J. Aust. Math. Soc. **107** (2019), 302–318
doi:10.1017/S1446788718000228

CHECKERBOARD COPULAS OF MAXIMUM ENTROPY WITH PRESCRIBED MIXED MOMENTS

JONATHAN BORWEIN and PHIL HOWLETT✉

(Received 13 April 2017; accepted 1 July 2018; first published online 26 November 2018)

Communicated by B. Sims

This work is dedicated to the memory of my friend and colleague Jonathan M. Borwein. The genesis for our paper was a suggestion by Jon that seemed so clear to him at the time but took me a while to digest. His contribution to the structural framework was—as always—both concise and astute.

Abstract

In modelling joint probability distributions it is often desirable to incorporate standard marginal distributions and match a set of key observed mixed moments. At the same time it may also be prudent to avoid additional unwarranted assumptions. The problem is to find the least ordered distribution that respects the prescribed constraints. In this paper we will construct a suitable joint probability distribution by finding the checkerboard copula of maximum entropy that allows us to incorporate the appropriate marginal distributions and match the nominated set of observed moments.

2010 *Mathematics subject classification*: primary 97K80; secondary 94A17, 90C46.

Keywords and phrases: joint probability distributions, maximum entropy, optimization, convex analysis, Fenchel duality.

1. Introduction

A *copula* is a continuous, increasing, grounded probability distribution on the unit hypercube with uniform marginals. The *mixed moments* are the expected values of products of natural powers of two or more of the marginal random variables. A *checkerboard* copula is one with the probability density defined by a step function on a uniform subdivision of the hypercube. Each such step function is uniquely defined by a multiply stochastic hypermatrix. A checkerboard copula constructed in this way can be used to model a joint probability distribution with known marginals and a finite set of prescribed mixed moments. We will solve the following problem.

PROBLEM 1.1. Find a checkerboard copula of maximum entropy subject to a given finite set of prescribed mixed moments.

We 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, for instance, in the books by Borwein and Lewis [5] and by Borwein and Vanderwerff [6].

2. Motivation

It is standard practice to model rainfall accumulation at a given location over a fixed time period—daily, monthly or yearly—as a random variable defined by a gamma distribution [8, 13, 15, 16]. If one wishes to model simultaneous rainfall accumulations at several sites within the same general locality or at the same site over consecutive time periods then it is necessary to construct a joint distribution with marginal gamma distributions and to incorporate various measures of dependence for accumulation of the separate individual totals. The checkerboard copula of maximum entropy [4, 7, 12, 13] and the checkerboard normal copula [7, 13] have both been used for this purpose.

3. Background

This paper is dedicated to the memory of my friend Jonathan Borwein. Jon was a generous and good man with a direct and forthright manner, a sharp wit and a real sense of community. He was a brilliant mathematician. His deep knowledge of fundamental mathematics was complemented by an acute sense of history, an insatiable appetite for real applications and a genuine love of numbers.

At the ANZIAM 2011 Conference in Queenstown, New Zealand, Julia Piantadosi and I presented our work on rainfall modelling using checkerboard copulas of maximum entropy to model joint distributions for seasonal rainfall with known marginal monthly distributions and prescribed covariances. Julia made the point during her talk that numerical solution of the corresponding constrained optimization problem was difficult because the Jacobian matrix was badly conditioned. Jon attended the talk but—in his own inimitable way—was enthusiastically engaged for the entire time with a set of unrelated calculations on a small portable electronic device. Nevertheless at the end of the talk he came over to us and said that our problem could be transformed into a much more tractable unconstrained optimization problem using the theory of Fenchel duality. This sage advice resulted in a much improved solution and ultimately led to a number of joint papers [7, 12, 13]. Some time later—in 2015, I think—Jon said to me that it might be a 'good idea' to extend our work to include models where higher-order moments were prescribed because 'maximum entropy methods were generally well suited to problems with moment constraints'.

I'm not sure why—perhaps I was just too busy with other things—but it was not until late July 2016 that I emailed Jon to say I was ready to tackle this extended problem and to ask whether he was still interested. His reply was typically short and sweet: 'Let's do it. JMB in Cyberspace.' A few days later I received the terrible news that he had died. This is my humble attempt to implement his good idea.

4. Notation

Let $m \in \mathbb{N} + 1$ 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 respectively for each $r = 1, 2, \ldots, m$ by $\boldsymbol{\pi}_r \boldsymbol{x} = x_r$ and

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

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 [11] 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. We will write $[\boldsymbol{0}, \boldsymbol{1}] = [0, 1]^m$ to denote the unit m-dimensional hypercube. Let $c : [\boldsymbol{0}, \boldsymbol{1}] \mapsto [0, \infty)$ be a joint probability density 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 $C : [\boldsymbol{0}, \boldsymbol{1}] \mapsto [0, 1]$ defined by

$$C(\boldsymbol{u}) = \int_{[\boldsymbol{0}, \boldsymbol{u}]} c(\boldsymbol{v})\, d\boldsymbol{v}$$

for all $\boldsymbol{u} \in [\boldsymbol{0}, \boldsymbol{1}]$ 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 hypercube $[\boldsymbol{0}, \boldsymbol{1}]$. 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 $\boldsymbol{F} = (F_1, \ldots, F_m) : \mathbb{R}^m \mapsto [\boldsymbol{0}, \boldsymbol{1}]$. 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

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

https://doi.org/10.1017/S1446788718000228 Published online by Cambridge University Press

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

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

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

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

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 $\boldsymbol{x} \in \mathbb{R}^m$. We will write

$$\boldsymbol{U} = \boldsymbol{F}(\boldsymbol{X}) \iff (U_1, \ldots, U_m) = (F_1(X_1), \ldots, F_m(X_m))$$

for the transformed random variables.

