Copulas with Maximum Entropy

Julia Piantadosi, Phil Howlett, Jonathan Borwein

Abstract

We shall find a multi-dimensional checkerboard copula of maximum entropy that matches an observed set of grade correlation coefficients. This problem is formulated as the maximization of a concave function on a convex polytope. Under mild constraint qualifications we show that a unique solution exists in the core of the feasible region. The theory of Fenchel duality is used to reformulate the problem as an unconstrained minimization which is well solved numerically using a Newton iteration. Finally, we discuss the numerical calculations for some hypothetical examples and describe how this work can be applied to the modelling and simulation of monthly rainfall.

1 Introduction

An m-dimensional *copula* where $m \geq 2$, is a continuous, m-increasing, grounded probability distribution $C : [0,1]^m \mapsto [0,1]$ on the unit m-dimensional hyper-cube with uniform marginal probability distributions. A *checkerboard* copula is a distribution with a corresponding density defined almost everywhere by a step function on an m-uniform subdivision of the hypercube. The *grade correlation coefficients* are the respective correlations of the marginal distributions. We shall solve the following constrained optimization problem.

Problem 1 (Preliminary statement). *Find a checkerboard copula of maximum entropy that matches a given set of grade correlation coefficients.*

We first formulate the primal problem as the maximum of a concave function on a convex polytope and show that subject to reasonable constraint qualifications there is a unique solution in the core of the feasible region. To calculate numerical solutions we shall use the theory of Fenchel duality to formulate an equivalent unconstrained minimization problem which we then solve using a Newton iteration. The formulation and solution of the primal and dual problems leans heavily on the established theory of convex analysis. The relevant theory can be found in the book by Borwein and Lewis [5] and elsewhere.

1.1 Motivation

The optimal management of urban storm-water systems can be demonstrated effectively using computer graphics simulation packages. To drive these simulations it is necessary to develop stochastic rainfall simulations that accurately reflect observed rainfall patterns. In a recent model of a storm-water management system at Parafield in South Australia we wished to simulate monthly rainfall [15] in successive months where the marginal distributions for each month were modelled as random variables with prescribed gamma distributions and where the monthly rainfall records for the past 100 years showed small but sometimes significant values of the grade correlation coefficients between various different months. Simulations in which monthly rainfall totals were modelled as mutually independent random variables showed a much higher standard deviation for the yearly totals than that obtained from the observed yearly totals.

We reasoned that the standard deviation could be reduced if we were able to incorporate an appropriate level of dependence amongst the individual monthly totals. Hence our interest in the use of an elementary copula to construct an appropriate joint distribution function which would preserve the prescribed marginal distributions and still match the observed grade correlation coefficients. Thus we pose the following general mathematical problem.

> Suppose we wish to model the joint distribution of a collection of random variables with prescribed marginal distributions and known grade correlation coefficients. Can we find a checkerboard copula of maximum entropy that matches the known grade correlation coefficients?

Since the entropy of the checkerboard copula measures the inherent disorder we are effectively finding the most disordered joint distribution with the prescribed marginal distributions that matches the observed second order statistics. Of course we expect a better solution if we increase the number of subdivisions in the checkerboard. A typical rainfall application will be discussed as a Case Study later in the paper.

1.2 Previous work

A *doubly stochastic* matrix $\boldsymbol{h} = [h_{ij}] \in \mathbb{R}^{n \times n}$ is a matrix with $h_{ij} \geq 0$ and with $\sum_j h_{ij} = 1$ for all $i = 1, 2, \ldots, n$ and $\sum_i h_{ij} = 1$ for all $j = 1, 2, \ldots, n$. In a recent paper Piantadosi *et al* [16] used a doubly stochastic matrix and a corresponding two-dimensional checkerboard copula to construct a joint distribution that preserved the prescribed marginal distributions and matched a known grade correlation coefficient in such a way that the entropy of the doubly stochastic matrix was maximized. We wish to extend those ideas to a multi–dimensional setting. Birkhoff [4, 5] proved that the set of doubly stochastic matrices is a convex polytope and that a doubly stochastic matrix is extremal if and only if it is a permutation matrix. Piantadosi *et al* used this idea to formulate the two dimensional problem as one of finding an appropriate convex combination of permutation matrices.

In this paper we consider the analogous problem in a general multi-dimensional setting. Can we find a checkerboard copula of maximum entropy to construct a multi-dimensional probability distribution that preserves prescribed marginal distributions and that matches the observed grade correlation coefficients? Although the set of multi-stochastic hyper-matrices is a convex polytope it is not possible when $m > 2$, to express each element as a convex combination of multi-stochastic matrices whose elements are either zero or one. That is there are some vertices that are not multi-dimensional permutation hyper-matrices. In this sense we could say there is no multi-dimensional analogue of the Birkhoff theorem. Thus we need to use a different method of analysis than the one proposed in [16].

1.3 Multi-dimensional copulas

Let $C : [0,1]^m \mapsto [0,1]$ be an m–dimensional copula for $m \geq 2$. If $F_r : \mathbb{R} \mapsto [0,1]$ are prescribed continuous distributions for the real valued random variable X_r for each $r = 1, \ldots, m$ then the function $G : \mathbb{R}^m \mapsto [0,1]$ defined by

$$G(x_1, \ldots, x_m) = C(F_1(x_1), \ldots, F_m(x_m))$$

is a joint probability distribution for the vector valued random variable $\boldsymbol{X} = (X_1, \ldots, X_m)$ with the marginal distribution for X_r defined by F_r for each $r = 1, 2, \ldots, m$. The joint density $g : \mathbb{R}^m \mapsto [0, \infty)$ is defined almost everywhere and is given by the formula

$$g(x_1, \ldots, x_m) = c(F_1(x_1), \ldots, F_m(x_m))f_1(x_1) \cdots f_m(x_m)$$

where $c : [0,1]^m \mapsto [0,\infty)$ is the density for the joint distribution defined by C and where $f_r : \mathbb{R} \mapsto [0,\infty)$ for each $r = 1, 2, \ldots, m$ are the densities for the prescribed marginal distributions.

If related real valued random variables $U_r = F_r(X_r)$ are defined for each $r = 1, 2, \ldots, m$ then each U_r is uniformly distributed on $[0, 1]$ and the copula C describes the distribution of the vector valued random variable $\boldsymbol{U} = (U_1, \ldots, U_m)$. When the joint distribution G is not known there are many standard choices for C but the best choice will depend on what else is known about the observed distribution.

It is often the case in practice that the grade correlation coefficients

$$\rho_{r,s} = \frac{E[(F_r(X_r) - 1/2)(F_s(X_s) - 1/2)]}{\sqrt{E[(F_r(X_r) - 1/2)^2] \cdot E[(F_s(X_s) - 1/2)^2]}} = \frac{E[(U_r - 1/2)(U_s - 1/2)]}{\sqrt{E[(U_r - 1/2)^2] \cdot E[(U_s - 1/2)^2]}}$$

for $1 \le r < s \le m$ are known. How should one choose C in order to satisfy these additional constraints? For each fixed $n \in \mathbb{N}$ we will consider a subdivision of the interval $[0, 1]$ into n equal length subintervals and a corresponding m-uniform subdivision of the unit hyper-cube $[0, 1]^m$ into n^m congruent hyper-cubes. For each sufficiently large $n \in \mathbb{N}$ we will show that an elementary copula $C = C_{\boldsymbol{h}}$ can be chosen by defining a density $c = c_{\boldsymbol{h}}$ in the form of a step function defined by an m-dimensional hyper-matrix $\boldsymbol{h}$ such that the density takes a constant non-negative value on each hyper-cube of the subdivision. The hyper-matrix will be chosen in such a way that the known grade correlations are imposed and the entropy of the hyper-matrix is maximized. Since entropy is a measure of disorder our solution for $c_{\boldsymbol{h}}$ can be interpreted as the *most disordered* or *least prescriptive* choice of step function for the selected value of n that satisfies the required correlation constraints. The corresponding checkerboard copula $C = C_{\boldsymbol{h}}$ is the most disordered such copula.

1.4 Literature review

The text by Nelsen [14] is a comprehensive reference for the general theory of copulas. Much of the work on copulas has been related to construction of general families with desirable properties. We refer to Nelsen [14] for a summary of the basic theory. Although copulas have been used in one form or another for many years much of the theory has been developed relatively recently. For instance Schmid and Schmidt [18] consider non-parametric estimation of multivariate measures of association. Genest, Ghoudi and Rivest [8] and Chen *et al* [7] consider semi-parametric estimation of dependence parameters in multivariate families of distributions, while Genest *et al* [9] review Goodness-of-fit tests for copulas and suggest some new procedures. Jaworski [11] looks at copulas where the *so-called* diagonal section is prescribed. More generally, Borwein, Lewis and Nussbaum [6] study the existence of distributions with prescribed marginals in an infinite dimensional setting.

There are other papers where the emphasis is on the application of copulas. For instance Venter *et al* [22] use multivariate copulas for financial modelling. Zakaria *et al* [24] use a bivariate skew *t*-distribution and an appropriate copula to model monthly rainfall at two sites in the Murray-Darling Basin. This work relies on general results obtained by Azzalini [1] and Sun *et al* [21] related to the multivariate skew *t*-distribution. Following earlier work by Holm and Alouini [10], Zakaria *et al* also propose rainfall models with correlated bivariate gamma distributions.

