DEPENDENCE IN STOCHASTIC SIMULATION MODELS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Soumyadip Ghosh

May 2004

DEPENDENCE IN STOCHASTIC SIMULATION MODELS

Soumyadip Ghosh, Ph.D.

Cornell University 2004

There is a growing need for the ability to model and generate samples of dependent random variables as primitive inputs to stochastic models. We consider the case where this dependence is modeled in terms of a partially-specified finite-dimensional random vector. A random vector sampler is commonly required to match a given set of distributions for each of its components (the *marginal* distributions) and values of their pairwise covariances. The NORTA method, which produces samples via a transformation of a joint-normal random vector sample, is considered the state-of-the-art method for matching this specification. We begin by showing that the NORTA method has certain flaws in its design which limit its applicability.

A covariance matrix is said to be feasible for a given set of marginal distributions if a random vector exists with these properties. We develop a computational tool that can establish the feasibility of (almost) any covariance matrix for a fixed set of marginals. This tool is used to rigorously establish that there are feasible combinations of marginals and covariance matrices that the NORTA method cannot match. We further determine that as the dimension of the random vector increases, this problem rapidly becomes acute, in the sense that NORTA becomes increasingly likely to fail to match feasible specifications. As part of this analysis, we propose a random matrix sampling technique that is possibly of wider interest.

We extend our study along two natural paths. First, we investigate whether NORTA can be modified to approximately match a desired covariance matrix that the original NORTA procedure fails to match. Results show that simple, elegant modifications to the NORTA procedure can help it achieve close approximations to the desired covariance matrix, and these modifications perform well with increasing dimension.

Second, the feasibility testing procedure suggests a random vector sampling technique that can exactly match (almost) any given feasible set of marginals and covariances, i.e., be free of the limitations of NORTA. We develop a strong characterization of the computational effort needed by this new sampling technique. This technique is computationally competitive with NORTA in low to moderate dimensions, while matching the desired covariances *exactly*.

BIOGRAPHICAL SKETCH

Banani and Kasi Nath gave birth to the author on the 11th of January, 1978 in Asansol (a border town of West Bengal, India) and, in keeping with the established social custom, got to name him Soumyadip. Noting with some concern their baby's apparent discomfort to the prevailing local culture, they decided to keep the family on the move, maintaining a remarkably consistent pace of a new location every 7 years. Their first sojourn also saw the happy occasion of the birth of their second (and, some maintain, a more sociable) son. After jaunting happily through 18 years of this, the author decided to branch out on his own and moved away, yet again, to Chennai (India) to join the Indian Institute of Technology there for a Bachelor's degree in Mechanical Engineering. Designing machinery did not turn out as inspiring as he had hoped, so he travelled to the University of Michigan to pursue a Doctoral degree in Operations Research. While there his thesis advisor, doubtless taking a cue from his family tradition, decided that a relocation to Cornell University would serve him good, and so here we finally come to the end of the trail.

To my parents, who set the ball rolling

and gave it quite some speed!

and to my brother

who chose never to interfere (though, mind you, he could have)

ACKNOWLEDGEMENTS

Foremost I wish to express my gratitude to my advisor, Prof Shane G Henderson, for doing such a sterling job of seeing me through the course of my graduate studies. Over the years, I have borrowed liberally from his wisdom, his sense of honor, balance and judgement (and humour), and his fortitude in pursuing academic research, in shaping my self, and I hope I do not leave him poorer from the experience.

I am also deeply indebted to Profs David Shmoys and Sid Resnick in helping make my transition from the University of Michigan a painless affair, and also for the many encouraging and helpful suggestions they provided in their capacity as members of my Special Committee.

I am grateful to the many professors who have patiently taught me all that I know and, more importantly, helping me understand how I can continue my learning on my own.

It will be futile to try to capture on paper a full measure of my appreciation for my family for all that they have done to get me where I am today. I owe much to my father, who introduced me to the wonders of the structured logic of science, my mother, who has been such a shining example of what a human can be, and my brother, who did so well in filling the void that my parents could not have filled; I can only hope they understand how important they are to me.

Life would be very hard to live if not for that vital component: friends. I have been very lucky in filling in my life's canvas with some of the most wonderful people one could hope to meet. My run of good fortune in this seems to have started early in life and continued almost unbroken till date. My most special and sincere thanks to those who have stood by, bore with and shared precious moments with me in

my three years here at Ithaca. Graduate School restrictions on name-dropping in thesis acknowledgements prevent me from quoting a complete list of friends I owe to (and I assure you, it would be *long*), but I am sure my memories of this place will always revolve around them.

My thanks go to the International Business Machines Corporation for chipping in with a fellowship that supported me for two years.

(*Legal reasons compel me to state that I have not actually found the aforementioned name-dropping rules codified anywhere, but that might probably be due to the scant time I could afford the search.*)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

RANDOM VECTOR GENERATION

1.1 Introduction

An important part of the setup of a stochastic simulation involves modeling the input environment of the system being studied. It is crucial that the uncertainty in this input be represented in an appropriate manner. There is a growing need for the ability to specify and generate dependent observations as primitive inputs to stochastic models. For example, in a manufacturing setting, the processing times of a single job at different stations may be correlated due to characteristics of the job such as size. In determining reservoir release rules, the inflows of water to different reservoirs are invariably correlated. Nelson (2003) report an agricultural insurance study where they find that dependence information between various factors like weather, crop yields etc. are important to the validity of the simulation study. In generating random test problems for a given algorithm, it is advantageous to ensure that some elements of the problem are correlated (Hill and Reilly 1994; 2000, Hodgson et al. 2000). Further applications have recently been reported in cost analysis (Lurie and Goldberg 1998), and in decision and risk analysis (Clemen and Reilly 1999).

Dependence information can be incorporated in an input process in two ways. The first captures "temporal" dependence that arises over time and is traditionally studied in terms of time series. The second kind can be characterized as consisting of a finite number of dependent random variables, jointly called a random vector. We examine the second case here; Biller and Ghosh (2005), Law and Kelton (2000) provide overviews of the first case.

The "ideal" approach to modeling a random vector is to specify its full joint distribution since a joint distribution completely determines all its properties. This approach works wonderfully in the univariate case, where many specialized methods take advantage of the availability of a univariate distribution function. These methods however do not extend well to higher dimensions, the culprit being the rather dramatically named "curse of dimensionality". The primary difficulty in this case is that a tremendous amount of information is typically required to specify (and fit) such a joint distribution. Furthermore, special methods must be devised to generate random vectors with the given joint distribution, and this can be a practically insurmountable problem for a model of even moderate complexity (Law and Kelton 2000, p. 479).

Another approach is to approximate the joint distribution using distributions from a chosen parametric family. Thus, given data or any relevant information, one then tries to estimate reasonable parameter values. This approach is also hampered by the dimensionality problem, and is typically limited to situations where the marginal distributions are all from the same parametric family. For methods of this type see, for example, Devroye (1986) and Johnson (1987). But the case where the marginals are not all from the same parametric family affords far greater modeling generality, and is perhaps the case of more interest from a practical standpoint.

A practical alternative is to partially specify the random vector. The most common situation is to require that a generation procedure match the univariate distributions of each of the components of the random vector (its *marginal distributions*) together with some dependence measure that, one hopes, captures the essence of the dependence structure while being convenient to work with. The

dependence measure is usually some form of covariance and could be Spearman's rank covariance, Pearson's product-moment covariance, Kendall's τ, or any other convenient covariance measure. We will focus on Pearson's product-moment covariance and Spearman's rank covariance because of their wide use and acceptance in application settings (they are defined below); Nelsen (1999, Chapter 5) defines Kendall's τ and other such measures.

Another argument in support of modeling random vectors using marginals and covariances relates to the use of diffusion approximations for modeling stochastic systems. In many cases the limiting diffusions depend only on the first two moments of the input distributions. Therefore, there is some insensitivity in performance measures computed from these models to the exact form of the input distributions. In general then, if a form of this insensitivity is present in a model, then the approach discussed here for modelling random vectors is quite reasonable.

Note that a partial specification does not necessarily uniquely specify the joint distribution. Indeed, the specification could even be inconsistent.

Definition 1.1.1 *We say that a correlation specification is* feasible *for a given set of marginal distributions if a random vector with the specified characteristics exists.*

Any method that generates random vectors from partial specifications should ideally be able to handle all feasible specifications.

The Pearson product-moment covariance between two random variables X and Y, defined when $E(X^2 + Y^2) < \infty$, is given by

$$\mathrm{Cov}(X, Y) = EXY - EXEY.$$

A related measure is the product-moment correlation defined as (let $\mathrm{Var}(X) = \sigma_X^2$

and $\mathrm{Var}(Y) = \sigma_Y^2$)

$$\mathrm{Cor}(X, Y) = \mathrm{Cov}(X, Y)/(\sigma_X \sigma_Y).$$

Specifying covariances is equivalent to specifying correlations, since the marginal distributions are also pre-specified. We shall henceforth use either term when describing specifications.

If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$ and $\mathrm{Cor}(X, Y) = 0$. On the other hand, zero covariance (or correlation) between X and Y does not generally imply that they are independent.

The product-moment correlation measure is easily shown to range in value between -1 and $+1$.

Correlations (and covariances) are measures of the degree of linear dependence between X and Y: If X is a linear function of Y so that $X = a+bY$ with probability one, then $\mathrm{Cor}(X, Y) = +1$ if $b > 0$ and $\mathrm{Cor}(X, Y) = -1$ if $b < 0$. Conversely, if the correlation between X and Y has magnitude 1, then a linear relationship between X and Y holds with probability one, and its sign reflects the direction of the linear dependence. This property limits the effectiveness of modeling using correlations since "non-linear" dependence information cannot be properly captured. Consider this simple example: let X and Y be two random variables defined such that the vector $(X, Y)'$ is uniformly likely to be any point on the $2-$dimensional unit circle. An easy calculation yields that $\mathrm{Cov}(X, Y) = \mathrm{Cor}(X, Y) = 0$ but they are clearly dependent since the (non-linear) relationship $X^2 + Y^2 = 1$ holds.

Another fact that is not given the attention it perhaps deserves is that though the possible values of correlation range in $[-1, +1]$, the actual set of values that can be achieved depends on the (univariate) distributions of the random variables and in general this set is a strict subset of $[-1, +1]$. For example, the correlation

between a random variable uniformly distributed on $(0, 1]$ and another exponentially distributed with mean 1 can be shown to range in $[-0.866, 0.866]$. It is thus not possible to find a bivariate distribution with these marginals and a correlation of 0.9.

The Pearson product-moment covariance structure of a $d-$dimensional random vector X consists of covariances of every pair of its components $\mathrm{Cov}(X_i, X_j),\ 1 \leq i, j \leq d$, collectively given as the $d \times d$ covariance matrix Σ. Since the definition of covariance is symmetric in its arguments, the covariance matrix is necessarily symmetric. These matrices are also necessarily positive semidefinite. To see why, note that $a^t X$ defines a valid random variable for any $d-$vector a, and has variance $a^t \Sigma a$. Since variances are non-negative and a could be any arbitrary vector we see that the matrix Σ is positive semidefinite.

Definition 1.1.2 *Any symmetric, positive semidefinite matrix with unit diagonal elements is called a correlation matrix.*

It is often mistakenly presumed that any arbitrary correlation matrix can be matched a random vector with the desired marginal distributions. Note however that positive semidefiniteness does not always guarantee feasibility of a correlation matrix for an arbitrary set of marginal distributions. Consider the example given above of the random vector consisting of a uniformly and an exponentially distributed component. Matrices of the form $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ where $|\rho| \leq 1$ are positive semidefinite. However, for this case, the matrix with $\rho = 0.9$ is not feasible as we have noted above.

Positive semidefiniteness can be sufficient for special cases of marginal distributions, for instance in the case of normal marginal distributions where a joint

(or multivariate) normal distribution can always be constructed given a positive semidefinite correlation matrix (see, for e.g., Tong 1990). Three dimensional random vectors with uniform marginals are also known to share this property (Joe 1997, Kurowicka and Cooke 2001), but it is not known if this extends to higher dimensions. In general then the set of feasible correlation matrices (for a given set of marginals) is a strict subset of the set of correlation matrices as defined by Definition 1.1.2. An exact description of this set for general marginals is unfortunately not known, and is a very interesting open question.

Spearman's rank covariance measure avoids certain limitations of product-moment covariances. The rank covariance between two random variables X and Y is defined to be

$$\mathrm{rcov}(X, Y) = EF(X)G(Y) - EF(X)EG(Y)$$

where F and G are the distribution functions of X and Y. (Passing from X to $F(X)$ is called the *probability transformation*.)

Rank correlations are defined as in the product-moment case: If $\sigma_F^2 = \mathrm{Var}(F(X))$ and $\sigma_G^2 = \mathrm{Var}(G(Y))$, then the rank correlation between X and Y is defined to be

$$\mathrm{rcor}(X, Y) = \mathrm{rcov}(X, Y)/(\sigma_F \sigma_G).$$

One can see from its definition that rank covariances are always well-defined, while product-moment covariances are defined only when the variances of X and Y are finite (thus, certain heavy-tailed distributions would be left out). This is because $F(X)$ and $G(Y)$ are bounded random variables. In fact, if F (G) is continuous, then $F(X)$ ($G(Y)$) is uniformly distributed on $(0, 1)$ (Billingsley 1995, p. 197).

Another important property of Spearman's rank covariance is that unlike Pearson's product-moment covariance, it is preserved under strictly increasing transformations of the random variables. This property is very important from a practical point of view, as it allows us to handle random vectors with continuous marginals and specified rank correlations easily. One simply generates a random vector with uniform marginals and the given rank correlations and then transforms each marginal using the inverse of the corresponding probability integral transform $F^{-1}(\cdot)$ (Equation (1.2) defines this transformation) to get the desired marginals. These transformations are strictly increasing and thus preserve the rank correlations.

The rank correlation structure of a random vector is defined by its rank correlation matrix. These matrices are again necessarily symmetric and positive semidefinite, but this condition is again not sufficient: the set of feasible rank correlation matrices is typically a strict subset of the set of correlation matrices as defined in Definition 1.1.2.

In light of the facts presented above, we shall now briefly review some of the more popular random vector sampling methods that match a marginals and correlation specification.

Hill and Reilly (1994) describe a method for generating random vectors with specified marginals and covariances through mixtures of *extremal covariance* distributions. They define extremal covariances as certain covariance matrices for which corresponding joint distribution functions are easily constructed and generated. The premise of this approach is since extremal covariances are easily sampled, one could try to match any arbitrary covariance matrix with a convex combination (probability mixture) of the distributions that produce these extremal covariances.

It is very effective for random vectors of low dimension ($d \leq 3$ say), but the computational requirements quickly become excessive for higher dimensional random vectors. There is another difficulty with this approach. The set of feasible covariance matrices is compact and convex (refer Proposition 2.2.6), but generally not a polytope and hence there are sets of marginals with feasible covariance matrix that cannot be matched using the technique developed by Hill and Reilly.

Meeuwissen and Cooke (1994) describe "tree-dependent" random vectors that can be rapidly generated, but cannot match all feasible covariance matrices. Cooke (1997) introduces a generalization of tree-dependent random vectors that is based on a "vine" representation of a joint distribution. These concepts arise from Bayesian and graph-theoretic literature. Such random vectors can be rapidly generated, but it is not yet clear whether they can be used to model *any* feasible covariance matrix. Other methods for tackling the problem of generating random vectors with specified marginals and covariance matrix have been developed; Nelson and Yamnitsky (1998), Vincent (1998) and the forthcoming Biller and Ghosh (2005) give a good survey.

Cario and Nelson (1997) described the "NORmal To Anything" (NORTA) method for generating random vectors with prescribed covariance matrix. The NORTA method basically involves a component-wise transformation of a multivariate normal random vector, and capitalizes on the fact that multivariate normal random vectors are easily generated; see e.g., Law and Kelton (2000), p. 480. Cario and Nelson traced the roots of the method back to Mardia (1970) who looked at bivariate distributions and product-moment covariances, and to Li and Hammond (1975) who concentrated on the case where all of the marginals have densities (with respect to Lebesgue measure). Iman and Conover (1982) implemented the same

transformation procedure to induce a given rank correlation in the output. Their method is only approximate, in that the output will have only very approximately the desired rank correlation.

The NORTA method is very efficient and easily implemented, and has seen use in a variety of contexts. Clemen and Reilly (1999) use the NORTA procedure to induce a desired rank correlation in the context of decision and risk analysis. Lurie and Goldberg (1998) implement a variant of the NORTA method for generating samples of a predetermined size, which they use in cost analysis. Henderson et al. (2000) adapt the NORTA method to generate samples of dependent quasi-random vectors. The NORTA method is also routinely used in portfolio models in industry.

So the NORTA procedure is often the method of choice for generating random vectors with prescribed marginals and correlation matrix. But can the NORTA procedure match *any* feasible covariance matrix for a given set of marginals? Both Li and Hammond (1975) and Lurie and Goldberg (1998) give examples where this does not appear to be the case. However, the random vectors that they propose as counterexamples were not proved to exist, and so the question was not completely settled.

For 2-dimensional random vectors, the NORTA method can match any feasible covariance matrix. This follows immediately from the characterizations in Whitt (1976). However, for dimensions 3 and greater, little is known.

This question formed the starting point from whence we began our investigation into random vector generation methodologies. Specifically, we were interested in determining whether there are feasible covariance matrices for a given set of marginals that the NORTA method cannot match. We were able to prove that this is indeed the case (refer Chapter 2).

Since the rest of this thesis presumes a good working knowledge of the NORTA method, we shall now precisely state the algorithm and the feasibility problem it faces. We will then conclude this chapter with a roadmap of the progress we have made on various fronts regarding random vector generation methodology. We believe that each chapter is substantially self-contained; thus, we shall postpone a survey of the relevant literature until the introductory part of each chapter.

1.2 The NORTA method

Cario and Nelson (1997) described the "NORmal To Anything" (NORTA) method for generating i.i.d. replicates of a random vector $X^* = (X_1^*, \ldots, X_d^*)$ say with prescribed marginal distributions and covariance structure. In this method, one starts by generating a random vector Z with a multivariate normal distribution and transforms Z to obtain a random vector $X = (X_1, \ldots, X_d)$. Let F_i be the desired marginal distribution function of X_i^*, for $i = 1, \ldots, d$.

The NORTA method generates i.i.d. replicates of X by the following procedure.

1. Generate an $\mathbb{R}^d$ valued multivariate normal random vector $Z = (Z_1, \ldots, Z_d)$ with mean vector 0 and covariance matrix $\Sigma_Z = (\Sigma_Z(i,j) : 1 \leq i, j \leq d)$, where $\Sigma_Z(i,i) = 1$ for $i = 1, \ldots, d$. (Refer Sections V.4 and XI.2 in Devroye (1986) for a generation procedure.)

2. Compute the vector $X = (X_1, \ldots, X_d)$ via

$$X_i = F_i^{-1}(\Phi(Z_i)), \tag{1.1}$$

for $i = 1, \ldots, d$, where Φ is the distribution function of a standard normal random variable, and

$$F_i^{-1}(u) = \inf\{x : F_i(x) \geq u\}. \tag{1.2}$$

The vector X generated by this procedure will have the prescribed marginal distributions. To see this, note that each Z_i has a standard normal distribution, so that $\Phi(Z_i)$ is uniformly distributed on $(0,1)$, and so $F_i^{-1}(\Phi(Z_i))$ will have the required marginal distribution.

The covariance matrix Σ_Z should be chosen so that it induces the required covariance structure on X. The NORTA method can use either Pearson's product-moment covariance or Spearman's rank covariance.

1.2.1 Pearson's Product-Moment Covariance

Suppose that we wish X^* to have Pearson product-moment covariance matrix Σ, where $\Sigma(i,j) = \mathrm{Cov}(X_i^*, X_j^*)$ for $1 \le i, j \le d$. This is the case that Cario and Nelson (1997) examined. To ensure that the required covariances are defined, we make the assumption that $E[(X_i^*)^2] < \infty$ for $i = 1, \ldots, d$. It turns out that choosing Σ_Z to arrive at the correct covariance matrix Σ is a nontrivial problem.

Let X be the random vector generated from (1.1) above and Σ_X denote its covariance matrix. As noted in Li and Hammond (1975) and Cario and Nelson (1997), each term $\Sigma_X(i,j) = \mathrm{Cov}(X_i, X_j)$ is a function of only $\mathrm{Cov}(Z_i, Z_j)$ (also refer Schmeiser 1990). To see this, note that when $\mathrm{Cor}(Z_i, Z_j) \neq \pm 1$,

$$\mathrm{Cov}(X_i, X_j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_i^{-1}(\Phi(z_i)) F_j^{-1}(\Phi(z_j)) \varphi_{ij}(z_i, z_j)\, dz_i\, dz_j - EX_i EX_j, \quad (1.3)$$

where φ_{ij} is the joint density of (Z_i, Z_j). The expression (1.3) depends only on the marginal distributions F_i and F_j, and the density φ_{ij}. The joint-normal density φ_{ij} depends only on the covariance between Z_i and Z_j. When $\mathrm{Cov}(Z_i, Z_j) = \pm 1$, the joint density φ_{ij} degenerates and the integral representation (1.3) is no longer valid. However, in this degenerate case the covariance between X_i and X_j is still a function only of the covariance between Z_i and Z_j. Hence, the relation (1.3)

defines a function $c_{ij} : [-1, 1] \to \mathbb{R}$ mapping $\mathrm{Cov}(Z_i, Z_j)$ to $\mathrm{Cov}(X_i, X_j)$, where X_i and X_j are defined via (1.1).

So the problem of matching a desired covariance matrix reduces to $d(d-1)/2$ separate root-finding problems of selecting $\mathrm{Cov}(Z_i, Z_j)$ to match $\mathrm{Cov}(X_i, X_j)$ to $\Sigma(i, j)$. Unfortunately, there is no general analytical expression for the function c_{ij}, and so we cannot determine the exact Σ_Z that is to be used analytically.

Cario and Nelson (1997) established that under very mild conditions, the function c_{ij} is a continuous non-decreasing function of $\Sigma_Z(i, j)$. This result allows us to perform an efficient numerical search for values $\Lambda_Z(i, j)$ that yield

$$c_{ij}(\Lambda_Z(i, j)) = \Sigma(i, j) \text{ for } i < j. \tag{1.4}$$

We take $\Lambda_Z(i, i) = 1$ for $i = 1, \ldots, d$, and for $i > j$, set $\Lambda_Z(i, j) = \Lambda_Z(j, i)$ to ensure that Λ_Z is symmetric. Alternatives to the numerical search suggested by Cario and Nelson (1997) include the use of a stochastic root-finding algorithm (Chen 2001), or polynomial expansions (van der Geest 1998). Unless otherwise stated, we henceforth assume that a solution to (1.4) exists, which is easily checked (Cario and Nelson 1997).

One might hope that if the matrix Λ_Z satisfies (1.4), then Λ_Z could be used in the NORTA method to generate i.i.d. replicates of X. Unfortunately, the results we state in Chapter 2 prove that this is not always the case. The problem arises when the matrix Λ_Z as determined from (1.4) is not positive semidefinite, in which case it is not a valid covariance matrix for a joint-normal vector.

Li and Hammond (1975) suggested the following example to illustrate this important fact. Let X_1^*, X_2^* and X_3^* be 3 random variables uniformly distributed

on $(0, 1]$ with covariance matrix

$$\Sigma = \frac{1}{12} \begin{pmatrix} 1 & -0.4 & 0.2 \\ -0.4 & 1 & 0.8 \\ 0.2 & 0.8 & 1 \end{pmatrix}.$$

In the special case when X^* has uniform marginals, the equations (1.4) can be solved analytically. In particular, Kruskal (1958) showed that the (unique) solution to (1.4) is given by

$$\Lambda_Z(i, j) = 2 \sin[2\pi\Sigma(i, j)]. \tag{1.5}$$

For the Li and Hammond example, the (unique) matrix Λ_Z found from (1.5) is not positive semidefinite. It is important to observe though, that this is a counterexample only if the postulated uniform random vector itself exists. Li and Hammond did not show this.

Lurie and Goldberg (1998) gave an example with nonuniform marginals and positive definite covariance matrix for which the solution to (1.4) is also not positive semidefinite. They did not establish that the postulated random vector exists.

When all of the marginals have continuous distribution functions, a natural alternative to the numerical search procedure mentioned earlier is to "work in Gaussian space" and thus avoid this problem altogether. In other words, given a set of data with known (or fitted) marginals with continuous distribution functions, we first transform the data set into normal random variates using the inverse of the transformation (1.1). We can then compute an empirical covariance matrix Σ_Z and use this covariance matrix in the NORTA procedure. (Note that $(\Phi^{-1}(F(X))$ is not normally distributed if the distribution function F is not continuous, since then $F(X)$ does not have a uniform distribution on $(0, 1)$. Therefore, continuity of the marginals is needed.)

This approach is certainly simpler than a numerical search procedure, but it has two important drawbacks. First, it requires a set of input data, which may not be available in general. But second, and perhaps more importantly, this procedure does not necessarily ensure that the resulting X variates will have the required covariance structure. To see why, observe that the transformed normal random variables mentioned above are unlikely to have a *joint* normal distribution. Therefore, Σ_Z, when used as the correlations of the jointly normal random variable in the NORTA procedure, will be unlikely to transform through (1.1) to yield the desired covariance matrix for X, as one might otherwise expect. This is a subtle point, but one that is worth bearing in mind.

1.2.2 Spearman's Rank Covariance

Suppose now that we wish X^* to have Spearman's rank covariance matrix Σ, where $\Sigma(i,j) = \text{rcov}(X_i^*, X_j^*)$ for $1 \leq i, j \leq d$. This is the case treated by Clemen and Reilly (1999). We have remarked earlier that in contrast to product-moment covariance, the rank covariance is always defined and is preserved under strictly increasing transformations of the random variables.

If all of the marginal distribution functions F_i are continuous then the NORTA transformation (1.1) is strictly increasing. In this case, the rank covariance is preserved by the NORTA transformation, and so if X is the NORTA generated random vector, then

$$\text{rcov}(X_i, X_j) = \text{Cov}(\Phi(Z_i), \Phi(Z_j)). \tag{1.6}$$

But (1.6) is precisely the quantity $\Sigma(i,j)$ in (1.5). (The rank covariance and product-moment covariance of uniform marginals are identical.) Therefore, given

a desired rank covariance matrix Σ, we simply compute the empirical covariance matrix $\Sigma_Z (= \Lambda_Z)$ via (1.5) and use this within the NORTA procedure.

Observe that if the random vector in the Li and Hammond example (given above) exists, then it is also an example showing that there are feasible rank covariance matrices for a given set of marginals that cannot be matched using a NORTA procedure.