5. An elementary form for the joint density

Let $m, n \in \mathbb{N} + 1$. A real m-dimensional hypermatrix $\boldsymbol{h} = [h_{\boldsymbol{i}}]$ of size $\ell = n^m$ is a mapping $\boldsymbol{h} : \{1, \ldots, n\}^m \to \mathbb{R}^\ell$ defined by $\boldsymbol{h}(\boldsymbol{i}) = h_{\boldsymbol{i}} \in \mathbb{R}$ for each $\boldsymbol{i} = (i_1, \ldots, i_m) \in \{1, \ldots, n\}^m$. Suppose $h_{\boldsymbol{i}} \geq 0$ for all $\boldsymbol{i} \in \{1, \ldots, n\}^m$. We write $\boldsymbol{h} \geq \boldsymbol{0}$ and say that $\boldsymbol{h}$ is nonnegative. For each $r \in \{1, \ldots, m\}$ define the *marginal sum* functions $\sigma_{\boldsymbol{h},r} : \{1, \ldots, n\} \to \mathbb{R}$ by the formulae

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

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

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

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

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

https://doi.org/10.1017/S1446788718000228 Published online by Cambridge University Press

and also that

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

for all $u_r \in [0, 1]$ and all $r = 1, 2, \ldots, m$. Therefore the step function $c_{\boldsymbol{h}} : [\mathbf{0}, \mathbf{1}] \mapsto [0, \infty)$ is a joint probability density function for a corresponding copula $C_{\boldsymbol{h}} : [\mathbf{0}, \mathbf{1}] \mapsto [0, 1]$ defined by

$$C_{\boldsymbol{h}}(\boldsymbol{u}) = \int_{[\mathbf{0}, \boldsymbol{u}]} c_{\boldsymbol{h}}(\boldsymbol{v})\, d\boldsymbol{v}$$

for all $\boldsymbol{u} \in [\mathbf{0}, \mathbf{1}]$. 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 $\boldsymbol{x} \in \mathbb{R}^m$. For each $\boldsymbol{q} = (q(1), \ldots, q(m)) \in (\mathbb{N} - 1)^m$ the moment $\mu_{\boldsymbol{h}}^{\boldsymbol{q}}$ of order $\boldsymbol{q}$ for $\boldsymbol{U} = \boldsymbol{F}(\boldsymbol{X})$ is given by the expected value

$$\begin{aligned}(\mu_{\boldsymbol{h}})^{\boldsymbol{q}} &= E[U^{\boldsymbol{q}}](c_{\boldsymbol{h}}) \\ &= E\Big[\prod_{r=1}^{m} U_r^{q(r)}\Big](c_{\boldsymbol{h}}) \\ &= \int_{[\mathbf{0},\mathbf{1}]} \Big[\prod_{r=1}^{m} u_r^{q(r)}\Big] c_{\boldsymbol{h}}(\boldsymbol{u})\, d\boldsymbol{u} \\ &= \sum_{\boldsymbol{i} \in \{1,\ldots,n\}^m} n^{m-1} h_{\boldsymbol{i}} \prod_{r=1}^{m} \Big[\int_{[(i_r-1)/n, i_r/n]} u_r^{q(r)}\, du_r\Big] \\ &= \sum_{\boldsymbol{i} \in \{1,\ldots,n\}^m} n^{-|\boldsymbol{q}|-1} h_{\boldsymbol{i}} \prod_{r=1}^{m} \Big[\frac{(i_r)^{q(r)+1} - (i_r - 1)^{q(r)+1}}{(q(r) + 1)}\Big],\end{aligned}$$

where we have written $|\boldsymbol{q}| = q(1) + \cdots + q(n)$. If there is some $s \in \{1, \ldots, m\}$ with $q(s) = q \in \mathbb{N}$ and $q(r) = 0$ for all $r \neq s$, then this is a pure moment for the random variable $U_s = F_s(X_s)$. Because the marginal distributions are uniform the pure moments are all

https://doi.org/10.1017/S1446788718000228 Published online by Cambridge University Press

fixed. Indeed an elementary calculation shows that

$$\begin{aligned}(\mu_{\boldsymbol{h}})_s^q &= E[U_s^q](c_{\boldsymbol{h}}) \\ &= \int_{[\mathbf{0},\,\mathbf{1}]} u_s^q\, c_{\boldsymbol{h}}(\boldsymbol{u})\, d\boldsymbol{u} \\ &= \int_{[0,1]} u_s^q \Big[\int_{[0,1]^{m-1}} c_{\boldsymbol{h}}(\boldsymbol{u})\, d\boldsymbol{\pi}_s^c \boldsymbol{u} \Big] du_s \\ &= \int_{[0,1]} u_s^q\, du_s \\ &= \frac{1}{(q+1)}\end{aligned}$$

for all $s = 1, \ldots, m$. If there is some set $\boldsymbol{s} = (s(1), \ldots, s(p)) \in \{1, \ldots, n\}^p$ with $s(1) < \cdots < s(p)$ such that $q(s(k)) \in \mathbb{N}$ for each $k \in \{1, \ldots, p\}$ and $q(r) = 0$ for $r \neq s(k)$, then

$$(\mu_{\boldsymbol{h}})_{\boldsymbol{s}}^{\boldsymbol{q}} = E[U_{\boldsymbol{s}}^{\boldsymbol{q}}](c_{\boldsymbol{h}}) = E[U_{s(1)}^{q(s(1))} \cdots U_{s(p)}^{q(s(p))}](c_{\boldsymbol{h}})$$

is a mixed moment. The previous general formula now becomes

$$(\mu_{\boldsymbol{h}})_{\boldsymbol{s}}^{\boldsymbol{q}} = \sum_{\boldsymbol{i} \in \{1,\ldots,n\}^m} n^{-|\boldsymbol{q}|-1} h_{\boldsymbol{i}} \prod_{k=1}^{p} \Big[\frac{(i_{s(k)})^{q(s(k))+1} - (i_{s(k)} - 1)^{q(s(k))+1}}{(q(s(k)) + 1)} \Big].$$

All moments are invariant under permutations of the coordinate indices.

6. Formulation of the primal problem

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

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

We wish to maximize the entropy subject to a finite set of prescribed mixed moment constraints in the general form $E[U_{\boldsymbol{s}}^{\boldsymbol{q}}] = m_{\boldsymbol{s}}^{\boldsymbol{q}}$ for all $(\boldsymbol{s}, \boldsymbol{q}) \in S$ where S is some specified set of indices and corresponding powers. Although there are some technical details that must be discussed in a moment we can now formulate the problem we wish to solve.

