, the uninteresting case of independent marginals). If we choose $a_i = 2/3^i, i = 1, 2, \ldots$ then $\underline{Y}$ will have a singular joint distribution with Cantor (and thus clearly singular) marginals. Thus we have a singular continuous example with Gaussian conditional structure. It is well known that sums of independent Cantor-like singular random variables can have non-singular (indeed uniform) distributions. Our present construction (using $a_i = 2/3^i$) allows us to give an example of dependent Cantor-like random variables whose sum is uniform. To do this, consider the special case $\underline{Y} = \sum_{i=1}^{\infty} \frac{2}{3^i}\underline{X}^{(i)}$ where the $\underline{X}^{(i)}$'s are Kwapien random vectors with $p = 2/3$. Here Y_1 and Y_2 are singular (Cantor) distributed on $(-1, 1)$ but $(Y_1 + Y_2)/2$ is uniform on $(-1, 1)$ (as is easily proved by looking at the convergent infinite product representation of its characteristic function obtained using the expression for the Kwapien characteristic function given in (3.3)).

5 FROM $GCS_2(k)$ TO CLASSICAL NORMAL

The examples of section 4 clearly indicate that additional conditions, besides appropriate behavior of conditional moments up to order 2 will be required to characterize the classical normal model. In this section we survey some known and some new results in this area. First a result due to Szablowski (1989).

Theorem 5.1: If $\underline{X} \in GCS_2(k)$ and if $\underline{X}$ is elliptically contoured then $\underline{X} \sim N^{(k)}(\underline{\mu}, \Sigma)$.

Next, we consider the generalized independence models described by Kagan (1988) classes.

Definition 5.2: A k-dimensional random vector $\underline{X}$ belongs to the Kagan class $D_{k,j}(loc), j = 1, 2, \ldots, k,\ k = 1, 2, \ldots$, if its characteristic function $\phi_{\underline{X}}$, in some neighborhood, V, of the origin in $\mathbf{R}^k$ has the form

$$\phi_{\underline{X}}(\underline{t}) = \prod_{1 \leq i_1 < \cdots < i_j \leq k} R_{i_1, \ldots, i_j}(t_{i_1} t_{i_2}, \ldots, t_{i_j})$$

where each $R_{\underline{i}}$ is a continuous complex function with $R_{\underline{i}}(\underline{0}) = 1,\ \forall \underline{i}$.

It is plausible that any $\underline{X}$ in a Kagan class $D_{k,j}(loc)$ that exhibits Gaussian conditional structure of the second order might be classical k-variate normal. Some progress towards proving this result is provided in the following result due to Wesolowski (1991).

Theorem 5.3: If $\underline{X} \in GCS_2(k)$ for some $k > 2$ and if $\underline{X} \in D_{k,2}(loc)$ then $\underline{X} \sim N^{(k)}(\underline{\mu}, \Sigma)$.

If we assume that $\underline{X}$, in addition to having Gaussian conditional classical structure of the second order, is infinitely divisible, then it must be classical normal. This result is due to Wesolowski (1993). We refer the reader to the original paper for a general proof. Here we provide a simple illuminating proof for the bivariate case only.

Theorem 5.4: If $\underline{X} \epsilon GCS_2(k)$ and if $\underline{X}$ is infinitely divisible then $\underline{X} \sim N^{(k)}(\underline{\mu}, \Sigma)$.

Proof: (in the bivariate case, i.e. $k = 2$, the case $k > 2$ was proved by another approach in Wesolowski (1993)). As usual, without loss of generality we assume zero means and unit variances. Since $\underline{X} = (X_1, X_2)$ is infinitely divisible, the logarithm of its joint characteristic function is of the form

$$\begin{aligned} \psi(\underline{t}) = \log \phi(\underline{t}) &= -\frac{1}{2}(t_1^2 + 2\rho t_1 t_2 + t_2^2) \\ &+ \int_{\mathbf{R}^2} [e^{i(t_1 x + t_2 y)} - 1 - i(t_1 x + t_2 y)] \frac{dK(x, y)}{x^2 + y^2} \end{aligned}$$

for some measure K. It then follows that

$$\frac{\partial^2}{\partial t_1^2} \psi(\underline{t})|_{t_1=0} = -1 - \int_{\mathbf{R}^2} e^{it_2 y} \frac{x^2}{x^2 + y^2} dK(x, y)$$

and

$$\frac{d^2}{dt_2^2}\psi(0,t_2) = -1 - \int_{\mathbf{R}^2} e^{it_{2y}} \frac{y^2}{x^2+y^2} dK(x,y) \ .$$