2 Notation

Let $m \in \mathbb{N}$ with $m \ge 2$ and let $\boldsymbol{X} = (X_1, \ldots, X_m) \in \mathbb{R}^m$ be a vector–valued random variable with joint probability density $g : \mathbb{R}^m \mapsto \mathbb{R}$. The corresponding marginal probability densities are

$$g_r(x_r) = \int_{\mathbb{R}^{m-1}} g(\boldsymbol{x}) \, d\boldsymbol{\pi}_r^c \boldsymbol{x}$$

for all $x_r \in \mathbb{R}$ and each $r = 1, \ldots, m$ where we write $\boldsymbol{x} = (x_1, \ldots, x_m) \in \mathbb{R}^m$ and where the projection $\boldsymbol{\pi}_r : \mathbb{R}^m \mapsto \mathbb{R}$ onto the x_r-axis and the complementary projection $\boldsymbol{\pi}_r^c : \mathbb{R}^m \mapsto \mathbb{R}^{m-1}$ are defined for each $r = 1, 2, \ldots, m$ by

$$\boldsymbol{\pi}_r \boldsymbol{x} = x_r \quad \text{and} \quad \boldsymbol{\pi}_r^c \boldsymbol{x} = \begin{cases} (x_2, \ldots, x_m) & \text{if} \quad r = 1 \\ (x_1, \ldots, x_{r-1}, x_{r+1}, \ldots, x_m) & \text{if} \quad r = 2, \ldots, m-1 \\ (x_1, \ldots, x_{m-1}) & \text{if} \quad r = m. \end{cases}$$

In simulation of random events it may be convenient to construct a joint probability distribution where the corresponding marginal distributions are already known. The method of copulas is one possible way. If the joint distribution is known and the marginal distributions are continuous then the copula is uniquely defined. We refer to the book by Nelsen [14] for the fundamental theory. In our discussion we assume the joint distribution is not completely known and so the question of uniqueness is not relevant. Nevertheless it is convenient to assume that the given marginal distributions are continuous. Let $c : [0,1]^m \mapsto [0, \infty)$ be a joint probability density on the unit m–dimensional hyper-cube with uniform marginal densities. That is, the marginal densities $c_r : [0,1] \mapsto [0, \infty)$, satisfy the conditions

$$c_r(u_r) = 1 \quad \Leftrightarrow \quad \int_{[0,1]^{m-1}} c(\boldsymbol{u}) \, d\boldsymbol{\pi}_r^c \boldsymbol{u} = 1$$

for all $u_r \in [0,1]$ and each $r = 1, \ldots, m$. The distribution $C : [0,1]^m \mapsto [0,1]$ defined by

$$C(\boldsymbol{u}) = \int_{\bigtimes_{i=1}^n [0, u_i]} c(\boldsymbol{v}) d\boldsymbol{v}$$

for all $\boldsymbol{u} \in [0,1]^m$ is an m-dimensional copula. The copula C defines a joint distribution for a vector valued random variable $\boldsymbol{U} = (U_1, \ldots, U_m)$ on the unit hyper-cube $[0,1]^m$. Let $f_s : \mathbb{R} \mapsto \mathbb{R}$ be a given probability density with corresponding cumulative distribution function $F_s : \mathbb{R} \mapsto [0,1]$ for each $s = 1, \ldots, m$. Write $\boldsymbol{f} = (f_1, \ldots, f_m) : \mathbb{R}^m \mapsto [0, \infty)^m$ and $\boldsymbol{F} = (F_1, \ldots, F_m) : \mathbb{R}^m \mapsto [0,1]^m$. The joint density $g : \mathbb{R}^m \mapsto [0, \infty)$ defined for the vector valued random variable $\boldsymbol{X} = (X_1, \ldots, X_m)$ by the formula

$$g(\boldsymbol{x}) = c(\boldsymbol{F}(\boldsymbol{x})) \prod_{s=1}^m f_s(x_s)$$

for $\boldsymbol{x} \in \mathbb{R}^m$ has prescribed marginal densities for the real valued random variables X_r given by

$$g_r(x_r) = f_r(x_r) \int_{\mathbb{R}^{m-1}} c(\boldsymbol{F}(\boldsymbol{x})) \prod_{s \neq r} f_s(x_s) \, d\boldsymbol{\pi}_r^c \boldsymbol{x} = f_r(x_r) \int_{[0,1]^{m-1}} c(\boldsymbol{u}) \, d\boldsymbol{\pi}_r^c \boldsymbol{u} = f_r(x_r)$$

for all $x_r \in \mathbb{R}$ and each $r = 1, \ldots, m$ where we have written

$$\boldsymbol{u} = \boldsymbol{F}(\boldsymbol{x}) \quad \Leftrightarrow \quad (u_1, \ldots, u_m) = (F_1(x_1), \ldots, F_m(x_m))$$

for each $\boldsymbol{x} = (x_1 \ldots, x_r) \in \mathbb{R}^m$. The corresponding m-dimensional distribution $G : \mathbb{R}^m \mapsto [0,1]$ is defined in terms of the copula C and the marginal distributions $\boldsymbol{F}$ by the formula

$$G(\boldsymbol{x}) = C(\boldsymbol{F}(\boldsymbol{x}))$$

for all $\boldsymbol{x} \in \mathbb{R}^m$.

3 An elementary form for the joint density

Let $n \in \mathbb{N}$ be a natural number and let $\boldsymbol{h}$ be a non-negative m–dimensional hyper–matrix given by $\boldsymbol{h} = [h_{\boldsymbol{i}}] \in \mathbb{R}^{\ell}$ where $\ell = n^m$ and $\boldsymbol{i} \in \{1, \ldots, n\}^m$ with $h_{\boldsymbol{i}} \in [0,1]$. Define the *marginal sums* $\sigma_r : \{1, \ldots, n\} \mapsto \mathbb{R}$ by the formulae

$$\sigma_r(i_r) = \sum_{\boldsymbol{\pi}_r^c \boldsymbol{i} \, \in \, \{1,2,\ldots,n\}^{m-1}} h_{\boldsymbol{i}}$$

for each $i_r = 1, 2, \ldots, m$. If $\sigma_r(i_r) = 1$ for all $r = 1, 2, \ldots, m$ then we say that $\boldsymbol{h}$ is *multiply stochastic*. Define the partition $0 = a(1) < a(2) < \cdots < a(n+1) = 1$ of the interval $[0,1]$ by setting $a(k) = (k-1)/n$ for each $k = 1, \ldots, n+1$ and define a step function $c_{\boldsymbol{h}} : [0,1]^m \mapsto \mathbb{R}$ almost everywhere by the formula

$$c_{\boldsymbol{h}}(\boldsymbol{u}) = n^{m-1} \cdot h_{\boldsymbol{i}} \quad \text{if} \quad \boldsymbol{u} \in I_{\boldsymbol{i}} = \bigtimes_{r=1}^{m} [a(i_r), a(i_r + 1)]$$

for each $\boldsymbol{i} = (i_1, \ldots, i_m) \in \{1, 2, \ldots, n\}^m$. Now it follows that

$$\int_{[0,1]^m} c_{\boldsymbol{h}}(\boldsymbol{u}) \cdot d\boldsymbol{u} = \sum_{\boldsymbol{i} \, \in \, \{1,\ldots,n\}^m} \int_{I_{\boldsymbol{i}}} c_{\boldsymbol{h}}(\boldsymbol{u}) \cdot d\boldsymbol{u} = \sum_{\boldsymbol{i} \, \in \, \{1,\ldots,n\}^m} n^{m-1} h_{\boldsymbol{i}} \cdot \frac{1}{n^m} = 1$$

and since

$$(c_{\boldsymbol{h}})_r(u_r) = \int_{[0,1]^{m-1}} c_{\boldsymbol{h}}(\boldsymbol{u}) \cdot d\boldsymbol{\pi}_r^c \boldsymbol{u} = \sum_{\boldsymbol{\pi}_r^c \boldsymbol{i} \, \in \, \{1,\ldots,n\}^{m-1}} n^{m-1} h_{\boldsymbol{i}} \cdot \frac{1}{n^{m-1}} = 1$$

for all $r = 1, 2, \ldots, m$ it follows that the step function $c_{\boldsymbol{h}} : [0,1]^m \mapsto [0,\infty)$ is a joint density function for a corresponding copula $C_{\boldsymbol{h}} : [0,1]^m \mapsto [0,1]$ defined by

$$C_{\boldsymbol{h}}(\boldsymbol{u}) = \int_{\bigtimes_{i=1}^{n} [0, u_i]} c_{\boldsymbol{h}}(\boldsymbol{v}) d\boldsymbol{v}$$

for all $\boldsymbol{u} \in [0,1]^m$. The proposed joint density $g_{\boldsymbol{h}} : \mathbb{R}^m \mapsto [0,\infty)$ for the random variable $\boldsymbol{X} = (X_1, \ldots, X_m)$ is defined by

$$g_{\boldsymbol{h}}(\boldsymbol{x}) = c_{\boldsymbol{h}}(\boldsymbol{F}(\boldsymbol{x})) \prod_{s=1}^{m} f_s(x_s)$$

for $\boldsymbol{x} \in \mathbb{R}^m$ and the corresponding distribution function $G_{\boldsymbol{h}} : \mathbb{R}^m \mapsto [0,1]$ is defined in terms of the copula $C_{\boldsymbol{h}}$ and the prescribed marginal distributions $\boldsymbol{F}$ by the formula

$$G_{\boldsymbol{h}}(\boldsymbol{x}) = C_{\boldsymbol{h}}(\boldsymbol{F}(\boldsymbol{x}))$$

for all $\boldsymbol{x} \in \mathbb{R}^m$.

4 The grade correlation coefficients

There exist several measures of association for joint distributions [14]. The most widely known are Kendall's tau and Spearman's rho, both of which measure a form of dependence known as concordance. Spearman's rho is often called the grade correlation coefficient. If x_r are observations from a real valued random variable X_r with cumulative distribution function F_r then the *grade* of x_r is given by $u_r = F_r(x_r)$. Note that the grade u_r can be regarded as an observation of the uniform random variable $U_r = F_r(X_r)$ on $[0,1]$ and that U_r has mean $1/2$ and variance $1/12$. The *grade correlation coefficient*

for the continuous random variables X_r and X_s where $r < s$ is defined as the correlation for the grade random variables $U_r = F_r(X_r)$ and $U_s = F_s(X_s)$ by the formula

$$\rho_{r,s} = \frac{E[(U_r - 1/2)(U_s - 1/2)]}{E[(U_r - 1/2)^2]^{1/2} E[(U_s - 1/2)^2]^{1/2}} = 12 \left(E[U_r U_s] - 1/4 \right).$$

We refer the reader to Nelsen [14] for further details. We will now calculate the grade correlation coefficients for the random variables X_r and X_s where $r < s$ when the joint distribution is defined by $G_{\boldsymbol{h}}(\boldsymbol{x}) = C_{\boldsymbol{h}}(\boldsymbol{F}(\boldsymbol{x}))$. To begin we define the projection $\boldsymbol{\pi}_{rs} : \mathbb{R}^m \mapsto \mathbb{R}^2$ onto the $u_r u_s$-plane and the complementary projection $\boldsymbol{\pi}_{rs}^c : \mathbb{R}^m \mapsto \mathbb{R}^{m-2}$ for $1 \le r < s \le m$ by the formulae

$$\boldsymbol{\pi}_{rs}\boldsymbol{u} = (u_r, u_s)$$

and

$$\boldsymbol{\pi}_{rs}^c \boldsymbol{u} = \begin{cases} (u_3, \ldots, u_m) & \text{if } r = 1, s = 2 \\ (u_2, \ldots, u_{s-1}, u_{s+1}, \ldots, u_m) & \text{if } r = 1, 2 < s < m \\ (u_2, \ldots, u_{m-1}) & \text{if } r = 1, s = m \\ (u_1, \ldots, u_{r-1}, u_{r+1}, \ldots, u_{s-1}, u_{s+1}, \ldots, u_m) & \text{if } 1 < r < s < m \\ (u_1, \ldots, u_{r-1}, u_{r+1}, \ldots, u_m) & \text{if } 1 < r < m - 1, s = m \\ (u_1, \ldots, u_{m-2}) & \text{if } r = m - 1, s = m. \end{cases}$$