Problem 6.1 (The primal problem). Let $m, n \in \mathbb{N} + 1$ and define $\ell = n^m$. Find the hypermatrix $\boldsymbol{h} \in \mathbb{R}^\ell$ to maximize the entropy $J(\boldsymbol{h})$ subject to the nonnegativity constraints

$$h_{\boldsymbol{i}} \geq 0 \tag{6.1}$$

https://doi.org/10.1017/S1446788718000228 Published online by Cambridge University Press

for all $\boldsymbol{i} \in \{1, \dots, n\}^m$, the multiply stochastic constraints

$$\sigma_{\boldsymbol{h},r}(i_r) = 1 \tag{6.2}$$

for all $i_r \in \{1, \dots, n\}$ and each $r = 1, \dots, m$, and the additional mixed moment constraints

$$(\mu_{\boldsymbol{h}})_{\boldsymbol{s}}^{\boldsymbol{q}} = m(\boldsymbol{s}, \boldsymbol{q}) \tag{6.3}$$

for some prescribed values $m(\boldsymbol{s}, \boldsymbol{q}) \in \mathbb{R}$ where $(\boldsymbol{s}, \boldsymbol{q}) \in S$ and S is a finite set of index–power pairs.

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 hypermatrices satisfying (6.1), (6.2) and (6.3) 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 nonempty then there must be a unique solution for $\boldsymbol{h}$ with strictly positive coordinates. The reader is referred to [1, 5, 6, 14] for a general discussion of the requisite convex analysis and nonlinear optimization.

7. Some remarks about the feasible set

If we omit the final constraint (6.3) then the set $\mathcal{E} \subseteq \mathbb{R}^\ell$ of all $\boldsymbol{h}$ satisfying (6.1) and (6.2) is simply the set of all multiply stochastic hypermatrices of size $\ell = n^m$. If we define $\boldsymbol{\delta} = [\delta_{\boldsymbol{i}}]$ with $\delta_{\boldsymbol{i}} = 1$ when $\boldsymbol{i} = (i, \dots, i)$ for all $i \in \{1, \dots, m\}$ and $\delta_{\boldsymbol{i}} = 0$ otherwise, then $\boldsymbol{\delta} \in \mathcal{E}$. Thus the set $\mathcal{E} \subseteq \mathbb{R}^\ell$ is a nonempty convex polytope. Nevertheless, in general, the set $\mathcal{E}$ is far from simple to describe. For instance, it is a decidedly nontrivial task even to list the vertices.

Example 7.1. Consider the triply stochastic matrices $\boldsymbol{h} \in \mathbb{R}^{3\times3\times3}$. We will write $\boldsymbol{h} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{h}_3]$ where the matrices $\boldsymbol{h}_j \in \mathbb{R}^{3\times3}$ for $j = 1, 2, 3$ represent the top, middle and bottom layers of the hypermatrix. The polytope of triply stochastic hypermatrices has $9 \times 4 \times 1 = 36$ primary vertices defined by the hypermatrices, with only one nonzero element in each of the nine planar sections. A typical representative is $\boldsymbol{v} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3]$ where

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{v}_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

To show that $\boldsymbol{v}$ is indeed a vertex suppose that $\boldsymbol{v} = \alpha\boldsymbol{a} + (1 - \alpha)\boldsymbol{b}$ for some $\alpha \in (0, 1)$ and two triply stochastic hypermatrices $\boldsymbol{a}$ and $\boldsymbol{b}$. From the $(1, 1, 1)$ element we have $1 = \alpha a_{111} + (1 - \alpha)b_{111}$ for some $\alpha \in (0, 1)$. Clearly the only solution is $a_{111} = b_{111} = 1$. Similarly, we argue that $a_{223} = b_{223} = 1$ and $a_{332} = b_{332} = 1$. Since $\boldsymbol{a}, \boldsymbol{b}$ are triply stochastic all remaining elements must be zero. Thus $\boldsymbol{a} = \boldsymbol{b} = \boldsymbol{v}$. Hence $\boldsymbol{v}$ is a vertex.

There are also less obvious vertices. Consider the hypermatrix defined by $\boldsymbol{w} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3]$ where

$$\boldsymbol{w}_1 = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{1}{6} & \frac{1}{6} \end{bmatrix}, \quad \boldsymbol{w}_2 = \begin{bmatrix} 0 & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & 0 & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & 0 \end{bmatrix}, \quad \boldsymbol{w}_3 = \begin{bmatrix} 0 & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & 0 \\ \frac{1}{6} & 0 & \frac{1}{6} \end{bmatrix}.$$

https://doi.org/10.1017/S1446788718000228 Published online by Cambridge University Press

Suppose that $\boldsymbol{w} = \alpha\boldsymbol{a} + (1-\alpha)\boldsymbol{b}$ for some $\alpha \in (0,1)$ and two triply stochastic hypermatrices $\boldsymbol{a}$ and $\boldsymbol{b}$. If $w_{ijk} = 0$ then $\alpha a_{ijk} + (1-\alpha)b_{ijk} = 0$. The only possible solution is $a_{ijk} = b_{ijk} = 0$. Thus $w_{ijk} = 0 \Rightarrow a_{ijk} = b_{ijk} = 0$. We have established that $\boldsymbol{w} = \alpha\boldsymbol{a} + (1-\alpha)\boldsymbol{b}$ for some $\alpha \in (0,1)$ only if the support sets $\operatorname{supp}\boldsymbol{a} = \{(i,j,k) \mid a_{ijk} > 0\}$ and $\operatorname{supp}\boldsymbol{b} = \{(i,j,k) \mid b_{ijk} > 0\}$ are subsets of $\operatorname{supp}\boldsymbol{w} = \{(i,j,k) \mid w_{ijk} > 0\}$. A simple extension of our argument will show that if $\boldsymbol{w} = \sum_{j=1}^{p} \alpha_j \boldsymbol{a}(j)$ for any finite collection of triply stochastic hypermatrices $\boldsymbol{a}(1), \ldots, \boldsymbol{a}(p)$ with $\alpha_1 + \cdots + \alpha_p = 1$ and $\alpha_j > 0$ for all $j \in \{1, \ldots, p\}$ then $\operatorname{supp}\boldsymbol{a}(j) \subseteq \operatorname{supp}\boldsymbol{w}$ for all $j \in \{1, \ldots, p\}$. Now consider the 36 vertices that we have already identified. If we define $\boldsymbol{\delta}(\boldsymbol{i}) \in \mathbb{R}^{3\times3\times3}$ as the hypermatrix with $\delta(\boldsymbol{i})_{\boldsymbol{j}} = 1$ when $\boldsymbol{j} = \boldsymbol{i}$ and $\delta(\boldsymbol{i})_{\boldsymbol{j}} = 0$ when $\boldsymbol{j} \neq \boldsymbol{i}$ then, in alphanumeric order, we have

$$\begin{aligned}
\boldsymbol{v}(1) &= \boldsymbol{\delta}(1,1,1) + \boldsymbol{\delta}(2,2,2) + \boldsymbol{\delta}(3,3,3),\\
\boldsymbol{v}(2) &= \boldsymbol{\delta}(1,1,1) + \boldsymbol{\delta}(2,2,3) + \boldsymbol{\delta}(3,3,2),\\
&\vdots\\
\boldsymbol{v}(36) &= \boldsymbol{\delta}(1,3,3) + \boldsymbol{\delta}(2,2,2) + \boldsymbol{\delta}(3,1,1).
\end{aligned}$$