In the case where F_i (say) is not continuous, (1.6) no longer holds. Therefore, the analytical expression (1.5) cannot be used. However, one could use a numerical search procedure as in Cario and Nelson (1997) to identify the covariance $\Sigma_Z(i, j)$ that yields the required rank covariance $\text{rcov}(X_i, X_j)$. This follows since the rank covariance between X_i and X_j is a nondecreasing continuous function of the covariance between Z_i and Z_j. The nondecreasing property follows immediately from the proof of Theorem 1 in Cario and Nelson (1997), and the fact that the function $F_i(F_i^{-1}(\Phi(\cdot)))$ is nondecreasing. Continuity follows from Theorem 2 of Cario and Nelson.

Thus, one might potentially run into a problem while using NORTA with either product-moment or rank correlation matrices. The existence of this problem had not been formally established since the postulated counterexamples were not shown to exist. Our first contribution was in establishing this.

1.3 A Roadmap

In Chapter 2, we rigorously establish that examples like those given by Li and Hammond and Lurie and Goldberg exist that show that NORTA fails for some feasible covariance matrices. We do this by devising a computational procedure based on linear programming that establishes whether or not a given covariance

matrix is feasible for a given set of marginals. If so, the method explicitly provides a joint distribution with the required properties. This then gives the required construction. To the best of our knowledge, this is the first example of such a procedure.

Our initial efforts were geared towards establishing that the 3−dimensional uniform random vector that forms the Li and Hammond example exists. We were not aware at that time that Joe (1997, on p. 137) had shown (and Kurowicka and Cooke (2001) had independently corroborated) that any correlation matrix is automatically feasible for the special case of a 3−dimensional random vector with uniform marginals, and thus the Li and Hammond example was valid. The power of our approach is however broader in that it ascertains feasibility of any $d-$dimensional random vector (with $d \geq 2$) with any arbitrary set of marginals. Thus, for instance, the Lurie and Goldberg example can also be established to exist by this approach.

Chapter 3 investigates how the feasibility problem behaves as the dimension of the random vector increases. Specifically, we determine that the probability that NORTA fails to work for (uniform marginals and) a correlation matrix chosen *uniformly* from the set of all correlation matrices increases with dimension, and NORTA almost never works in dimensions 20 and higher. This clearly is of concern when continuous marginal distributions and rank correlations are specified. This also has important implications for NORTA when used with product-moment correlations and non-uniform marginals, since uniform random vectors are generated as an intermediary in the NORTA transformation (1.1).

An important part of our analysis in Chapter 3 is the development of a method for sampling uniformly from the set of all correlation matrices of a given dimen-

sion. As discussed in Chapter 2, this set can be viewed as a closed, bounded and full-dimensional subset of a real vector space, and we sample uniformly over this bounded set. We choose to call our sampling method the *onion* method for reasons that will be clear once the working of the method is explained. This sampling technique can be employed more generally in situations where positive semidefinite matrices with fixed diagonal entries are to be sampled uniformly.

The philosophy of specifying marginals and covariances to model dependent random variates is clearly an approximate one, since the joint distribution is not completely specified. Therefore, one should be willing to live with reasonable (this is, of course, a relative term) discrepancies in the covariance matrix from that desired. In Chapter 4 we study certain modifications to the initialization phase of NORTA that can achieve the desired marginals and a covariance matrix that approximates the desired one. The results indicate that some of these augmented NORTA procedures can typically get very close to a target covariance matrix, even in very high dimensions. So in high dimensions, while NORTA is (very) unlikely to be able to exactly match a desired covariance matrix, some of these methods allow us to get very close. So, NORTA does perform reasonably in higher dimensions.

NORTA, both in its original and modified forms, can only approximately match feasible covariance matrices in higher dimensions. But the feasibility testing technique introduced in Chapter 2 can construct random vectors (in the form of "chessboard" distributions) for almost any feasible covariance matrix. Now, chessboards turn out to be easy to generate from. This suggests that a chessboard based generation technique should be of substantial interest. Its ability to more closely match certain covariance matrices than NORTA should be seen as a clear advantage of this approach. Moreover, the user retains greater control on the constructed

distribution since various other dependence characteristics can be matched using chessboards (Ghosh and Henderson 2001).

We develop this idea in Chapter 5, where we provide a new method for generating random vectors with specified marginals and covariance matrix based on chessboard distributions. Constructing chessboards requires a nontrivial amount of computational effort during initialization. We provide a strong characterization of the effort needed to construct chessboards. We believe this method will typically be quite competitive with the NORTA method in low to moderate dimensions while exactly matching the required characteristics of the random vectors.

CHAPTER 2

CHESSBOARD DISTRIBUTIONS AND FEASIBILITY

This chapter describes a computational procedure for determining whether a given covariance matrix is feasible or not for a given set of marginal distributions. We have remarked in Section 1.1 that there is no known characterization of the set of feasible covariance matrices for any arbitrarily chosen set of marginals. Thus, short of constructing a joint distribution with these properties, there is no easy method to check feasibility.

Our procedure, which is based on this observation, attempts to find suitable distributions using a linear programming approach. We call the constructed distributions *chessboard* distributions because of their structure; see Section 2.1.

The procedure can also identify specifications that are infeasible. It works for almost all covariance matrices, and the set over which it does not work has been exactly identified; see Section 2.2. To the best of our knowledge, this is the first example of such a procedure.

We first specialize to the uniform $(0, 1]$ marginals case. (Joint distributions with uniform $(0, 1]$ marginals are known as *copulas*. The term was coined in Sklar (1959), and Nelsen (1999) is a useful recent reference.) Note that in this case the rank covariance and product-moment covariance are identical. Hence this procedure can equivalently check whether rank covariance matrices can be matched to arbitrary marginals with continuous distributions. Section 2.2 describes the various properties of this method that help rigorously establish the feasibility, or not, of a given matrix. This, in turn, provides a way to demonstrate that there are feasible covariance matrices that cannot be matched using the NORTA method.

We extend this computational procedure in Section 2.3 to provide a method for

determining whether or not a given Pearson product-moment covariance matrix is feasible for a given set of marginal distributions that are not necessarily uniform. We require that the marginals have densities with respect to Lebesgue measure, but this is just for convenience. Again our procedure either constructs a distribution with the required properties, or proves that such a distribution does not exist. And again, the procedure works for almost all covariance matrices.

2.1 Chessboard Copulas

Chessboard distributions are perhaps closest in nature to the "piecewise-uniform copulae" developed in Mackenzie (1994). Mackenzie attempts to identify a piecewise-uniform copula that matches a given set of rank covariances. He assumes that such copulae exist, and then selects the one with maximum entropy. In contrast, we do not assume this feasibility, and indeed develop the theoretical properties of the approach to check this property.

For notational simplicity we confine our attention to the three-dimensional case. The extension to higher dimensions is straightforward. Our goal is to construct the density of a random vector X with uniform marginals on (0,1] and product-moment covariance matrix $\Sigma = (\Sigma_{ij} : 1 \leq i, j \leq 3)$. (Recall that the rank covariances and product-moment covariances of uniform random variables on $(0, 1]$ are identical.)

The chessboard density we construct has a simple structure. We divide $(0, 1]^3$ into a large grid of rectangular regions (cells) with sides parallel to the coordinate axes. Let $n \geq 1$ be an integral parameter that determines the level of division that is performed. The range $(0, 1]$ of the ith variable is divided into n equal-length sub-intervals by the set of points $y_{i,k} = \frac{k}{n},\ k = 0, \ldots, n$. Denote the cells

as $C(j_1, j_2, j_3)$, indexed by $j_1, j_2, j_3 = 1, \ldots, n$. Thus,

$$C(j_1, j_2, j_3) = \{(x_1, x_2, x_3) : y_{i,j_i-1} < x_j \leq y_{i,j_i},\ i = 1, 2, 3\},$$

for $1 \leq j_1, j_2, j_3 \leq n$. The density f of X is piecewise constant, taking the value

$$n^3 q(j_1, j_2, j_3) \tag{2.1}$$

in the cell $C(j_1, j_2, j_3)$, so that

$$P(X \in C(j_1, j_2, j_3)) = q(j_1, j_2, j_3). \tag{2.2}$$

(We shall occasionally use f^n to denote this chessboard density to emphasize that it is constructed from an $n-$level of discretization.) Note that the density f is such that, conditional on lying in a fixed cell C, each component is conditionally independent of the others, with marginal distributions given by the uniform distribution restricted to the cell. In a sense, the value $q(j_1, j_2, j_3)$ has been "smeared" uniformly over the cell $C(j_1, j_2, j_3)$.

To ensure that the $q(j_1, j_2, j_3)$s define a proper density f with uniform marginals, we require that

$$\begin{aligned}
\sum_{j_2,j_3=1}^{n} q(j_1, j_2, j_3) &= P(X_1 \in (y_{1,j_1-1},\ y_{1,j_1}]) = \frac{1}{n}, \quad \forall j_1 = 1, \ldots, n, \\
\sum_{j_1,j_3=1}^{n} q(j_1, j_2, j_3) &= P(X_2 \in (y_{2,j_2-1},\ y_{2,j_2}]) = \frac{1}{n}, \quad \forall j_2 = 1, \ldots, n, \\
\sum_{j_1,j_2=1}^{n} q(j_1, j_2, j_3) &= P(X_3 \in (y_{3,j_3-1},\ y_{3,j_3}]) = \frac{1}{n}, \quad \forall j_3 = 1, \ldots, n, \\
q(j_1, j_2, j_3) &\geq 0 \quad \forall j_1, j_2, j_3 = 1, \ldots, n.
\end{aligned} \tag{2.3}$$

Assuming (2.3) holds, it then follows that X has the desired uniform marginals and f defines a copula.

Theorem 2.1.1 *If X is distributed as the density f, defined by (2.1), with cell probabilities q satisfying the constraints (2.3), then X has uniform marginals.*

Proof: Let the marginal distribution function of X_1 be denoted by $G_1(\cdot)$. We have to show that $G_1(x) = x$ for $x \in (0, 1]$. For any $x \in (y_{1,i-1},\ y_{1,i}]$, we have that

$$\begin{aligned} G_1(x) &= \sum_{j_1 \le i-1} \sum_{j_2,j_3=1}^{n} q(j_1, j_2, j_3) + \\ &\quad \sum_{j_2,j_3=1}^{n} P(y_{1,i-1} < X_1 \le x | X \in C(i, j_2, j_3)) \cdot q(i, j_2, j_3) \\ &= \frac{i-1}{n} + \sum_{j_2,j_3=1}^{n} \frac{\int_{y_{1,i-1}}^{x} 1dy}{\int_{y_{1,i-1}}^{y_{1,i}} 1dy} \cdot q(i, j_2, j_3) \\ &= \frac{i-1}{n} + \sum_{j_2,j_3=1}^{n} \frac{(x - \frac{i-1}{n})}{1/n} q(i, j_2, j_3) \\ &= \frac{i-1}{n} + x - \frac{i-1}{n} = x \end{aligned}$$

as required. The first equation follows by conditioning on the cell in which the random vector lies. The second holds because of (2.3), and (2.1) which gives that, conditional on X lying in the cell $C(j_1, j_2, j_3)$, the components X_1, X_2 and X_3 are conditionally independent of each other and the marginal distribution of each component is uniform conditional on lying in this cell. A similar result holds for the marginals of X_2 and X_3, and so X has the right marginals. □

Remark 2.1.2 *The name "chessboard" distribution is motivated by the form of (2.1) in a 2 dimensional problem. In this case, the unit square is broken down in n^2 squares, and the density f is constant on each square, with value $n^2 q(j_1, j_2)$.*

There is no need for the cells used in the above construction to be of equal size. Indeed, Theorem 2.1.1 remains true for more general discretizations; see Theorem 2.3.1 in Section 2.3.

Recall that our goal is to match the covariance matrix Σ. We do this using a linear program. If $\Sigma^X_{ij} = \mathrm{Cov}(X_i, X_j)$ gives the covariances of the random vector X with density f, then we want to minimize the distance $r(\Sigma^X, \Sigma)$ between Σ^X and Σ, where

$$r(\Sigma^X, \Sigma) = \sum_{1 \le i < j \le 3} \left| \Sigma^X_{ij} - \Sigma_{ij} \right|.$$

Now, X has uniform marginals so $EX_i = 1/2$ for $i = 1, 2, 3$. Also, by conditioning on the cell containing X we see that

$$\begin{aligned} EX_1X_2 &= \sum_{j_1,j_2,j_3} q(j_1, j_2, j_3) E[X_1X_2 | X \in C(j_1, j_2, j_3)] \\ &= \sum_{j_1,j_2,j_3} q(j_1, j_2, j_3)\ \mu_{1,j_1}\ \mu_{2,j_2}, \end{aligned} \tag{2.4}$$

where

$$\mu_{\ell,i} = E[X_\ell | X \in (y_{\ell,i-1}, y_{\ell,i}]] = \frac{2i-1}{2n}$$

is the conditional mean of X_ℓ given that it lies in the ith subinterval. In (2.4) we used the conditional independence of the components of X given that X lies in one of the cells.

It follows that Σ^X_{12} is a linear function of the $q(j_1, j_2, j_3)$'s, as is Σ^X_{13} and Σ^X_{23}. Using a standard trick in linear programming, we can represent $|\Sigma^X_{12} - \Sigma_{12}|$ and the other terms in $r(\Sigma^X, \Sigma)$ in a linear fashion as follows.

Define Z^+_{ij} and Z^-_{ij} to be the positive and negative parts of the difference $\Sigma^X_{ij} - \Sigma_{ij}$, i.e.,

$$Z^+_{ij} = (\Sigma^X_{ij} - \Sigma_{ij})^+ = \max\{\Sigma^X_{ij} - \Sigma_{ij}, 0\}, \text{ and } (\Sigma^X_{ij} - \Sigma_{ij})^- = -\min\{\Sigma^X_{ij} - \Sigma_{ij}, 0\}.$$

We can now attempt to match Σ^X to Σ using the LP

$$\min \qquad \sum_{i=1}^{2} \sum_{j=i+1}^{3} (Z^+_{ij} + Z^-_{ij}) \tag{2.5}$$

$$\text{subject to} \quad \Sigma^X_{ij} - \Sigma_{ij} = Z^+_{ij} - Z^-_{ij},\ i = 1 \text{ to } 2 \text{ and } j = i+1 \text{ to } 3$$

$$Z^+_{ij} \geq 0, Z^-_{ij} \geq 0, \text{ together with constraints (2.3).}$$

Note that only one of either Z^+_{ij} or Z^-_{ij} can be non-zero since they represent the positive and negative parts respectively of the same variable. This constraint however need not be imposed on the program (2.5) because the convexity of the objective function and other constraints in the program automatically ensures that it is satisfied at an optimal solution.

This LP is always feasible since we can set $q(j_1, j_2, j_3) = n^{-3}$ for all j_1, j_2, j_3. Also, the objective function of the LP is bounded below by 0, so an optimal solution exists.

If the optimal objective value for the LP is 0, then the solution gives a distribution with the desired marginals and covariance structure, i.e., $\Sigma^X = \Sigma$. This provides the desired construction.

Recall that we also want a test that can establish that certain matrices Σ cannot be matched. To this end we develop bounds on the Z^+_{ij} and Z^-_{ij} variables. These additional bounds are important, because if they cannot be satisfied by any feasible solution to the LP then a random vector with the given covariance matrix and uniform marginals does not exist, as discussed further in Section 2.2.

The bounds are developed by assuming that a random vector $\tilde{X}$ with uniform marginals and covariance matrix Σ exists, and modifying the distribution of $\tilde{X}$ to that of a random vector X that has a chessboard distribution. The modification consists of keeping the total mass within each cell constant, but making the conditional distribution within the cell uniform. The distribution of X then gives a feasible solution to the LP (2.5). We can bound the change in the covariances resulting from this operation on the distribution.

Let

$$\tilde{q}(j_1, j_2, j_3) = P(X \in C(j_1, j_2, j_3)) = P(\tilde{X} \in C(j_1, j_2, j_3)).$$

Observe that

$$\begin{aligned} \mathrm{Cov}(X_1, X_2) - \Sigma_{12} &= EX_1X_2 - E\tilde{X}_1\tilde{X}_2 \\ &= \sum_{j_1,j_2,j_3=1}^{n} \Big\{ (\mu_{1,j_1}\mu_{2,j_2} - E[\tilde{X}_1\tilde{X}_2 | \tilde{X} \in C(j_1, j_2, j_3)]) \cdot \\ &\qquad \tilde{q}(j_1, j_2, j_3) \Big\}. \end{aligned} \tag{2.6}$$

But

$$y_{1,j_1-1}\ y_{2,j_2-1} \leq E[\tilde{X}_1\tilde{X}_2 | X \in C(j_1, j_2, j_3)] \leq y_{1,j_1}\ y_{2,j_2}. \tag{2.7}$$

Combining (2.6) with (2.7) we see that

$$\mathrm{Cov}(X_1, X_2) - \Sigma_{12} \leq \sum_{j_1,j_2,j_3=1}^{n} \tilde{q}(j_1, j_2, j_3)(\mu_{1,j_1}\ \mu_{2,j_2} - y_{1,j_1-1}\ y_{2,j_2-1}), \tag{2.8}$$

$$\mathrm{Cov}(X_1, X_2) - \Sigma_{12} \geq \sum_{j_1,j_2,j_3=1}^{n} \tilde{q}(j_1, j_2, j_3)(\mu_{1,j_1}\ \mu_{2,j_2} - y_{1,j_1}\ y_{2,j_2}). \tag{2.9}$$

Equation (2.8) gives an upper bound on Z_{12}^{+}, and (2.9) gives an upper bound on Z_{12}^{-}. Similar bounds may be obtained for the other covariances. After substituting in the explicit expressions for $y_{i,k}$ and $\mu_{i,k}$, these bounds simplify to

$$Z_{ij}^{+} \leq \frac{1}{2n} - \frac{1}{4n^2} \quad \text{and} \quad Z_{ij}^{-} \leq \frac{1}{2n} + \frac{1}{4n^2} \quad 1 \leq i < j \leq 3. \tag{2.10}$$

The optimal solution of the linear program (2.5) has to necessarily satisfy the bounds (2.10) if the desired covariance matrix Σ is feasible for a uniform random vector. This then gives us a way to check the feasibility of covariance matrices, as we shall see in the next section.

2.2 Testing Feasibility

We shall now derive some important properties of the chessboard construction procedure given in the previous section and describe how these help determine

the feasibility of a given matrix Σ. We can (and do) easily state and prove these results for a general dimension d (i.e., not just $d = 3$) without any notational difficulty. Thus, a $d-$dimensional chessboard density is constructed with an expression analogous to (2.1), and we let a $d-$dimensional cell $C(j_1, \ldots, j_d)$ be represented compactly by $C(j; d)$, and similarly $q(j_1, \ldots, j_d)$ by $q(j; d)$.

For the purpose of proving the results in this section, we assume that the linear program (2.5) includes the bounds (2.10), and they are collectively referred to as the *augmented LP*.

Once the LP procedure is augmented with the bounds (2.10), it is no longer guaranteed to be feasible. In fact, Theorem 2.2.1 below establishes that if the augmented LP is infeasible for *any* value of $n \geq 1$, then the covariance matrix Σ is not feasible for uniform marginals. The proof is basically a summary of the above discussion, and is given to help clarify these ideas.

Theorem 2.2.1 *If the augmented LP is infeasible for some $n \geq 1$, then there cannot exist a random vector X with uniform marginals and the desired covariance matrix Σ.*

Proof: Suppose there exists a random vector X with uniform marginals and covariance matrix Σ. Then, as above, we can construct a solution $\tilde{q}$ by discretizing X that satisfies all of the constraints, including the bounds (2.10). Thus the augmented LP is feasible, which is a contradiction. $\square$

In fact, one can prove a converse to Theorem 2.2.1.

Theorem 2.2.2 *If the covariance matrix Σ is not feasible for uniform $(0, 1]$ marginals, then there exists an $n \geq 1$ such that the augmented LP is infeasible.*

Proof: On the contrary, suppose that the augmented LP is feasible for all $n \geq 1$. Let q^n denote an optimal solution to the nth augmented LP, and let μ_n denote the probability measure corresponding to the density resulting from the operation (2.1) applied to q^n. Then each μ_n is the distribution of a random vector with support contained in $(0,1]^3$ with uniform$(0,1]$ marginals. Hence, the sequence $(\mu_n : n \geq 1)$ is tight, and by Theorem 29.3 on p. 392 of Billingsley (1986), it possesses a weakly convergent subsequence $(\mu_{n(k)} : k \geq 1)$, converging to μ say.

Now, μ has uniform $(0,1]$ marginals. This follows from Theorem 29.2, p. 391 of Billingsley (1986) since each $\mu_{n(k)}$ has uniform$(0,1]$ marginals, $\mu_{n(k)} \Rightarrow \mu$ as $k \to \infty$, and the projection map $\pi_j : \mathbb{R}^{d(d-1)/2} \to \mathbb{R}$ that returns the jth coordinate of a vector in $\mathbb{R}^{d(d-1)/2}$ is continuous.

If C^n is the covariance matrix of the distribution μ^n, then

$$\sum_{i=1}^{d-1} \sum_{j=i+1}^{d} |C_{ij}^n - \Sigma_{ij}| \leq \frac{d(d-1)}{4n} + \frac{d(d-1)}{8n^2} \to 0 \tag{2.11}$$

as $n \to \infty$. This follows from the bounds (2.10), and the fact that in any optimal solution, it is not the case that both Z_{ij}^+ and Z_{ij}^- are strictly positive.

Finally, if $X^{n(k)}$ has distribution $\mu_{n(k)}$, then $(X_i^{n(k)} X_j^{n(k)} : k \geq 1)$ is a uniformly bounded sequence of random variables, and therefore uniformly integrable. It immediately follows that the covariance matrix Λ of μ is given by

$$\Lambda = \lim_{k \to \infty} C^{n(k)} = \Sigma.$$

Thus, μ has the required marginals and covariance matrix, which is a contradiction, and the result is proved. □

Combining Theorems 2.2.1 and Theorem 2.2.2, we see that

Theorem 2.2.3 *A covariance matrix is infeasible for uniform marginals if, and only if, the augmented LP is infeasible for some $n \geq 1$.*

Given this very sharp characterization of infeasible covariance matrices, it is natural to ask whether a similar result holds for feasible covariance matrices. We would then have the result that a covariance matrix is feasible for a given set of marginals if and only if there is some finite n such that the optimal objective value of the augmented LP is zero. Unfortunately, this conjecture is false.

Example 1 *Suppose that Z is a $2-$dimensional random vector with uniformly distributed components $Z_1 = Z_2$ on $(0, 1]$, so that the*

$$Cov(Z_1, Z_2) = Var(Z_1) = 1/12.$$

For a bivariate chessboard random vector X of a given size n, the covariance between X_1 and X_2 is maximized by concentrating all mass on the cells (i, i), and so $q(i, i) = n^{-1}$ for $1 \leq i \leq n$. In that case, we have that

$$Cov(X_1, X_2) = \frac{1}{12} - \frac{1}{12n^2}.$$

Therefore, $Cov(X_1, X_2) < 1/12 = Cov(Z_1, Z_2)$ for all finite n.

Notice that the covariance matrix in this example is singular. This counterexample is a special case of the following result.

Theorem 2.2.4 *All chessboard densities have nonsingular covariance matrices.*

Proof: On the contrary, suppose that f is a chessboard density with singular covariance matrix Σ, and let X have density f. Since Σ is singular, there exists a nonzero vector α such that $\Sigma\alpha = 0$. Hence, $\mathrm{Var}(\alpha' X) = \alpha'\Sigma\alpha = 0$, and so $\alpha' X = \alpha' EX$ a.s. Since α is nonzero, we may, by relabelling variables if necessary, write X_1 as a linear function of the other components, say $X_1 = \sum_{k=2}^{d} \beta_k X_k$ a.s.

This equality must also hold conditional on X being in any cell with $q(j;d) > 0$. But the components of X are conditionally independent given that $X \in C(j;d)$ because f is a chessboard density, which is the required contradiction. $\square$

The importance of Theorem 2.2.4 is that if Σ is feasible for the given marginals and singular, then no matter how large n may be, the optimal objective value of the LP will always be > 0, i.e., we cannot exactly match the covariance matrix Σ. However, we can come arbitrarily close, as the following result shows.

Theorem 2.2.5 *Suppose that the covariance matrix Σ is feasible for uniform $(0,1]$ marginals. Then for all $n \geq 1$, the augmented LP is feasible, and if $z(n)$ is the optimal objective value of the nth LP, then $z(n) \to 0$ as $n \to \infty$.*

Proof: Since Σ is feasible for uniform marginals, the augmented LP is feasible for all $n \geq 1$. (This is just the contrapositive of Theorem 2.2.1.) Let q^n denote an optimal solution to the nth LP, and let f^n be the corresponding chessboard density. If C^n is the covariance matrix corresponding to f^n, then the bounds (2.10) imply that

$$z(n) = \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} |C_{ij}^n - \Sigma_{ij}| \leq \frac{d(d-1)}{4n} + \frac{d(d-1)}{8n^2} \to 0$$

as $n \to \infty$. $\square$

Therefore, chessboard densities can come arbitrarily close to any required Σ that is feasible for uniform marginals. In fact, one can prove that chessboard densities can exactly match a (very) slightly restricted class of feasible covariance matrices. To state this result we need some notation.

Any covariance matrix Σ of a d dimensional random vector with uniform$(0,1]$ marginals can be characterized by $d(d-1)/2$ covariances, since the diagonal entries

are determined by the marginals, and the matrix is symmetric. Hence we can, with an abuse of notation, think of Σ as a $(d(d-1)/2)-$dimensional vector in some contexts, and as a $d \times d$ matrix in others.

Let $\Omega \subset [-1/12, 1/12]^{d(d-1)/2}$ denote the space of *feasible* covariance matrices, so that $\Sigma \in \Omega$ implies that there exists a random vector with uniform$(0, 1]$ marginals and covariance matrix Σ. We will show below that Ω is nonempty and convex (this is well-known), but also closed and full-dimensional (this appears to be new). In particular then, any covariance matrix on the boundary of Ω is feasible. We will also show that Σ is contained in the interior of Ω if, and only if, there is some finite n for which the LP has objective value 0. The collective implications of this and our previous results will be discussed after the statement and proof of these results.

Proposition 2.2.6 *The set Ω is nonempty, convex, closed and full-dimensional.*

Proof: If the components of X are independent, then the covariance matrix Σ is diagonal, and so Ω contains the zero vector, and is therefore nonempty.

It is well-known that Ω is convex. For if $\Sigma_1, \Sigma_2 \in \Omega$, then there exist random vectors X, Y with uniform$(0, 1]$ marginals, and covariance matrices Σ_1 and Σ_2 respectively. For $\lambda \in (0, 1)$, let Z be given by X with probability λ, and Y with probability $1 - \lambda$. Then Z has covariance matrix $\lambda\Sigma_1 + (1 - \lambda)\Sigma_2$.