We have

$$\begin{aligned} E[U_r U_s] &= \int_{[0,1]^m} u_r u_s \cdot c_{\boldsymbol{h}}(\boldsymbol{u}) \cdot d\boldsymbol{u} \\ &= \sum_{\boldsymbol{i} \in \{1,\ldots,n\}^m} \int_{I_{\boldsymbol{\pi}_{rs}\boldsymbol{i}}} u_r u_s \left[\int_{I_{\boldsymbol{\pi}_{rs}^c \boldsymbol{i}}} c_{\boldsymbol{h}}(\boldsymbol{u}) \cdot d\boldsymbol{\pi}_{rs}^c \boldsymbol{u} \right] d\boldsymbol{\pi}_{rs}\boldsymbol{u} \\ &= \sum_{\boldsymbol{i} \in \{1,\ldots,n\}^m} n h_{\boldsymbol{i}} \left(\int_{a(i_r)}^{a(i_r+1)} u_r \cdot du_r \right) \left(\int_{a(i_s)}^{a(i_s+1)} u_s \cdot du_s \right) \\ &= \frac{1}{n^3} \sum_{\boldsymbol{i} \in \{1,\ldots,n\}^m} h_{\boldsymbol{i}}(i_r - 1/2)(i_s - 1/2). \end{aligned}$$

Thus the grade correlation coefficient is given by

$$\rho_{r,s} = 12 \left(\frac{1}{n^3} \sum_{\boldsymbol{i} \in \{1,\ldots,n\}^m} h_{\boldsymbol{i}}(i_r - 1/2)(i_s - 1/2) \; - \; \frac{1}{4} \right). \tag{1}$$

The maximum possible value for the grade correlation coefficient occurs if

$$\sum_{\boldsymbol{\pi}_{rs}^c \boldsymbol{i} \in \{1,\ldots,n\}^{m-2}} h_{\boldsymbol{i}} = \begin{cases} 1 & \text{if } i_r = i_s \\ 0 & \text{otherwise} \end{cases}$$

in which case we have $\rho_{r,s} = 12 \cdot (E[U_r U_s] - 1/4) = 1 - 1/n^2$. The minimum possible value for the covariance occurs if

$$\sum_{\boldsymbol{\pi}_{rs}^c \boldsymbol{i} \in \{1,\ldots,n\}^{m-2}} h_{\boldsymbol{i}} = \begin{cases} 1 & \text{if } i_r + i_s = n + 1 \\ 0 & \text{otherwise} \end{cases}$$

in which case we have $\rho_{r,s} = 12 \cdot (E[U_r U_s] - 1/4) = -1 + 1/n^2$. Thus,

$$-1 + \frac{1}{n^2} \le \rho_{r,s} \le 1 - \frac{1}{n^2}. \tag{2}$$

5 Entropy

Let $\boldsymbol{h} \in \mathbb{R}^\ell$ be a multiply stochastic hyper–matrix and let $c_{\boldsymbol{h}} : [0,1]^m \to \mathbb{R}$ be the associated elementary joint density defined previously. The *entropy* of $\boldsymbol{h}$ is defined by

$$
\begin{aligned}
J(\boldsymbol{h}) &= (-1) \cdot \int_{[0,1]^m} c_{\boldsymbol{h}}(\boldsymbol{u}) \log_e c_{\boldsymbol{h}}(\boldsymbol{u}) \cdot d\boldsymbol{u} \\
&= (-1) \cdot \sum_{\boldsymbol{i} \in \{1,\ldots,n\}^m} (n^{m-1} h_{\boldsymbol{i}}) \cdot \log_e (n^{m-1} h_{\boldsymbol{i}}) \cdot \frac{1}{n^m} \\
&= (-1) \cdot \frac{1}{n} \sum_{\boldsymbol{i} \in \{1,\ldots,n\}} h_{\boldsymbol{i}} \left[\log_e h_{\boldsymbol{i}} + (m-1) \log_e n \right] \\
&= (-1) \left[\frac{1}{n} \sum_{\boldsymbol{i} \in \{1,\ldots,n\}^m} h_{\boldsymbol{i}} \log_e h_{\boldsymbol{i}} + (m-1) \log_e n \right].
\end{aligned}
\tag{3}
$$

We wish to select a multiply stochastic hyper–matrix $\boldsymbol{h} = [h_{\boldsymbol{i}}] \in \mathbb{R}^\ell$ to match known grade correlation coefficients $\rho_{r,s}$ for all $r < s$ in such a way that the entropy is maximized.

6 Formulation and solution of the primal problem

We can now formulate the problem we wish to solve.

Problem 2 (The primal problem). *Find the hyper-matrix $\boldsymbol{h} \in \mathbb{R}^\ell$ to maximize the entropy*

$$
J(\boldsymbol{h}) = (-1) \left[\frac{1}{n} \sum_{\boldsymbol{i} \in \{1,\ldots,n\}^m} h_{\boldsymbol{i}} \log_e h_{\boldsymbol{i}} + (m-1) \log_e n \right]
\tag{4}
$$

subject to the constraints

$$
\sum_{\boldsymbol{\pi}_r^c \boldsymbol{i} \in \{1,\ldots,n\}^{m-1}} h_{\boldsymbol{i}} = 1
\tag{5}
$$

for all $i_r \in \{1,\ldots,n\}$ and each $r = 1,\ldots,m$ and

$$
h_{\boldsymbol{i}} \geq 0
\tag{6}
$$

for all $\boldsymbol{i} \in \{1,\ldots,n\}^m$ and the additional grade correlation coefficient constraints

$$
12 \left[\frac{1}{n^3} \cdot \sum_{\boldsymbol{i} \in \{1,\ldots,n\}^m} h_{\boldsymbol{i}}(i_r - 1/2)(i_s - 1/2) - 1/4 \right] = \rho_{r,s}
\tag{7}
$$

for $1 \leq r < s \leq m$ where $\rho_{r,s}$ is known for all $1 \leq r < s \leq m$.

In general terms the problem is well posed. There are a finite number of linear constraints on $\boldsymbol{h}$ and so the *feasible* set $\mathcal{F}$ of hyper–matrices satisfying (5,6,7) is a bounded (closed) convex set in $\mathbb{R}^\ell$. The function $J : \mathcal{F} \mapsto [0,\infty)$ is strictly concave. If the interior or core of $\mathcal{F}$ is non-empty then there must be a unique solution for $\boldsymbol{h}$ with strictly positive coordinates. The reader is referred to [2, 5] for a general discussion of the requisite convex analysis and nonlinear optimization.

6.1 Solution of the primal problem

To solve the problem we begin by forming a Lagrangian function

$$
\begin{aligned}
\mathcal{J}(\boldsymbol{h}) \;=\;& n \cdot J(\boldsymbol{h}) + \sum_{r=1}^{m}\sum_{i_r=1}^{n} \lambda_{r,i_r} \left[\sum_{\boldsymbol{\pi}_r^c \boldsymbol{i} \,\in\, \{1,\dots,n\}^{m-1}} h_{\boldsymbol{i}} - 1 \right] \\
&+ \sum_{\boldsymbol{i} \,\in\, \{1,\dots,n\}^m} \mu_{\boldsymbol{i}} h_{\boldsymbol{i}} \\
&+ \sum_{s=1}^{m-1}\sum_{t=s+1}^{m} \nu_{s,t} \left[\sum_{\boldsymbol{i} \,\in\, \{1,\dots,n\}^m} h_{\boldsymbol{i}}(i_s - 1/2)(i_t - 1/2) - \sigma_{s,t} \right]
\end{aligned}
\tag{8}
$$

where we have written $\sigma_{s,t} = n^3(\rho_{s,t}+3)/12$ for each $1 \leq s < t \leq m$ and where $\lambda_{r,i_r} \geq 0$ for each $r = 1,\dots,m$ and $\mu_{\boldsymbol{i}} \geq 0$ for each $\boldsymbol{i} \in \{1,\dots,n\}^m$ and $\nu_{s,t} \geq 0$ for each $1 \leq s < t \leq m$ are Lagrange multipliers.

We now apply the Karush–Kuhn–Tucker (KKT) conditions Our derivation is legitimated by the dual development described in Section 7 and Proposition 2; and refer the reader to [2, 5] for further details. For each fixed $\boldsymbol{j} \in \{1,\dots,n\}^m$ we have the necessary condition

$$
\frac{\partial \mathcal{J}}{\partial h_{\boldsymbol{j}}} = 0
\tag{9}
$$

from which it follows that

$$
-\log_e h_{\boldsymbol{j}} - 1 + \sum_{r=1}^{m} \lambda_{r,j_r} + \mu_{\boldsymbol{j}} + \sum_{s=1}^{m-1}\sum_{t=s+1}^{m} \nu_{s,t}(j_s - 1/2)(j_t - 1/2) = 0.
\tag{10}
$$

By taking an exponential and rearranging the terms we can write each optimal $h_{\boldsymbol{j}}$ in the form

$$
h_{\boldsymbol{j}} = \alpha_{\boldsymbol{j}} \cdot \prod_{r=1}^{m} \beta_{r,j_r} \cdot \prod_{s=1}^{m-1}\prod_{t=s+1}^{m} [\gamma_{s,t}]^{j_s j_t}
\tag{11}
$$

where $\alpha_{\boldsymbol{j}}$, β_{r,j_r} and $\gamma_{s,t}$ are positive constants defined by

$$
\alpha_{\boldsymbol{j}} \;=\; \exp\left[\mu_{\boldsymbol{j}}\right]
$$

$$
\beta_{r,j_r} \;=\; \exp\left[\lambda_{r,j_r} - \frac{1}{m} + \frac{1}{4m}\sum_{s=1}^{m-1}\sum_{t=s+1}^{m} \nu_{s,t} - \frac{1}{2}\left(\sum_{s=1}^{r-1} \nu_{s,r} + \sum_{t=r+1}^{m} \nu_{r,t} \right) j_r \right]
$$

$$
\gamma_{s,t} \;=\; \exp\left[\nu_{s,t}\right]
$$

for each $\boldsymbol{j} \in \{1,\dots,n\}^m$, each $r = 1,\dots,m$ and each $1 \leq s < t \leq m$. In the case where the grade correlation coefficients are zero we can expect that the entropy will be maximized when all the elements $h_{\boldsymbol{j}}$ are equal. Note that in general any solution to the equations (10) requires $h_{\boldsymbol{j}} > 0$ for all $j \in \{1,\dots,m\}$ since all terms in these equations must be finite. Therefore, to solve the equations we may assume that $h_{\boldsymbol{j}} > 0$ for all $\boldsymbol{j} \in \{1,\dots,\}^m$. Consequently it follows from the KKT conditions that $\mu_{\boldsymbol{j}} = 0$ and hence that $\alpha_{\boldsymbol{j}} = 1$ for all $\boldsymbol{j} \in \{1,\dots,n\}^m$. Thus the expression for the optimal form of $h_{\boldsymbol{j}}$ simplifies to

$$
h_{\boldsymbol{j}} = \prod_{r=1}^{m} \beta_{r,j_r} \cdot \prod_{s=1}^{m-1}\prod_{t=s+1}^{m} [\gamma_{s,t}]^{j_s j_t}
\tag{12}
$$

for all $\boldsymbol{j} \in \{1,\dots,n\}^m$.