The only hypermatrices on this list with $\operatorname{supp}\boldsymbol{\delta}(\boldsymbol{i}) \subseteq \operatorname{supp}\boldsymbol{w}$ are

$$\begin{aligned}
\boldsymbol{v}(18) &= \boldsymbol{\delta}(1,2,2) + \boldsymbol{\delta}(2,1,3) + \boldsymbol{\delta}(3,3,1),\\
\boldsymbol{v}(19) &= \boldsymbol{\delta}(1,2,2) + \boldsymbol{\delta}(2,3,1) + \boldsymbol{\delta}(3,1,3),\\
\boldsymbol{v}(22) &= \boldsymbol{\delta}(1,2,3) + \boldsymbol{\delta}(2,1,2) + \boldsymbol{\delta}(3,3,1),\\
\boldsymbol{v}(23) &= \boldsymbol{\delta}(1,2,3) + \boldsymbol{\delta}(2,3,1) + \boldsymbol{\delta}(3,1,2),\\
\boldsymbol{v}(30) &= \boldsymbol{\delta}(1,3,2) + \boldsymbol{\delta}(2,1,3) + \boldsymbol{\delta}(3,2,1),\\
\boldsymbol{v}(31) &= \boldsymbol{\delta}(1,3,2) + \boldsymbol{\delta}(2,2,1) + \boldsymbol{\delta}(3,1,3),\\
\boldsymbol{v}(34) &= \boldsymbol{\delta}(1,3,3) + \boldsymbol{\delta}(2,1,2) + \boldsymbol{\delta}(3,2,1),\\
\boldsymbol{v}(35) &= \boldsymbol{\delta}(1,3,3) + \boldsymbol{\delta}(2,2,1) + \boldsymbol{\delta}(3,1,2).
\end{aligned}$$

It is an elementary exercise to see that for all eight hypermatrices on this list we have $v(j)_{222} = v(j)_{223} = v(j)_{232} = v(j)_{233} = v(j)_{322} = v(j)_{323} = v(j)_{332} = v(j)_{333} = 0$. However $w_{223} > 0$, $w_{232} > 0$, $w_{322} > 0$ and $w_{333} > 0$ so it is not possible to write $\boldsymbol{w} = \sum_{k=1}^{8} \alpha_k \boldsymbol{v}(j_k)$ with $\sum_{k=1}^{8} \alpha_k = 1$ and $\alpha_k > 0$ for each $k \in \{1, \ldots, 8\}$ where $j_1 = 18, j_2 = 19, j_3 = 22, \ldots, j_8 = 35$. Thus we conclude that either $\boldsymbol{w}$ is a vertex or else there exists at least one other vertex not yet identified.

Because the vertices of $\mathcal{E}$ are not easy to identify, the feasible set $\mathcal{F}$ is also difficult to describe. Despite this pessimistic outlook it is nevertheless true that the set $\mathcal{F}$ will be nonempty if we impose realistic mixed moment constraints. What do we mean by this statement? It was shown in [12] that for a checkerboard copula with $\ell = n^m$ subdivisions we have

$$\frac{1}{6} + \frac{1}{12n^2} \leq (\mu_{\boldsymbol{h}})^{(1,1)}_{(r,s)} \leq \frac{1}{3} - \frac{1}{12n^2}$$

for $1 \leq r < s \leq m$. If we allow all possible copulas—not just checkerboard copulas with $\ell = n^m$ subdivisions—then the above inequality is replaced by a slightly more

https://doi.org/10.1017/S1446788718000228 Published online by Cambridge University Press

relaxed inequality,

$$\tfrac{1}{6} \le (\mu_{\boldsymbol{h}})^{(1,1)}_{(r,s)} \le \tfrac{1}{3},$$

for $1 \le r < s \le m$. Similar constraints apply to other mixed moments. A complete theoretical understanding of the moment constraints demands a comprehensive knowledge of the vertices of $\mathcal{E}$ and $\mathcal{F}$. In practice this is not usually an issue since the moments are calculated from observed data and hence are inherently feasible. The moments are linear functions and so they take extreme values at the vertices of $\mathcal{E}$. We can estimate the extreme values for a particular moment and find corresponding vertices by solving a linear programming problem. The complexity of $\mathcal{E}$ means that these problems may not be easy to solve in practice.

Problem 7.1 (Extreme moment problem). Let $m, n \in \mathbb{N} + 1$ and define $\ell = n^m$. Find a multiply stochastic hypermatrix $\boldsymbol{h} \in \mathbb{R}^\ell$ to maximize (or minimize) the mixed moment $(\mu_{\boldsymbol{h}})^{\boldsymbol{q}}_{\boldsymbol{s}}$ for some given $\boldsymbol{s} = (s(1), \ldots, s(p)) \in \{1, \ldots, n\}^p$ with $s(1) < \cdots < s(p)$ and $\boldsymbol{q} \in (\mathbb{N} - 1)^m$ such that $q(s(k)) \in \mathbb{N}$ for each $k \in \{1, \ldots, p\}$ with $q(r) = 0$ for all $r \neq s(k)$.

The mixed moments are linear functions on the closed convex set $\mathcal{E}$ and so the set $\mathcal{M} = \boldsymbol{\mu}(\mathcal{E})$ of all possible mixed moments is the set of all convex combinations of mixed moments at the vertices $V(\mathcal{E})$. We have an elementary consequence.