The proof that Ω is closed is virtually identical to that of Theorem 2.2.2 and is omitted.

We use the NORTA method to prove that Ω is full-dimensional. We will show that each of the vectors $\pm e_k/12$ are contained in Ω, where e_k is the vector whose components are all 0 except for a 1 in the kth position, for $k = 1, \ldots, d(d-1)/2$.

The convexity of Ω then ensures that Ω is full-dimensional.

Let Z be a multivariate normal random vector with mean 0 and covariance matrix consisting of 1's on the diagonal, and also in the (i,j)th and (j,i)th position $(i \neq j)$, with the remaining components being 0. That is, Z consists of 2 perfectly correlated standard normal random variables Z_i and Z_j, and $d-2$ independent standard normal random variables. Now let U be the random vector with uniform $(0,1)$ marginals obtained by setting $U_m = \Phi(Z_m)$ for $m = 1, \ldots, d$. Then U_i and U_j are perfectly correlated, and independent of all of the remaining components of U. Thus, U has covariance matrix whose components are all 0 except for the diagonal elements, and the (i,j), and (j,i)th elements, which are equal to $1/12$. Thus, $e_k/12$ lies in Ω, where k corresponds to the position (i,j). A similar argument with perfectly negatively correlated Z_i and Z_j shows that $-e_k/12 \in \Omega$. Since $i \neq j$ were arbitrary, the proof is complete. $\square$

In Theorem 2.2.4 we showed that all chessboard densities have nonsingular covariance matrices. This is almost sufficient to establish that all boundary points of Ω do not have chessboard densities. However, it is certainly conceivable that the boundary of Ω contains nonsingular, as well as singular, covariance matrices. So we strengthen Theorem 2.2.4 with the following result.

Theorem 2.2.7 *If f^n is a chessboard density with covariance matrix Σ, then Σ is contained in the interior of Ω.*

Proof: Let X have density f^n. We will show that we can both increase, and decrease, the covariance between X_1 and X_2 while keeping all other covariances constant. Symmetry then allows us to conclude that the same result holds for X_i and X_j with $i \neq j$. The convexity of Ω then completes the proof.

Let q be the discretization of f^n into its n^3 cells, and let $C(j;d)$ be a cell with $q(j;d) > 0$. Divide the cell $C(j;d)$ into 4 (equal size) sub-cells,

$$\begin{aligned} C_{ab}(j;d) &= \{x \in C(j;d) : \frac{j_1-1}{n} + \frac{a}{2n} < x_1 \leq \frac{j_1-1}{n} + \frac{a+1}{2n}, \\ &\qquad \frac{j_2-1}{n} + \frac{b}{2n} < x_2 \leq \frac{j_2-1}{n} + \frac{b+1}{2n}\}, \end{aligned}$$

for $0 \leq a, b \leq 1$.

Generate a new density g by the usual relation (2.1) in all cells except $C(j;d)$. Within the cell $C(j;d)$, assign a mass of $q(j;d)/2$ to each of the cells $C_{11}(j;d)$, and $C_{22}(j;d)$, and then define the density within these cells by (2.1). In other words, for a point x contained in these two cells, set $g(x) = 2n^d q(j;d)$ and set g to be 0 in the cells $C_{ab}(j;d)$ for $a \neq b$. Then it is straightforward to show that g has uniform marginals, that the $(1,2)$th covariance is strictly increased, and that the other covariances remain unchanged.

A similar argument placing the mass in the cells $C_{ab}(j;d)$ with $a \neq b$ shows that the covariance can be strictly decreased, and so the proof is complete. □

We have thus far shown that if a covariance matrix Σ is not in Ω, then the augmented LP will be infeasible for some $n \geq 1$, and if Σ is on the boundary of Ω, then the LP approach will yield distributions with covariance matrices that arbitrarily closely approximate Σ, but never actually achieve it. Our final result shows that if Σ is contained in the interior of Ω, then there is some $n \geq 1$ for which the optimal objective value of the augmented LP is 0, and so one can exactly match Σ using a chessboard density. Before proving this result, we need the following lemma. This lemma basically states that given a fixed vector x, we can choose certain other vectors arbitrarily close to x, so that x is a convex combination of these "close" vectors, and if we perturb the close vectors slightly, then x is still a

convex combination of the perturbed vectors.

For $x \in \mathbb{R}^m$ and $\epsilon > 0$, let $B(x, \epsilon)$ denote the (open) set of vectors $\{y \in \mathbb{R}^m : \rho(x, y) < \epsilon\}$, where ρ is the L_1 distance

$$\rho(x, y) = \sum_{i=1}^{m} |x_i - y_i|.$$

Lemma 2.2.8 *Let $x \in I\!R^m$, and let $\epsilon > 0$ be arbitrary. There exist $m + 1$ points $x_1, \ldots, x_{m+1} \in B(x, \epsilon)$, and a $\delta > 0$ such that if*

$$\rho(x_i, x_i') < \delta \quad \forall i = 1, \ldots, m + 1,$$

then x may be written as a convex combination of $x_1', \ldots, x_{m+1}'$.

Proof: Basically, one chooses the x_i's to be the vertices of a simplex centered at x. To be precise, let $r > 0$ be a parameter, and set

$$\begin{array}{lcccccccc}
x_1 & = & (& -a_1 & -a_2 & \cdots & -a_{m-1} & -a_m &)' + x \\
x_2 & = & (& a_1 & -a_2 & \cdots & -a_{m-1} & -a_m &)' + x \\
x_3 & = & (& 0 & 2a_2 & \cdots & -a_{m-1} & -a_m &)' + x \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
x_m & = & (& 0 & 0 & \cdots & (m-1)a_{m-1} & -a_m &)' + x \\
x_{m+1} & = & (& 0 & 0 & \cdots & 0 & ma_m &)' + x,
\end{array}$$

where

$$a_i = r\sqrt{\frac{m}{m+1}}\sqrt{\frac{1}{i(i+1)}}.$$

Then (Dantzig 1991) the x_i's define the vertices of an equilateral simplex whose center is x, and whose vertices are a (Euclidean) distance $rm/(m + 1)$ from x. Choose r so that $x_i \in B(x, \epsilon)$ for all i.

Observe that the average of the x_i's is x. In fact, it is easy to show that the $(m+1) \times (m+1)$ matrix B consisting of the x_i's in columns, supplemented with a row of 1's is nonsingular, and so

$$y = B^{-1}x = (m+1)^{-1}(1, 1, \ldots, 1)'.$$

Now, observe that B^{-1} is a continuous function of B, at least in a neighbourhood of B, and so $y = B^{-1}x$ is locally a continuous function of $x_1, \ldots, x_{m+1}$. Hence, there is a $\delta > 0$ such that if $\rho(x_i, x_i') < \delta$ for all $i = 1, \ldots, m+1$, and D consists of the vectors x_i' in columns supplemented with a row of 1's, then $y = D^{-1}x$ consists of all positive components, and the elements of y sum to 1. $\square$

We are now ready to state the final result of this section.

Theorem 2.2.9 *If Σ is contained in the interior of Ω, then there exists an $n \geq 1$ such that the optimal objective value of the augmented LP is 0.*

Proof: Let $m = d(d-1)/2$, and for now, consider Σ as an m-vector. Let $\epsilon > 0$ be such that $B(\Sigma, \epsilon) \subseteq \Omega$, and choose $\Sigma_1, \Sigma_2, \ldots, \Sigma_{m+1} \in B(\Sigma, \epsilon)$ and δ as in Lemma 2.2.8.

Since $\Sigma_i \in \Omega$, from Theorem 2.2.5 there exists an $n(i)$ such that the augmented LP with target covariance matrix Σ_i has optimal objective value smaller than δ, for each $i = 1, \ldots, m+1$. Let $n = n(1)n(2)\cdots n(m+1)$, and let q^i denote a solution to the augmented LP with target matrix Σ_i and discretization level n for $i = 1, \ldots, m+1$. Then the optimal objective value corresponding to q^i is also less than δ. (Note that if $k, n \geq 1$ are integers, then the optimal objective values $z(n)$ and $z(kn)$ satisfy the relationship $z(kn) \leq z(n)$, since the chessboard density obtained from the solution to the nth LP can also be obtained from the (kn)th LP.)

Let Σ'_i denote the covariance matrix corresponding to the chessboard density f^i for the solution q^i, for $i = 1, \ldots, m+1$. Then, by Lemma 2.2.8, there exist nonnegative multipliers $\lambda_1, \lambda_2, \ldots, \lambda_{m+1}$ summing to 1 such that

$$\Sigma = \sum_{i=1}^{m+1} \lambda_i \Sigma'_i. \tag{2.12}$$

If we set

$$f = \sum_{i=1}^{m+1} \lambda_i f^i,$$

then f is also a chessboard density with discretization level n, and from (2.12), its covariance matrix is exactly Σ. $\square$

In summary then, we have shown that if Σ is infeasible for uniform marginals, then the augmented LP will be infeasible for some $n \geq 1$. This includes the case where Σ is singular and infeasible for uniform marginals. Furthermore, we have shown that if Σ is contained in the interior of Ω, then the augmented LP will have optimal objective value 0 for some $n \geq 1$, and so one can construct a chessboard density from the solution to the augmented LP with the required marginals and covariance matrix. So if Σ is not contained in the boundary of Ω, then we have an algorithm for determining, in finite time, whether Σ is feasible for the given marginals or not. One simply solves the augmented LP for $n = 1, 2, 3, \ldots$ until the augmented LP is either infeasible, or has an optimal objective value of 0. In the latter case, we can deliver an explicit construction of the desired distribution.

The case where Σ lies on the boundary of Ω is more problematical. We have shown that in this case, Σ is feasible for uniform marginals, but that a chessboard density *cannot* be constructed with uniform marginals and covariance matrix Σ. Therefore, for such matrices, the algorithm outlined above will not terminate in finite time. However, a chessboard distribution can come arbitrarily close to the

required covariance matrix. Furthermore, in Chapter 5, we shall prove a result on how close a chessboard distribution of discretization size n can come to matrices from the boundary.

Proposition 2.2.6 establishes that the set Ω is a compact set with a non-empty interior, and thus has a non-zero finite Lebesgue measure (in $\mathbb{R}^{d(d-1)/2}$), while its boundary is a zero Lebesgue measure set (this can be proved, for instance, using Theorem 10.2(iv) in Billingsley 1995). Thus, chessboard distributions can model almost any (in a Lebesgue measure sense) feasible covariance matrix from Ω.

An application of the theory developed in this section easily shows that the NORTA method fails to match uniform marginals with all feasible covariances matrices Σ from the interior of Ω. The steps needed to determine this are straightforward. First, we sample the set of all possible covariance matrices for uniform marginals (symmetric positive semidefinite matrices with 1/12 on the diagonal) and check each chosen matrix for feasibility using the LP-based procedure outlined here. For matrices that turn out feasible, we next check if (1.5) yields a positive semidefinite estimate Λ_Z for the covariance matrix of the joint normal random vector that forms the base of the NORTA procedure. Any feasible matrix that fails the second check is an example where NORTA fails. The Li and Hammond example can be validated in this fashion, and one can find other examples in three and higher dimensions.

We note here that though this section assumed that the chessboard construction procedure consists of solving the "augmented LP" for increasing values of the discretization parameter n, in practice the procedure is implemented in a different, but equivalent, way: the original LP (2.5) is solved first and then the optimal solution to this minimization problem is checked to see whether it satisfies the upper

bounds (2.10). We do this for two reasons. First, in its original form, the linear program possesses a special structure (in the two-dimensional case it resembles an assignment problem) and hence is conceivably easier to solve without the bounds (2.10) imposed as constraints. This alternative approach aims to preserves this advantage. Second, this implementation faithfully identifies every infeasible matrix that the "augmented LP"-solving approach would. In addition it also automatically provides a distribution with a covariance matrix approximately close to the desired (but infeasible) one, *with* a guaranteed upper bound (based on (2.10)) on the distance between them.

2.3 Chessboards with General Marginals

In this section we generalize the uniform marginals assumption to consider random vectors with marginal distributions that have densities and finite variance. The requirement that the marginal distributions have densities is again for convenience. It allows us to be less stringent with endpoints of intervals than we would otherwise need to be. The method provided here can determine whether a specified product-moment covariance matrix is feasible for any set of marginals (with densities). The theory developed in the earlier sections of this chapter can be translated to this case in a straightforward manner for the most part.

We shall illustrate this procedure for a $3-$dimensional random vector. The general $d-$dimensional case is virtually identical. Suppose we wish to construct a chessboard distribution for $X = (X_1, X_2, X_3)$ such that each of the marginals has a density f_i and finite variance. Furthermore, let F_i represent the corresponding ith marginal distribution function and $\mathrm{dom}(F_i)$ its domain. Let

$$\{y_{i,j_i} : i = 1, 2, 3,\ j_i = 0, \ldots, n\}$$

be the set of points that divide the range of the ith variable $\mathrm{dom}(F_i)$ into n subintervals. The range could be infinite, in which case we allow the corresponding endpoint to be $\pm\infty$. Let M_i^- and M_i^+ represent the leftmost and rightmost finite points respectively. Thus, if X_i were exponentially distributed, $M_i^- = y_{i,0} = 0, y_{i,n} = \infty$ and $M_i^+ = y_{i,n-1}$. The range can be divided in any manner, as long as the points satisfy two conditions. Firstly, the internal mesh becomes dense, i.e.,

$$\lim_{n\to\infty} \sup_{i,j_i} |y^n_{i,j_i} - y^n_{i,j_i-1}| = 0, \tag{2.13}$$

where the sup excludes infinite endpoints. Secondly, $\min_i |M_i^{\pm}| \to \infty$ as $n \to \infty$. These conditions are satisfied, for example, if we take the points to be of the form

$$-\infty, -\sqrt{n}, -\sqrt{n} + \frac{2}{\sqrt{n}}, -\sqrt{n} + \frac{4}{\sqrt{n}}, \ldots, \sqrt{n}, +\infty.$$

These conditions avoid complications that might arise if the intervals are simply assumed to be equi-probable. For example, if one of the marginal distribution functions has a "flat patch", i.e., there is some i and $x < y$ such that $F_i(x) = F_i(y) \in (0, 1)$, then a technique we shall propose later in this section to derive bounds similar to (2.10) fails to work for chessboard densities that have intervals of equal probability.

For $1 \le j_1, j_2, j_3 \le n$ define the cell $C(j_1, j_2, j_3)$ to be the (j_1, j_2, j_3)th rectangular region

$$\{x = (x_1, x_2, x_3) : y_{i,j_i-1} < x_i \le y_{i,j_i}\, i = 1, 2, 3\} \cap \Re^3.$$

Define

$$q(j_1, j_2, j_3) = P(X \in C(j_1, j_2, j_3))$$

to be the probability that the constructed random vector appears in the (j_1, j_2, j_3)th cell. The chessboard distribution is defined so that within each cell the components of X are independent, and are distributed according to the desired marginals

f_1, f_2, f_3 restricted to the cell $C(j_1, j_2, j_3)$. Let $p_{i,k} = P(X_i \in (y_{i,k-1}, y_{i,k}])$ be the probability that the ith marginal random variable lies in the kth subinterval. The density $f(x)$ of X evaluated at $x \in C(j_1, j_2, j_3)$ is then given by

$$q(j_1, j_2, j_3)\frac{f_1(x_1)}{p_{1,j_1}}\frac{f_2(x_2)}{p_{2,j_2}}\frac{f_3(x_3)}{p_{3,j_3}}. \tag{2.14}$$

To be consistent with the given marginals, the $q(j_1, j_2, j_3)$ values must satisfy the constraints

$$\begin{aligned}
\sum_{j_2,j_3=1}^{n} q(j_1, j_2, j_3) &= p_{1,j_1}, \quad j_1 = 1, \ldots, n \\
\sum_{j_1,j_3=1}^{n} q(j_1, j_2, j_3) &= p_{2,j_2}, \quad j_2 = 1, \ldots, n \\
\sum_{j_1,j_2=1}^{n} q(j_1, j_2, j_3) &= p_{3,j_3}, \quad j_3 = 1, \ldots, n \\
q(j_1, j_2, j_3) \geq 0 & \qquad 1 \leq j_1, j_2, j_3 \leq n.
\end{aligned} \tag{2.15}$$

An argument along the lines of the proof of Theorem 2.1.1 shows that if a chessboard distribution is constructed via (2.14) then it has the correct marginals. We shall give the proof here because it is useful in understanding the nature of these chessboard distributions.

Theorem 2.3.1 *If q satisfies the constraints (2.15), and X is constructed with density f as given in (2.14), then X has the desired marginals.*

Proof: Let the marginal distribution function of X_1 be denoted by $G_1(\cdot)$. We have to show that $G_1(x) = F_1(x)$ for $x \in \mathrm{dom}(F_1)$. For any $x \in (y_{1,i-1},\ y_{1,i}]$, we have that

$$G_1(x) = \sum_{j_1 \leq i-1} \sum_{j_2,j_3=1}^{n} q(j_1, j_2, j_3) +$$

$$\begin{aligned}
&\sum_{j_2,j_3=1}^{n} P(y_{1,i-1} < X_1 \leq x | X \in C(i,j_2,j_3)) \cdot q(i,j_2,j_3) \\
= &\sum_{j_1 \leq i-1} p_{1,j_1} + \sum_{j_2,j_3=1}^{n} \frac{\int_{y_{1,i-1}}^{x} f_1(y)dy}{\int_{y_{1,i-1}}^{y_{1,i}} f_1(y)dy} \cdot q(i,j_2,j_3) \\
= &\sum_{j_1 \leq i-1} p_{1,j_1} + \sum_{j_2,j_3=1}^{n} \frac{\int_{y_{1,i-1}}^{x} f_1(y)dy}{p_{1,i}} q(i,j_2,j_3) \\
= &\sum_{j_1 \leq i-1} p_{1,j_1} + \int_{y_{1,i-1}}^{x} f_1(y)dy = F_1(x)
\end{aligned}$$

as required. The first equation follows by conditioning on the cell in which the random vector lies. The second holds because of (2.15), and (2.14) which gives that, conditional on X lying in the cell $C(j_1, j_2, j_3)$, X_1 is independent of the components X_2 and X_3 and distributed according to F_1 conditional on lying in the j_1th sub-interval corresponding to this cell. A similar result holds for the marginals of X_2 and X_3, and so X has the right marginals. □

Our goal is to match the product-moment covariance matrix of the chessboard density (2.14), denoted by Σ^X, to the desired product-moment covariance matrix Σ. The diagonal elements of the covariance matrices Σ and Σ^X are determined by the marginal distributions, and so our objective is to minimize the difference $r(\Sigma, \Sigma_X)$, where once again we have

$$r(\Sigma^X, \Sigma) = \sum_{1 \leq i < j \leq 3} \left| \Sigma_{ij}^X - \Sigma_{ij} \right|.$$

Reasoning similar to the uniform marginals case, we find that for $i \neq k$,

$$\Sigma_X(i,k) = \sum_{j_1,j_2,j_3} \gamma_i(j_i)\gamma_k(j_k)q(j_1,j_2,j_3) - EX_iEX_k, \tag{2.16}$$

where, for $1 \leq i \leq 3$ and $1 \leq m \leq n$,

$$\gamma_i(m) = E[X_i | X_i \in (y_{i,m-1}, y_{im}]]$$

is the conditional mean of X_i given that it lies in the mth subinterval (which is determined by its marginal density f_i.)

Thus, we have again expressed $\Sigma_X(i,k)$ as a linear function of q. To match the desired covariance matrix Σ, we solve the linear program

$$\begin{aligned} \min \quad & \sum_{i=1}^{2} \sum_{j=i+1}^{3} (Z_{ij}^{+} + Z_{ij}^{-}) \qquad (2.17) \\ \text{subject to} \quad & \Sigma_{ij}^{X} - \Sigma_{ij} = Z_{ij}^{+} - Z_{ij}^{-}, \ i = 1 \text{ to } 2 \text{ and } j = i+1 \text{ to } 3 \\ & Z_{ij}^{+} \geq 0, Z_{ij}^{-} \geq 0, \text{ together with constraints (2.15).} \end{aligned}$$

where the variables $Z_{ij}^{\pm}$ are again defined as they were in Section 2.1.

The properties derived in Section 2.2 for the chessboard copula matching technique hinge on the crucial fact that one is able to obtain bounds on the objective function of the LPs (2.5) that vanish as the discretization parameter $n \to \infty$. These results can be proved for general marginals with densities and finite variances if a similar vanishing bound can be identified. The technique employed in the uniform marginal distribution case (Section 2.1) does not carry over to this case since it depends on the support being finite. We present a technique to derive such bounds under the assumption that the marginal distributions have finite variance. We can assume this at no additional cost since it is needed to ensure the existence of product-moment covariances. This then extends the power of the chessboard-based covariance matching technique to the general marginals case.

We shall restrict our description to the case where the support of the marginal densities are all $(-\infty, +\infty)$. We assume this only to keep the explanation simple; the method itself is applicable generally. We take $|M_i^{-}| = |M_i^{+}| \stackrel{\triangle}{=} M_i$. Suppose there exists a random vector $\tilde{X}$ with the prescribed covariance matrix Σ. We again redistribute the probability mass of its distribution within cells (thus keeping the

cell probability masses constant) so that the conditional density given a cell is one of independent random variables with the desired marginals. Let X denote a random vector with the redistributed probability mass. We provide a bound on the change in covariance due to this redistribution.

Let $\mathcal{S}$ represent the the part of the support of X (and $\tilde{X}$) bounded by the rectangle $[-M_1, M_1] \times [-M_2, M_2]$. As per our notation, $\mathcal{S}$ is given by the collection of cells $C(j_1, j_2, j_3)$ with indices j_1 and j_2 ranging over $2, \ldots, n-1$. The absolute change in covariance due to the redistribution operation, $|EX_1X_2 - E\tilde{X}_1\tilde{X}_2|$, can be split into two terms:

$$\begin{aligned} & |EX_1X_2 - E\tilde{X}_1\tilde{X}_2| \\ & \leq \left|E[X_1X_2 I\{(X_1, X_2) \in \mathcal{S}\}] - E[\tilde{X}_1\tilde{X}_2 I\{(\tilde{X}_1, \tilde{X}_2) \in \mathcal{S}\}]\right| + \\ & \quad \left|E[X_1X_2 I\{(X_1, X_2) \in \mathcal{S}^c\}] - E[\tilde{X}_1\tilde{X}_2 I\{(\tilde{X}_1, \tilde{X}_2) \in \mathcal{S}^c\}]\right| \end{aligned} \tag{2.18}$$

where $I\{A\}$ is the indicator function taking the value 1 on the event and 0 otherwise.

The first term in (2.18) represents the change due to the redistribution operation in a compact part $\mathcal{S}$ of the support. This can thus be bounded in a fashion similar to that used in Section 2.1. Assumption (2.13) ensures that this bound goes to 0 as $n \to \infty$.

Now, consider the second term in (2.18), which includes cells of infinite length. This term can be bounded by

$$|E[X_1X_2 I\{(X_1, X_2) \in \mathcal{S}^c\}]| + \left|E[\tilde{X}_1\tilde{X}_2 I\{(\tilde{X}_1, \tilde{X}_2) \in \mathcal{S}^c\}]\right|. \tag{2.19}$$

The first of these terms can be rewritten as

$$E[X_1X_2 I\{(X_1, X_2) \in \mathcal{S}^c\}] \;=\; E[X_1X_2 \; I\{ \; \{|X_1| > M_1\} \cup \{|X_2| > M_2\} \; \}],$$

$$\leq \quad E[X_1 I\{|X_1| > M_1\} X_2] + E[X_1 X_2 I\{|X_2| > M_2\}]. \tag{2.20}$$

Since the variances of all components of X are finite, the first expression in (2.20) can be bounded, with help from the Cauchy-Schwarz inequality, as

$$\begin{aligned} |E[X_1 I\{|X_1| > M_1\} X_2]| &\leq E[\ |\ X_1 I\{|X_1| > M_1\}\ |^2\ ]^{1/2} E[X_2^2]^{1/2} \\ &= E[X_1^2\ I\{|X_1| > M_1\}\ ]^{1/2} E[X_2^2]^{1/2}. \end{aligned} \tag{2.21}$$

The second term in (2.21) is a constant that depends on the marginal distribution of X_2. As for the first term, observe that $(X_1^2\ I\{|X_1| > M_1\}) \to 0$ almost surely as $M_1 \to \infty$ since X_1 has finite second moments. Moreover, it is also bounded by X_1^2, which has a finite mean EX_1^2, and so the Dominated Convergence Theorem (Billingsley 1995, Theorem 16.4), implies that $E[X_1^2\ I\{|X_1| > M_1\}]$ converges to 0 as $M_1 \to \infty$. The same holds for the second term in (2.20).

Putting together the bound in (2.21) with equations (2.20), (2.19) and finally (2.18) then gives us a bound on the objective function of the LP (2.17). These bounds are similar to those obtained for the objective function of the chessboard copula LP in Section 2.1 in the sense that they converge to 0 as $n \to \infty$. This is all that was required to help prove the results in Section 2.2. For the uniform marginals case, we were also able to simplify the expressions for the bounding value to determine that it was of the order of magnitude n^{-1}. The corresponding value here depends intimately on the tail behaviour of the marginal distributions, which determine the rate at which $E[X_1^2\ 1\{|X_1| > M_1\}] \to 0$.

We will now show how these bounds help us prove results analogous to those in Section 2.2 for this case. We shall state all the results, but will omit the proofs that are virtually identical to the uniform marginals case. Again, following Section 2.2,

we shall present the following results for the case of a general dimension d.

Theorem 2.3.2 *A covariance matrix is infeasible for the given marginals if, and only if, the chessboard LP (2.17) augmented with the appropriate bounds is infeasible for some $n \geq 1$.*

Proof: The "if" part of the proof follows from the preceding discussion: if a random vector X exists with the covariance matrix, then we can obtain a feasible solution $\tilde{q}$ for the augmented LP from its distribution. This makes the LP feasible, which is a contradiction.

For the "only if" part, let us assume the contrary, that is, suppose the augmented LP is feasible for all $n \geq 1$. Let q^n denote an optimal solution to the nth augmented LP, and let μ_n denote the probability measure corresponding to the density resulting from the operation (2.14) applied to q^n. Then each μ_n is the distribution of a random vector with marginals possessing finite second moments. Hence, the sequence $(\mu_n : n \geq 1)$ is tight, and by Theorem 29.3 on p. 392 of Billingsley (1986), it possesses a weakly convergent subsequence $(\mu_{n(k)} : k \geq 1)$, converging to μ say.