6.2 A solution scheme for the primal problem

One possible method of numerical solution would be to proceed as follows.

> Choose fixed values for the constants $\gamma_{s,t} > 0$ and substitute the expressions (12) into the key equations (5). Now use a Newton iteration to solve for the unknowns $\beta_{r,j_r} > 0$ and hence calculate $\rho_{s,t}$.

In the case where $m = 3$ and $n = 4$ we have 64 variables $h_{\boldsymbol{j}}$ for $\boldsymbol{j} \in \{1, \ldots, 4\}^3$ and the optimal form of $h_{\boldsymbol{j}}$ is given by

$$h_{\boldsymbol{j}} = \beta_{1,j_1}\beta_{2,j_2}\beta_{3,j_3}[\gamma_{1,2}]^{j_1 j_2}[\gamma_{1,3}]^{j_1 j_3}[\gamma_{2,3}]^{j_2 j_3}$$

where $\beta_{1,j_1} > 0$, $\beta_{2,j_2} > 0$ and $\beta_{3,j_3} > 0$ for each $j_1, j_2, j_3 \in \{1, 2, 3, 4\}$ and $\gamma_{1,2} > 0$, $\gamma_{1,3} > 0$ and $\gamma_{2,3} > 0$ are unknown constants. We fix values for $\gamma_{1,2}, \gamma_{1,3}, \gamma_{2,3}$ and solve for the 12 unknowns β_{1,j_1}, β_{2,j_2} and β_{3,j_3}, substituting the optimal expression for $h_{\boldsymbol{j}}$ in the equations (5) to give the 12 key equations

$$\beta_{1,j_1}\delta_{1,j_1} = 1, \quad \beta_{2,j_2}\delta_{2,j_2} = 1, \quad \beta_{3,j_3}\delta_{3,j_3} = 1 \tag{13}$$

where we have defined

$$\begin{aligned}
\delta_{1,j_1} &= \sum_{j_2=1}^{4} \beta_{2,j_2}[\gamma_{1,2}]^{j_1 j_2} \sum_{j_3=1}^{4} \beta_{3,j_3}[\gamma_{1,3}]^{j_1 j_3}[\gamma_{2,3}]^{j_2 j_3} \\
\delta_{2,j_2} &= \sum_{j_1=1}^{4} \beta_{1,j_1}[\gamma_{1,2}]^{j_1 j_2} \sum_{j_3=1}^{4} \beta_{3,j_3}[\gamma_{1,3}]^{j_1 j_3}[\gamma_{2,3}]^{j_2 j_3} \\
\delta_{3,j_3} &= \sum_{j_1=1}^{4} \beta_{1,j_1}[\gamma_{1,3}]^{j_1 j_3} \sum_{j_2=1}^{4} \beta_{2,j_2}[\gamma_{1,2}]^{j_1 j_2}[\gamma_{2,3}]^{j_2 j_3}
\end{aligned}$$

for each $j_1, j_2, j_3 \in \{1, 2, 3, 4\}$. The key equations (13) are written in vector form

$$\boldsymbol{p}(\boldsymbol{\beta}) = \boldsymbol{0} \tag{14}$$

where $\boldsymbol{p} = \boldsymbol{\beta} . * \boldsymbol{\delta} - \boldsymbol{1} \in \mathbb{R}^{12}$ and where the MATLAB notation $\boldsymbol{\beta} . * \boldsymbol{\delta}$ has been used to denote the *Hadamard product* vector whose components are the products of the corresponding components from the vectors $\boldsymbol{\beta} \in \mathbb{R}^{12}$ and $\boldsymbol{\delta} \in \mathbb{R}^{12}$. We have written $\boldsymbol{1} \in \mathbb{R}^{12}$ and $\boldsymbol{0} \in \mathbb{R}^{12}$ respectively for the vectors with all components equal to 1 and all components equal to 0. Now the Newton iteration can be written as

$$\boldsymbol{\beta}^{(j+1)} = \boldsymbol{\beta}^{(j)} - J^{\dagger}[\boldsymbol{\beta}^{(j)}]\boldsymbol{p}[\boldsymbol{\beta}^{(j)}] \tag{15}$$

where $J^{\dagger}$ denotes the MATLAB *Moore-Penrose inverse* [3, 13] of the Jacobian matrix $J \in \mathbb{R}^{12 \times 12}$. The Moore-Penrose inverse is used to avoid, where possible, unstable calculations. It should be noted that the optimal form for the elements $h_{\boldsymbol{j}}$ can give both very small and very large values for relatively modest values of the parameters. Thus it is common to find MATLAB will warn that the inverse matrix J^{-1} is undefined. The proposed procedure formally ignores the constraints (7). Instead we nominate values for the constants $\gamma_{s,t}$ for $1 \leq s < t \leq m$. The corresponding grade correlations can be calculated from the formula (1). To match the given grade correlation coefficients it is necessary to repeatedly adjust the selected values for $\gamma_{s,t}$ and then solve the corresponding key equations (13). Although we have used this method successfully in small problems the Newton iteration proposed here becomes unstable for larger problems, and so we prefer the dual method developed later in the paper.

6.3 A fundamental correlation constraint

We have assumed that the grade correlations are known but in practice they are estimated from observed data. Thus, before applying our suggested numerical calculation procedure, it would be prudent to check that the given grade correlations are feasible. There is a well-known necessary and sufficient condition of which we should be aware. The matrix

$$R = [\rho_{i,j}] \in \mathbb{R}^{m \times m}$$

where $\rho_{i,i} = 1$ and $\rho_{i,j} \in [-1, 1]$ when $i \neq j$ is symmetric and non-negative. That is the matrix must satisfy the conditions $R^* = R$, where R^* denotes the adjoint matrix, and $\langle \boldsymbol{x}, R\boldsymbol{x} \rangle \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^m$. The corresponding restrictions on R are not always obvious. For an intuitive view of the implications suppose X_r, X_s and X_t are jointly distributed random variables. The grade correlations are essentially the cosines of the angles between each pair of elements in a Hilbert space of random variables. Thus if $\gamma_{r,s} = \cos\theta$, $\gamma_{r,t} = \cos\varphi$ and $\gamma_{s,t} = \cos\psi$ then we can regard θ, φ and ψ as the angles between the respective random variable pairs, (X_r, X_s), (X_r, X_t) and (X_s, X_t). Some simple geometry tells us that

$$\varphi - \psi \leq \theta \leq \varphi + \psi$$

after which an elementary algebraic argument gives

$$\cos^2\theta + \cos^2\varphi + \cos^2\psi \leq 1 + 2\cos\theta\cos\varphi\cos\psi.$$

Thus we obtain a fundamental grade correlation constraint

$$\rho_{1,2}^2 + \rho_{1,3}^2 + \rho_{2,3}^2 \leq 1 + 2\rho_{1,2}\rho_{1,3}\rho_{2,3}. \tag{16}$$

This shows us, for instance, that we cannot have $\rho_{1,2} = 0.7$, $\rho_{1,3} = 0.5$ and $\rho_{2,3} = -0.3$.

6.4 Some remarks about the feasible set

If we omit the final constraint (7) then the set $\mathcal{E}$ of all $\boldsymbol{h}$ satisfying (5) and (6) is simply the set of all multiply stochastic hyper-matrices in $\mathbb{R}^\ell$. The set $\mathcal{E}$ is a non-empty convex polytope but in general it is far from simple to describe the set. For instance it is a decidedly non-trivial task even to list the vertices. Hence the set $\mathcal{F}$ is also difficult to describe. Despite this somewhat pessimistic outlook it is nevertheless true that the set $\mathcal{F}$ will be non-empty if we impose realistic correlation constraints. What do we mean by this statement? It is certainly true that the matrix $R = [\rho_{r,s}]$ of imposed grade correlation coefficients must be non-negative but we also know from an earlier calculation that

$$-1 + \frac{1}{n^2} \leq \rho_{r,s} \leq 1 - \frac{1}{n^2}$$

for all $1 \leq r < s \leq m$. We could consider a different viewpoint. Formula (1) shows that the grade correlation coefficients are linear functions on the set $\mathcal{E}$. Thus, for each $1 \leq r < s \leq m$, we define $\rho_{r,s} : \mathcal{E} \mapsto [-1 + 1/n^2, 1 - 1/n^2]$ by the formula

$$\rho_{r,s}(\boldsymbol{h}) = 12 \left(\frac{1}{n^3} \sum_{\boldsymbol{i} \,\in\, \{1,\dots,n\}^m} h_{\boldsymbol{i}} \left(i_r - \frac{1}{2} \right) \left(i_s - \frac{1}{2} \right) - \frac{1}{4} \right)$$

for each $\boldsymbol{h} \in \mathcal{E}$. Since these are linear functions they take extreme values at the vertices of the set $\mathcal{E}$. Thus the set $\mathcal{R} = R(\mathcal{E})$ of feasible grade correlation matrices is the set of all convex combinations of the grade correlation matrices at the vertices $V(\mathcal{E})$ of $\mathcal{E}$. We have the following proposition.

Proposition 1. *If*

$$R = \sum_{\boldsymbol{h} \in V(\mathcal{E})} \alpha_{\boldsymbol{h}} R_{\boldsymbol{h}}$$

where $\sum_{\boldsymbol{h} \in V(\mathcal{E})} \alpha_{\boldsymbol{h}} = 1$ *with* $\alpha_{\boldsymbol{h}} \geq 0$ *then* $R = R_{\boldsymbol{k}}$ *where*

$$\boldsymbol{k} = \sum_{\boldsymbol{h} \in V(\mathcal{E})} \alpha_{\boldsymbol{h}} \boldsymbol{h} \in \mathcal{F}.$$

If $R = R_{\boldsymbol{k}}$ where $\boldsymbol{k} \in \mathcal{F}$ and $k_{\boldsymbol{j}} > 0$ for all $\boldsymbol{j} \in \{1, 2, \ldots, m\}^n$ then Problem 2 has a unique solution $\boldsymbol{h}$ with $h_{\boldsymbol{j}} > 0$ for all $\boldsymbol{j} \in \{1, 2, \ldots, m\}^n$. This relies on the fact that the entropy has compact upper level sets [5, p. 47]. Since the feasible set is non-empty there must be a solution point $\boldsymbol{h} \in \mathcal{F}$. If $h_{\boldsymbol{j}} > 0$ for all $\boldsymbol{j}$ then $\boldsymbol{h}$ is the desired unique solution. If not then define

$$\boldsymbol{h}(\alpha) = (1 - \alpha)\boldsymbol{h} + \alpha\boldsymbol{k}$$

for $0 \leq \alpha \leq 1$. The feasible set is convex and so $\boldsymbol{h}(\alpha) \in \mathcal{F}$ for all $0 \leq \alpha \leq 1$ and we have

$$J(\alpha) = (-1) \left[\frac{1}{n} \sum_{\boldsymbol{i} \in \{1,\ldots,n\}^m} [(1 - \alpha)h_{\boldsymbol{j}} + \alpha k_{\boldsymbol{j}}] \log_e[(1 - \alpha)h_{\boldsymbol{j}} + \alpha k_{\boldsymbol{j}}] + (m - 1)\log_e n \right]$$

from which it follows that

$$J'(\alpha) = \frac{1}{n} \sum_{\boldsymbol{i} \in \{1,\ldots,n\}^m} [h_{\boldsymbol{j}} - k_{\boldsymbol{j}}] \left\{ \log_e[(1 - \alpha)h_{\boldsymbol{j}} + \alpha k_{\boldsymbol{j}}] - 1 \right\}.$$

If we let $\alpha \downarrow 0$ then those terms where $h_{\boldsymbol{j}} > 0$ approach a finite limit $[h_{\boldsymbol{j}} - k_{\boldsymbol{j}}] \log_e h_{\boldsymbol{j}}$. All other terms have the form $(-1)k_{\boldsymbol{j}} \log_e[\alpha k_{\boldsymbol{j}}]$ and approach $+\infty$. Therefore the right hand derivative $J'_+(0)$ at $\alpha = 0$ takes the value $+\infty$. Hence $J(0) < J(\alpha)$ for all sufficiently small $\alpha > 0$. Thus $\boldsymbol{h}$ is not a solution. This is a contradiction. Hence $h_{\boldsymbol{j}} > 0$ for all $\boldsymbol{j} \in \{1, 2, \ldots, m\}$. We have therefore established the following proposition.