Proposition 7.1. *If* $\boldsymbol{\mu} = \sum_{\boldsymbol{h} \in V(\mathcal{E})} \alpha_{\boldsymbol{h}} \boldsymbol{\mu}_{\boldsymbol{h}} \in \mathcal{M}$ *where* $\sum_{\boldsymbol{h} \in V(\mathcal{E})} \alpha_{\boldsymbol{h}} = 1$ *and* $\alpha_{\boldsymbol{h}} \ge 0$ *then* $\boldsymbol{\mu} = \boldsymbol{\mu}_{\boldsymbol{k}}$ *where* $\boldsymbol{k} = \sum_{\boldsymbol{h} \in V(\mathcal{E})} \alpha_{\boldsymbol{h}} \boldsymbol{h} \in \mathcal{E}$.

Any convex combination of feasible points is also feasible. Suppose there exists $\boldsymbol{k} \in \mathcal{F}$ with $k_{\boldsymbol{j}} > 0$ for all $\boldsymbol{j} \in K^n_m = \{1, 2, \ldots, m\}^n$. Since the feasible set is nonempty there must be a solution point $\boldsymbol{h} \in \mathcal{F}$ for Problem 6.1. 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 $0 \le \alpha \le 1$. The feasible set is convex and so $\boldsymbol{h}(\alpha) \in \mathcal{F}$ for all $0 \le \alpha \le 1$ and we have

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

for some constant $c \in \mathbb{R}$, from which it follows that

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

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 $+\infty$ as $\alpha \downarrow 0$. Hence the right-hand derivative $J'_+(0)$ at $\alpha = 0$ takes the value $J'_+(0) = +\infty$. 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 $\boldsymbol{j} \in K^n_m$. We have therefore established another useful proposition.

https://doi.org/10.1017/S1446788718000228 Published online by Cambridge University Press

Proposition 7.2. *If there exists* $\boldsymbol{k} \in \mathcal{F}$ *with* $k_{\boldsymbol{j}} > 0$ *for all* $\boldsymbol{j} \in K_m^n$ *then Problem 6.1 has a unique solution* $\boldsymbol{h}$ *with* $h_{\boldsymbol{j}} > 0$ *for all* $\boldsymbol{j} \in K_m^n$.

In theory it is perfectly reasonable to attempt a direct numerical solution of the primal problem. However, as shown for the less general problem in [12], there are significant numerical difficulties that may occur—even for small values of $\ell = n^m$. The same issues arise here. We will not discuss these matters but rather move on to discuss a much more tractable solution procedure. Readers are referred to [12] for an extended discussion of the direct method in problems where the mixed moment constraints are restricted to those defined by a prescribed covariance matrix.

8. Formulation and solution of the Fenchel dual problem

The explanations in this section parallel those given for the similar but less general problem in [12]. However, for the convenience of readers who may not be familiar with the line of argument and in deference to the beauty of the relevant convex analysis [5, 6], we have chosen to present the more general argument in full. 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 notation and conventions in the book by Borwein and Lewis [5]. With appropriate definitions we can write the constraints (6.2) and (6.3) in the form $A\boldsymbol{h} = \boldsymbol{b}$ where $A \in \mathbb{R}^{k \times \ell}$ and $\boldsymbol{b} \in \mathbb{R}^k$ 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 (6.1) from our statement of Problem 6.1. We can now write a mathematical statement for the primal problem in standard form.

Problem 8.1 (Mathematical statement of the primal problem). Find

$$\inf_{\boldsymbol{h} \in \mathbb{R}^\ell} \{g(\boldsymbol{h}) \mid A\boldsymbol{h} = \boldsymbol{b}\}. \tag{8.1}$$

If we assume that (8.1) 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 follows from [5, Corollary 3.3.11, page 53] and [5, Exercise 7, page 56]. We observe that the *Fenchel conjugate* of the function g is the function $g^* : \mathbb{R}^\ell \mapsto \mathbb{R} \cup \{-\infty\}$ defined by

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

https://doi.org/10.1017/S1446788718000228 Published online by Cambridge University Press

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

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

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

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

Using [5, Corollary 3.3.11, page 53], we can now write a mathematical statement of the dual problem in standard form.

Problem 8.2 (Mathematical statement of the dual problem). Find

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

Let

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

for all $\boldsymbol{\varphi} \in \mathbb{R}^k$. We have used the notation $A^* = [a^*_{ij}] \in \mathbb{R}^{\ell \times k}$. We can use elementary calculus once again to show that if the maximum of $H(\boldsymbol{\varphi})$ occurs when $\boldsymbol{\varphi} = \overline{\boldsymbol{\varphi}}$ then

$$\sum_{i=1}^{\ell} a^*_{ir} \exp\Big[n \cdot \sum_{j=1}^{k} a^*_{ij} \overline{\varphi}_j\Big] = b_r \tag{8.2}$$

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

8.1. 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 that $\overline{k}_{\boldsymbol{j}} > 0$ for all $\boldsymbol{j} \in \{1, 2, \ldots, m\}^n$. Then the unique solution to the primal problem (Problem 8.1) is given by

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

The underlying analysis is described in [5, Theorem 3.3.5, page 52] and [5, Exercise 17, page 82]. The necessary and sufficient conditions (8.2) produce an everywhere defined smooth set of equations which can be solved by a variety of methods.

8.2. A solution scheme for the dual problem. We can use a pure Newton iteration to solve the key equations (8.2) very well. The key equations are written in the form

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

https://doi.org/10.1017/S1446788718000228 Published online by Cambridge University Press

where

$$q_r(\boldsymbol{\varphi}) = \sum_{i=1}^{\ell} a_{ir}^* \exp\Big[n \cdot \sum_{j=1}^{k} a_{ij}^* \varphi_j\Big] - b_r = \sum_{i=1}^{\ell} a_{ir}^* \exp[n\langle \boldsymbol{a}_i, \boldsymbol{\varphi}\rangle] - b_r$$

for each $r = 1, 2, \ldots, k$. In the above expression we have used the notation $A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_\ell] \in \mathbb{R}^{k\times\ell}$. Hence we have

$$\frac{\partial q_r}{\partial \varphi_s}(\boldsymbol{\varphi}) = n \sum_{i=1}^{\ell} a_{ir}^* a_{is}^* \exp[n\langle \boldsymbol{a}_i, \boldsymbol{\varphi}\rangle] = n \sum_{i=1}^{\ell} a_{ri} \exp[n\langle \boldsymbol{a}_i, \boldsymbol{\varphi}\rangle] a_{is}^*.$$

If we define the diagonal matrix $D(\boldsymbol{\varphi}) = \operatorname{diag}(\exp[n\langle \boldsymbol{a}_i, \boldsymbol{\varphi}\rangle]) \in \mathbb{R}^{\ell\times\ell}$ then the Jacobian matrix can be written as

$$J(\boldsymbol{\varphi}) = \Big[\frac{\partial q_r}{\partial \varphi_s}(\boldsymbol{\varphi})\Big] = n\, A D(\boldsymbol{\varphi}) A^* \in \mathbb{R}^{k\times k}.$$

Now the Newton iteration is given by

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

for each $p \in \mathbb{N}$ where we use the Matlab inverse of the Jacobian matrix $J \in \mathbb{R}^{k\times k}$.