Now, μ has the right marginals. This follows from the Mapping theorem (Theorem 29.2, p. 391 of Billingsley 1986) since each $\mu_{n(k)}$ has the marginals we desire, $\mu_{n(k)} \Rightarrow \mu$ as $k \to \infty$, and the projection map $\pi_j : \mathbb{R}^{d(d-1)/2} \to \mathbb{R}$ that returns the jth coordinate of a vector in $\mathbb{R}^{d(d-1)/2}$ is continuous.

If C^n is the covariance matrix of the distribution q^n, then

$$\sum_{i=1}^{d-1} \sum_{j=i+1}^{d} |C_{ij}^n - \Sigma_{ij}| \to 0$$

as $n \to \infty$. This follows from the bounds derived in the prelude to this theorem.

Finally, if $X^{n(k)}$ has distribution $\mu_{n(k)}$, then $(X_i^{n(k)} X_j^{n(k)} : k \geq 1)$ is uniformly integrable. To see this, note that (let $m = n(k)$)

$$\begin{aligned} &\sup_m E\left[|X_1^m X_2^m I\left\{|X_1^m X_2^m| > K\right\}|\right] \\ &\leq \sup_m E\left[|X_1^m X_2^m| I\left\{|X_1^m| > \sqrt{K}\right\} + |X_1^m X_2^m| I\left\{|X_2^m| > \sqrt{K}\right\}\right] \end{aligned}$$

This holds because, for any two positive numbers x and y, $\{xy > K\} \subseteq \{\max\{x, y\} > \sqrt{K}\} \subseteq \{x > \sqrt{K}\} \cup \{y > \sqrt{K}\}$. An argument along the lines of those given in the context of the bounding expression in (2.21) shows that the expression on the right side converges to 0 as $K \to \infty$. This establishes the uniform integrablity result.

It immediately follows (Theorem 25.12 in Billingsley 1995) that the covariance matrix Λ of μ is given by

$$\Lambda = \lim_{k \to \infty} C^{n(k)} = \Sigma.$$

Thus, μ has the required marginals and covariance matrix, which is a contradiction, and the result is proved. □

The following result follows from the availability of the bounds:

Theorem 2.3.3 *Suppose that Σ is feasible. Then for all $\epsilon > 0$, there exists a chessboard distribution with covariance matrix Λ with the property that $r(\Sigma, \Lambda) < \epsilon$.*

Thus, chessboards can come arbitrarily close to any required Σ that is feasible for the desired marginals. One can also prove that chessboards can match almost any feasible product-moment covariance matrix. As before, we define Ω to be the set of feasible covariance matrices, viewed as a subset of $(d(d-1)/2)-$dimensional Euclidean space. We immediately have that

Proposition 2.3.4 *The set Ω is nonempty, convex, closed and full-dimensional.*

The proof is virtually identical to Proposition 2.2.6; one just needs to substitute the maximum and minimum covariance values achievable between any two components in the NORTA part of the proof.

The following result is the final property of importance that we shall need. It carries over almost exactly from the uniform marginals case. Again, one only needs to pay attention to the fact that the marginal distributions are potentially different here when reworking the proofs in Theorem 2.2.7 and Theorem 2.2.9.

Theorem 2.3.5 *There is a chessboard distribution of the form (2.14) with the desired marginals and covariance matrix* Σ *if, and only if,* $\Sigma \in \Omega^{\circ}$.

We have now constructed a procedure that can determine whether a given set of marginal distributions and product-moment covariance matrix is infeasible, or not. And if feasible, the procedure returns a joint distribution that has the given properties. This procedure works for almost all covariance matrices, in the sense that it does not terminate in finite time for covariance matrices on the boundary of Ω. For such covariance matrices, the procedure gets arbitrarily close, but never exactly matches the desired covariance matrix.

This procedure, as described here, is not suited for adoption as a practical method to match covariance matrices with general marginals (with densities). It has been provided here more in the spirit of a "proof-of-concept" to establish that such a procedure indeed exists. Significant wrinkles remain to be ironed out before this can be successfully implemented. For instance, it is not clear what the best choice of the set of dividing points $\{y_{i,j_i}\}$ should be, and how it interacts with the rate at which $E[X_i^2\ 1\{|X_i| > M_i\}] \to 0$, which is determined by the tail behaviour of the ith marginal distribution. We plan to pursue these issues further.

The procedure can conceivably be extended to the case where the marginal distributions do not necessarily have densities with respect to the Lebesgue measure. While we have unfortunately not been able to devote much attention to this, we conjecture that this should be a relatively straightforward exercise.

CHAPTER 3

NORTA METHOD IN HIGHER DIMENSIONS

The NORTA method for modeling partially specified random vectors faces a limitation in matching a given set of marginals with a (product-moment or rank) correlation matrix. Recall that the problem arises because the correlation matrix of the base joint normal random vector is constructed in a way that does not necessarily ensure that it is positive semidefinite. It might indeed turn out to be indefinite, in which case it cannot be a valid covariance matrix for a joint normal distribution, and NORTA will fail. Chapter 2 discusses how valid examples that demonstrate this possibility can be identified.

The focus of this chapter will be on how this flaw in NORTA affects its applicability as the dimension of the desired random vector increases. Section 3.1 reviews some characteristics of this problem when matching 3−dimensional random vectors. Motivated by some observations in this section, Section 3.2 proposes a simulation experiment to test NORTA in higher dimensions. The results are discussed in Section 3.4.

The methodological meat of this chapter lies in Section 3.3 where we develop a new method for sampling uniformly from the set of all correlation matrices of a given dimension. As discussed later, this set can be viewed as a closed, bounded and full-dimensional subset of a real vector space, and we sample uniformly over this bounded set. We choose to call our sampling method the *onion* method for reasons that will be clear once the working of the method is explained. The method is easily generalized to generate from the set of symmetric positive definite matrices with arbitrary (fixed) positive diagonal entries. Thus a possible use of an appropriately modified version of the sampling scheme might be to study the

performance of algorithms that operate on such matrices.

A note on the terminology in this chapter: the earlier, and subsequent, chapters discuss matching a *covariance* matrix using NORTA. We shall refer to the equivalent problem of matching *correlation* matrices in this chapter. Specifying a covariance matrix is equivalent to specifying a correlation matrix since the marginal distributions are also being specified simultaneously. The problem we shall investigate in this chapter affects NORTA's performance with any marginal distributions, and using correlation matrices helps us cleanly define it independent of the particular desired marginal distributions.

3.1 The Story in 3 Dimensions

The NORTA method models a random vector X by matching its marginal distributions F_i with a desired correlation matrix Σ. Recall that X is obtained from a joint normal random vector Z via the transformation (1.2). This transformation procedure can occasionally fail to match a feasible marginal and correlation specification when the correlation matrix Λ_Z estimated for the base joint normal random vector turns out to be indefinite. Each component of Λ_Z is estimated independently from the corresponding component of the target correlation matrix Σ using the relation (1.4). The problem then happens because the joint normal correlation matrix Λ_Z is estimated in a way that does not necessarily ensure that it is positive semidefinite.

Li and Hammond (1975) first proposed an example of a $3-$dimensional random vector with uniform marginals and a specifically chosen correlation matrix to demonstrate this flaw in NORTA. For the special case of uniform marginals, the

relation (1.4) is known to be the closed form (1.5), which we reproduce here:

$$\Lambda_Z(i,j) = 2\sin[\frac{\pi}{6}\Sigma(i,j)]. \tag{3.1}$$

Hence, it is easy to check whether any feasible correlation matrix (for a uniform random vector) fails with NORTA. Li and Hammond had not however established that a uniform random vector with their proposed correlation matrix exists, i.e., is feasible. Joe (1997, (on p. 137)) asserts that for 3-dimensional random vectors with uniform marginals, as in the Li and Hammond example, any correlation matrix is feasible. Hence the example exists. Furthermore, Chapter 2 develops a computational procedure that can determine, for *almost* any (in a Lebesgue measure sense) given correlation matrix, whether it is feasible for a given set of marginal distributions (assumed to be continuous and have bounded support) or not. Applying this procedure to the Li and Hammond example gives a construction of the random vector.

Let us call feasible correlation matrices that cannot be matched using the NORTA method *NORTA defective* matrices. It is interesting to question why the NORTA defective matrices are, in fact, NORTA defective, and how likely is this failure.

We will look to answer this question in the case of a $3-$dimensional uniform random vector. This case is particularly easy to study now given the result in Joe (1997). We simply choose samples of correlation matrices (symmetric, positive semidefinite matrices with ones on the diagonal) and check them for NORTA feasibility. We generate a number of such feasible matrices for three-dimensional uniform random vectors that are NORTA defective. The numerical results suggest a structure to the failure of NORTA. To explain this observation more carefully we need to reuse some notation introduced in Chapter 2.

With an abuse of notation, we can view a $d \times d$ correlation matrix as an element of a $d(d-1)/2$ dimensional vector space, since there are $d(d-1)/2$ independent elements above the diagonal, the matrix is symmetric, and the diagonal elements are equal to 1. Let Ω denote the set of correlation matrices, i.e., symmetric positive semidefinite matrices with unit diagonal elements. We view this set as a subset of the real vector space $\mathbb{R}^{d(d-1)/2}$. This set has been established to be nonempty, convex, closed and full-dimensional; see Vandenberghe and Boyd (1996).

Note that Chapter 2 defines Ω to be the set of *feasible* correlation matrices for $d-$dimensional random vectors with a fixed set of arbitrary marginals. Thus, Ω as defined there depends on whether we consider product-moment or rank correlation. Our definition of Ω here is independent of this distinction. For the special case of $3-$dimensional uniform random vectors, these two definitions coincide.

A 3×3 correlation matrix Σ is represented by the vector $(\Sigma_{12}, \Sigma_{13}, \Sigma_{23})$ in Ω, and Ω is a proper subset of the cube $[-1, 1]^3$. We examined all correlation matrices with off-diagonal components in the set {-1.0, -0.9, ..., -0.1, 0, 0.1, ..., 0.9, 1.0} for NORTA defectiveness. Taking account of the symmetry of the correlations in terms of its components (for example $(0.5, -0.5, 0.5)$ and $(-0.5, 0.5, 0.5)$ represent the same uniform vector), we identified 31 such matrices.

Table 3.1 tabulates these NORTA defective matrices. An inspection of the NORTA defective matrices in Table 3.1 shows that the determinants and the smallest eigenvalues of all these matrices are quite close to zero in magnitude. This means that the corresponding points in Ω lie either on the boundary or in its close proximity, i.e., the NORTA defective matrices lie close to, or on, the boundary of the set of achievable correlations (Vandenberghe and Boyd (1996) establish that the boundary of Ω consists of correlation matrices where the (continuous)

Table 3.1: Matching Chessboards to Correlations in 3 dimensions

Correlations Σ_{12}	Σ_{13}	Σ_{23}	Discretization Level n	Determinant $\lvert\Sigma\rvert$	Smallest Eigenvalue of Σ
-0.9	-0.6	0.2	18	0.0060	0.0034
-0.9	-0.5	0.1	11	0.0200	0.0105
-0.9	-0.2	0.6	18	0.0060	0.0034
-0.9	-0.1	0.5	11	0.0200	0.0105
-0.8	-0.8	0.3	12	0.0140	0.0087
-0.8	-0.5	-0.1	11	0.0200	0.0097
-0.8	-0.4	-0.2	10	0.0320	0.0151
-0.8	-0.3	-0.3	10	0.0360	0.0169
-0.8	-0.3	0.8	12	0.0140	0.0087
-0.8	0.1	0.5	11	0.0200	0.0097
-0.8	0.2	0.4	10	0.0320	0.0151
-0.8	0.3	0.3	10	0.0360	0.0169
-0.7	-0.7	0.0	12	0.0200	0.0101
-0.7	0.0	0.7	12	0.0200	0.0101
-0.6	-0.2	0.9	18	0.0060	0.0034
-0.5	-0.1	0.9	11	0.0200	0.0105
-0.5	0.1	0.8	11	0.0200	0.0097
-0.4	0.2	0.8	10	0.0320	0.0151
-0.3	0.3	0.8	10	0.0360	0.0169
-0.2	0.4	0.8	10	0.0320	0.0151
-0.1	0.5	0.8	11	0.0200	0.0097
0.0	0.7	0.7	12	0.0200	0.0101
0.1	0.5	0.9	11	0.0200	0.0105
0.2	0.6	0.9	18	0.0060	0.0034
0.3	0.8	0.8	12	0.0140	0.0087
-0.8	-0.6	0.0	–	0	0
-0.8	0.0	0.6	–	0	0
-0.6	0.0	0.8	–	0	0
-0.5	-0.5	-0.5	–	0	0
-0.5	0.5	0.5	–	0	0
0.0	0.6	0.8	–	0	0

determinant function takes a value zero.)

Table 3.1 also provides the *level of discretization* n required by the computational procedure of Section 2.1 to construct a chessboard distribution for these matrices. For 25 of the 31 matrices that are not on Ω's boundary, chessboard distributions that exactly match Σ were constructed with a discretization level $n \leq 18$. Chessboard distributions could not exactly match Σ in the remaining 6 cases, but this is to be expected from Theorem 2.2.4 (it establishes that correlation matrices of chessboard distributions are non-singular) since in all of these cases Σ was singular.

These results seem to suggest that NORTA defective Σ matrices are those that are near-singular, and perhaps are then relatively rare. However, Lurie and Goldberg (1998) believe that singular and near-singular correlation matrices actually represent a common situation in cost analysis for example. This is because correlations between cost elements are typically estimated from unbalanced data sets. This is likely to lead to indefinite target correlation matrices, so that any small adjustment to them is almost certainly going to result in an adjusted target matrix that is singular, or very nearly so.

We can expect to find the same pattern for the case for more general distributions. We reason heuristically as follows. The set Ω also represents the set of feasible correlation matrices for a standard joint normal random vector (any positive semidefinite matrix with unit diagonal entries is feasible for joint normals.) The NORTA method transforms each element of this set back into another element. The fact that NORTA fails therefore implies the set Ω is mapped into a subset of itself. Assuming the transformation (1.4) is continuous, which indeed it is in great generality (see Cario and Nelson 1997), and not "too nonlinear", it is

reasonable to expect that any elements of Ω that are not covered by the transformation will be those that are close to the boundary. Indeed, if one plots the set of NORTA defective vectors using a three dimensional plotting package, this is exactly what we see. We note here that the indefinite matrices Λ_Z of these 31 cases were observed to lie very close to (and outside) the set Ω.

3.2 In Higher Dimensions

In three dimensions, NORTA appears to fail most often when the correlation matrix is close to the boundary of the set Ω for 3-dimensional random vectors. Now, in a sense that can be made precise, "most" points in certain sets (compact sets with non-empty interiors) in high dimensions lie close to the boundary. For example, consider the interior of the unit hypercube $[-\frac{1}{2}, \frac{1}{2}]^m$ in $\mathbb{R}^m$ represented by the hypercube $[-\frac{1-\epsilon}{2}, \frac{1-\epsilon}{2}]^m$, where $\epsilon \in (0, 1)$. The ratio of the volume of the interior to that of the whole set is $(1-\epsilon)^m$, which decreases rapidly to 0 as m increases.

This suggests that matrices within the set Ω may be increasingly likely to fail with NORTA as the dimension of the problem increases, so that the problem that a user faces in using NORTA may become increasingly acute as the dimension increases.

Continuing the example considered in Section 3.1, let us consider this possibility in the context of generating samples of a uniform random vector. This case has special significance to the NORTA method because, by construction, the method has to generate a uniform random vector $(\Phi(Z_1), \ldots, \Phi(Z_d))$ as an intermediary step. Furthermore, the rank correlation matrix of a NORTA-generated vector with continuous marginal distributions coincides with the product-moment correlation

matrix for the intermediate uniform random vector.

We shall test whether randomly chosen correlation matrices are NORTA defective. In particular, we shall determine the probability of NORTA defectiveness of randomly chosen correlation matrices for uniform marginals (the probability space we use is explained shortly). This special case has the advantage that the explicit form of the function c_{ij} (3.1) helps us easily determine NORTA defectiveness. Again, one simply computes the (symmetric) matrix Λ_Z as given by (3.1) and checks whether it is positive semidefinite or not. Note that if Λ_Z is positive semidefinite then a joint normal random vector Z with this correlation matrix exists, and this vector transforms through NORTA to a random vector with the desired correlation matrix. On the other hand a correlation matrix must be positive semidefinite, so if Λ_Z is indefinite, then it is not a correlation matrix and NORTA will fail.

We should note here that NORTA might fail for two reasons. First, a given correlation matrix might not be feasible so that the uniform random vector as specified does not exist, in which case no generation method can achieve the given correlation matrix. Second, a given correlation matrix might be feasible, but might be NORTA defective. Joe (1997) established feasibility of all correlation matrices only for 3−dimensional uniform random vectors. In general, there is no clear characterization of the set of feasible correlation matrices for uniform marginal distributions in dimensions four and above. Therefore, it is unreasonable to expect a practitioner to specify a *feasible* correlation matrix, but we can reasonably expect the practitioner to specify a correlation matrix. The probability we estimate is thus over all correlation matrices, and not necessarily *feasible* correlation matrices.

Thus the problem of estimating the probability that NORTA fails reduces to

the following algorithm.

1. Let $n \geq 1$ be given.

2. Let $\Sigma_X(1), \ldots, \Sigma_X(n)$ be an i.i.d. sample chosen uniformly from

$$\Omega = \{\Sigma : \Sigma = \Sigma^T, \Sigma \succeq 0, \Sigma_{jj} = 1 \quad j = 1, \ldots, d\}. \tag{3.2}$$

3. For each $i = 1, \ldots, n$ let $\Lambda_Z(i)$ be obtained from $\Sigma_X(i)$ using the component-wise relation (1.5).

4. Estimate the probability that NORTA fails by the proportion of matrices in $\{\Lambda_Z(i) : i = 1, \ldots, n\}$ that are not positive semidefinite.

(The matrix inequality $A \succeq 0$ signifies a constraint that the matrix A be positive semidefinite.)

Note that in estimating the probability of NORTA defectiveness we have had to choose a probability distribution on Ω. Recall that Ω can be seen as a convex, compact and full-dimensional (i.e., with a non-empty interior) subset of the real vector space $\mathbb{R}^{d(d-1)/2}$. The uniform distribution (with respect to Lebesgue measure) on Ω is thus a natural choice, and is the one we choose to work with.

Kurowicka and Cooke (2001) report simulation results on a similar problem, and find that the probability the NORTA procedure fails to work grows rapidly with dimension. They provide these results to help motivate their copula-vine method for modeling random vectors. The probability distribution they used is not uniform over the set of all correlation matrices, so one possible explanation for their results prior to our study is that their probability distribution gives excessive weight to matrices for which NORTA fails. However, our results using uniform

distributions confirm their finding, suggesting that perhaps the NORTA procedure is unlikely to be effective in high-dimensional problems.

We now come to the question of sampling uniformly from the set Ω to estimate this probability. A straightforward approach (and one that we adopted early on) is to combine three well-known methods in simulation estimation: acceptance-rejection, importance sampling and ratio estimation. We used importance sampling and acceptance rejection on the hypercube $H = [-1, 1]^{\frac{d(d-1)}{2}}$ (Ω is a strict subset of H) to choose correlation vectors from Ω. We then used ratio estimation (Henderson 2001, see, e.g.,) to estimate the probability that NORTA fails. The probability we aim to estimate can be rewritten as

$$P(\Lambda_Z \not\succeq 0 | \Sigma_X \succeq 0) = \frac{P(\Lambda_Z \not\succeq 0, \Sigma_X \succeq 0)}{P(\Sigma_X \succeq 0)},$$

where Σ_X is uniformly distributed on H and Λ_Z is computed from (1.5). An estimator of this probability is therefore of the form

$$\frac{\sum_{i=1}^{n} [I(\Sigma_X(i) \succeq 0, \Lambda_Z(i) \not\succeq 0) \ \frac{2^{-d(d-1)/2}}{\phi(\Sigma_X(i))}]}{\sum_{i=1}^{n} [I(\Sigma_X(i) \succeq 0) \ \frac{2^{-d(d-1)/2}}{\phi(\Sigma_X(i))}]}, \tag{3.3}$$

where I represents the indicator function that equals 1 if its argument is true, and 0 otherwise, and the matrices $\Sigma_X(i)$ are chosen independently with density ϕ from the hypercube H. We chose the density ϕ in a heuristic fashion.

This method of estimation works well in lower dimensions but we found that it became excessively slow as the dimension increased. Indeed, it took more than two days to generate on the order of a thousand samples of positive-definite matrices even for a dimension as low as $d = 12$. With a better choice of ϕ the algorithm would presumably be much faster, but it is not clear how to choose ϕ. A better sampling technique is needed.

In related work, Marsaglia and Olkin (1984) survey methods for sampling random correlation matrices, but none of the methods they mention samples *uniformly* over the set of all correlation matrices (of fixed dimension). Theorems 3.1 and 3.2 of Edelman (1989) show how the distribution of a symmetric positive definite matrix can be expressed as a function of the distributions of the matrices of its eigenvalue decomposition. One could conceivably use this result in obtaining a sampling procedure, though the technique would work on the space of eigenvalues and eigenvectors of correlation matrices. Our representation of Ω, through its unique upper-diagonal elements, as a subset of a real space fits more naturally with the analysis in the other chapters, and hence we choose not to follow this approach in our analysis.

Another approach to sampling uniformly from the compact and convex Ω would be to use a Markov-Chain based random walk sampler, for instance the *Hit-n-Run Algorithm* presented by Smith (1984). These samplers set up a random walk on a convex set such that the stationary distribution of the Markov chain is uniform over the set. This approach presents two difficulties for our case. Firstly, random walk samplers require that the boundary of the convex set be easily described. The boundary of Ω is represented in terms of polynomial functions of the components of the matrix Σ, and in general not easy to compute. This is seen by noting that the boundary of Ω consists of matrices with zero determinants, and the determinant is polynomial in the components of a matrix. Secondly, the distribution of the samples obtained from a Markov-Chain based sampler is only *asymptotically* uniform, and moreover subsequent samples are usually not independent. This implies that the probability estimated from any finite sample obtained from such a sampler would be biased.

We present a new method to sample *exactly* and *independently* from the uniform distribution on Ω. The *onion method* is simple to implement since it uses nothing more than standard tools from the simulation input modelling toolkit, and sample generation is very fast. Indeed, the most complex and computationally demanding part of the method involves sampling from univariate beta distributions, which is a very well-studied problem with many efficient algorithms available (see Law and Kelton 2000, p. 467).

3.3 The Onion Method

Our goal is to construct a method that samples *exactly*, and very quickly, from the uniform distribution on the set Ω_d as defined in (3.2), when viewed as a subset of $\mathbb{R}^{d(d-1)/2}$. We use the suffix d to emphasize the dependence on the dimension d. We thus have to construct a procedure that samples uniformly from the convex, closed, compact and full-dimensional set Ω_d, i.e., generate samples from the density

$$f(\Sigma) \;\propto\; 1\,, \quad \text{for any} \quad \Sigma \in \Omega_d, \tag{3.4}$$

where f is a function of the $d(d-1)/2$ upper-diagonal elements of Σ.

For the random matrix Σ let $\Sigma_{\mathbf{k}}$ represent its $k \times k$ dimensional principal leading minor (i.e., the upper-left $k \times k$ sub-matrix of Σ), and $f_{\mathbf{k}}$ represent the marginal density of $\Sigma_{\mathbf{k}}$ when Σ has the joint density (3.4). Let $\mathbf{q}$ be the vector such that

$$\Sigma_{\mathbf{k}} = \begin{bmatrix} \Sigma_{\mathbf{k-1}} & \mathbf{q} \\ \mathbf{q}^t & 1 \end{bmatrix}.$$

We call $\mathbf{q}$ the *completion* of $\Sigma_{\mathbf{k-1}}$ in $\Sigma_{\mathbf{k}}$.

The onion method is iterative in that it starts with the one-dimensional matrix 1 and then "grows out" the matrix to the dimension desired by successively adding

an extra row (and the corresponding mirrored column) chosen from an appropriate distribution. This successive layering approach is the inspiration behind its name. Marsaglia and Olkin (1984) use a similar matrix-growing approach in their algorithm to sample correlation matrices with a given set of eigenvalues, but they apply it to transform diagonal elements of arbitrary positive definite matrices to 1 in order to form correlation matrices from them. Ouellette (1981) points out some other uses of the layering approach, notable among them being the numerical method proposed by Guttman (1946) to compute inverses of large non-singular matrices.

To be more precise the onion method is as follows.

1. Let $\Sigma_{\mathbf{1}}$ be the 1×1 matrix 1.

2. For $k = 2, \ldots, d$

   (a) Let $\mathbf{q}$ be a column vector in $\mathbb{R}^{k-1}$ sampled, independently of all else, from density $\varphi_k(\cdot; \Sigma_{\mathbf{k-1}})$ say.

   (b) Set $\Sigma_{\mathbf{k}} = \begin{pmatrix} \Sigma_{\mathbf{k-1}} & \mathbf{q} \\ \mathbf{q^t} & 1 \end{pmatrix}$.

   (c) Next k.

The densities φ_k, which determine the kth layer, are conditional densities that depend on the partial matrix $\Sigma_{\mathbf{k-1}}$ constructed thus far. We now state the key result that motivates the iterative sampling scheme, and in particular provides the form of the φ_ks.

Proposition 3.3.1 *Let $f_{\mathbf{k}}$ be the marginal density of $\Sigma_{\mathbf{k}}$ when Σ is distributed as in (3.4). Then*

$$f_{\mathbf{k}}(\Sigma_{\mathbf{k}}) \propto \ (\det(\Sigma_{\mathbf{k}}))^{\frac{d-k}{2}} \qquad \forall \ \Sigma_{\mathbf{k}} \in \Omega_k \, , \ \forall \, 2 \leq k \leq d.$$

The marginal density $f_{\mathbf{k}}$ represents the *joint* marginal density of the components $\Sigma_{\mathbf{k-1}}$ and $\mathbf{q}$ of $\Sigma_{\mathbf{k}}$, and Proposition 3.3.1 provides an expression for $f_{\mathbf{k}}$ in terms of $\Sigma_{\mathbf{k-1}}$ and $\mathbf{q}$ (through $\Sigma_{\mathbf{k}}$). The density φ_k of the completion $\mathbf{q}$ can then be obtained from $f_{\mathbf{k}}$ by conditioning on a fixed $\Sigma_{\mathbf{k-1}}$, for each $k = 2, \ldots, d$. The key to the generation scheme is the fact that the expression obtained for φ_k by this conditioning argument can be unravelled into separate parts that contain $\Sigma_{\mathbf{k-1}}$ and $\mathbf{q}$ in a way that allows for easy generation of $\mathbf{q}$ for a fixed $\Sigma_{\mathbf{k-1}}$.