Proposition 2. *If there exists* $\boldsymbol{k} \in \mathcal{F}$ *with* $k_{\boldsymbol{j}} > 0$ *for all* $\boldsymbol{j} \in \{1, 2, \ldots, m\}^n$ *then Problem 2 has a unique solution* $\boldsymbol{h}$ *with* $h_{\boldsymbol{j}} > 0$ *for all* $\boldsymbol{j} \in \{1, 2, \ldots, m\}^n$.

6.5 Numerical examples for the primal problem

To obtain a direct numerical solution of the primal problem we select certain parameter values to begin and calculate the remaining parameters by finding a numerical solution to the key equations (14). We then calculate the corresponding grade correlation coefficients. To match given grade correlations we need to repeatedly adjust the initial choice of parameter values so that the corresponding grade correlation coefficients converge to the given values. We assume this can be done in an heuristic manner.

Example 1. Let $m = 3$ and $n = 4$ and take $\gamma_{1,2} = 1.15$, $\gamma_{1,3} = 0.93$ and $\gamma_{2,3} = 0.9$. We write $\boldsymbol{p} = (p_1, \ldots, p_{12}) \in \mathbb{R}^{12}$. The numerical calculations are routine. We take the initial value $\boldsymbol{\beta}^{(0)} = \boldsymbol{1}$. After 18 iterations the solution, shown to four decimal place accuracy, is given by

$$\begin{aligned}
\overline{\boldsymbol{\beta}} &\approx [0.5300, \ 0.4634, \ 0.3917, \ 0.3202, \ 0.4402, \ 0.4208, \\
&\qquad\qquad 0.3861, \ 0.3401, \ 0.2360, \ 0.3770, \ 0.5882, \ 0.8965], \\
\overline{\boldsymbol{\delta}} &\approx [1.8870, \ 2.1581, \ 2.5528, \ 3.1233, \ 2.2715, \ 2.3763, \\
&\qquad\qquad 2.5897, \ 2.9401, \ 4.2366, \ 2.6524, \ 1.7000, \ 1.1155]
\end{aligned}$$

11

and

$$\overline{h}_1 \approx \begin{bmatrix} 0.0530 & 0.0709 & 0.0925 & 0.1180 \\ 0.0524 & 0.0631 & 0.0742 & 0.0851 \\ 0.0498 & 0.0539 & 0.0571 & 0.0589 \\ 0.0454 & 0.0443 & 0.0421 & 0.0392 \end{bmatrix}, \; \overline{h}_2 \approx \begin{bmatrix} 0.0496 & 0.0616 & 0.0748 & 0.0888 \\ 0.0564 & 0.0631 & 0.0690 & 0.0736 \\ 0.0616 & 0.0620 & 0.0610 & 0.0586 \\ 0.0646 & 0.0585 & 0.0518 & 0.0448 \end{bmatrix},$$

$$\overline{h}_3 \approx \begin{bmatrix} 0.0448 & 0.0518 & 0.0585 & 0.0646 \\ 0.0586 & 0.0610 & 0.0620 & 0.0616 \\ 0.0736 & 0.0690 & 0.0631 & 0.0564 \\ 0.0888 & 0.0748 & 0.0616 & 0.0496 \end{bmatrix}, \; \overline{h}_4 \approx \begin{bmatrix} 0.0392 & 0.0421 & 0.0443 & 0.0454 \\ 0.0589 & 0.0571 & 0.0539 & 0.0498 \\ 0.0851 & 0.0742 & 0.0631 & 0.0524 \\ 0.1180 & 0.0925 & 0.0709 & 0.0530 \end{bmatrix},$$

where we have used the notation $\overline{h}_i = [\overline{h}_{ijk}]$. The correlations, shown to four decimal place accuracy, are $\overline{\rho}_{1,2} \approx 0.1707$, $\overline{\rho}_{1,3} \approx -0.1050$ and $\overline{\rho}_{2,3} \approx -0.1359$. The MATLAB calculations show that

$$|\overline{p}_r| \leq 3 \times 10^{-16}$$

for each $r = 1, \ldots, 12$ and that the value of the objective function, shown to four decimal place accuracy, is given by $\overline{J} \approx -0.0312$. $\qquad\square$

Example 2. Let $m = 3$ and $n = 4$ and set $\gamma_{1,2} = 8$, $\gamma_{1,3} = 4$ and $\gamma_{2,3} = 0.3$. The values for $\boldsymbol{\gamma}$ are intended to generate very large and very small parameter values. Once again we write $\boldsymbol{p} = (p_1, \ldots, p_{12}) \in \mathbb{R}^{12}$. The Newton iteration is modified to take the form

$$\boldsymbol{\beta}^{(j+1)} = \boldsymbol{\beta}^{(j)} - J^{\dagger}[\boldsymbol{\beta}^{(j)}]\boldsymbol{f}[\boldsymbol{\beta}^{(j)}]/500.$$

We take the initial value $\boldsymbol{\beta}^{(0)} = \mathbf{1}$. After $30{,}000$ iterations the solution for both $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, shown to four significant figures, is given by

$$\overline{\boldsymbol{\beta}} \approx [2.9403, \, 0.0037, \, 0.0000, \, 0.0000, \, 0.0000, \, 0.0000,$$
$$0.0000, \, 0.0000, \, 0.0000, \, 0.0000, \, 0.0000, \, 0.0000] \times 10^5,$$

$$\overline{\boldsymbol{\delta}} \approx [0.0000, \, 0.0000, \, 0.0000, \, 6.6104, \, 0.0024, \, 0.0041,$$
$$0.0369, \, 1.7324, \, 0.7747, \, 0.6306, \, 0.9947, \, 3.0408] \times 10^5$$

while the solution for $\boldsymbol{h}$, shown to four decimal place accuracy, is

$$\overline{h}_1 \approx \begin{bmatrix} 0.1494 & 0.2203 & 0.1676 & 0.0658 \\ 0.2115 & 0.0935 & 0.0213 & 0.0025 \\ 0.0569 & 0.0075 & 0.0005 & 0.0000 \\ 0.0029 & 0.0001 & 0.0000 & 0.0000 \end{bmatrix}, \; \overline{h}_2 \approx \begin{bmatrix} 0.0061 & 0.0359 & 0.1092 & 0.1715 \\ 0.0689 & 0.1219 & 0.1113 & 0.0524 \\ 0.1483 & 0.0787 & 0.0216 & 0.0030 \\ 0.0606 & 0.0097 & 0.0008 & 0.0000 \end{bmatrix},$$

$$\overline{h}_3 \approx \begin{bmatrix} 0.0000 & 0.0008 & 0.0097 & 0.0606 \\ 0.0030 & 0.0216 & 0.0787 & 0.1483 \\ 0.0524 & 0.1113 & 0.1219 & 0.0689 \\ 0.1715 & 0.1092 & 0.0359 & 0.0061 \end{bmatrix}, \; \overline{h}_4 \approx \begin{bmatrix} 0.0000 & 0.0000 & 0.0001 & 0.0029 \\ 0.0000 & 0.0005 & 0.0075 & 0.0569 \\ 0.0025 & 0.0213 & 0.0935 & 0.2115 \\ 0.0658 & 0.1676 & 0.2203 & 0.1494 \end{bmatrix},$$

where we have again written $\overline{h}_i = [\overline{h}_{ijk}]$. The correlations are shown to four decimal place accuracy as $\overline{\rho}_{1,2} = 0.6610$, $\overline{\rho}_{1,3} = 0.3508$ and $\overline{\rho}_{2,3} = -0.1226$ and the MATLAB calculations show that

$$|\overline{p}_r| \leq 7 \times 10^{-14}$$

for each $r = 1, \ldots, 12$. The value of the objective function, shown to four decimal place accuracy, is given as $\overline{J} \approx -0.6206$. $\qquad\square$

7 Formulation and solution of the Fenchel dual problem

Let us define $g : \mathbb{R}^\ell \mapsto [0, \infty) \cup \{+\infty\}$ by setting

$$g(\boldsymbol{h}) = \begin{cases} (-1)J(\boldsymbol{h}) & \text{if } h_{\boldsymbol{j}} \geq 0 \text{ for all } \boldsymbol{j} \in \{1, 2, \ldots, m\}^n \\ +\infty & \text{otherwise} \end{cases}$$

where we have used the convention that $h \log_e h = 0$ when $h = 0$ and where we will allow functions to take values in an extended set of real numbers. Unless otherwise stated we follow the notations and conventions in the book by Borwein and Lewis [5]. With appropriate definitions we can write the constraints (5) and (7) in the form $A\boldsymbol{h} = b$ where $A \in \mathbb{R}^{k \times \ell}$ and $b \in \mathbb{R}^k$ and where k is the collective rank of the coefficient matrix defining the two sets of linear constraints. In particular we note that the definition of g allows us to omit the restriction (6) from our statement of Problem 2. Indeed we can now write a mathematical statement for Problem 2 in standard form.