8.3. An elementary numerical example for the dual problem. We consider an elementary example to illustrate the basic principles of the calculation. Readers are referred to previous papers on rainfall modelling [4, 7, 12, 13] for more complex examples where checkerboard copulas of maximum entropy have been used to model joint rainfall distributions with known marginals and prescribed covariance matrices at various Australian locations.

Example 8.1. Suppose $m = 2$ and $n = 4$ and that we are given the mixed moment constraints

$$(\mu_{\boldsymbol{h}})^{(1,1)} = 0.262, \quad (\mu_{\boldsymbol{h}})^{(2,1)} = 0.169, \quad (\mu_{\boldsymbol{h}})^{(1,2)} = 0.164.$$

The objective function is

$$g^*(\boldsymbol{k}) = \frac{1}{4} \sum_{(r,s)\in\{1,\ldots,4\}} \exp[4k_{(r,s)}] - \log_e 4.$$

If we write $\boldsymbol{h} \cong [h_i] \in \mathbb{R}^{16}$, where the elements appear in alphanumeric order $h_1 \cong h_{(1,1)} \prec h_2 \cong h_{(1,2)} \prec \cdots \prec h_{15} \cong h_{(4,3)} \prec h_{16} \cong h_{(4,4)}$, then we can write the relevant

https://doi.org/10.1017/S1446788718000228 Published online by Cambridge University Press

constraints in the form $A\boldsymbol{h} = \boldsymbol{b}$ where the constraint matrix $A \in \mathbb{R}^{10\times 16}$ is given by

$$\left[\begin{array}{cccccccccccccccc}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1\\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0\\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0\\
\hline
\frac{1}{L} & \frac{3}{L} & \frac{5}{L} & \frac{7}{L} & \frac{3}{L} & \frac{9}{L} & \frac{15}{L} & \frac{21}{L} & \frac{5}{L} & \frac{15}{L} & \frac{25}{L} & \frac{35}{L} & \frac{7}{L} & \frac{21}{L} & \frac{35}{L} & \frac{49}{L}\\
\frac{1}{M} & \frac{3}{M} & \frac{5}{M} & \frac{7}{M} & \frac{21}{M} & \frac{35}{M} & \frac{49}{M} & \frac{19}{M} & \frac{57}{M} & \frac{95}{M} & \frac{133}{M} & \frac{37}{M} & \frac{111}{M} & \frac{21}{M} & \frac{185}{M} & \frac{259}{M}\\
\frac{1}{M} & \frac{7}{M} & \frac{19}{M} & \frac{37}{M} & \frac{3}{M} & \frac{21}{M} & \frac{37}{M} & \frac{111}{M} & \frac{5}{M} & \frac{35}{M} & \frac{95}{M} & \frac{185}{M} & \frac{7}{M} & \frac{49}{M} & \frac{133}{M} & \frac{259}{M}
\end{array}\right]$$

and where $L = 256$ and $M = 1536$. We have omitted the final column sum constraint because the sum of the column sums is equal to the sum of the row sums and so we know that one of the row and column sum constraints is redundant. The constraint vector $\boldsymbol{b} \in \mathbb{R}^{10}$ is given by

$$\boldsymbol{b} = \left[\begin{array}{c} \mathbf{1} \\ \hline 0.262 \\ 0.169 \\ 0.164 \end{array}\right],$$

where we have written $\mathbf{1} = [1] \in \mathbb{R}^7$. After 10 iterations starting with $\boldsymbol{\varphi}^{(0)} = \mathbf{0}$ we obtained the approximate solution

$$\boldsymbol{\varphi} = \begin{bmatrix} -1.8157 \\ -3.2843 \\ -4.4209 \\ -5.3320 \\ 1.2101 \\ -1.2906 \\ -1.5258 \\ 173.1362 \\ -48.3632 \\ -130.6556 \end{bmatrix} \quad \text{and} \quad \boldsymbol{h} = \begin{bmatrix} 0.8324 & 0.0009 & 0.0010 & 0.1658 \\ 0.1245 & 0.0237 & 0.0851 & 0.7667 \\ 0.0330 & 0.2580 & 0.6415 & 0.0674 \\ 0.0101 & 0.7174 & 0.2724 & 0.0001 \end{bmatrix}$$

with $\|A\boldsymbol{h} - \boldsymbol{b}\|_\infty < 10^{-14}$ and $J(\boldsymbol{h}) \approx -0.6879$.

In this problem the vertices of $\mathcal{E}$ are known from the Birkhoff theorem [3] as the set $V(\mathcal{E}) = \mathcal{P} \subseteq \mathbb{R}^{4\times 4}$ of all $4! = 24$ permutation matrices. If $\{\boldsymbol{e}_i\}_{i=1}^4 \in \mathbb{R}^4$ are the usual unit vectors then the vertices $V(\mathcal{E}) = \mathcal{P}$ are the matrices $\boldsymbol{h}_{pqrs} = [\boldsymbol{e}_p, \boldsymbol{e}_q, \boldsymbol{e}_r, \boldsymbol{e}_s]$. It is convenient to write these vertices as vectors in $\mathbb{R}^{16}$ and list them in alphanumeric order as $\boldsymbol{v}(1) \cong \boldsymbol{h}_{1234} \prec \boldsymbol{v}(2) \cong \boldsymbol{h}_{1243} \prec \cdots \prec \boldsymbol{v}(23) \cong \boldsymbol{h}_{4312} \prec \boldsymbol{v}(24) \cong \boldsymbol{h}_{4321}$. If we write

$$A_\mu = \begin{bmatrix} \boldsymbol{u}_8 \\ \boldsymbol{u}_9 \\ \boldsymbol{u}_{10} \end{bmatrix} \in \mathbb{R}^{3\times 16},$$