We describe an efficient scheme to sample $\mathbf{q}$ from φ_k after we prove Proposition 3.3.1. We need two preliminary results for the proof.

Lemma 3.3.2 *If $m \geq 0$ and A is some symmetric p.d. matrix in Ω_d, then*

$$C \stackrel{\triangle}{=} \int_{I\!R^d} I(x^t A x \leq 1)\ (1 - x^t A x)^m dx = L(m,d) \cdot \det(A)^{-\frac{1}{2}},$$

where $0 < L(m,d) < \infty$ does not depend on A.

Proof: Since A is symmetric and positive definite, it has a unique upper triangular Cholesky factor $A^{1/2}$ say, so that $A = (A^{\frac{1}{2}})^t A^{\frac{1}{2}}$. Applying the linear change of variables $w = A^{\frac{1}{2}} x$ gives

$$\begin{aligned} C &= |\det(A)^{-\frac{1}{2}}| \cdot \int_{\mathbb{R}^d} I(w^t w \leq 1)\ (1 - w^t w)^m dw \\ &= \det(A)^{-\frac{1}{2}} \cdot L(m,d). \end{aligned}$$

The function $g(w) = I(w^t w \leq 1)\ (1 - w^t w)^m$ is non-negative, bounded and non-zero only over the compact region that forms the unit ball in $\mathbb{R}^d$. Hence $0 \leq L(m,d) < \infty$. Since $g(0) = 1$ and g is continuous about 0, $L(m,d) > 0$. $\square$

For the second result that we use in the proof of Proposition 3.3.1, first note that any positive definite symmetric $d \times d$ matrix A can be written as a product of two

matrices, akin to a first step in an LU factorization of A, as

$$A = \begin{pmatrix} B & b \\ b^t & 1 \end{pmatrix} = \begin{pmatrix} B & 0 \\ b^t & 1 \end{pmatrix} \begin{pmatrix} I_{d-1} & B^{-1}b \\ 0 & 1 - b^t B^{-1} b \end{pmatrix}, \tag{3.5}$$

where B is a $(d-1) \times (d-1)$ matrix, b is a $(d-1)-$vector and I_{d-1} is the $(d-1)$ dimensional square identity matrix. The quantity $(1 - b^t B^{-1} b)$ is called the *Schur complement* of B in A. Ouellette (1981) is a useful source of literature on Schur complements. In particular, Ouellette (1981) points to the result obtained in Guttman (1946) that $\text{rank}(A) = \text{rank}(B) + I(1 - b^t B^{-1} b > 0)$ (The indicator function $I(A)$ has value 1 if the event A is true and 0 otherwise.) This immediately gives us the following lemma.

Lemma 3.3.3 *A necessary and sufficient condition for A to be positive definite is that B be positive definite, and $1 - b^t B^{-1} b > 0$.*

Ouellette (1981) also describes a result that follows from (3.5) and was first shown by Frobenius (1968):

$$\det(A) = \det(B)\,(1 - b^t B^{-1} b). \tag{3.6}$$

We are now ready to prove Proposition 3.3.1.

Proof of Proposition 3.3.1: We use induction on k from d to 2 to complete the proof. The result is immediate for $k = d$, since the density (3.4) in this case is the density we are aiming for in the first place. This establishes the base case.

Let $\Psi_k = \{q \in \mathbb{R}^k |\ q^t (\Sigma_{\mathbf{k}})^{-1} q < 1\}$. Then, by Lemma 3.3.3, Ψ_k represents the set of all completion vectors $\mathbf{q}$ of $\Sigma_{\mathbf{k}}$ in $\Sigma_{\mathbf{k+1}}$. For any general k, assuming that the induction hypothesis holds for $k+1$, we get

$$f_{\mathbf{k}}(\Sigma_{\mathbf{k}}) \;\; = \;\; \int_{\Psi_k} f_{\mathbf{k+1}}(\Sigma_{\mathbf{k+1}})\, dq \tag{3.7}$$

$$\begin{aligned}
&\propto \int_{\Psi_k} \det(\Sigma_{\mathbf{k+1}})^{\frac{d-k-1}{2}}\, dq && (3.8)\\
&= \int_{\Psi_k} \det(\Sigma_{\mathbf{k}})^{\frac{d-k-1}{2}}\ (1 - q^t \Sigma_{\mathbf{k}}^{-1} q)^{\frac{d-k-1}{2}}\, dq && (3.9)\\
&\propto \det(\Sigma_{\mathbf{k}})^{\frac{d-k-1}{2}} \cdot \det(\Sigma_{\mathbf{k}})^{\frac{1}{2}} && (3.10)\\
&= \det(\Sigma_{\mathbf{k}})^{\frac{d-k}{2}}.
\end{aligned}$$

The first step (3.7) above expresses the marginal density of $\Sigma_{\mathbf{k}}$ as the function that results from integrating out the $(k+1)$st column, the completion vector $\mathbf{q}$, from the marginal density of $\Sigma_{\mathbf{k+1}}$ over the set Ψ_k. The inductive hypothesis gives (3.8). The equality (3.9) uses (3.6), and (3.10) follows from Lemma 3.3.2.
Thus the induction hypothesis holds for k, and hence is true for all k from d to 2.
□

We now determine the densities φ_k used in the iterative generation procedure from the marginal densities of Proposition 3.3.1. As mentioned before, the densities $f_{\mathbf{k}}$ represent the joint densities of $\Sigma_{\mathbf{k-1}}$ and its corresponding completion vector $\mathbf{q}$ in $\Sigma_{\mathbf{k}}$. Hence, if $\Sigma_{\mathbf{k-1}}$ were fixed at A say, we would have that for a $q \in \Psi_{k-1}$,

$$\begin{aligned}
\varphi_k(q) &= f_{\mathbf{k}}(\Sigma_{\mathbf{k}}) | \{\Sigma_{\mathbf{k-1}} = A\}\\
&\propto \det\left(\begin{pmatrix} A & q \\ q^t & 1 \end{pmatrix}\right)^{\frac{d-k}{2}}\\
&= \det(A)^{\frac{d-k}{2}} \cdot (1 - q^t A^{-1} q)^{\frac{d-k}{2}}\ .
\end{aligned}$$

Therefore, given $\Sigma_{\mathbf{k-1}}$, the conditional density for its completion vector $\mathbf{q}$ is

$$\varphi_k(q) \ \propto\ (1 - q^t \Sigma_{\mathbf{k-1}}^{-1} q)^{\frac{d-k}{2}}\ \forall\ q \in \Psi_{k-1}. \qquad (3.11)$$

Next comes the question of generating from densities of the form (3.11). For this we employ a sequence of variable transformations. First we apply the linear

transformation $w = \Sigma_{\mathbf{k-1}}^{-1/2} q$, where $\Sigma_{\mathbf{k-1}}^{-1/2}$ represents the (unique) upper triangular Cholesky factor of $\Sigma_{\mathbf{k-1}}^{-1}$, to get that

$$\varphi_k(q)dq \;\propto\; \tilde{\varphi}_k(w)dw,$$

where $\tilde{\varphi}_k(w) \propto (1-w^t w)^{(d-k)/2}$, and $w \in \mathbb{B}^{k-1}$, the unit ball in $\mathbb{R}^{k-1}$ (the constant Jacobian term that arises out of the transformation is included in the proportionality constant). Hence, to sample $\mathbf{q}$ from φ_k we could equivalently generate a $\mathbf{w}$ from $\tilde{\varphi}_k$ and set $\mathbf{q}$ to be the appropriate linear transformation of $\mathbf{w}$.

Now, $\tilde{\varphi}_k$ is radially symmetric, as is the set $\mathbb{B}^{k-1}$. Thus if we apply a polar transformation $w = (r, \theta)$, where r is the L_2-norm of w and $\theta = (\theta_1, \ldots, \theta_{k-2})$ represents the angles of the polar transformation (refer Kendall (1961) p. 15 for a treatment of polar transformations in higher dimensions), then

$$\begin{aligned}
\tilde{\varphi}_k(w)dw &\propto (1-r^2)^{\frac{d-k}{2}}\; J(r,\theta)\; dr d\theta_1 \ldots d\theta_{k-2} \\
&= (1-r^2)^{\frac{d-k}{2}}\; r^{k-2} dr\; (\cos\theta_1)^{k-3} (\cos\theta_2)^{k-4} \ldots \cos\theta_{k-3}\; d\theta_1 \ldots d\theta_{k-2} \\
&\propto h(r)dr
\end{aligned}$$

where $J(r, \theta)$ represents the Jacobian term of the variable transformation and expands out as given in the second equation, and $h(r) = (1-r^2)^{\frac{d-k}{2}} r^{k-2}$. The second equation implies that the distribution of r is independent of the distributions of the angles θ_i. Moreover, the radial symmetry of the integrand also gives us that $\tilde{\varphi}_k(w)$ affects only the distribution of r, and the angles need to be sampled such that a point is chosen uniformly on the surface of the unit hyper-ball $\mathbb{B}^{k-1}$.

This suggests that we can sample a $\mathbf{w}$ from $\tilde{\varphi}_k$ by instead first sampling a radius from a normalized version of h and then multiplying it by a point chosen *uniformly* over the surface of the unit ball $\mathbb{B}^{k-1}$. Such a point can be generated by normalizing a joint-normal independent random vector (i.e., one with the identity

matrix as its correlation matrix) to have unit norm. The radius has to be sampled from h, but note that under yet another change of variable $y = r^2$, we have that

$$h(r)dr \;\propto\; y^{\alpha_1 - 1}(1-y)^{\alpha_2 - 1}dy,$$

which (after normalization) is simply a univariate beta density function with parameters α_1 and α_2. For our case, the parameters α_1 and α_2 turn out to be $(k-1)/2$ and $(d-k)/2$ respectively. Law and Kelton (2000, (p. 467)) points to extensive literature on generating from beta distributions. In our study, we implement the algorithm given by Schmeiser and Babu (1980; 1983).

To recap, at the kth stage of the iterative generating procedure, to generate a realization q of $\mathbf{q}$ from φ_k given the matrix $\Sigma_{\mathbf{k-1}}$ already constructed, we do the following:

- Sample y from a beta distribution with $\alpha_1 = (k-1)/2$ and $\alpha_2 = (d-k)/2$,
- Set $r = \sqrt{y}$,
- Sample a unit vector θ uniformly from the surface of $\mathbb{B}^{k-1}$,
- Set $w = r\theta$, and finally
- Set $q = \Sigma_{\mathbf{k-1}}^{\frac{1}{2}} w$.

This completes the description of the onion method.

This exact sampling method is very competitive when estimating statistical properties of the set Ω_d when compared to methods like the one described in Section 3.2. First, since sampling from φ_i can be reduced to the problem of sampling from a univariate beta distribution and a joint-normal independent random vector, the method scales very well with dimension. In our study we were able to generate

samples consisting of many thousands of matrices up to dimension $d = 25$ in a matter of hours. Second, this method does not involve a ratio-estimation step, which means that the estimation is more straightforward to implement. Third, for a given sample size, we also found the results to be more accurate, in the sense that confidence-interval widths are smaller for this sampling method.

As noted before, this method can be generalized and applied very easily to generate uniformly from sets of symmetric positive-definite matrices with any arbitrary (fixed) positive diagonal elements. One simply has to modify the method by substituting the diagonal values of 1 assumed in this section with the corresponding positive values at the appropriate places (the definition of Ψ_k in the proof of Proposition 3.3.1 is one such place). The constants of the variate generation method would be affected accordingly (for instance, the beta variate generation would not be over $(0, 1]$).

One can also use this method to sample from any bounded non-uniform density f defined on a set of symmetric positive definite matrices of the kind mentioned above. One simply uses the acceptance-rejection framework of random variate generation to do this, namely by first generating a point s uniformly from the set and then checking whether $f_{\max} * U \leq f(s)$ (where U is an independent uniform random variable, and $f_{\max} = \max_x f(x)$) in order to accept s as a sample.

The beta distribution used to sample the polar variable r above can be replaced with any distribution over the positive real line. Thus the onion method can also sample from any member of the family of distributions on the set of symmetric p.s.d. random matrices that are radially symmetric under an affine transformation (the first transformation in the sequence above).

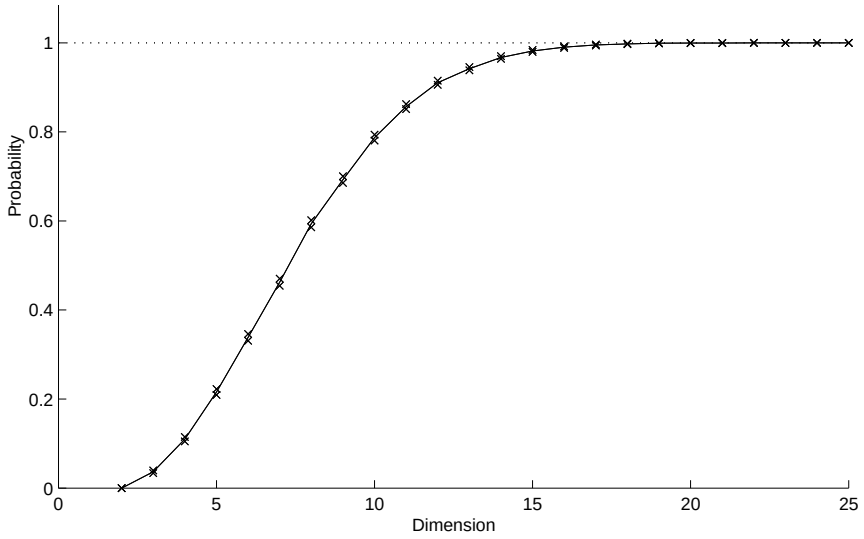

Figure 3.1: Probability of NORTA defectiveness

3.4 NORTA's Performance: A Plot

We used the exact sampling approach of Section 3.3 to estimate the probability that NORTA fails for various dimensions. For each dimension, we sampled $15,000$ correlation matrices uniformly from Ω_d, and estimated the probability that NORTA fails in dimension d via the procedure outlined in Section 3.2. Figure 3.1 plots the estimated probabilities against dimension. Also shown are 95% confidence intervals.

The plot establishes that the problem rapidly becomes acute as the dimension increases, and NORTA becomes very hard to work with when generating even moderately sized random vectors of dimensions 20 and above.

We stress here that the increasing rate of failure of correlation matrices with the NORTA procedure observed in Figure 3.1 could be the combined effect of two causes. First, the chosen correlation matrix might not be feasible, in which case this

matrix fails not only with NORTA but with any method. Second, the correlation matrix might be feasible but NORTA defective. One needs a characterization of feasible correlation matrices to be able to differentiate between these two effects. In the absence of any such characterization, the set of correlation matrices remains the natural set for a user to choose from, and so the probability we calculate is important from the practical standpoint.

It would thus seem that NORTA can only be successful in low-dimensional problems. NORTA's transformation-based generation approach is very simple and easily implemented. One would thus like to see if the advantages of this approach can be preserved while compromising slightly on the accuracy, for instance by requiring that correlations be matched approximately well. Recall that in Section 3.1, the indefinite matrices Λ_Z were observed to lie very close to the set of feasible correlation matrices for joint normal random vectors (i.e., the set of positive semidefinite matrices with ones on the diagonal). This fact suggests the modified versions of NORTA that we study in the next chapter.

CHAPTER 4

AUGMENTED NORTA PROCEDURES

To recoup, we saw in Chapter 2 that NORTA fails with certain specified marginals and correlation matrices. The reason these matrices are NORTA defective is that the correlation matrix Λ_Z determined for the base joint normal random vector of NORTA turns out to be indefinite, and hence infeasible. We have established in Chapter 3 that the NORTA procedure is increasingly unlikely to work with (uniform marginals and) any matrix chosen uniformly from the set of correlation matrices as the dimension of the random vector increases. It would thus seem that NORTA can only be successful in low-dimensional problems. NORTA is very general in its approach, and quite easily implemented, and so we may wish to employ the method to generate random vectors with the required marginals and, at least approximately, the right correlation matrix. In this chapter, we shall discuss some approaches that make this possible.

Recall that in Chapter 2, we had observed that the indefinite matrices Λ_Z lie very close to the set of feasible correlation matrices for joint normal random vectors. For joint normals, this is the set of positive semidefinite matrices with ones on the diagonal, which means that all correlation matrices are feasible. This suggests that the setup stage of NORTA should be augmented with an extra step that is used, if Λ_Z turns out indefinite, to find a correlation matrix Σ_Z that produces a NORTA correlation matrix Σ_X for the desired marginals that is "close" to the desired Σ. The augmented NORTA procedure can thus be described as:

1. Given the desired marginals F_i and covariance matrix Σ, estimate the joint-normal covariance matrix Λ_Z using numerical search/closed form expression.

2. If Λ_Z is positive semidefinite, then set $\Sigma_Z = \Lambda_Z$.

3. If Λ_Z is indefinite, heuristically determine an approximately close correlation matrix Σ_Z.

4. Run NORTA with Σ_Z.

Most of the approaches we describe here tackle the problem in Step 2 by finding a correlation matrix Σ_Z that is "close" to the indefinite Λ_Z. In other words, these methods work in the "Gaussian space". Why is this approach reasonable? In Theorem 2 of Cario and Nelson (1997), it is shown that under a certain moment condition, the output covariance matrix is a continuous function of the input covariance matrix Σ_Z used in the NORTA procedure. So if Σ_Z is "close" to Λ_Z, then we can expect the covariance matrix of the NORTA generated random vectors to be close to the desired matrix Σ. The moment condition always holds when we are attempting to match rank covariances, and we can expect it to hold almost invariably when matching product-moment correlations. Therefore, it is eminently reasonable to try and minimize some measure of distance $r(\Lambda_Z, \Sigma_Z)$ say, between Λ_Z and Σ_Z.

Lurie and Goldberg (1998) described a method for identifying a positive semidefinite covariance matrix Σ_Z for use within the NORTA method that yields approximately the desired product-moment covariance matrix Σ. Their approach involves a complicated nonlinear optimization, and must be specialized for approximating the rank correlation or product-moment correlation, depending on the case desired. Furthermore, although they report that their optimization procedure always converges in practice, they do not have a proof of this result. Finally, their approach appears to be limited to fixed sample sizes. We present alternative methods based on semidefinite programming and matrix algebra that do not share these limitations.

4.1 SDP-augmented NORTA

We do not distinguish between the cases where Λ_Z is chosen to induce a given rank, product-moment, or other correlation in the output random vector X. If Λ_Z is indefinite, then we use a semidefinite program (SDP) to find a matrix Σ_Z that is "close" to Λ_Z and is positive semidefinite and has ones on the diagonal. The SDP falls under the broad class of matrix completion problems; see Alfakih and Wolkowicz (2000), or Johnson (1990). For this case, given Λ_Z as data, we wish to solve the problem Σ_Z to

$$\begin{aligned} \text{minimize} \quad & r(\Sigma_Z, \Lambda_Z) \\ \text{subject to} \quad & \Sigma_Z \succeq 0, \\ & \Sigma_Z(i,j) = \Sigma_Z(j,i), \quad \text{and} \\ & \Sigma_Z(i,i) = 1. \end{aligned} \tag{4.1}$$

where the matrix inequality $A \succeq 0$ signifies a constraint that the matrix A be positive semidefinite.

The metric $r(\cdot,\cdot)$ can be chosen as desired. In particular, given that we have previously considered correlation matrices to belong to the real space $\mathbb{R}^{d(d-1)/2}$, we choose either the l_1 metric

$$r(A,B) = \sum_{i>j} |A_{ij} - B_{ij}|$$

or the l_∞ metric

$$r(A,B) = \max_{i>j} |A_{ij} - B_{ij}|$$

as defined for $\mathbb{R}^{d(d-1)/2}$. Either of these metrics make the minimization problem an SDP-constrained problem with a linear objective function. To see why, we shall use the same trick used in formulating the LPs in Chapter 2. We define variables

Z_{ij}^+ and Z_{ij}^- to be the positive and negative parts respectively of the difference $\Sigma_Z(i,j) - \Lambda_Z(i,j)$, with $Z_{ij}^+, Z_{ij}^- \geq 0$. In particular, we have that for $1 \leq i < j \leq d$,

$$\begin{aligned} \Sigma_Z(i,j) = \Lambda_Z(i,j) &+ Z_{ij}^+ - Z_{ij}^-, \text{ and} \\ |\Sigma_Z(i,j) - \Lambda_Z(i,j)| &= Z_{ij}^+ + Z_{ij}^-. \end{aligned}$$

Thus, if r is set to be either the l_1 or l_∞ norm, we can write the objective function of (4.1) as a linear function of the Z_{ij}^+s and Z_{ij}^-s.

The matrix inequality (semidefiniteness) constraint in (4.1) can also be rewritten in the $Z_{ij}^\pm$s. The diagonal elements of Σ_Z are set to have value 1. This leaves the non-negativity constraints on the variables $Z_{ij}^\pm$. It is well known (Wolkowicz et al. 2000) in SDP formulations that any set of linear inequalities of the form $Ax+b \geq 0$ can be transformed to a matrix inequality of the form $\mathbf{diag}(Ax+b) \succeq 0$ (for a vector y, let $\mathbf{diag}(y)$ denote the diagonal matrix with diagonal entries equal to the elements of y.) The non-negativity constraints can be easily handled in this manner, and the problem is thus reformulated as an SDP-constrained problem with a linear objective problem (The set of constraints that ensure that only one of Z_{ij}^+ or Z_{ij}^- be non-zero are automatically satisfied since the objective function and constraints are convex.) This is a convex optimization problem and therefore any local minimum is, in fact, a global minimum. Efficient algorithms, and public domain codes implementing them, are available for solving semidefinite programs of this type to optimality; see Wolkowicz et al. (2000).

It is worth noting that Σ_X, the NORTA generated covariance matrix from Σ_Z, may not be the closest NORTA feasible covariance matrix to Σ, because the optimization was performed "in Gaussian space". This is in contrast to the Lurie and Goldberg (1998) procedure. But, preliminary tests in 3 dimensions indicate that this SDP augmentation yields NORTA generated random vectors with corre-

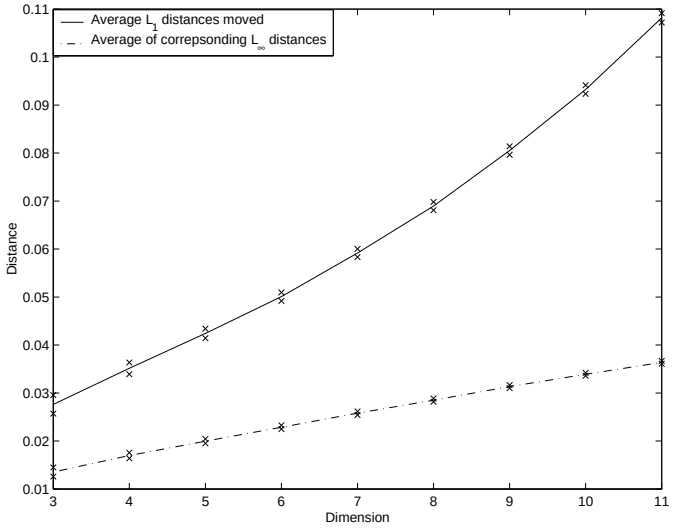

Figure 4.1: l_1 Performance of the SDP-augmented NORTA method

lation matrices that are close to the desired ones. We study the performance of this augmented NORTA method as the random vector dimension increases, using a setting identical to that in Chapter 3. We choose our measure of performance to be the expected l_1 distance that we have to move from the desired correlation matrix to reach a NORTA feasible one. For each dimension, 15,000 matrices were generated uniformly from the set of correlation matrices and the semidefinite program, with r taken as the l_1 distance, solved for the cases where NORTA was found to fail. Figure 4.1 plots the results. The solid line gives the expected l_1 distance $||\Sigma_X - \Sigma||_{l_1}$ with 95% confidence intervals as marked, with the average taken only over matrices that fail to work with NORTA. The dotted line gives the corresponding expected distance as measured in the l_∞ metric.

We see that the expected l_1 distance, which represents the total absolute change in the correlation values, increases as the dimension d increases at what might be perceived as a linear rate, although a super-linear rate seems more likely. If the rate of increase is indeed linear, then, since there are $d(d-1)/2$ matrix entries

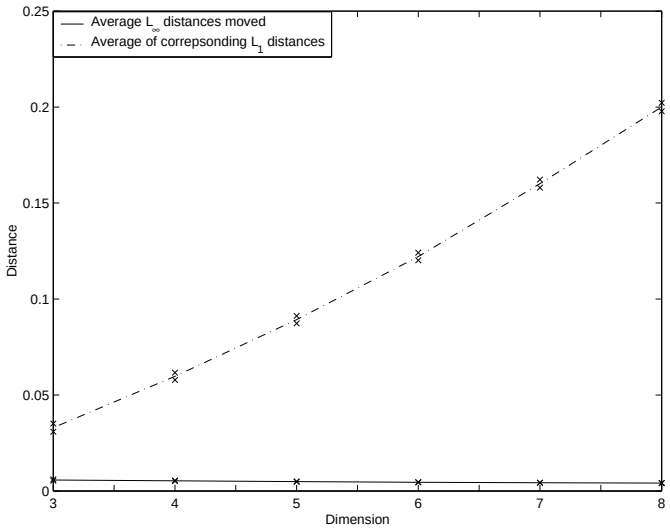

Figure 4.2: l_∞ performance of the SDP-augmented NORTA method

above the diagonal, the *average* change per entry is (eventually) decreasing with dimension. Of course, it is possible that a small number of entries change by a large amount. The l_∞ distance is also shown, and we see that indeed, at least one entry is changed by an increasing amount as the dimension increases.

It might be preferable from a modelling standpoint to instead minimize the l_∞ distance, so that one tries to minimize the maximum deviation between the achievable and target correlations. The results in this case are shown in Figure 4.2. The experiment performed is identical to the earlier case except for the change in metrics. The solid line gives the expected l_∞ distance with 95% confidence intervals as marked, with the average taken only over matrices that fail with NORTA. The dotted line gives the corresponding expected distance as measured in the l_1 metric. We see that the expected l_∞ distance appears to remain constant at around 0.005 or even decrease with dimension. The corresponding l_1 distance seems to grow at a super-linear rate.

While the total absolute change in correlations seems to grow in either case,

Figure 4.2 suggests that one could attempt a hybrid of the l_1 and l_∞ approaches by, for instance, minimizing the l_1 distance subject to an upper bound on the l_∞ distance, and thus avoid changing any single component of the correlation matrix by too large a value while keeping the total change within reasonable limits.

We remark here that the SDP framework used here in searching for a "close" positive definite matrix in the SDP problem (4.1) allows us a certain degree of control on how the search is performed. For instance, we can also restrict the change in certain components by adding additional constraints to the SDP.

In summary, the SDP-augmented NORTA problem performs well in moderate dimensions: It generates random vectors with correlation matrices that are close to the desired ones, while keeping changes to the individual correlations within reasonable limits.