Problem 3 (Mathematical statement of the primal problem). *Find*

$$\inf_{\boldsymbol{h} \in \mathbb{R}^\ell} \left\{ g(\boldsymbol{h}) \mid A\boldsymbol{h} = b \right\}. \tag{17}$$

If we assume that (17) has a unique solution $\boldsymbol{h} \in \mathcal{F}$ with $h_{\boldsymbol{j}} > 0$ for all $\boldsymbol{j} \in \{1, 2, \ldots, m\}^n$ then the Fenchel dual problem is an unconstrained maximization and the solution to the primal problem can be recovered from the solution to the dual problem. The necessary justification for this statement can be found from Corollary 3.3.11 and Exercise 7 on page 56 of the book by Borwein and Lewis [5]. We observe that the *Fenchel conjugate* of the function g is the function $g^* : \mathbb{R}^\ell \mapsto \mathbb{R} \cup \{-\infty\}$ defined by

$$g^*(\boldsymbol{k}) = \sup_{\boldsymbol{h} \in \mathbb{R}^\ell} \left\{ \langle \boldsymbol{k}, \boldsymbol{h} \rangle - g(\boldsymbol{h}) \right\} \tag{18}$$

and refer to Borwein and Lewis [5] for further details. For each fixed $\boldsymbol{k} \in \mathbb{R}^\ell$ we define

$$G(\boldsymbol{h}) = \sum_{\boldsymbol{i} \in \{1, \ldots, n\}^m} k_{\boldsymbol{i}} h_{\boldsymbol{i}} - \frac{1}{n} \sum_{\boldsymbol{i} \in \{1, \ldots, n\}^m} \left(h_{\boldsymbol{i}} \log_e h_{\boldsymbol{i}} - h_{\boldsymbol{i}} \right) - (m - 1) \log_e n$$

where we note that $\sum_{\boldsymbol{i} \in \{1, 2, \ldots, m\}^n} h_{\boldsymbol{i}} = n$. We can now use elementary calculus to show that $G(\boldsymbol{h})$ is maximized when $h_{\boldsymbol{i}} = \exp[nk_{\boldsymbol{i}}]$ and hence find that

$$g^*(\boldsymbol{k}) = \frac{1}{n} \sum_{\boldsymbol{i} \in \{1, \ldots, n\}^m} \exp[nk_{\boldsymbol{i}}] - (m - 1) \log_e n.$$

Note that $A^* \in \mathbb{R}^{\ell \times k}$. Using Corollary 3.3.11 from [5] we can now write a mathematical statement of the dual problem in standard form.

Problem 4 (Mathematical statement of the dual problem). *Find*

$$\sup_{\boldsymbol{\varphi} \in \mathbb{R}^k} \left\{ \langle \boldsymbol{b}, \boldsymbol{\varphi} \rangle - g^*(A^* \boldsymbol{\varphi}) \right\}. \tag{19}$$

Let

$$H(\boldsymbol{\varphi}) = \sum_{j=1}^{k} b_j \varphi_j - \frac{1}{n} \sum_{i=1}^{\ell} \exp\left[n \cdot \sum_{j=1}^{k} a_{ij}^* \varphi_j \right] + (m - 1) \log_e n$$

and use elementary calculus once again to show that if the maximum of $H(\boldsymbol{\varphi})$ occurs when $\boldsymbol{\varphi} = \overline{\boldsymbol{\varphi}}$ then

$$\sum_{i=1}^{\ell} a_{ir}^* \exp\left[n \cdot \sum_{j=1}^{k} a_{ij}^* \overline{\varphi}_j\right] = b_r \tag{20}$$

for all $r = 1, 2, \ldots, k$.

7.1 Recovering the primal solution

In general there is a closed form for the primal solution $\overline{\boldsymbol{h}}$. Let $\overline{\boldsymbol{k}} = A^* \overline{\boldsymbol{\varphi}}$ and suppose $\overline{k}_{\boldsymbol{j}} > 0$ for all $\boldsymbol{j} \in \{1, 2, \ldots, m\}^n$. Then the unique solution to the primal problem (3) is given by

$$\overline{\boldsymbol{h}} = \nabla g^*(A^* \overline{\boldsymbol{\varphi}}). \tag{21}$$

The underlying analysis is described in [5] where we refer to Theorem 3.3.5 and to Exercise 17 on page 82. The key equations (20) produce an every where defined smooth set of equations and can be solved by many methods.

7.2 A solution scheme for the dual problem

We can use a pure Newton iteration to solve the key equations (20) very well. The key equations are written in the form

$$\boldsymbol{q}(\boldsymbol{\varphi}) = \boldsymbol{0}$$

where

$$q_r(\boldsymbol{\varphi}) = \sum_{i=1}^{\ell} a_{ir}^* \exp\left[n \cdot \sum_{j=1}^{k} a_{ij}^* \overline{\varphi}_j\right] - b_r$$

for each $r = 1, 2, \ldots, k$. Now the Newton iteration can be written as

$$\boldsymbol{\varphi}^{(j+1)} = \boldsymbol{\varphi}^{(j)} - J^{-1}[\boldsymbol{\varphi}^{(j)}]\boldsymbol{q}[\boldsymbol{\varphi}^{(j)}]$$

where we use the MATLAB inverse of the Jacobian matrix $J \in \mathbb{R}^{k \times k}$.

7.3 Numerical examples for the dual problem

We consider several examples to illustrate different aspects of the calculation. The first example is chosen so that we can display the constraint $A\boldsymbol{h} = \boldsymbol{b}$. In larger examples the matrix A, although relatively sparse, is too large to display conveniently. The second example is chosen to show that solution *via* the dual problem gives the same result as solution of the primal problem. The third example is larger and shows that the solution scheme for the dual problem is significantly more stable.

Example 3. In the case where $m = 2$ and $n = 3$ the objective function is given by

$$g^*(\boldsymbol{k}) = \frac{1}{3} \sum_{r=1}^{3} \sum_{s=1}^{3} \exp\left[3k_{rs}\right] - \log_e 3.$$

and the constraints (5) and (7) can be written in the form $A\boldsymbol{h} = \boldsymbol{b}$ where

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ \frac{1}{9} & \frac{1}{3} & \frac{5}{9} & \frac{1}{3} & 1 & \frac{5}{3} & \frac{5}{9} & \frac{5}{3} & \frac{25}{9} \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \rho + 3 \end{bmatrix}.$$

If we set $\rho = 0.7$ and let $\boldsymbol{\varphi}^{(0)} = \boldsymbol{0} \in \mathbb{R}^6$ then after 8 iterations the solution, shown to four decimal place accuracy, is given by

$$\overline{\boldsymbol{\varphi}} \approx \begin{bmatrix} -0.8682 \\ -2.4632 \\ -0.2825 \\ -1.1506 \\ -2.7457 \\ 1.8474 \end{bmatrix} \quad \text{and} \quad \overline{\boldsymbol{h}} \approx \begin{bmatrix} 0.7933 & 0.2010 & 0.0058 \\ 0.2010 & 0.5980 & 0.2010 \\ 0.0058 & 0.2010 & 0.7933 \end{bmatrix}$$

where we have written $\overline{\boldsymbol{h}} = [\overline{h}_{ij}]$ for convenience. The MATLAB calculations show that

$$\|A\overline{\boldsymbol{h}} - \boldsymbol{b}\| < 8 \times 10^{-15}$$

and the value of the objective function, shown to four decimal place accuracy, is given by $g^*(\overline{\boldsymbol{k}}) \approx -0.5761$. The *duality gap* satisfies the inequality

$$J(\overline{\boldsymbol{h}}) - g^*(\overline{\boldsymbol{k}}) < 3 \times 10^{-16}$$

$\square$

Example 4. In the case where $m = 3$ and $n = 4$ the objective function is given by

$$g^*(\boldsymbol{k}) = \frac{1}{4} \sum_{r=1}^{4} \sum_{s=1}^{4} \sum_{t=1}^{4} \exp\left[4k_{rst}\right] - 2\log_e 4$$

and the constraints (5) and (7) can be written in matrix form $A\boldsymbol{h} = \boldsymbol{b}$ where $A \in \mathbb{R}^{13 \times 64}$ and $\boldsymbol{b} \in \mathbb{R}^{13}$. If the grade correlation coefficients are given by

$$\rho_{1,2} = 0.1707, \quad \rho_{1,3} = -0.1050 \quad \text{and} \quad \rho_{2,3} = -0.1359$$

and if we let $\boldsymbol{\varphi}^{(0)} = \boldsymbol{0} \in \mathbb{R}^{13}$ then after 15 iterations the approximate solution, shown to four decimal place accuracy, is given by

$$\begin{aligned} \overline{\boldsymbol{\varphi}} \approx \ & [-0.0251, \ -0.0587, \ -0.1007, \ -0.0070, \ -0.0242, \ -0.0516, \ -0.7320, \\ & -0.6371, \ -0.5482, \ -0.4650, \ 0.1863, \ -0.0968, \ -0.1405]. \end{aligned}$$

and

$$\overline{\boldsymbol{h}}_1 \approx \begin{bmatrix} 0.0530 & 0.0709 & 0.0925 & 0.1180 \\ 0.0524 & 0.0631 & 0.0742 & 0.0851 \\ 0.0498 & 0.0539 & 0.0571 & 0.0590 \\ 0.0454 & 0.0443 & 0.0421 & 0.0392 \end{bmatrix}, \ \overline{\boldsymbol{h}}_2 \approx \begin{bmatrix} 0.0496 & 0.0616 & 0.0748 & 0.0888 \\ 0.0564 & 0.0631 & 0.0690 & 0.0736 \\ 0.0616 & 0.0620 & 0.0610 & 0.0586 \\ 0.0646 & 0.0585 & 0.0518 & 0.0448 \end{bmatrix},$$

$$\overline{\boldsymbol{h}}_3 \approx \begin{bmatrix} 0.0448 & 0.0518 & 0.0585 & 0.0646 \\ 0.0586 & 0.0610 & 0.0620 & 0.0616 \\ 0.0736 & 0.0690 & 0.0631 & 0.0564 \\ 0.0888 & 0.0748 & 0.0616 & 0.0496 \end{bmatrix}, \ \overline{\boldsymbol{h}}_4 \approx \begin{bmatrix} 0.0392 & 0.0421 & 0.0443 & 0.0454 \\ 0.0590 & 0.0571 & 0.0539 & 0.0498 \\ 0.0851 & 0.0742 & 0.0631 & 0.0524 \\ 0.1180 & 0.0925 & 0.0709 & 0.0530 \end{bmatrix},$$

where we have written $\overline{\boldsymbol{h}}_i = [\overline{h}_{ijk}]$. The MATLAB calculations show that

$$\|A\overline{\boldsymbol{h}} - \boldsymbol{b}\| < 2 \times 10^{-15}$$

and the value of the objective function, shown to four decimal place accuracy, is given by $g^*(\overline{\boldsymbol{k}}) \approx -0.0312$. The *duality gap* satisfies the inequality

$$J(\overline{\boldsymbol{h}}) - g^*(\overline{\boldsymbol{k}}) < 9 \times 10^{-16}.$$

This solution confirms the answer obtained earlier when we solved the corresponding primal problem in Example 1. $\square$