where $\boldsymbol{u}_i \in \mathbb{R}^{1\times 16}$ for each $i = 1, \ldots, 10$ are the rows of A, then the mixed moments at the vertices are given by $\boldsymbol{\mu}(j) = A_{\mu}\boldsymbol{v}(j)$ for each $j = 1, \ldots, 24$. The relevant values are displayed in Table 1. Now we can show that

$$\boldsymbol{h} \approx \sum_{k=1}^{10} \alpha_k \boldsymbol{v}(j_k) = \begin{bmatrix} 0.8324 & 0.0009 & 0.0010 & 0.1658 \\ 0.1245 & 0.0237 & 0.0851 & 0.7667 \\ 0.0330 & 0.2580 & 0.6415 & 0.0674 \\ 0.0101 & 0.7174 & 0.2724 & 0.0001 \end{bmatrix},$$

where $j_1 = 1,\ j_2 = 4,\ j_3 = 5,\ j_4 = 6,\ j_5 = 10,\ j_6 = 16,\ j_7 = 19,\ j_8 = 21,\ j_9 = 23$ and $j_{10} = 24$ and

$$\boldsymbol{\alpha} = \begin{bmatrix} 0.0001 \\ 0.0656 \\ 0.1252 \\ 0.6414 \\ 0.0009 \\ 0.0010 \\ 0.1245 \\ 0.0227 \\ 0.0103 \\ 0.0083 \end{bmatrix} \geq \mathbf{0}.$$

Note that $\sum_{k=1}^{10} \alpha_k = 1.0000$ and also that

$$\boldsymbol{\mu}_{\boldsymbol{h}} = \sum_{k=1}^{10} \alpha_k \boldsymbol{\mu}(j_k) = \begin{bmatrix} 0.2620 \\ 0.1690 \\ 0.1640 \end{bmatrix}$$

as required.

We describe briefly the elementary method used to determine the representation $\boldsymbol{h} = \sum_{k=1}^{10} \alpha_k \boldsymbol{v}(j_k)$. It is convenient for the following reduction to write all vectors as row vectors. We have $\boldsymbol{h}^{(0)} = \boldsymbol{h}$ given by

$$\boldsymbol{h}^{(0)} = [0.8324, 0.0009, 0.0010, 0.1658, 0.1245, 0.0237, 0.0851, 0.7667,\\ 0.0330, 0.2580, 0.6415, 0.0674, 0.0101, 0.7174, 0.2724, 0.0001].$$

We see that $\min\{h(1), h(8), h(11), h(14)\} = 0.6415$. These elements correspond to the nonzero elements for the vertex $\boldsymbol{v}(6) = [\boldsymbol{e}_1, \boldsymbol{e}_4, \boldsymbol{e}_3, \boldsymbol{e}_2]$. Thus we reduce $\boldsymbol{h}$ by subtracting an appropriate multiple of $\boldsymbol{v}(6)$ to give

$$\begin{aligned}\boldsymbol{h}^{(1)} &= \boldsymbol{h}^{(0)} - 0.6415\boldsymbol{v}(6) \\ &= [0.1909, 0.0009, 0.0010, 0.1658, 0.1245, 0.0237, 0.0851, 0.1252,\\ &\qquad 0.0330, 0.2580, 0.0000, 0.0674, 0.0101, 0.0757, 0.2724, 0.0001].\end{aligned}$$

The process is repeated a second time by subtracting an appropriate multiple of another vertex. Indeed, we see that $\min\{h(1), h(8), h(10), h(15)\} = 0.1252$. These elements

https://doi.org/10.1017/S1446788718000228 Published online by Cambridge University Press

TABLE 1. Table of mixed moments at the vertices for Example 8.1.

$\boldsymbol{h}$	$(\mu_{\boldsymbol{h}})^{(1,1)}$	$(\mu_{\boldsymbol{h}})^{(2,1)}$	$(\mu_{\boldsymbol{h}})^{(1,2)}$
$\boldsymbol{v}(1) = \boldsymbol{h}_{1234}$	0.3281	0.2448	0.2448
$\boldsymbol{v}(2) = \boldsymbol{h}_{1243}$	0.3125	0.2214	0.2214
$\boldsymbol{v}(3) = \boldsymbol{h}_{1324}$	0.3125	0.2292	0.2292
$\boldsymbol{v}(4) = \boldsymbol{h}_{1342}$	0.2812	0.1823	0.1901
$\boldsymbol{v}(5) = \boldsymbol{h}_{1423}$	0.2812	0.1901	0.1823
$\boldsymbol{v}(6) = \boldsymbol{h}_{1432}$	0.2656	0.1667	0.1667
$\boldsymbol{v}(7) = \boldsymbol{h}_{2134}$	0.3125	0.2370	0.2370
$\boldsymbol{v}(8) = \boldsymbol{h}_{2143}$	0.2969	0.2135	0.2135
$\boldsymbol{v}(9) = \boldsymbol{h}_{2314}$	0.2812	0.2057	0.2135
$\boldsymbol{v}(10) = \boldsymbol{h}_{2341}$	0.2344	0.1354	0.1667
$\boldsymbol{v}(11) = \boldsymbol{h}_{2413}$	0.2500	0.1667	0.1667
$\boldsymbol{v}(12) = \boldsymbol{h}_{2431}$	0.2188	0.1198	0.1432
$\boldsymbol{v}(13) = \boldsymbol{h}_{3124}$	0.2812	0.2135	0.2057
$\boldsymbol{v}(14) = \boldsymbol{h}_{3142}$	0.2500	0.1667	0.1667
$\boldsymbol{v}(15) = \boldsymbol{h}_{3214}$	0.2656	0.1979	0.1979
$\boldsymbol{v}(16) = \boldsymbol{h}_{3241}$	0.2188	0.1276	0.1510
$\boldsymbol{v}(17) = \boldsymbol{h}_{3412}$	0.2031	0.1198	0.1198
$\boldsymbol{v}(18) = \boldsymbol{h}_{3421}$	0.1875	0.0964	0.1120
$\boldsymbol{v}(19) = \boldsymbol{h}_{4123}$	0.2344	0.1667	0.1354
$\boldsymbol{v}(20) = \boldsymbol{h}_{4132}$	0.2188	0.1432	0.1138
$\boldsymbol{v}(21) = \boldsymbol{h}_{4213}$	0.2188	0.1510	0.1276
$\boldsymbol{v}(22) = \boldsymbol{h}_{4231}$	0.1875	0.1042	0.1042
$\boldsymbol{v}(23) = \boldsymbol{h}_{4312}$	0.1875	0.1120	0.0964
$\boldsymbol{v}(24) = \boldsymbol{h}_{4321}$	0.1719	0.0883	0.0883

correspond to the nonzero elements of the vertex $\boldsymbol{v}(5) = [\boldsymbol{e}_1, \boldsymbol{e}_4, \boldsymbol{e}_2, \boldsymbol{e}_3]$. Thus we apply a further reduction,

$$\begin{aligned}\boldsymbol{h}^{(2)} &= \boldsymbol{h}^{(1)} - 0.1252\boldsymbol{v}(5)\\ &= [0.0657, 0.0009, 0.0010, 0.1658, 0.1245, 0.0237, 0.0851, 0.0000,\\ &\qquad 0.0330, 0.1328, 0.0000, 0.0674, 0.0101, 0.0757, 0.1472, 0.0001].\end{aligned}$$