Computational results show that the SDP problem in the SDP-augmented method is solved within a reasonable amount of time for dimensions less than 10. However we find that the SDP problem (4.1), especially when the l_∞ metric is used as r, becomes increasingly harder to solve as the dimension increases. In the instances where l_∞ SDP problems were solved for $d = 10$, the publicly available implementations of the solvers took, on an average, around 10 minutes to terminate for each sample problem! This indicates that the setup time for the SDP-augmented NORTA might be excessive in dimensions 10 and higher.

While tweaking the SDP formulations might possibly reduce the time taken by the solvers, we shall take a different approach. We have noted before that the SDP-based augmentation optimizes in "Gaussian space", and the non-linearity of the transformation (1.1) makes it possible that the resulting NORTA correlation matrix might not be the NORTA feasible correlation matrix closest to the desired

Σ. Fast heuristics that provide a reasonably close approximation to the desired Σ might thus be equally effective as a NORTA augmentation step. We shall study three such heuristics.

4.2 The "Just-Use-Sigma" (JUS) Method

As the name suggests, this method uses the desired correlation matrix Σ as the normal correlation matrix Σ_Z. This obviously does not have any computational overhead, and hence this method can be used in any dimension. It is always assured to work, assuming that the desired matrix Σ is indeed a correlation matrix. Moreover, one can exactly predict the output correlations the NORTA method produces if the function c_{ij} defined via (1.3), which relates the NORTA output correlations with those of the joint normal, is known in closed form.

This approach has another characteristic in its favour, which is that one can estimate an upper bound on the l_∞ or maximum absolute change in correlations. To see why, recall that the function c_{ij} has been shown to be continuous and non-decreasing under certain mild moment conditions (Cario and Nelson 1997). So, assuming that this condition is satisfied, $\sup_{\rho\in[-1,1]} |c_{ij}(\rho) - \rho|$ can be estimated and thus gives an upper bound on the l_∞ change in correlations. For instance, in the case where the target marginals are all uniform, c_{ij} is given by (1.5) and the l_∞ change bound can be calculated to be 0.01808.

4.3 The Eigenvalue Correction (EC) Method

This heuristic aims to provide a good feasible approximation to the optimal solution of the SDP optimization problem (4.1). The premise of this approach is quite straightforward. The matrix Λ_Z is indefinite, and hence has a few negative

eigenvalues. This heuristic identifies a positive semidefinite matrix by replacing these negative eigenvalues.

The matrix Λ_Z is factorized as UD_0U^t, where D_0 is a diagonal matrix consisting of the eigenvalues $\{\lambda_i^0,\, i = 1, \ldots, d\}$ of Λ_Z and U is an orthogonal matrix whose columns are the eigenvectors of Λ_Z (a matrix A is orthogonal if $A^tA = AA^t = I$, the identity matrix.) Since Λ_Z is indefinite, one or more of its eigenvalues λ_i^0 are negative. We shall replace any negative eigenvalue by 0, and let $\bar{\Sigma}_Z$ represent the matrix with this new set of eigenvalues together with the same eigenvectors U. Now, the "zero-out" operation guarantees that $\bar{\Sigma}_Z$ is a positive semidefinite matrix, but $\bar{\Sigma}_Z$ is unlikely to have ones on its diagonal as we require. This is easily fixed by scaling the rows and columns of $\bar{\Sigma}_Z$ by appropriate (non-zero) constants. This then gives us the correlation matrix Σ_Z we desire.

The procedure is easily implemented and scales very well with dimension. It dispenses with the need for optimization solvers. The eigenvalue factorization step is essentially the only limitation, but this matrix algebra procedure is well studied, easily coded and very fast practical routines are available (see, for instance Golub and Van Loan 1996, Chapter 8).

We note that in practice the negative eigenvalues are replaced with a positive number ϵ smaller than all the positive eigenvalues, rather than 0. This will ensure that the NORTA covariance matrix Σ_Z constructed is positive definite, and thus avoids numerical instability issues that arise while finding the Cholesky factorization of a positive semidefinite Σ_Z (the Cholesky factor is needed to generate the joint normal base random vector.) While we discuss zeroing out negative eigenvalues here mainly to keep the discussion clean, all the results hold equally well for when they are replaced by ϵ.

The EC method identifies a matrix that is optimal for a modified version of the SDP problem (4.1). To see why, suppose r is the Frobenius norm:

$$||A||_F = (\sum_{i,j} A_{ij}^2)^{1/2}.$$

This norm is equivalent to the l_2 norm used on the matrix when considered to be an element in the vector space $\mathbb{R}^{d\times d}$. It is also *orthogonal invariant* (Golub and Van Loan 1996, define a norm r to be orthogonal invariant if $r(\Sigma_1, \Sigma_2) = r(P\Sigma_1 Q, P\Sigma_2 Q)$ for any orthogonal matrices P and Q.).

Next, suppose the constraints on the diagonal elements in (4.1) are relaxed, and the matrix variable Σ_Z is constrained to be of type UDU^t, where U is the eigenvector matrix of Λ_Z and D is a variable diagonal matrix. Then, SDP (4.1) can be simplified to

$$\min \qquad r(D, D_0) \text{ s.t. } D \succeq 0,$$

where the orthogonal invariance of r is used to ensure that $r(\Sigma_Z, \Lambda_Z) = r(D, D_0)$. But, the Frobenius norm on diagonal matrices is simply $r(D, D_0) = ||\lambda - \lambda^0||_2$, where $\lambda = De$ (e is a vector of ones), a vector of the diagonal elements of the diagonal matrix D and $||\cdot||_2$ is the ordinary l_2 vector-norm. Moreover, the constraint $D \succeq 0$ is equivalent to $\lambda \geq 0$. The problem thus reduces to a rather simple form that one can solve using the KKT conditions to show that the optimal solution λ^* is λ^0 with the negative elements zeroed out.

The solution $\bar{\Sigma}_Z = UD^*U^t$ will however probably not have ones on the diagonal, and thus not be a correlation matrix, since we relaxed the corresponding constraint. This is rectified by pre- and post-multiplying $\bar{\Sigma}_Z$ with a diagonal scaling matrix O with diagonal elements $O(i,i) = 1/\sqrt{\bar{\Sigma}_Z(i,i)}$. We set the final output $\Sigma_Z = O\bar{\Sigma}_Z O$.

This method thus searches for a feasible solution for SDP (4.1) in a restricted portion of the feasible set of the SDP problem. In the original problem, one has the freedom to change the matrix variables by changing both their eigenvalues and eigenvectors to find the optimal solution to the problem. In this scheme, we do vary both, but in a very specific way, and hence the Σ_Z value output by this approach is only a feasible solution to the SDP constraints.

To summarize, the EC method finds a correlation matrix Σ_Z thus:

1. Factorize Λ_Z into its eigenvalue and eigenvector matrices D_0 and U.
2. Replace all negative values in D_0 with a small positive constant ϵ. Let D^* represent the resulting diagonal matrix.
3. Let $\bar{\Sigma}_Z = UD^*U^t$,
4. Define a diagonal matrix O, with $O(i,i) = 1/\sqrt{\bar{\Sigma}_Z(i,i)}$.
5. Set $\Sigma_Z = O\bar{\Sigma}_Z O$.

4.4 The Least Squares (LS) Based Method

This final approach is similar to the EC method in that it solves a restricted version of the SDP problem (4.1) to find a good feasible solution to the original problem. The EC method searches for a good approximation to the optimal solution of the SDP (4.1) in a restricted space of eigenvalues and eigenvector combinations. This approach, on the other hand, completely fixes the eigenvectors and optimizes over only the eigenvalues.

We again let r be the Frobenius norm. The matrix Σ_Z is again fixed to be of form UDU^t, where U is the eigenvector matrix of Λ_Z and D is a variable diagonal

matrix. However, unlike the earlier case, we shall not drop the "ones-on-diagonal" constraint. The optimization problem reduces to

$$\begin{aligned} \min \quad & ||D_0 - D||_F && (4.2) \\ \text{s.t.} \quad & D(i,i) \geq 0 \\ \text{and} \quad & (UDU^t)(i,i) = 1, && (4.3) \end{aligned}$$

where D_0 denote the diagonal matrix of eigenvalues of Λ_Z.

As before, let $\lambda = De$ be the vector of diagonal elements of D. Constraint (4.3) is a set of d linear equations in d variables (λ_is), and one might thus expect that $\lambda^0 (= D_0 e)$, the eigenvalues of Λ_Z, is the unique solution to (4.3) since Λ_Z has ones on its diagonal. This is however not the case, and this set of constraints is under-determined. Firstly, the equations (4.3) can be rewritten as $A\lambda = e$ where A is a $d \times d$ matrix defined as $A(i,j) = U(i,j)^2$. Now, since the eigenvector matrix U is orthogonal, the vector e is a right-eigenvector for A, i.e., $Ae = e$. But, Λ_Z is already known to satisfy 4.3, i.e., $A\lambda^0 = e$. Thus, $A(\lambda^0 - e) = 0$. The vector λ^0 however cannot be e since Λ_Z is indefinite. So, $(\lambda^0 - e)$ is non-zero, implying that A is rank-deficient. The constraint $A\lambda = e$ thus has multiple solutions.

The problem (4.2) can be recast as:

$$\min ||y||_2 \text{ s.t. } Ay = 0,\ y + \lambda^0 \geq 0,$$

where y is a $d-$vector such that $y = \lambda - \lambda^0$. Notice that this version of the problem resembles a Least Squares minimization problem. We know that A is a rank-deficient matrix, i.e., $\text{rank}(A) = a < d$. So, the $Ay = 0$ constraint implies that we optimize over the null space of A, which has dimension $(d - a)$. In other words, we can use $Ay = 0$ to "factor out" as many as a variables out of the LS optimization problem (Lawson and Hanson 1974). This is done by first orthogonally factorizing

A as $A = HRK^t$ where H and K are orthogonal matrices such that of dimension $d \times d$, and $R = \begin{pmatrix} R_{11} & 0 \\ 0 & 0 \end{pmatrix}$ where R_{11} is an $a \times a$ matrix of full rank (One instance of such a factorization is the singular value decomposition of A.) We define $K^t y = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = u$, where u_1 and u_2 are vectors of dimension a and $(d-a)$ respectively. Then, $Ay = 0$ becomes $R_{11}u_1 = 0$, independent of u_2. Now, since R_{11} is of full rank, $u_1 = 0$. Thus, the problem above simplifies to an LS problem on a $(d-a)$-dimensional subspace:

$$\min ||u_2||_2 \text{ s.t. } K \begin{pmatrix} 0 \\ u_2 \end{pmatrix} + \lambda^0 \geq 0. \tag{4.4}$$

This problem is commonly termed a linear-inequality constrained LS problem, and efficient numerical procedures are available to solve this exactly; see, e.g., Lawson and Hanson (1974, Pg 165, Chapter 23, where they call it the LDP problem).

The correlation matrix to be output Σ_Z is calculated from the optimal solution of the LS problem (4.4). To summarize this approach (given an indefinite Λ_Z with ones on its diagonal):

1. Factorize Λ_Z into its eigenvalue and eigenvector matrices D_0 and U. Let $\lambda^0 = D_0 e$.

2. Define a matrix A as $A(i,j) = U(i,j)^2$.

3. Let $A = HRK^t$ be an orthogonal factorization of A.

4. Solve the Least Squares problem (4.4) using standard numerical procedures. Let $u^* = \begin{pmatrix} 0 \\ u_2^* \end{pmatrix}$, where u_2^* is the optimal solution obtained for (4.4).

5. Define $y^* = Ku^*$ and $\lambda^* = y^* + \lambda^0$.

6. Set $\Sigma_Z = UD^*U^t$, where $D^*e = \lambda^*$ is a diagonal matrix.

We had seen earlier that the EC method searches for a good approximation to the optimal solution of the SDP (4.1) in a restricted space of eigenvalues and eigenvector combinations. The LS approach, on the other hand, completely fixes the eigenvectors U and optimizes over only the d variables λ_i. This reduces the number of variables from an order $O(d^2)$ in the original problem to $O(d)$. Thus this scheme might produce approximations that are not quite as tight as the EC method.

4.5 A Comparison

We gauge the performance of the methods proposed above in a setup identical to that used for the SDP-augmented NORTA: we specialize to the case where NORTA is used to generate uniform random vectors, and the expected l_1 and l_∞ change in correlations serve as the performance measure. For each dimension, 15,000 matrices were chosen uniformly from the set of correlation matrices and each of the three heuristics described above, along with the l_1 and l_∞ SDP-augmented approaches, were utilized to augment the NORTA setup in the cases where NORTA was found to fail. (The l_1 SDP-augmentation method solves the SDP (4.1) with the l_1 metric and similarly for the l_∞ SDP-augmentation.) As usual, we measure the performance by the expected distance, as measured by a suitably chosen metric, that we have to move from the desired correlation matrix Σ to reach a NORTA feasible one Σ_X.

Figure 4.3 plots the observed expected l_1 change in correlations against dimension for each method. In each case, the expected l_1 distance moved seems to change slightly super-linearly with dimension. We notice that the l_1 SDP-augmented

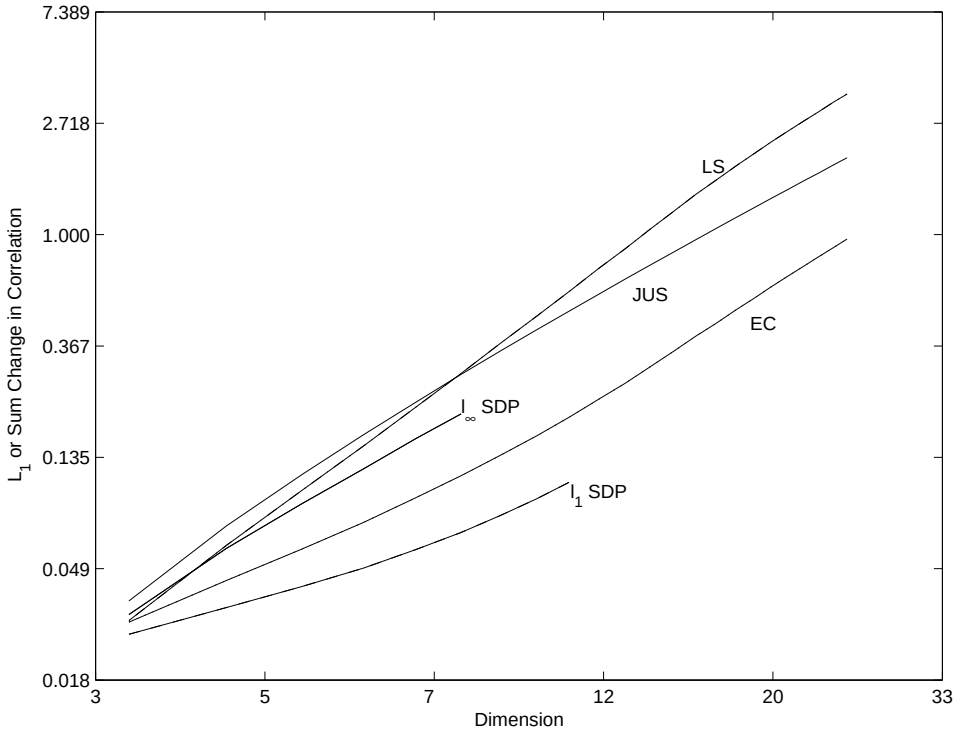

Figure 4.3: l_1 performance of all NORTA augmentation methods

method outperforms the rest. One might expect this since the SDP (4.1) is solved to optimality with r as l_1. Of the rest the Eigenvalue Correction (EC) method clearly performs better. The "Just-Use-Sigma" (JUS) method seems to fare about as well as, or better than, the Least Squares (LS) procedure, which is perhaps a disappointing outcome. The SDP approaches' increasing computational difficulty with dimensionality is also illustrated by the fact that these methods could only be tracked till dimension 11.

Figure 4.4 presents the l_∞ results, where the expected l_∞ change in correlations is plotted against dimension for each method. As one might expect, in this case the l_∞ SDP augmented method performs the best. The JUS method is clearly bounded by 0.01808 as predicted. The interesting fact one notices here is that the EC method again outperforms the rest of the methods. Moreover the average

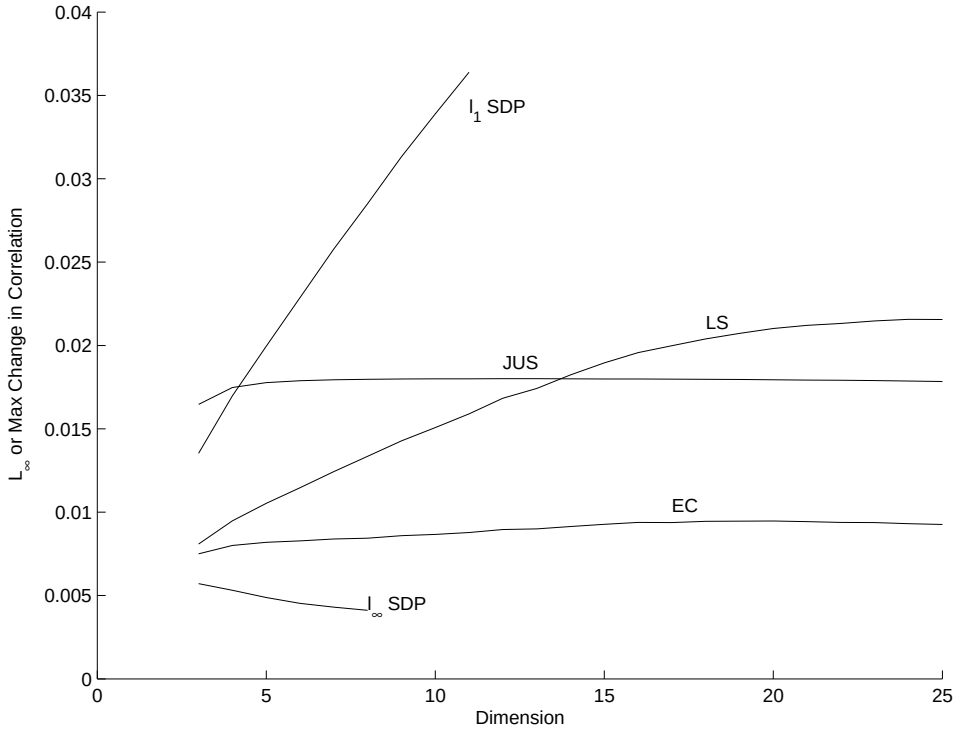

Figure 4.4: l_∞ performance of all NORTA augmentation methods

change by the EC method seems to be bounded by 0.01. This would suggest that the EC method provides a nice compromise between computational tractability and accuracy in approximation.

The fact that the expected absolute change in the correlations for the EC method seems bounded by 0.01 is an unexpected boon. We do not know of a theoretical reason for this behaviour of the EC method, but we experimentally study this property further. Specifically, we again sample 15,000 correlation matrices in dimensions 10 to 100 and find the *maximum* l_∞ change induced by the EC method among the NORTA defective samples in each dimension. A plot of this result is given in Figure 4.5. We see that the maximum l_∞ change actually falls with dimension. This can be heuristically explained by an observation we had made earlier, in that though almost all correlation matrices fail in higher dimensions,

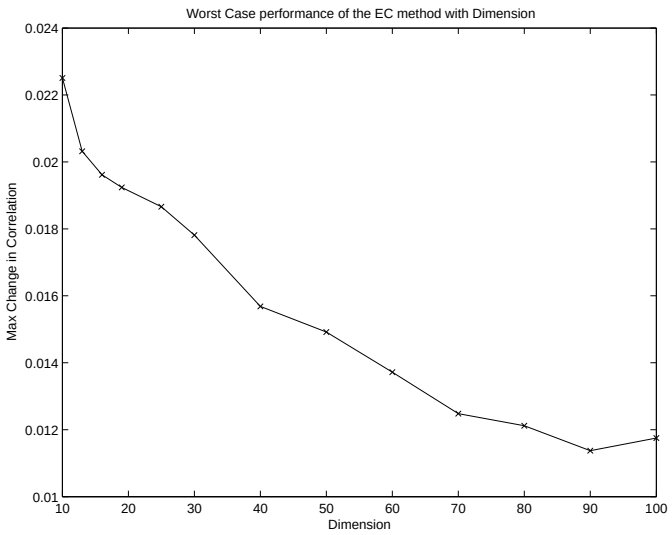

Figure 4.5: Maximum l_∞ change observed for the EC method

almost all the indefinite Λ_Z probably lie close to the set of correlation matrices. Thus a change of the same magnitude in the correlations suffices to find a correlation matrix for most of the indefinite Λ_Z. We also note here that the EC method worked in dimensions as high as 100 in a matter of seconds, which indicates that it is computationally quite agreeable.

We conclude that despite the feasibility problem, the NORTA method *is* a viable method even in high-dimensional problems. We recommend augmenting NORTA with the EC heuristic as:

1. Given the desired marginals F_i and covariance matrix Σ_X, estimate the covariance matrix Λ_Z using numerical search/closed form expression.

2. If Λ_Z is positive semidefinite, then set $\Sigma_Z = \Lambda_Z$.

3. If Λ_Z is indefinite, use the EC algorithm in Section 4.3 to find an "approximately close" correlation matrix Σ_Z.

4. Run NORTA with Σ_Z.

CHAPTER 5

GENERATION USING CHESSBOARD DISTRIBUTIONS

The previous chapter argues that though NORTA fails to work for a large portion of feasible covariance specifications in higher dimensions, a modified NORTA procedure can be successfully adopted to produce random vectors with the right marginals and a covariance matrix that approximates the desired one. This, we feel, is essentially the best option for generating random vectors of high dimensionality.

We can, however, do a better job of matching feasible covariances in moderate dimensions. This chapter will discuss how chessboard distributions can be used to model partially specified random vectors. Our motivation for studying chessboard distributions comes from a broader need in stochastic simulation for an easily applied class of distributions that can capture a range of features of a desired distribution. Indeed, researchers in a variety of fields have sought such a class. See, for example, Devroye (1986), Johnson (1987) and Biller and Ghosh (2005) for a survey of these efforts. We shall limit the discussion here to the random vector specification we have studied throughout this thesis, the case where one specifies the (one dimensional) marginal distributions and some measure of dependence that is usually the (Spearman) rank covariance matrix or the (Pearson) product moment covariance matrix, but we do note that one can consider other features such as joint probabilities of certain regions and so forth (Ghosh and Henderson 2001) while modeling using chessboard distributions.

The chief argument in favour of adopting a chessboard-based generation approach is the fact that chessboards can match almost any (in a precise sense, see Section 5.2) feasible covariance specification. But, the setup time needed in solving

the linear programs that define a chessboard distribution can be expensive, and hence their practical viability might be limited to modeling moderate dimensional random vectors. The new results we develop (Section 5.3) on their modeling power show that they can perform quite satisfactorily in moderate dimensions.

We believe that for non-Gaussian marginals, it is more appropriate to use rank covariance than product-moment covariance. Recall that the product-moment covariance between two random variables X and Y with distribution functions F and G respectively is given by

$$EXY - EXEY,$$

and the rank covariance is given by

$$EF(X)G(Y) - EF(X)EG(Y).$$

The product moment covariance is well-defined when X and Y have finite second moment, while the rank covariance is always defined. In the case where F and G are continuous, $F(X)$ and $G(Y)$ are uniformly distributed. Hence, one can reduce a study of rank covariance of random vectors with arbitrary continuous distributions to one of product moment covariance of uniform random variables on $(0, 1]$. (We adopt the convention of open intervals on the left and closed on the right. The choice is immaterial for absolutely continuous distributions.)

We therefore focus on the case of generating a random vector with uniform marginals and a desired product moment covariance matrix. The distribution function of a random vector with uniform marginals on $(0, 1]$ is known as a *copula*. The term was coined in Sklar (1959), and Nelsen (1999) is a useful recent reference.

We describe chessboard distributions as a subclass of a new class of distributions that we call *replicated copulas*. Our interest in replicated copulas lies primarily in

their use in furthering our understanding of chessboard distributions, but they are interesting in their own right. In Section 5.1 we review some high-level facts about chessboard distributions and introduce replicated copulas. Then, in Section 5.2 we review what we feel are the main results of Chapter 2 that are of interest to us here. Many of these results are extended to replicated copulas without any additional work. We shall then present such results in terms of replicated copulas, but omit giving the full proofs. We shall also shed further light on the class of distributions that cannot be exactly matched by chessboard distributions and certain replicated copulas (Theorem 5.2.4). We continue in Section 5.3 by developing some new results that help determine the computational effort needed in modeling partial random vectors with chessboard distributions. In particular, we provide bounds on the relative size (in a certain precise sense) of the subset of feasible covariance specifications that can be matched to a chessboard distribution discretized to a level n.

5.1 Chessboard Distributions and Replicated Copulas

A chessboard copula is a member of a class of copulas with a special structure. We call this the family of *replicated* copulas. The structure of a replicated copula for a random vector X is easily described. For notational convenience we confine our discussion to the three-dimensional case; the d dimensional case is similar. We divide $(0,1]^3$ into a large grid of rectangular regions (cells) with sides parallel to the coordinate axes. Let $n \geq 1$ be an integral parameter that determines the level of division that is performed. The range $(0,1]$ of the ith variable is divided into n equal-length sub-intervals by the set of points $\{y_{i,k} = \frac{k}{n},\ k = 0, \ldots, n\}$. Denote the cells as $C(j_1, j_2, j_3)$, indexed by $j_1, j_2, j_3 = 1, \ldots, n$. Within each cell $C(j_1, j_2, j_3)$

the replicated distribution follows an appropriately scaled and translated version of a copula $\mathcal{C}(j_1, j_2, j_3)$. We call this copula the (j_1, j_2, j_3)th *base* copula. Suppose $Z = (Z_1, Z_2, Z_3)$ is a random vector distributed as $\mathcal{C}(j_1, j_2, j_3)$. Then, conditional on being in cell $C(j_1, j_2, j_3)$, the replicated random vector X can be obtained from Z as

$$X_1 = \frac{Z_1}{n} + y_{1,j_1-1}, \quad \text{and similarly for } X_2, X_3. \tag{5.1}$$

The base copulas could be the same or could vary over different cells. We limit this discussion to replicated copulas that have the same base copula in each cell. So, in essence we divide the region $(0, 1]^3$ into non-overlapping cells and replicate a given copula within the cells. Theorem 5.1.1 below shows that a function created by this replication operation is a valid copula.