Example 5. In the case where $m = 4$ and $n = 4$ the objective function is given by

$$g^*(\boldsymbol{k}) = \frac{1}{4}\sum_{r=1}^{4}\sum_{s=1}^{4}\sum_{t=1}^{4}\sum_{u=1}^{4} \exp\left[4k_{rstu}\right] - 3\log_e 4.$$

and the constraints (5) and (7) can be written in matrix form $A\boldsymbol{h} = \boldsymbol{b}$ where $A \in \mathbb{R}^{19\times256}$ and $\boldsymbol{b} \in \mathbb{R}^{19}$. If the grade correlation coefficients are given by

$$\rho_{1,2} = 0.8250, \ \rho_{1,3} = 0.1050, \ \rho_{1,4} = 0.6372, \ \rho_{2,3} = -0.1359, \ \rho_{2,4} = 0.3450, \ \rho_{3,4} = 0.1556$$

and if we let $\boldsymbol{\varphi}^{(0)} = \boldsymbol{0} \in \mathbb{R}^{19}$ then after 15 iterations the approximate solution, shown to four decimal place accuracy, is given by

$$\begin{aligned}
\overline{\boldsymbol{\varphi}} \approx [\ &{-1.6101}, \ -7.7873, \ -17.5404, \ -30.8695, \ -0.3976, \ -3.1928, \ -8.3856,\\
&0.2452, \ 0.3785, \ 0.3998, \ 0.0999, \ -0.3473, \ -1.3414,\\
&15.5038, \ 3.1601, \ 7.3445, \ -2.7066, \ -5.3433, \ -0.8088]
\end{aligned}$$

and

$$\overline{\boldsymbol{h}}_{11} \approx \begin{bmatrix} 0.0398 & 0.0927 & 0.0242 & 0.0007 \\ 0.0928 & 0.1180 & 0.0168 & 0.0003 \\ 0.1384 & 0.0959 & 0.0075 & 0.0001 \\ 0.1319 & 0.0498 & 0.0021 & 0.0000 \end{bmatrix}, \quad \overline{\boldsymbol{h}}_{12} \approx \begin{bmatrix} 0.1326 & 0.0056 & 0.0000 & 0.0000 \\ 0.0407 & 0.0009 & 0.0000 & 0.0000 \\ 0.0080 & 0.0001 & 0.0000 & 0.0000 \\ 0.0010 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix},$$

$$\overline{\boldsymbol{h}}_{13} \approx \begin{bmatrix} 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix}, \quad \overline{\boldsymbol{h}}_{14} \approx \begin{bmatrix} 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix},$$

$$\overline{\boldsymbol{h}}_{21} \approx \begin{bmatrix} 0.0000 & 0.0000 & 0.0005 & 0.0034 \\ 0.0000 & 0.0001 & 0.0035 & 0.0138 \\ 0.0000 & 0.0009 & 0.0166 & 0.0357 \\ 0.0001 & 0.0048 & 0.0503 & 0.0590 \end{bmatrix}, \quad \overline{\boldsymbol{h}}_{22} \approx \begin{bmatrix} 0.0048 & 0.0498 & 0.0584 & 0.0077 \\ 0.0156 & 0.0890 & 0.0569 & 0.0041 \\ 0.0327 & 0.1017 & 0.0354 & 0.0014 \\ 0.0438 & 0.0742 & 0.0141 & 0.0003 \end{bmatrix},$$

$$\overline{\boldsymbol{h}}_{23} \approx \begin{bmatrix} 0.1218 & 0.0232 & 0.0005 & 0.0000 \\ 0.0524 & 0.0054 & 0.0001 & 0.0000 \\ 0.0144 & 0.0008 & 0.0000 & 0.0000 \\ 0.0025 & 0.0001 & 0.0000 & 0.0000 \end{bmatrix}, \quad \overline{\boldsymbol{h}}_{24} \approx \begin{bmatrix} 0.0002 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix},$$

$$\overline{\boldsymbol{h}}_{31} \approx \begin{bmatrix} 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0002 \end{bmatrix}, \quad \overline{\boldsymbol{h}}_{32} \approx \begin{bmatrix} 0.0000 & 0.0000 & 0.0001 & 0.0025 \\ 0.0000 & 0.0000 & 0.0008 & 0.0144 \\ 0.0000 & 0.0001 & 0.0054 & 0.0524 \\ 0.0000 & 0.0005 & 0.0232 & 0.1218 \end{bmatrix},$$

$$\overline{\boldsymbol{h}}_{33} \approx \begin{bmatrix} 0.0003 & 0.0141 & 0.0742 & 0.0438 \\ 0.0014 & 0.0354 & 0.1017 & 0.0327 \\ 0.0041 & 0.0569 & 0.0890 & 0.0156 \\ 0.0077 & 0.0584 & 0.0498 & 0.0048 \end{bmatrix}, \quad \overline{\boldsymbol{h}}_{34} \approx \begin{bmatrix} 0.0590 & 0.0503 & 0.0048 & 0.0001 \\ 0.0357 & 0.0166 & 0.0009 & 0.0000 \\ 0.0138 & 0.0035 & 0.0001 & 0.0000 \\ 0.0034 & 0.0005 & 0.0000 & 0.0000 \end{bmatrix},$$

$$\overline{\boldsymbol{h}}_{41} \approx \begin{bmatrix} 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix}, \quad \overline{\boldsymbol{h}}_{42} \approx \begin{bmatrix} 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix},$$

$$\overline{\boldsymbol{h}}_{43} \approx \begin{bmatrix} 0.0000 & 0.0000 & 0.0000 & 0.0010 \\ 0.0000 & 0.0000 & 0.0001 & 0.0080 \\ 0.0000 & 0.0000 & 0.0009 & 0.0407 \\ 0.0000 & 0.0000 & 0.0056 & 0.1326 \end{bmatrix}, \quad \overline{\boldsymbol{h}}_{44} \approx \begin{bmatrix} 0.0000 & 0.0021 & 0.0498 & 0.1319 \\ 0.0001 & 0.0075 & 0.0959 & 0.1384 \\ 0.0003 & 0.0168 & 0.1180 & 0.0928 \\ 0.0007 & 0.0242 & 0.0927 & 0.0398 \end{bmatrix},$$

where we have written $\overline{\boldsymbol{h}}_{ij} = [\overline{h}_{ijk\ell}]$. The MATLAB calculations show that

$$\|A\overline{\boldsymbol{h}} - \boldsymbol{b}\| < 2 \times 10^{-14}$$

and the optimal value of the objective function, shown to four decimal place accuracy, is $g^*(\overline{\boldsymbol{k}}) \approx -5.3829$. The *duality gap* satisfies the inequality

$$J(\overline{\boldsymbol{h}}) - g^*(\overline{\boldsymbol{k}}) < 10^{-16},$$

as intended. $\qquad\square$

Example 6. In the case where $m = 5$ and $n = 4$ the objective function is given by

$$g^*(\boldsymbol{k}) = \frac{1}{4} \sum_{r=1}^{4} \sum_{s=1}^{4} \sum_{t=1}^{4} \sum_{u=1}^{4} \sum_{v=1}^{4} \exp\left[4k_{rstuv}\right] - 4\log_e 4.$$

and the constraints (5) and (7) can be written in matrix form $A\boldsymbol{h} = \boldsymbol{b}$ where $A \in \mathbb{R}^{26 \times 1024}$ and $\boldsymbol{b} \in \mathbb{R}^{26}$. If the grade correlation coefficients are given by

$$\rho_{1,2} = 0.597, \ \rho_{1,3} = -0.122, \ \rho_{1,4} = 0.346, \ \rho_{1,5} = 0.225, \ \rho_{2,3} = -0.175,$$
$$\rho_{2,4} = 0.304, \ \rho_{2,5} = 0.406, \ \rho_{3,4} = 0.272, \ \rho_{3,5} = 0.113, \ \rho_{4,5} = 0.324$$

and if we let $\boldsymbol{\varphi}^{(0)} = \boldsymbol{0} \in \mathbb{R}^{26}$ then after 15 iterations we obtain an approximate solution $\overline{\boldsymbol{\varphi}}$ given to four decimal place accuracy by

$$\overline{\boldsymbol{\varphi}} \approx [\ -0.2642, \ -0.6818, \ -1.2529, \ -0.3254, \ -0.8557, \ -1.5908, \ -0.0034, \ -0.0562,$$
$$-0.1586, \ -0.4323, \ -0.9512, \ -1.5567, \ -0.9273, \ -1.2144, \ -1.5697, \ -1.9933,$$
$$0.9942, \ -0.1539, \ 0.4320, \ -0.1586, \ -0.4030, \ 0.1831, \ 0.6397, \ 0.5000, \ 0.1979, \ 0.2686].$$

Similar calculations to those used earlier show the optimal value of the objective function, shown to four decimal place accuracy, is $g^*(\overline{\boldsymbol{k}}) \approx -0.4311$ and that the *duality gap* satisfies the inequality

$$J(\overline{\boldsymbol{h}}) - g^*(\overline{\boldsymbol{k}}) < 3 \times 10^{-15},$$

again as intended. $\qquad\square$

Table 1: Monthly means and standard deviations (SD) for Sydney

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
mean	102.5	117.9	129.9	126.2	102.7	130.6	98.1	82.1	69.6	76.8	83.7	78.1
SD	76.4	109.6	103.4	112.0	110.7	115.8	82.2	83.6	60.0	66.4	75.5	63.4

8   Case Study

We have available 150 years of official monthly rainfall records supplied by the Australian Bureau of Meteorology for Sydney, in NSW, Australia, during the period 1859-2008. The rainfall is measured in millimetres (mm). Table 1 shows the monthly statistics for Sydney. September has the lowest average of 69 mm and June has the highest of 130 mm.

The distributions for individual months can be modelled effectively using a gamma distribution [20, 12, 23, 19, 17, 15]. The gamma distribution is used to model strictly positive random variables. There have never been any months with an observed zero rainfall total. The gamma distribution is defined on $(0, \infty)$ by the formula

$$F[\alpha, \beta](x) = \int_0^x \frac{\xi^{\alpha-1} e^{-\xi/\beta}}{\beta^\alpha \Gamma(\alpha)} d\xi$$

where $\alpha > 0$ and $\beta > 0$ are parameters. The parameters $\alpha = \alpha[t]$ and $\beta = \beta[t]$ for month t were determined from the observed non-zero records by the method of maximum likelihood. The calculated values are

$$\boldsymbol{\alpha} = (1.817, 1.359, 1.741, 1.333, 1.258, 1.338, 1.202, 1.039, 1.412, 1.468, 1.461, 1.777)$$

and

$$\boldsymbol{\beta} = (56.40, 86.75, 74.60, 94.70, 95.97, 97.64, 81.56, 79.06, 49.33, 52.31, 57.29, 43.92).$$

The distributions appear to be weakly correlated. Table 2 shows the grade correlation coefficients for all monthly pairs. The correlation for (Oct, Nov) is significant at the 0.01 level (2-tailed) and the correlations for (Jan,Feb), (Jan,Apr), (Jan,Oct), (Mar,Jun), (Apr,May), (Jun,Sep) are significant at the 0.05 level (2-tailed). The significant correlations are shown in bold print.