However since $\underline{X} \epsilon GCS_2(2)$, we know that (4.8) holds. Consequently we have

$$\int_{\mathbf{R}^2} \frac{x^2}{x^2+y^2} dK(x,y) = \rho^2 \int \frac{y^2}{x^2+y^2} dK(x,y) \qquad (1)$$

(where $\rho^2 < 1$). Analogously, by considering $\frac{\partial^2}{\partial t_2^2}\psi(\underline{t})|_{t_2=0}$ and $\frac{d^2}{dt_1^2}\psi(t_1,0)$ we find that (5.1) again holds with the roles of x and y interchanged. Summing we conclude that for $\rho^2 < 1$,

$$\int_{\mathbf{R}^2} dK(x,y) = \rho^2 \int_{\mathbf{R}^2} dK(x,y) \ ,$$

i.e. $dK \equiv 0$. Consequently $\underline{X}$ must be classical bivariate normal.

6 REMARKS

Progress towards understanding the class of distributions with Gaussian conditional structure is accelerating. Many interesting questions remain open. Perhaps the most frustrating lacuna in the current inventory of examples involves the absence of any discrete example with Gaussian conditional structure of dimension greater than 2 (as discussed in Section 3). Theorem 4.3 together with the celebrated Kwapien example permits construction of a plethora of two dimensional discrete distributions with Gaussian conditional structure. The elusive 3 dimensional examples should appear soon.

REFERENCES

1. Arnold, B.C., Castillo, E. and Sarabia, J.M., Conditionally Specified Distributions, *Lecture Notes in Statistics*, Vol. 73, Springer-Verlag, Berlin, (1992).

2. Arnold, B.C., Castillo, E. and Sarabia, J.M., A conditional characterization of the multivariate normal distribution, *Statistics and Probability Letters*, 19, 313-315, (1994).

3. Bhattacharya, A., On some sets of sufficient conditions leading to the normal bivariate distribution, *Sankhya*, 6, 399-406, (1943).

4. Bryc, W., Some remarks on random vectors with nice enough behavior of conditional moments, *Bull. Polish Acad. Sci. Math.*, 33, 677-684, (1985).

5. Bryc, W. and Plucinska, A., A characterization of infinite Gaussian sequences by conditional moments, *Sankhya*, A47, 166-173, (1985).

6. Kagan, A.M., New classes of dependent random variables and a generalization of the Darmois-Skitovich to several forms, *Theory of Probability and Applications*, 33, 286-295, (1988).

7. Kagan, A.M., Linnik, J.V. and Rao, C.R., *Characterization Problems of Mathematical Statistics,* Wiley, New York, (1973).

8. Nguyen, T.T., Rempala, G. and Wesolowski, J., Non-Gaussian measures with Gaussian structure, to appear in *Probability and Mathematical Statistics*, (1994).

9. Plucinska, A., On a stochastic process determined by the conditional expectation and the conditional variance, *Stochastics*, 10, 115-129, (1983).

10. Szablowski, P.J., Can the first two conditional moments identify a mean square differentiable process?, *Comput. Math. Appl.*, 18, 329-348, (1989).

11. Wesolowski, J., A characterization of the Gaussian process based on properties of conditional moments, *Demonstratio Math.*, 18, 795-808, (1984).

12. Wesolowski, J., Gaussian conditional structure of the second order and the Kagan classification of multivariate distributions, *Journal of Multivariate Analysis*, 39, 79-86, (1991).

13. Wesolowski, J., Multivariate infinitely divisible distributions with the Gaussian conditional structure of the second order. In *Stability Problems for Stochastic Models* (Kalashnikov, V.V. and Zolotarev, V.M. eds). *Lecture Notes in Mathematics*, Vol. 1546, 180-183, Springer-Verlag, Berlin, (1993).