Johnson and Kotz (2004) study similar replicated bivariate distributions, which they term *cloned* distributions. They also assume the mass assigned to each cell to be the same, while we allow it to vary. To be more precise, let $q(j_1, j_2, j_3) = P(X \in C(j_1, j_2, j_3))$ be the mass assigned to the copula replicated at cell $C(j_1, j_2, j_3)$. Then, the cloned distributions of Johnson and Kotz (2004) assume that each $q(j_1, j_2, j_3) = 1/n^3$. Our copula construction technique allows the $q(j_1, j_2, j_3)$ values to vary. This helps us match desired dependence measures like the covariance matrix subject to the constraint that the constructed copula is a valid joint distribution function.

Theorem 5.1.1 shows that the copula created for the replicated random vector is a valid distribution function. To ensure this, we will require the $q(j_1, j_2, j_3)$ satisfy

$$\sum_{j_2,j_3=1}^{n} q(j_1, j_2, j_3) \;=\; P(X_1 \in (y_{1,j_1-1},\, y_{1,j_1}]) = \frac{1}{n}, \quad \forall j_1 = 1, \ldots, n,$$

$$\begin{aligned}
\sum_{j_1,j_3=1}^{n} q(j_1,j_2,j_3) &= P(X_2 \in (y_{2,j_2-1},\, y_{2,j_2}]) = \frac{1}{n}, \quad \forall j_2 = 1,\ldots,n, \qquad (5.2)\\
\sum_{j_1,j_2=1}^{n} q(j_1,j_2,j_3) &= P(X_3 \in (y_{3,j_3-1},\, y_{3,j_3}]) = \frac{1}{n}, \quad \forall j_3 = 1,\ldots,n,\\
q(j_1,j_2,j_3) &\geq 0 \quad \forall j_1,j_2,j_3 = 1,\ldots,n.
\end{aligned}$$

Theorem 5.1.1 *If the distribution of X is a replicated copula with cell probabilities q satisfying the constraints (5.2), then X has uniform marginals.*

A proof of this result for the chessboard copula case can be found in Theorem 2.1.1. We present this slightly generalized version because it is useful in understanding the nature of replicated distributions.

Proof: Let the marginal distribution function of X_1 be denoted by $F_1(\cdot)$. We have to show that $F_1(x) = x$ for $x \in (0,1]$. Let Z represent a random vector corresponding to the base copula $\mathcal{C}$. The components of X and Z are related as in (5.1). For any $x \in (y_{1,i-1},\ y_{1,i}]$, we have that

$$\begin{aligned}
F_1(x) &= \sum_{j_1 \leq i-1} \sum_{j_2,j_3=1}^{n} q(j_1,j_2,j_3) + \\
&\quad \sum_{j_2,j_3=1}^{n} P(y_{1,i-1} < X_1 \leq x | X \in C(i,j_2,j_3)) \cdot q(i,j_2,j_3)\\
&= \frac{i-1}{n} + \sum_{j_2,j_3=1}^{n} P(0 < Z_1 \leq n(x - y_{1,i-1})) \cdot q(i,j_2,j_3)\\
&= \frac{i-1}{n} + \sum_{j_2,j_3=1}^{n} n(x - \frac{i-1}{n}) q(i,j_2,j_3)\\
&= \frac{i-1}{n} + x - \frac{i-1}{n} = x
\end{aligned}$$

as required. The first equation follows by conditioning on the cell in which the random vector lies, and the second is obtained from (5.2) and the transformation that relates X with Z. The third equation uses the fact that Z_1 is uniformly

distributed and the final equation again uses (5.2). A similar result holds for X_2 and X_3, and so the joint distribution has uniform marginals. □

Replicated copulas can be used to achieve desired covariance matrices by solving a linear program formulated with the $q(j_1, j_2, j_3)$s as variables. The LP formulation is similar to that in the case of chessboard distributions, and is given in Section 5.2. Much of the theory developed for covariance matching with chessboard copulas extends to the replicated case in a straightforward fashion, as we shall see shortly. So, this technique concludes either by finding $q(j_1, j_2, j_3)$s that give a joint distribution for X with the desired properties, or by determining that no joint distribution can be constructed for X with these properties.

Chessboard copulas, as introduced in Chapter 2, are replicated copulas where the base copula is that of independent uniform random variables. Chessboard distributions are essentially the "piecewise-uniform copulae" that Mackenzie (1994) developed. Mackenzie (1994) identifies chessboard copulas with maximum entropy that match a given covariance matrix, assuming that the covariance matrix can be matched. We did not make this assumption in Chapter 2, and developed the chessboard family partly to provide a procedure to check whether given covariance matrices can be matched. The product-form copula has a density such that each component is independent of the other, and hence its replication in cells makes the components of X conditionally independent (conditional on being in the cell) with marginal distributions given by the uniform distribution restricted to the cell. This special form has an advantage in that it leads to a simple scheme for generating samples from the chessboard copula.

There are many possible methods for generating random vectors with replicated copulas. The methods vary in terms of their time and storage requirements

for setup, and for generating random vectors once the setup is complete. We shall now describe a generic approach that requires a moderate amount of time and storage for setup, but once the setup is complete requires very little time to generate random vectors. Let d denote the dimension of the random vector X with marginal distribution functions $F_1, \ldots, F_{d-1}$ and F_d. Suppose that $q(\cdot; d)$ and $\mathcal{C}$ together represent the replicated distribution constructed for X. (We use $q(\cdot; d)$ to compactly denote $q(\cdot, \ldots, \cdot)$; similarly, we use $C(j; d)$ for the cell $C(j_1, \ldots, j_d)$, where j denotes the vector of indices $(j_1, \ldots, j_d)$.) The algorithm is as follows.

1. Generate the index vector j of the cell containing X from the probabilities $q(\cdot; d)$.

2. Generate X from its conditional distribution given that it lies in the cell $C(j; d)$: an appropriately scaled version of $\mathcal{C}$.

The first step can be performed efficiently using, for example, the alias method. The alias method, developed by Walker (1977) and discussed in detail in Law and Kelton (2000), can generate the appropriate cell in constant time, and requires $O(m)$ storage and $O(m)$ setup time, where m is the number of positive $q(j; d)$ values. If $q(\cdot; d)$ is an extreme-point solution to the linear programs similar to those developed in Chapter 2, then there are on the order of nd strictly positive cell probabilities. This follows from a standard result in linear programming that any extreme point solution to a system of m linear equalities in nonnegative variables has at most m strictly positive values. The exact number of positive values depends on the number of equality constraints in the LP and the degree to which the extreme-point solution is degenerate. (A degenerate extreme-point solution is one with less than m strictly positive values.)

The fact that $m = O(nd)$ is relatively small can be viewed as an advantage with respect to variate generation since it reduces the setup time required to implement the alias method. However, it can also be viewed as a disadvantage in terms of modeling power. For a given dimension d and discretization level n there are n^d cells. Of these, $O(nd)$ receive strictly positive probabilities $q(j; d)$. So as the dimension d increases, the fraction of cells receiving positive probabilities is vanishingly small. This means that the set of values that the random vector X can assume is somewhat limited.

Mackenzie (1994) avoids this problem by maximizing the entropy of the discrete distribution $q(\cdot; d)$. In this case, all of the cells receive positive probability. However, the problem of maximizing the entropy of q subject to linear constraints is a convex optimization problem that is more difficult to solve than the LPs discussed in this thesis.

Suppose cell $C(j; d)$ is chosen in Step 1 above. Conditional on X lying in this cell, the components $(X_1, \ldots, X_d)$ of X are jointly distributed according to a transformed version of $\mathcal{C}$. Suppose $Z = (Z_1, \ldots, Z_d)$ is a random vector distributed as $\mathcal{C}$. A sample of X can be obtained by first sampling a Z and then transforming Z as in (5.1). Thus, for instance, if $\mathcal{C}$ is the product-form copula, as in Chapter 2, then in Step 2, we can independently generate each component from its respective conditional (marginal) distribution. The efficiency of this step clearly depends on the form of $\mathcal{C}$. The product-form base copula requires d independent uniform random variables to generate a sample of Z. On the other hand, if the base copula were the *maximally-correlated* copula, where $Z_1 = Z_2 = \ldots = Z_d$, then only one uniform random variable need be generated to get a sample of Z.

Thus, in general, we can sample rapidly from a replicated copula once such a

copula is constructed to match a given covariance matrix. One important issue that needs to be addressed is the time it takes for the replicated copula to be set up. We shall address this aspect with respect to chessboard copulas in Section 5.3 after summarizing and extending some of the key results of Chapter 2 to replicated copulas.

5.2 Extending Chessboard Results to Replicated Copulas

Replicated copulas can be constructed by formulating covariance matching linear programs similar to those introduced for constructing chessboard distributions with uniform marginals and covariance matrix.

If Σ^X be the covariance matrix of a replicated random vector X, then we want to minimize the distance $r(\Sigma^X, \Sigma)$ between Σ^X and the desired Σ, where

$$r(\Sigma^X, \Sigma) = \sum_{1 \le i < j \le 3} \left| \Sigma^X_{ij} - \Sigma_{ij} \right|.$$

Now, X has uniform marginals so $EX_i = 1/2$ for $i = 1, 2, 3$. Also, by conditioning on the cell containing X we see that

$$\begin{aligned} EX_1X_2 &= \sum_{j_1,j_2,j_3} q(j_1, j_2, j_3) E[X_1X_2 | X \in C(j_1, j_2, j_3)] \\ &= \sum_{j_1,j_2,j_3} q(j_1, j_2, j_3)\ \mu_{12}(j_1, j_2), \end{aligned}$$

where

$$\mu_{lm}(i, k) = E\left[\ X_lX_m \mid X_l \in (y_{l,i-1}, y_{l,i}] \text{ and } X_m \in (y_{m,k-1}, y_{m,k}]\ \right]$$

is the conditional joint moment of X_l and X_m given that X_l lies in the ith subinterval and X_m in the kth one.

In general then, the quantity EX_1X_2 is a weighted sum of (and hence linear in) $q(j_1, j_2, j_3)$s, as is Σ^X_{13} and Σ^X_{23}. Using the standard trick introduced in Section 2.1,

we can attempt to match Σ^X to Σ using the LP

$$\begin{aligned} \min \quad & \sum_{i=1}^{2} \sum_{j=i+1}^{3} (Z_{ij}^{+} + Z_{ij}^{-}) \qquad (5.3) \\ \text{subject to} \quad & \Sigma_{ij}^{X} - \Sigma_{ij} = Z_{ij}^{+} - Z_{ij}^{-}, \ i = 1 \text{ to } 2 \text{ and } j = i+1 \text{ to } 3 \\ & Z_{ij}^{+} \geq 0, Z_{ij}^{-} \geq 0, \text{ together with constraints (5.2)}, \end{aligned}$$

where the variables $Z_{ij}^{\pm}$ are again defined as in Section 2.1. The LP is always feasible since $q(j_1, j_2, j_3) = n^{-3}$, for all j_1, j_2, j_3, represents a feasible solution. Also, the objective function of the LP is bounded below by 0, so an optimal solution exists.

The properties derived in Section 2.2 for chessboard LPs depend on the bounds derived in (2.10) on the objective function (2.5). These bounds can be modified for the objective function of the LP (5.3) above.

The bounds are developed by assuming that a random vector $\tilde{X}$ with uniform marginals and covariance matrix Σ exists, and modifying the distribution of $\tilde{X}$ to that of a random vector X that has a replicated structure. The modification consists of keeping the total mass within each cell constant, but making the conditional distribution within the cell that of the (scaled version of) the base copula. The distribution of X then gives a feasible solution to the LP (minus the bounds), and we can bound the change in the covariances resulting from this modification of the distribution.

Let

$$\tilde{q}(j_1, j_2, j_3) = P(X \in C(j_1, j_2, j_3)) = P(\tilde{X} \in C(j_1, j_2, j_3)).$$

Observe that

$$\begin{aligned} \text{Cov}(X_1, X_2) - \Sigma_{12} &= EX_1X_2 - E\tilde{X}_1\tilde{X}_2 \\ &= \sum_{j_1, j_2, j_3 = 1}^{n} (\mu_{12}(j_1, j_2) - E[\tilde{X}_1\tilde{X}_2 | \tilde{X} \in C(j_1, j_2, j_3)]) \cdot \end{aligned}$$

$$\tilde{q}(j_1, j_2, j_3). \tag{5.4}$$

But

$$y_{1,j_1-1}\ y_{2,j_2-1} \leq E[\tilde{X}_1\tilde{X}_2|\tilde{X} \in C(j_1, j_2, j_3)] \leq y_{1,j_1}\ y_{2,j_2}. \tag{5.5}$$

Combining (5.4) with (5.5) we see that

$$\mathrm{Cov}(\tilde{X}_1, \tilde{X}_2) - \Sigma_{12} \leq \sum_{j_1,j_2,j_3=1}^{n} \tilde{q}(j_1, j_2, j_3)(\mu_{12}(j_1, j_2) - y_{1,j_1-1}\ y_{2,j_2-1}) \tag{5.6}$$

$$\mathrm{Cov}(\tilde{X}_1, \tilde{X}_2) - \Sigma_{12} \geq \sum_{j_1,j_2,j_3=1}^{n} \tilde{q}(j_1, j_2, j_3)(\mu_{12}(j_1, j_2) - y_{1,j_1}\ y_{2,j_2}). \tag{5.7}$$

The bounding expressions can be simplified, once the μ_{12}s are determined from the base copula, to a form similar to that of (2.10). Note that each difference term in the summations above is bounded by $1/n$. These bounds can thus be shown to converge to 0 as $n \to \infty$ at the rate $1/n$.

The question that naturally arises is whether the replicated copula construction technique is as effective as the chessboard technique of Section 2.2 is. This is indeed the case, as we shall see now. We summarize some key results from Chapter 2.2 that detail the power of the chessboard construction technique to match covariance specifications and extend them to replicated copulas. In giving these results, we allow the random vector to have arbitrary, but finite, dimension $d > 1$.

Theorem 5.2.1 *A covariance matrix is infeasible for uniform marginals if, and only if, the replicated-copula LP (5.3), augmented with the bounds (5.6) and (5.7), is infeasible for some $n \geq 1$.*

Theorem 2.2.3 proves the chessboard case, and the same proof holds for replicated copulas, where the LP (2.5) is supplemented with the appropriate bounds. This result establishes that if one of the augmented LPs is infeasible for *any* discretization level n, then the proposed covariance matrix is infeasible. Furthermore,

the theorem establishes that if a covariance matrix is infeasible, then one will eventually discover this by solving an LP with n sufficiently large.

A more positive characterization of the modeling power of chessboards, to the effect that "chessboards can match all feasible covariance matrices", however does not hold. Recall from Example 1 in Section 2.2 that chessboard distributions cannot exactly match all feasible covariance matrices. This is potentially a limitation that even replicated copulas face. Theorem 5.3.3 at the end of Section 5.3 shows that replicated copulas constructed with the same base copula replicated over all cells cannot match all feasible covariances for a finite level of discretization n.

Theorem 2.2.5, though, says that the error in the covariance matrix when using chessboard distributions can be made arbitrarily small. This extends to replicated copulas, so that they can arbitrarily-closely approximate any feasible covariance matrix. The proof of Theorem 5.2.2 is virtually identical to that of Theorem 2.2.5 and omitted.

Theorem 5.2.2 *Suppose that Σ is feasible. Then for all $\epsilon > 0$, there exists a replicated distribution with covariance matrix Λ with the property that $r(\Sigma, \Lambda) < \epsilon$.*

Theorem 2.2.7 shows that chessboard distributions can not only closely approximate any feasible covariance matrix, but they can *exactly* match "almost all" feasible covariance matrices. To state this result precisely, we reuse some terminology and a definition stated in Chapter 2. We can view a $d \times d$ covariance matrix as an element of the real vector space $\mathbb{R}^{d(d-1)/2}$. This follows because there are $d(d-1)/2$ elements above the diagonal, the matrix is symmetric, and the diagonals are determined by the marginal distributions. Let Ω denote the set of feasible covariance matrices. We view Ω as a subset of $d(d-1)/2$ dimensional Euclidean space, and Ω is also a subset of the hypercube $[-1/12, 1/12]^{d(d-1)/2}$.

Proposition 2.2.6 in Chapter 2 had established that the set Ω is nonempty, convex, closed and full-dimensional.

Let A° and ∂A denote the interior and boundary of the set A respectively. Theorem 2.2.7 and Theorem 2.2.9 from Chapter 2 can be combined to give that

Theorem 5.2.3 *There is a chessboard distribution with covariance matrix Σ if, and only if, $\Sigma \in \Omega^\circ$.*

Thus, chessboards can match any covariance matrix that fall within the in (non-empty) interior of the set Ω, but cannot match those on the boundary. The "if" part of Theorem 5.2.3 remains true for replicated copulas and is shown by the same arguments as in Theorem 2.2.7, but it is also possible for replicated copulas to achieve some points on the boundary of Ω. For instance, continuing Example 1, suppose that the base copula corresponds to a perfectly correlated pair of uniform random variables. Then one can achieve a covariance of 1 with $n = 1$.

Proposition 2.2.6 establishes that the set Ω has a non-zero finite Lebesgue measure (in $\mathbb{R}^{d(d-1)/2}$), while $\partial\Omega$ is a zero Lebesgue measure set. It follows from Theorem 5.2.3 that chessboard distributions can model almost any (in a Lebesgue measure sense) feasible covariance matrix from Ω, and replicated copulas can do at least as well. Given a feasible covariance matrix, the procedure to determine a corresponding replicated distribution is then straightforward: one solves the augmented LP based on (2.5) for a chosen level of discretization n and if the optimal objective value is greater than 0, the parameter n is increased successively till the objective value drops to 0 or an acceptable value.

Theorem 5.2.3 establishes that a distribution F that achieves a covariance matrix $\Sigma \in \partial\Omega$ will not have a chessboard distribution structure. We can prove a slightly stronger result regarding the structure of any such distribution F that

has a covariance matrix in $\partial\Omega$ (and uniform marginals). This will have a bearing on the power of replicated distributions in modeling covariance matrices from the boundary $\partial\Omega$.

The distribution F can be decomposed into a singular part F_s and an absolutely continuous part F_{ac} with respect to Lebesgue measure restricted to $(0,1]^3$ (the *Lebesgue Decomposition*; see Billingsley 1995, p. 414). Thus,

$$F = F_{ac} + F_s.$$

Moreover, the absolutely continuous part has a density f_{ac} in the sense of the *Radon-Nikodym derivative* of F_{ac}.

Theorem 5.2.4 *Suppose that f_{ac} is defined as above where $\Sigma \in \partial\Omega$. There cannot exist an open set G such that*

$$f_{ac}(x) \geq \phi > 0 \qquad a.e.\ in\ G. \tag{5.8}$$

A property holds *almost everywhere (a.e.)* on the set G if it is true for all $x \in G$ except over a subset of Lebesgue measure 0.

Proof: For notational ease we give a proof in the 3-dimensional case. The general case is virtually identical. Suppose such a G exists. We can reassign f_{ac} to have value ϕ over any measurable subset of measure zero where the f_{ac} cannot be bounded away from 0, without changing the function F_{ac}. Thus, we assume that f_{ac} is bounded away from 0 by at least ϕ over all $x \in G$. We can choose an open ball $B(x,\epsilon)$ within G and an open cubical region C with sides aligned to the axes within $B(x,\epsilon)$ such that the interior of C is non-empty. Split f_{ac} into two parts f_C and $f_{\bar{C}}$ defined as:

$$f_C(x) = \begin{cases} \phi & x \in C \\ 0 & \text{elsewhere} \end{cases} \quad \text{and} \quad f_{\bar{C}}(x) = \begin{cases} f_{ac}(x) - \phi & x \in C \\ f_{ac}(x) & \text{elsewhere} \end{cases}.$$

Let u and v be the endpoints that define C so that

$$C = \{(x, y, z) \in (0, 1]^3 : u_1 < x \leq v_1, u_2 < y \leq v_2, u_3 < z \leq v_3\}.$$

Divide the cell C into 4 (equal size) sub-cells,

$$\begin{aligned} C_{ab} &= \{(x, y, z) \in C : u_1 + (a-1)\frac{v_1 - u_1}{2} < x \leq u_1 + a\frac{v_1 - u_1}{2}, \\ &\quad u_2 + (b-1)\frac{v_2 - u_2}{2} < x \leq u_2 + b\frac{v_2 - u_2}{2}\} \end{aligned}$$

for $1 \leq a, b \leq 2$.

Define a new distribution H from F as follows. The singular parts H_s and F_s coincide, as do the $h_{\bar{C}}$ and $f_{\bar{C}}$ parts respectively of the absolutely continuous density. The density h_C is defined such that it assigns a value 2ϕ to each of the cells C_{11}, and C_{22}, and set h_C to be 0 in the cells C_{ab} for $a \neq b$. Then it is straightforward to show that H has uniform marginals, that the $(1, 2)$th covariance is strictly increased, and that the other covariances remain unchanged. A similar argument increasing the density in the cells C_{ab} with $a \neq b$ shows that the covariance can be strictly decreased.

Convexity of Ω then implies that Σ must lie in the interior Ω° which is a contradiction, and the proof is complete.□

It is conceivable that the support of distributions that match matrices from $\partial\Omega$ could consist of sets of zero Lebesgue measure in $\mathbb{R}^d$ or exotic sets like Cantor sets with no interior but non-zero measure. Generating random vectors with such distributions could thus prove difficult.

Theorem 5.2.4 also tells us that if one uses a base copula with sets as described by (5.8) in its support, then one cannot construct replicated copulas to exactly match covariance matrices from $\partial\Omega$.

In summary then, replicated distributions

- can detect if a given matrix is infeasible,
- can arbitrarily closely approximate any feasible covariance matrix,
- can exactly match any feasible covariance matrix in the interior of the set of feasible covariance matrices, but
- *might not* exactly match any covariance matrix on the boundary of the set of feasible covariance matrices. (Replicated copulas might match some covariances on the boundary.)

5.3 More on Modeling Power

In Section 5.1, we propose a method to generate from chessboard distributions (and in general all replicated distributions). We have posited that once the chessboard distribution is set up, generation should be fast. However, for the method to be viable, one should be able to set up the distribution in a reasonable amount of time. Critical to this is the time it takes to obtain a solution for the linear programs based on (2.5). Efficient algorithms are available to solve linear programs, and they are theoretically known to be solvable in time which is at worst a polynomial function of the size (in binary representation of all the data) of the problem. The setup time thus depends on the size n of the discretization that is used. We now turn to the question of how large n needs to be to match covariance matrices in Ω° for a fixed dimension d of the random vector.

Let Ω^n represent the set of covariance matrices that chessboard distributions of size n can match. That Ω^n is non-empty is easily seen: an argument along the lines of the proof of Theorem 2.2.7 or Theorem 5.2.4 shows that a non-empty convex full-dimensional set of covariances centered around the origin of $\mathbb{R}^{d(d-1)/2}$ (which

corresponds to a random vector with independent components) is contained within Ω^n. It is also convex since for any two matrices in Ω^n, any convex combination of them can be achieved by the corresponding convex combination of their chessboard distributions. We have shown in Theorem 2.2.9 and in Theorem 5.2.3 that chessboards can achieve any feasible covariance matrix in the interior of Ω for some finite n. Thus, in a sense the sequence $\{\Omega^n, n \geq 1\}$ should grow to cover the whole of Ω as $n \to \infty$; we shall establish this rigorously. Let $\mathbb{L}$ represent the Lebesgue measure on the real vector space $\mathbb{R}^{d(d-1)/2}$. Then, the theorem below gives the main result we shall establish in this section:

Theorem 5.3.1 *Suppose Ω^n represents the set of covariance matrices that chessboards of size n match. Let $m(d) = d(d-1)/2$ and $\ell = \sqrt{m(d)}$. Then,*

$$a) \qquad \Omega^n \subseteq \mathcal{U}(\frac{1}{n^2}), \qquad \textit{and}$$

$$b) \qquad \mathcal{L}(\frac{\ell}{n}) \subseteq \Omega^n.$$

The sets $\mathcal{L}(\cdot)$ and $\mathcal{U}(\cdot)$ will be defined shortly. But, we shall first state a corollary to this theorem that is of practical interest:

Corollary 5.3.2

$$\mathbb{L}(\Omega) - \frac{K(d)}{n} \leq \mathbb{L}(\Omega^n) \leq \left(1 + \frac{1}{n^2}\right)^{-m(d)} \mathbb{L}(\Omega),$$

where $K(d)$ is some positive value that depends on d.

The implications of Corollary 5.3.2 are clear. The rate at which chessboard distributions can cover the set Ω of feasible covariance matrices is at least of the order $\frac{1}{n}$, but can be no faster than a factor of the order $\left(1 + \frac{1}{n^2}\right)^{-m(d)}$ ($\approx \left(1 - \frac{m(d)}{n^2}\right)$ when n is large). Of course, the lower bound is of interest only when the expression evaluates to a positive value. This result then establishes the efficiency of

the chessboard copula-based random vector modeling technique. Recall that no complete characterization of the set Ω is currently known, and hence the position of a specific covariance matrix relative to Ω can not be easily given. Thus, it might still be hard to predict the size of the chessboard distribution needed to achieve a given covariance matrix.

We shall prove each set inclusion assertion of Theorem 5.3.1 separately. We need some notation and definitions first. We write $B(x, \epsilon) = \{y : \ ||x - y||_2 < \epsilon\}$ to represent the (open) $\epsilon-$ball centered at x, defined under the l_2 metric on the space $\mathbb{R}^{m(d)}$. $B(0, 1)$, the unit open ball centered at the origin, is simply denoted B. Thus, $B(x, \epsilon) = x + \epsilon B$, where the notation vM denotes the set $\{vx : x \in M\}$ for any scalar v, and $y + M = \{y + x : x \in M\}$.

Call any compact, convex set with a non-empty interior a *convex body*. The *Minkowski subtraction* set operation on two convex bodies M and N can be defined (Schneider 1993a, Chapter 3) as

$$M \sim N \stackrel{\triangle}{=} \{x \in M : x + N \subset M\}.$$

This set operation is given various other names in the literature, but we shall follow the conventions used in Schneider (1993a). A convex body E is said to be *centered* if it contains the origin as an interior point. Sangwine-Yager (1988) defines, for an $\epsilon > 0$, the ϵth *relative inner parallel body* of a convex body M with respect to a centered convex body E to be $M \sim \epsilon E$. The *relative inradius* $r(M; E)$ is defined to be

$$r(M; E) = \sup\{\epsilon : x + \epsilon E \subset M, \text{ for some } x \in M\}.$$

By this definition, we have that the set $M \sim \epsilon E$ has a non-empty interior for any $0 < \epsilon < r(M; E)$, and hence would have a positive Lebesgue measure. Sangwine-Yager (1988) notes that the set $M \sim (r(M; E)E)$ has zero Lebesgue measure.

When the ball B is used as E, the sets $M \sim \epsilon B$ are simply called the ϵth inner parallel body, and $r(M; B)$ the inradius of M.