Example 7. Consider the months Mar-Jun-Sep. The observed correlations are $\rho_{1,2} = 0.19$, $\rho_{1,3} = -0.12$ and $\rho_{2,3} = -0.17$. For $n = 4$ the copula of maximum entropy is defined by

$$\overline{\boldsymbol{h}}_1 \approx \begin{bmatrix} 0.0493 & 0.0702 & 0.0965 & 0.1283 \\ 0.0459 & 0.0604 & 0.0768 & 0.0943 \\ 0.0410 & 0.0499 & 0.0586 & 0.0665 \\ 0.0351 & 0.0395 & 0.0429 & 0.0450 \end{bmatrix}, \quad \overline{\boldsymbol{h}}_2 \approx \begin{bmatrix} 0.0505 & 0.0630 & 0.0758 & 0.0881 \\ 0.0549 & 0.0632 & 0.0703 & 0.0755 \\ 0.0572 & 0.0609 & 0.0626 & 0.0621 \\ 0.0572 & 0.0562 & 0.0534 & 0.0490 \end{bmatrix},$$

$$\overline{\boldsymbol{h}}_3 \approx \begin{bmatrix} 0.0490 & 0.0534 & 0.0562 & 0.0572 \\ 0.0621 & 0.0626 & 0.0609 & 0.0572 \\ 0.0755 & 0.0703 & 0.0632 & 0.0549 \\ 0.0881 & 0.0758 & 0.0630 & 0.0505 \end{bmatrix}, \quad \overline{\boldsymbol{h}}_4 \approx \begin{bmatrix} 0.0450 & 0.0429 & 0.0395 & 0.0351 \\ 0.0665 & 0.0586 & 0.0499 & 0.0410 \\ 0.0943 & 0.0768 & 0.0604 & 0.0459 \\ 0.1283 & 0.0965 & 0.0702 & 0.0493 \end{bmatrix}$$

where we have written $\overline{\boldsymbol{h}}_i = [\overline{h}_{ijk}]$ and

$$\overline{\varphi} \approx [\ -0.0272, \ -0.0647, \ -0.1127, \ 0.0038, \ -0.0064, \ -0.0305, \ -0.7490,$$
$$-0.6340, \ -0.5276, \ -0.4299, \ 0.2055, \ -0.1053, \ -0.1784].$$

Table 2: Grade correlation coefficients for all monthly pairs

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Jan		**0.18**	-0.06	**-0.19**	-0.01	-0.02	-0.02	0.13	0.09	**-0.16**	0.05	-0.04
Feb	**0.18**		-0.03	-0.08	-0.09	0.05	-0.01	0.10	0.09	-0.05	0.08	-0.07
Mar	-0.06	-0.03		0.11	0.04	**0.19**	-0.14	-0.15	-0.12	0.15	-0.05	-0.01
Apr	**-0.19**	-0.08	0.11		**0.18**	0.05	0.13	0.12	-0.08	0.11	0.09	-0.03
May	-0.01	-0.09	0.04	**0.18**		0.05	-0.02	-0.05	-0.08	-0.07	0.05	-0.06
Jun	-0.02	0.05	**0.19**	0.05	0.05		-0.04	-0.07	**-0.17**	0.02	0.05	-0.05
Jul	-0.02	-0.01	-0.14	0.13	-0.02	-0.04		0.11	0.12	0.08	-0.08	-0.02
Aug	0.13	0.10	-0.15	0.12	-0.05	-0.07	0.11		0.13	0.13	0.12	-0.09
Sep	0.09	0.09	-0.12	-0.08	-0.08	**-0.17**	0.12	0.13		0.04	0.07	-0.01
Oct	**-0.16**	-0.05	0.15	0.11	-0.07	0.02	0.08	0.13	0.04		**0.22**	-0.03
Nov	0.05	0.08	-0.05	0.09	0.05	0.05	-0.08	0.12	0.07	**0.22**		0.08
Dec	-0.04	-0.07	-0.01	-0.03	-0.06	-0.05	-0.02	-0.09	-0.01	-0.03	0.08	

Example 8. Consider the months Jan-Oct-Nov. The observed correlations are $\rho_{1,2} = -0.16$, $\rho_{1,3} = 0.05$ and $\rho_{2,3} = 0.22$. For $n = 4$ the copula of maximum entropy is defined by

$$\overline{h}_1 \approx \begin{bmatrix} 0.0737 & 0.0501 & 0.0322 & 0.0197 \\ 0.0855 & 0.0631 & 0.0441 & 0.0292 \\ 0.0961 & 0.0770 & 0.0584 & 0.0420 \\ 0.1046 & 0.0910 & 0.0750 & 0.0585 \end{bmatrix}, \ \overline{h}_2 \approx \begin{bmatrix} 0.0737 & 0.0619 & 0.0492 & 0.0371 \\ 0.0729 & 0.0664 & 0.0574 & 0.0469 \\ 0.0698 & 0.0691 & 0.0648 & 0.0575 \\ 0.0647 & 0.0696 & 0.0708 & 0.0683 \end{bmatrix},$$

$$\overline{h}_3 \approx \begin{bmatrix} 0.0683 & 0.0708 & 0.0696 & 0.0647 \\ 0.0575 & 0.0648 & 0.0691 & 0.0698 \\ 0.0469 & 0.0574 & 0.0664 & 0.0729 \\ 0.0371 & 0.0492 & 0.0619 & 0.0737 \end{bmatrix}, \ \overline{h}_4 \approx \begin{bmatrix} 0.0585 & 0.0750 & 0.0910 & 0.1046 \\ 0.0420 & 0.0584 & 0.0770 & 0.0961 \\ 0.0292 & 0.0441 & 0.0631 & 0.0855 \\ 0.0197 & 0.0322 & 0.0501 & 0.0737 \end{bmatrix}.$$

where we have written $\overline{h}_i = [\overline{h}_{ijk}]$ and

$$\overline{\varphi} \approx [\ 0.0468, \ 0.0857, \ 0.1167, \ -0.0063, \ -0.0319, \ -0.0768, \ -0.6603,$$
$$-0.7935, \ -0.9403, \ -1.1007, \ -0.2134, \ 0.1097, \ 0.2817].$$

□

9 Conclusions

We have proposed a checkerboard copula of maximum entropy to construct a joint probability density that preserves the prescribed marginal distributions and matches known grade correlation coefficients. We showed firstly that solution of the primal optimization problem allows us to find a general parametric form for the optimal solution but we also found that direct numerical calculations using this formula become unstable for larger problems. By reformulating the problem as an *unconstrained* optimization problem using the theory of Fenchel duality we were able to show that solution of the dual problem and subsequent recovery of the primal solution is a much more tractable procedure. The underlying entropy model assures that the dual problem has many attractive features both theoretically and numerically. Lastly, our theoretical ideas were illustrated by calculating copulas of maximum entropy in a stochastic rainfall modelling setting.

References

[1] A. Azzalini (2003), Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution, *J. R. Statist. Soc. B*, **65**(2), 367-389.

[2] M. S. Bazaraa, C. M. Shetty (1979), *Nonlinear Programming: Theory and Algorithms*, John Wiley & Sons Inc., New York.

[3] A. Ben-Israel, T. N. E. Greville (1974), *Generalized Inverses, Theory and Applications*, Pure & Applied Mathematics, John Wiley & Sons Inc., New York.

[4] G. Birkhoff (1946), Tres observaciones sobre el algebra lineal, *Univ. Nac. Tucumán Rev*, Ser. A, **5**, 147-151.

[5] Jonathan M. Borwein, Adrian S. Lewis (2006), *Convex Analysis and Nonlinear Optimization, Theory and Examples*, Second Edition, CMS Books in Mathematics, Springer-Verlag New York, Inc.

[6] J. M. Borwein, A. S. Lewis, R. Nussbaum (1994), Entropy minimization, DAD problems and stochastic kernels, *J. Functional Analysis*, **123**, 264-307.

[7] X. Chen, Y. Fan, V. Tsyrennikov (2006), Efficient estimation of semiparametric multivariate copula models, *Journal of the American Statistical Association*, **101**(475), 1228-1240.

[8] C. Genest, K. Ghoudi, L. P. Rivest (1995), A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika*, **82**(3), 543-552.

[9] C. Genest, B. Remillard, D. Beaudoin (2009), Goodness-of-fit tests for copulas: A review and a power study, *Insurance: Mathematics and Economics*, **44**, 199-213.

[10] H. Holm and M. S. Alouini (2004), Sum and difference of two squared correlated Nakagami variates in connection with the Mckay distribution, *IEEE Trans. Commun.*, **52**(8), 1367-1376.

[11] Piotr Jaworski (2009), On copulas and their diagonals, *Information Sciences*, **179**(17), 2863-2871.

[12] R. W. Katz, M. B. Parlange (1998), Overdispersion phenomenon in stochastic modelling of precipitation, *J. Climate*, **11**, 591-601.

[13] Mathworks, "MATLAB" Mathworks version 6.5., 2003.

[14] R. B. Nelsen (1999), *An Introduction to Copulas*. Springer Lecture Notes in Stat., New York.

[15] J. Piantadosi, J. W. Boland, P. G. Howlett (2009), Simulation of rainfall totals on various time scales Daily, Monthly and Yearly, *Environmental Modeling and Assessment*, **14**(4), 431-438.

[16] J. Piantadosi, P. G. Howlett (2007), J. W. Boland, Matching the grade correlation coefficient using a copula with maximum disorder. *Journal of Industrial and Management Optimization*, **3**(2), 305-312.

[17] K. Rosenberg, J. W. Boland, P. G. Howlett (2004), Simulation of monthly rainfall totals, *ANZIAM J.*, **46**(E), E85-E104.

[18] F. Schmid, R. Schmidt (2007), Multivariate extensions of Spearman's rho and related statistics, *Statistics & Probability Letters*, **77**(4), 407-416.

[19] R. Srikanthan, T. A. McMahon (2001), Stochastic generation of annual, monthly and daily climate data: A review, *Hydr. and Earth Sys. Sci.*, **5**(4), 633-670.

[20] R.D. Stern, R. Coe (1984), A model fitting analysis of daily rainfall, *J. Roy. Statist. Soc. A*, **147**, Part 1, 1-34.

[21] W. Sun, S. Rachev, S. Stoyanov, F. Fabozzi (2008), Multivariate Skewed Students t Copula in Analysis of Nonlinear and Asymmetric Dependence in German Equity Market, *Studies in Nonlinear Dynamics & Econometrics*, **12**2/3, 1-35.

[22] G. Venter, J. Barnett, R. Kreps, J. Major (2007), Multivariate Copulas for Financial Modeling, *Variance Casualty Actuarial Society - Arlington, Virginia*, **1**(1), 103-119.

[23] D. S. Wilks, R. L. Wilby (1999), The weather generation game: a review of stochastic weather models, *Prog. Phys. Geog.*, **23**(3), 329-357.

[24] R. Zakaria, A. V. Metcalfe, P. G. Howlett, J. Piantadosi, J. Boland (2010), Using the skew *t*-copula to model bivariate rainfall distribution, *ANZIAM J.*, **51**, C231-C246.