The families of sets $\mathcal{U}(\epsilon)$ and $\mathcal{L}(\epsilon)$, used in Theorem 5.3.1, are indexed by ϵ and defined as

$$\mathcal{U}(\epsilon) \stackrel{\triangle}{=} \Omega \sim \epsilon\Omega, \quad \text{and} \tag{5.9}$$

$$\mathcal{L}(\epsilon) \stackrel{\triangle}{=} \Omega \sim \epsilon B. \tag{5.10}$$

Recall from Section 2.2 that the origin is contained in the interior of Ω (see, for example, the proof of Theorem 2.2.7.) A matrix z belongs to $\mathcal{U}(\epsilon)$ if the set $z + \epsilon\Omega$ also belongs to $\mathcal{U}(\epsilon)$. The sets $\mathcal{U}(\epsilon)$ have a non-empty interior for all $0 < \epsilon < 1$ (by the definition of the relative inradius, $r(\Omega; \Omega) = 1$.)

Similarly, a matrix z belongs to $\mathcal{L}(\epsilon)$ if the $\epsilon-$ball $B(z, \epsilon) \subset \Omega$. This has a simple interpretation, in that $\mathcal{L}(\epsilon)$ is the subset of points in Ω that are at least an ϵ l_2-distance away from the boundary $\partial\Omega$. Again, the sets $\mathcal{L}(\epsilon)$ can be empty for large ϵ, but by the preceding discussion are non-empty for $0 < \epsilon < r(\Omega; B)$.

Brannen (1997), Sangwine-Yager (1988), Schneider (1993a;b) give various bounds on the Lebesgue measures of these relative inner parallel bodies. These bounds show that $\mathbb{L}(\mathcal{U}(\epsilon)) \to \mathbb{L}(\Omega)$ and $\mathbb{L}(\mathcal{L}(\epsilon)) \to \mathbb{L}(\Omega)$ as $\epsilon \to 0$ (recall that the bounded convex set Ω is of finite non-zero Lebesgue measure.) We shall use a simple version of these bounds in the proof of Corollary 5.3.2.

We are now ready to prove the first part of Theorem 5.3.1.

Proof of Theorem 5.3.1(a): For notational ease we prove the result for $d = 3$. The case $d > 3$ is proved similarly. We establish the result by showing that a certain operation on any n-sized chessboard distribution, which has covariance matrix $\Sigma \in \Omega^n$, constructs a distribution function that represents a new covariance

matrix in Ω not too far from Σ. One can obtain a bound on the distance between these matrices, which then gives the result.

Let $\{q(j_1, j_2, j_3)\}$ represent an LP solution that constructs a chessboard distribution corresponding to covariance matrix Σ. Then

$$\begin{aligned}\Sigma_{12} &= EX_1X_2 - EX_1EX_2 \\ &= \sum_{j_1,j_2,j_3=1}^{n} E[X_1X_2|X \in C(j_1, j_2, j_3)] \cdot q(j_1, j_2, j_3) - \frac{1}{4}. \end{aligned} \tag{5.11}$$

Let $Z = (Z_1, Z_2, Z_3)$ be a random vector endowed with the base copula being replicated within the cells and $\Sigma^Z \in \Omega$ be its covariance matrix. In our case of a chessboard distribution, Z is a vector of independent uniform random variables and $\Sigma^Z = (0, 0, 0)$. Let y_{i,j_i}, $i = 1, 2, 3,\ j_i = 0, \ldots, n$ be the points that define the grid as in (5.1). Since $EZ_i = 1/2$, $i = 1, 2, 3$, we see that

$$\begin{aligned} E[X_1X_2|X \in C(j_1, j_2, j_3)] &= E\left[\left(\frac{Z_1}{n} + y_{1,j_1-1}\right)\left(\frac{Z_2}{n} + y_{2,j_2-1}\right)\right] \\ &= \frac{EZ_1Z_2}{n^2} + \frac{EZ_1\, y_{2,j_2-1} + EZ_2\, y_{1,j_1-1}}{n} + y_{1,j_1-1}\, y_{2,j_2-1} \\ &= \frac{EZ_1Z_2}{n^2} + \frac{y_{2,j_2-1} + y_{1,j_1-1}}{2n} + y_{1,j_1-1}\, y_{2,j_2-1} \\ &= \frac{EZ_1Z_2}{n^2} + t(j_1, j_2), \end{aligned} \tag{5.12}$$

where $t(j_1, j_2)$ is a function only of the indices j_1 and j_2.

Suppose now that we replace the product-form copula in each cell of the chessboard distribution with another copula represented by the random vector Z'. The result is still a valid replicated copula because of Theorem 5.1.1, and represents the distribution of a random vector X' say. If Σ' is the covariance matrix of X', then

$$\Sigma'_{12} = \sum_{j_1,j_2,j_3=1}^{n} E[X'_1X'_2|X' \in C(j_1, j_2, j_3)] \cdot P(X' \in C(j_1, j_2, j_3)) - \frac{1}{4}$$

$$= \sum_{j_1,j_2,j_3=1}^{n} \left(\frac{EZ_1'Z_2'}{n^2} + t(j_1,j_2) \right) \cdot q(j_1,j_2,j_3) - \frac{1}{4}. \tag{5.13}$$

Let $\Sigma^{Z'}$ be the covariance matrix of Z'. The net change in covariance due to the replacement operation is, from (5.11), (5.12) and (5.13),

$$\begin{aligned} \Sigma'_{12} - \Sigma_{12} &= \sum_{j_1,j_2,j_3=1}^{n} \frac{1}{n^2}(EZ_1'Z_2' - EZ_1Z_2) \cdot q(j_1,j_2,j_3) \\ &= \frac{1}{n^2}(\Sigma_{12}^{Z'} - \Sigma_{12}^{Z}) \\ &= \frac{1}{n^2}\Sigma_{12}^{Z'}. \end{aligned} \tag{5.14}$$

We have used the fact that $\Sigma_{12}^{Z} = 0$, since Z represents the product-form copula. Equation (5.14) holds for every component of the covariance matrix. Hence,

$$\Sigma' = \Sigma + \frac{1}{n^2}\Sigma^{Z'}.$$

Observe that Z' can have any arbitrary covariance matrix in Ω, including those from the boundary $\partial\Omega$. Thus, the set $\Sigma + \frac{1}{n^2}\Omega \subset \Omega$, and we have established that for any $\Sigma \in \Omega^n$, $\Sigma \in \mathcal{U}(\frac{1}{n^2})$. This gives us the result. $\square$

The proof shows that Ω_n is a subset of $\mathcal{U}(\frac{1}{n^2})$. This result is tight in a certain sense. From Example 1, a chessboard of size n can come to within $\frac{1}{12n^2}$ of achieving a covariance value of $\frac{1}{12}$. Thus, since $\Sigma = \frac{1}{12} - \frac{1}{12n^2}$ belongs to the boundary of $\mathcal{U}(\frac{1}{n^2})$, Ω^n can have some points in common with the boundary $\partial\mathcal{U}(\frac{1}{n^2})$. Whether $\Omega^n = \mathcal{U}(\frac{1}{n^2})$ is however unknown.

We have shown that the set Ω^n can be "bounded above" by the set $\mathcal{U}(\frac{1}{n^2})$. We will now show that the Ω^n can be lower-bounded by sets from the $\mathcal{L}$ family.

We need a preliminary result for the proof below. Note that all norms in a real vector space are equivalent; refer, for example, Pg. 53 of Golub and Van Loan (1996). Thus, we have that for any $x \in \mathbb{R}^{m(d)}$

$$||x||_\infty \ \leq \ ||x||_2 \ \leq \ell \ ||x||_\infty. \tag{5.15}$$

Proof of Theorem 5.3.1(b): The result is trivial if $\mathcal{L}(\frac{\ell}{n})$ is empty. We shall thus assume it is non-trivial. In Section 2.1, we had introduced a technique to derive bounds for the objective function of the chessboard LP (2.5), and these led to the bounds in (2.10). Part of our argument here is identical to that given while deriving (2.10).

Choose a $\Sigma \in \partial\Omega$. Since Ω is closed, there exists a joint distribution function (a copula), call it F, that achieves this covariance matrix. We modify F to construct a chessboard copula F^n. The modification consists of keeping the total mass assigned to each cell by F constant, but making the conditional distribution within the cell uniform. The distribution function constructed by this process is a valid chessboard distribution since it satisfies the constraints (5.2). Let Σ^n represent the covariance matrix of F^n. We showed in Section 2.1 how a bound can be derived for the change in the individual covariances from Σ to Σ^n resulting from this modification of the distribution F. Specifically, Equation (2.10) derives the bounds to be

$$\begin{aligned}(\Sigma(i,j) - \Sigma^n(i,j))^+ &\leq \frac{1}{2n} - \frac{1}{4n^2} \quad \text{and} \\ (\Sigma(i,j) - \Sigma^n(i,j))^- &\leq \frac{1}{2n} + \frac{1}{4n^2} \quad 1 \leq i < j \leq 3,\end{aligned}$$

where $(x)^+$ refers to the positive part of x, $\max\{x, 0\}$, and $(x)^-$ refers to its negative part, $-\min\{x, 0\}$. We thus have that

$$\begin{aligned}|\Sigma(i,j) - \Sigma^n(i,j)| &= (\Sigma(i,j) - \Sigma^n(i,j))^+ + (\Sigma(i,j) - \Sigma^n(i,j))^- \\ &\leq \frac{1}{n} \ \forall\ 1 \leq i < j \leq 3. \qquad (5.16)\end{aligned}$$

(The bound can actually be tightened because only one of either the positive or the negative part is non-zero, but this looser bound is sufficient for our purpose.)

Equation (5.16) essentially tells us that the maximum difference between the components of Σ^n and Σ is less than $\frac{1}{n}$. In other words, Σ^n is within an $l_\infty-$distance

$\frac{1}{n}$ from Σ. From (5.15), we then have that $\Sigma^n \in B(\Sigma, \frac{\ell}{n})$. Hence, for any $\Sigma \in \partial\Omega$, we can pick a matrix $\Sigma^n \in \Omega^n$ such that $\Sigma^n \in B(\Sigma, \frac{\ell}{n})$.

Now, suppose the assertion in the theorem is false, and there exists a $\Lambda \in \mathcal{L}(\frac{\ell}{n})$ that does not belong to Ω^n. Since the set Ω^n is convex, the celebrated Separating Hyperplane Theorem (for e.g., Luenberger 1969, Theorem 3, Section 5.12) gives us a hyperplane $\mathcal{H}$ through Λ that separates the point Λ from Ω^n.

Consider a line $\mathcal{N}$ passing through Λ orthogonal to the hyperplane $\mathcal{H}$. Busemann (1958, Chapter 1) tells us that since Λ is in the interior of Ω, this line intersects the boundary $\partial\Omega$ of the convex Ω at exactly two points, say Σ^1 and Σ^2. By definition, the point Λ from within $\mathcal{L}(\frac{\ell}{n})$ does not belong to the sets $B(\Sigma^i, \frac{\ell}{n}),\ i = 1, 2$. Moreover, since the line $\mathcal{N}$ is orthogonal to the hyperplane $\mathcal{H}$, we have from the Projection Theorem (Luenberger 1969, Pg. 51) that Λ is the unique point in $\mathcal{H}$ that minimizes the l_2 distance between the points $\Sigma^i,\ i = 1, 2$ and points in $\mathcal{H}$. Thus, $\mathcal{H}$ separates the sets $B(\Sigma^i, \frac{\ell}{n}),\ i = 1, 2$ from Λ. Moreover, the sets lie on opposite sides of $\mathcal{H}$ since $\Lambda \in \Omega^\circ$. Thus, at least one ball is separated from Ω^n by the hyperplane $\mathcal{H}$. But this contradicts the earlier observation that one can always choose a point that belongs to Ω^n from each ball $B(\Sigma^i, \frac{\ell}{n}),\ i = 1, 2$. This completes the proof. $\square$

We will now prove Corollary 5.3.2, the final result in this chapter. We need the following result in its proof. Brannen (1997, Theorem 1) give a lower bound (which they quote from Sangwine-Yager 1988) for the Lebesgue measure of a relative inner parallel body $M \sim \epsilon E$:

$$\mathbb{L}(M \sim \epsilon E) \geq \mathbb{L}(M) - \epsilon S(M; E) + R(m(d), \epsilon),$$

where $S(M; E)$ represents the *relative surface area* of M with respect to E, and the

function $R(m(d), \epsilon)$ consists of non-negative terms in ϵ. Both Brannen (1997) and Sangwine-Yager (1988) show that $R(m(d), \epsilon)$ can be expressed as a polynomial in ϵ of degree two or higher. So, if $S(M; E)$ is a finite positive value and $\epsilon < 1$, then the Lebesgue measure of the inner parallel set grows at least at the rate ϵ, i.e.,

$$\mathbb{L}(M \sim \epsilon E) \geq \mathbb{L}(M) - k\epsilon \tag{5.17}$$

for some positive k, where k possibly depends on the dimension $m(d)$.

Equation (5.17) requires that the function $S(M; E)$ be positive. Brannen (1997), Sangwine-Yager (1988) give detailed expressions for this function, but we shall not require them here except to note that this function is positive if, and only if, a collection of $m(d)$ line segments can be chosen, where $(m(d) - 1)$ of them are from the set M and one from E, with linearly independent directions (this is from Schneider 1993a, Theorem 5.1.7, Pg. 277). This condition holds, for instance, if the convex bodies M and E have non-empty interiors.

Proof of Corollary 5.3.2:

(5.17) gives that for an n large enough such that $\left(\frac{\ell}{n}\right) < 1$,

$$\mathbb{L}(\Omega) - k(d)\left(\frac{\ell}{n}\right) \leq \mathbb{L}(\mathcal{L}\left(\frac{\ell}{n}\right)),$$

where $k(d)$ is a positive value that depends on d. This equation, along with Theorem 5.3.1(b), thus gives us the lower bound in the statement of the result with $K(d) = k(d)\ell$.

For the upper bound, note that

$$\mathcal{U}(\frac{1}{n^2}) \subseteq \left(1 + \frac{1}{n^2}\right)^{-1} \Omega.$$

This is easily seen to be true: if $A \in \mathcal{U}(\frac{1}{n^2})$, then it follows from the definition of the sets $\mathcal{U}$ (5.9) that $A(1 + \frac{1}{n^2}) \in \Omega$. Hence, $A \in (1 + \frac{1}{n^2})^{-1}\Omega$.

Now, the Lebesgue measure of the linearly scaled set $(1+\frac{1}{n^2})^{-1}\Omega$ is given by $(1+\frac{1}{n^2})^{-m(d)}\mathbb{L}(\Omega)$ (see Billingsley 1995, Theorem 12.2). This, along with Theorem 5.3.1(a), establishes the upper bound on the Lebesgue measure of Ω^n and we are done. □

An intermediary step in the proof above shows that chessboard distributions with discretization level n cannot then come closer than within a factor $(1+\frac{1}{n^2})^{-m(d)}$ ($\approx (1-\frac{m(d)}{n^2})$ for large n) of the matrices in $\partial\Omega$. This reaffirms the fact that no chessboard can match a matrix in $\partial\Omega$ (Theorem 5.2.3).

Analogous results can be derived for replicated copulas. These results shall not be discussed here in detail, but we conclude this chapter with an observation about the modeling power of replicated copulas. Suppose $\Sigma^{\mathcal{C}}$ is the covariance matrix of the base copula $\mathcal{C}$ of the replicated copulas to be used in modeling uniform random vectors with specified feasible covariance matrices. Theorem 5.3.1(a) can be reworked with essentially the same steps till the second expression in (5.14). In the chessboard case, the Σ^Z in this expression is 0, but in the replicated case $\Sigma^Z = \Sigma^{\mathcal{C}}$. The rest of the proof works out in a similar fashion, but with slightly different definitions. To be concrete, the new covariance matrix Σ' produced by replacing the base copula with a copula of covariance $\Sigma^{Z'}$ will be given by

$$\Sigma' = \Sigma + \frac{1}{n^2}(\Sigma^{Z'} - \Sigma^{\mathcal{C}}).$$

Let Ω^n be the set of covariances achievable by $n-$sized replicated copulas with base copula $\mathcal{C}$. This leads to the fact that

$$\Omega^n \subseteq \{\Sigma : \Sigma + \frac{1}{n^2}(\Lambda - \Sigma^{\mathcal{C}}) \in \Omega,\ \forall \Lambda \in \Omega\} = \Omega \sim \frac{1}{n^2}(\Omega - \Sigma^{\mathcal{C}}) \stackrel{\triangle}{=} S(n),$$

where we use the convention $(A-b) = \{x-b : x \in A\}$. The set $(\Omega - \Sigma^{\mathcal{C}})$ is just the set Ω translated such that $\Sigma^{\mathcal{C}}$ now coincides with the origin.

The implications of this are as follows: if $\Sigma^{\mathcal{C}} \in \Omega^{\circ}$, then the compact, convex set $(\Omega - \Sigma^{\mathcal{C}})$ still contains the origin in its interior. In this case, the set $S(n)$ will again be strictly contained within Ω, and hence the $\Sigma^{\mathcal{C}}-$based replicated copulas can not match any matrix from $\partial\Omega$ for finite n, i.e., the "only if" part of Theorem 5.2.3 will hold for these replicated copulas.

On the other hand, if $\Sigma^{\mathcal{C}} \in \partial\Omega$, then the set Ω^n (and thus $S(n)$) would touch $\partial\Omega$ at least at the point $\Sigma^{\mathcal{C}}$ because a trivial replicated copula with $n = 1$ achieves this covariance. But, even in this case, Ω^n cannot match all of Ω for any finite n. This is easily proved by a counterexample. Suppose it did match every $\Sigma \in \Omega$ for some $n \geq 1$. Consider a line $\mathcal{N}$ through $\Sigma^{\mathcal{C}}$ and the origin. Since Ω is compact, convex and the origin is in its interior, this line intersects $\partial\Omega$ at exactly one other point, say $\bar{\Sigma}$. By the supposition, a $\Sigma^{\mathcal{C}}$-based replicated copula, say of size n_0, achieves this value. Then, by the argument above, a Σ' of value

$$\Sigma' = \bar{\Sigma} + \frac{1}{n_0^2}(\bar{\Sigma} - \Sigma^{\mathcal{C}})$$

can also be achieved by replacing the base copula with a copula generating $\bar{\Sigma}$. This point is however outside Ω. This proves that the supposition was wrong in the first place, and we have that:

Theorem 5.3.3 *For a fixed base copula, replicated copulas cannot achieve all covariances in the feasible set Ω for any finite level of discretization n.*

Whether replicated copulas with different base copulas in each cell can match all of Ω for some finite n is an open problem. We conjecture that this is impossible.

BIBLIOGRAPHY

A. Alfakih and H. Wolkowicz. Matrix completion problems. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming: Theory, Algorithms and Applications*, pages 533–545, Boston, 2000. Kluwer.

B. Biller and S. Ghosh. Multivariate input processes. In B. L. Nelson and S. G. Henderson, editors, *Handbooks in Operations Research and Management Science: Simulation*, To Appear. Elsevier Science, Amsterdam, 2005.

P. Billingsley. *Probability and Measure*. Wiley, New York, 2nd edition, 1986.

P. Billingsley. *Probability and Measure*. Wiley, New York, 3rd edition, 1995.

N. S. Brannen. The Wills conjecture. *Transactions of the American Mathematical Society*, 349:3977–3987, 1997.

H. Busemann. *Convex Surfaces*. Number 6 in Interscience Tracts in Pure and Applied Mathematics. Interscience, New York, 1958.

M. C. Cario and B. L. Nelson. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Technical report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois, 1997.

H. Chen. Initialization for NORTA: Generation of random vectors with specified marginals and correlations. *INFORMS Journal on Computing*, 13(4):312–331, 2001.

R. T. Clemen and T. Reilly. Correlations and copulas for decision and risk analysis. *Management Science*, 45:208–224, 1999.

R. M. Cooke. Markov and entropy properties of tree- and vine-dependent variables. In *Proceedings of the ASA Section on Bayesian Statistical Science*, 1997.

G. B. Dantzig. Converting a converging algorithm into a polynomially bounded algorithm. Technical Report 91–5, Systems Optimization Laboratory, Dept of Operations Research, Stanford University, Stanford, California, 1991.

L. Devroye. *Non-uniform Random Variate Generation*. Springer-Verlag, New York, 1986.

A. Edelman. *Eigenvalues and condition numbers of random matrices*. PhD thesis, Department of Mathematics, MIT, Cambridge, MA, 1989.

F. G. Frobenius. *Gessammelte Abhandlungen*, volume 3. Springer, Berlin, 1968.

S. Ghosh and S. G. Henderson. Chessboard distributions. In B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, editors, *Proceedings of the 2001 Winter Simulation Conference*, pages 385–393, Piscataway NJ, 2001. IEEE.

G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, M.D., 3rd edition, 1996.

L. Guttman. Enlargement methods for computing the inverse matrix. *Annals of Mathematical Statistics*, 17:336–343, 1946.

S. G. Henderson. Mathematics for simulation. In B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, editors, *Proceedings of the 2001 Winter Simulation Conference*, pages 83–94, Piscataway NJ, 2001. IEEE.

S. G. Henderson, B. A. Chiera, and R. M. Cooke. Generating "dependent" quasi-random numbers. In J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, editors, *Proceedings of the 2000 Winter Simulation Conference*, pages 527–536,

Piscataway NJ, 2000. IEEE.

R. R. Hill and C. H. Reilly. Composition for multivariate random variables. In *Proceedings of the 26th conference on Winter simulation*, pages 332–339. Society for Computer Simulation International, 1994. ISBN 0-7803-2109-X.

R. R. Hill and C. H. Reilly. The effects of coefficient correlation structure in two-dimensional knapsack problems on solution procedure performance. *Management Science*, 46(2):302–317, 2000.

T. J. Hodgson, J. A. Joines, S. D. Roberts, K. A. Thoney, and J. R. Wilson. Satisfying due-dates in large job shops: Characteristics of "real" problems. Preprint, 2000.

R. Iman and W. Conover. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics: Simulation and Computation*, 11:311–334, 1982.

H. Joe. *Multivariate Models and Dependence Concepts*. Chapman & Hall, London, 1997.

C. R. Johnson. Matrix completion problems: a survey. *Proceedings of Symposia in Applied Mathematics*, 40:171–198, 1990.

M. E. Johnson. *Multivariate Statistical Simulation*. Wiley, New York, 1987.

N. L. Johnson and S. Kotz. Cloning of distributions. Preprint, 2004.

M. G. Kendall. *A course in the geometry of n dimensions*. Charles Griffin and Co., London, 1961.

W. Kruskal. Ordinal measures of associaton. *Journal of the American Statistical*

Association, 53:814–861, 1958.

D. Kurowicka and R. M. Cooke. Conditional, partial and rank correlation for the elliptical copula; dependence modelling in uncertainty analysis. Technical report, Delft University of Technology, Mekelweg 4, 2628CD Delft, Netherlands, 2001.

A. M. Law and W. D. Kelton. *Simulation Modeling and Analysis*. McGraw-Hill, New York, 3rd edition, 2000.

C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1974.

S. T. Li and J. L. Hammond. Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. *IEEE Transactions on Systems, Man, and Cybernetics*, 5:557–561, 1975.

D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley and Sons, Inc., New York, 1969.

P. M. Lurie and M. S. Goldberg. An approximate method for sampling correlated random variables from partially-specified distributions. *Management Science*, 44:203–218, 1998.

G. R. Mackenzie. *Approximately Maximum-Entropy Multivariate Distributions with Specified Marginals and Pairwise Correlations*. PhD thesis, Department of Decision Sciences, University of Oregon, Eugene OR, 1994.

K. V. Mardia. A translation family of bivariate distributions and Fréchet's bounds. *Sankhya*, A32:119–122, 1970.

G. Marsaglia and I. Olkin. Generating correlation matrices. *SIAM Journal of*

Scientific and Statistical Computing, 5:470–475, 1984.

A. M. H. Meeuwissen and R. M. Cooke. Tree dependent random variables. Technical Report 94-28, Department of Mathematics, Delft University of Technology, Delft, The Netherlands, 1994.

R. B. Nelsen. *An Introduction to Copulas*. Number 139 in Lecture Notes in Statistics. Springer-Verlag, New York, 1999.

B. L. Nelson. The NORTA method: An application in agricultural insurance. Personal Communication, 2003.

B. L. Nelson and M. Yamnitsky. Input modeling tools for complex problems. In *Proceedings of the 30th conference on Winter simulation*, pages 105–112. IEEE Computer Society Press, 1998. ISBN 0-7803-5134-7.

D. V. Ouellette. Schur complements and statistics. *Linear Algebra and Its Applications*, 36:187–295, 1981.

J. R. Sangwine-Yager. Bonnesen-style inequalities for Minkowski relative geometry. *Transactions of the American Mathematical Society*, 307:373–382, 1988.

B. W. Schmeiser. Simulation experiments. In D. P. Heyman and M. J. Sobel, editors, *Stochastic Models*, volume 2 of *Handbooks in Operations Research and Management Science*, pages 295–330. Elsevier Science, North-Holland, Amsterdam, 1990.

B. W. Schmeiser and A. J. G. Babu. Beta variate generation via exponential majorizing functions. *Operations Research*, 28:917–926, 1980.

B. W. Schmeiser and A. J. G. Babu. Errata: Beta variate generation via exponential majorizing functions. *Operations Research*, 31:802, 1983.

R. Schneider. *Convex bodies: The Brunn-Minkowski Theory*. Cambridge University Press, Cambridge, U.K., 1993a.

R. Schneider. Convex surfaces, curvature and surface area measures. In P. M. Gruber and J. M. Wills, editors, *Handbook of Convex Geometry*, volume A, pages 273–299. Elsevier Science, Amsterdam, 1993b.

A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8:229–231, 1959.

R. L. Smith. Efficient Monte Carlo procedures for generating random feasible solutions over bounded regions. *Operations Research*, 32:1296–1308, 1984.

Y. L. Tong. *The Multivariate Normal Distribution*. Springer–Verlag, New York, 1990.

P. A. G. van der Geest. An algorithm to generate samples of multi-variate distributions with correlated marginals. *Computational Statistics and Data Analysis*, 27:271–289, 1998.

L. Vandenberghe and S. Boyd. Semidenite programming. *SIAM Review*, 38:49–95, 1996.

S. Vincent. Input data analysis. In J. Banks, editor, *The Handbook of Simulation*, pages 55–91. John Wiley & Sons, New York, 1998.

A. J. Walker. An efficient method for generating discrete random variables with general distributions. *ACM Transactions on Mathematical Software*, 3:253–256, 1977.

W. Whitt. Bivariate distributions with given marginals. *The Annals of Statistics*, 4:1280–1289, 1976.

H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors. *Handbook of Semidefinite Programming: Theory, Algorithms and Applications*. Kluwer, Boston, 2000.