The process continues until we reach $\boldsymbol{h}^{(9)} \approx \boldsymbol{0}$. To check the reduction, we define

$$B = [\boldsymbol{v}(4), \boldsymbol{v}(5), \boldsymbol{v}(6), \boldsymbol{v}(10), \boldsymbol{v}(16), \boldsymbol{v}(19), \boldsymbol{v}(21), \boldsymbol{v}(23), \boldsymbol{v}(24)] \in \mathbb{R}^{16\times 9},$$

where the columns of B are simply the vertices used in the reduction. Then we solve $B\boldsymbol{\alpha} = \boldsymbol{h}$ to find $\boldsymbol{\alpha} = B^{\dagger}\boldsymbol{h} \in \mathbb{R}^9$, where $B^{\dagger} \in \mathbb{R}^{9\times 16}$ denotes the MATLAB Moore–Penrose inverse [2]. Although the representation $\boldsymbol{h}_{\text{calc}} = B\boldsymbol{\alpha}$ obtained by this method was very close to the correct value, we noted that $\|\boldsymbol{h}_{\text{calc}} - \boldsymbol{h}\|_\infty = 0.0001$. In particular, we

https://doi.org/10.1017/S1446788718000228 Published online by Cambridge University Press

obtained $(h_{\rm calc})_{16} = 0$ rather than the correct value $h_{16} = 0.0001$. We then realized that the nine vertices $\boldsymbol{v}(4), \ldots, \boldsymbol{v}(24)$ used in the reduction all had $v(j_k)_{16} = 0$. Consequently, a linear combination of these vertices cannot possibly represent an interior point with $h_{16} = 0.0001$. Thus we decided (arbitrarily) to add the vertex $\boldsymbol{v}(1) = \boldsymbol{h}_{1234}$ to the list and replace our original definition of B with the revised definition,

$$B = [\boldsymbol{v}(1), \boldsymbol{v}(4), \boldsymbol{v}(5), \boldsymbol{v}(6), \boldsymbol{v}(10), \boldsymbol{v}(16), \boldsymbol{v}(19), \boldsymbol{v}(21), \boldsymbol{v}(23), \boldsymbol{v}(24)] \in \mathbb{R}^{16\times 10}.$$

Now we solved the revised equation $B\boldsymbol{\alpha} = \boldsymbol{h}$ to find $\boldsymbol{\alpha} = B^\dagger \boldsymbol{h} \in \mathbb{R}^{10}$ where $B^\dagger \in \mathbb{R}^{10\times 16}$. The revised answer $\boldsymbol{h}_{\rm calc} = B\boldsymbol{\alpha}$ agreed precisely with the answer obtained by the Newton iteration.

9. Conclusions

This article extends previous work on checkerboard copulas of maximum entropy to allow additional constraints that require matching of observed values for a finite set of mixed moments. This extension may have useful application to topics such as rainfall modelling. In strict mathematical terms the theoretical justification is a straightforward extension of the methods used by Piantadosi *et al.* in [12]. More generally, the problem is interesting because the set $\mathcal{E}$ of multiply stochastic hypermatrices with dimension higher than 2 is a complex polytope for which no complete theoretical description is known [9, 10]. One may speculate that imposition of a single mixed moment constraint and judicious variation of that constraint may allow us to estimate the extreme values and at the same time to find approximate locations for the vertices where those extreme values occur.

References

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

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

[3] G. Birkhoff, 'Tres observaciones sobre el algebra lineal', *Univ. Nac. Tucumán Rev, Ser.* A **5** (1946), 147–151; (in Spanish).

[4] J. Boland, P. Howlett, J. Piantadosi and R. Zakaria, 'Modelling and simulation of volumetric rainfall for a catchment in the Murray–Darling basin', *ANZIAM J.* **58**(2) (2016), 119–142.

[5] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization, Theory and Examples*, 2nd edn, CMS Books in Mathematics (Springer, New York, 2006).

[6] J. M. Borwein and J. D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples*, Encyclopedia of Mathematics and its Applications, 109 (Cambridge University Press, Cambridge, 2010).

[7] J. Borwein, P. Howlett and J. Piantadosi, 'Modelling and simulation of seasonal rainfall using the principle of maximum entropy', *Entropy* **16**(2) (2014), 747–769.

[8] R. W. Katz and M. B. Parlange, 'Overdispersion phenomenon in stochastic modelling of precipitation', *J. Clim.* **11** (1998), 591–601.

[9] X. Li, P. Mikusinski, H. Sherwood and M. D. Taylor, 'In quest of Birkhoff's theorem in higher dimensions', in: *Distributions with Fixed Marginals and Related Topics*, IMS Statistics Lecture Notes – Monograph Series, 28 (Institute of Mathematical Statistics, Hayward, CA, 1996), 187–197.

[10] Z. Li, F. Zhang and X.-D. Zhang, 'On the number of vertices of the stochastic tensor polytope', *Linear Multilinear Algebra* **65**(10) (2017), 2064–2075.
[11] R. B. Nelsen, *An Introduction to Copulas*, Lecture Notes in Statistics, 139 (Springer, New York, 1999).
[12] J. Piantadosi, P. Howlett and J. Borwein, 'Copulas of maximum entropy', *Opt. Lett.* **6**(1) (2012), 99–125.
[13] J. Piantadosi, P. Howlett, J. Borwein and J. Henstridge, 'Maximum entropy methods for generating simulated rainfall', *Numer. Algebra Control Optim.* **2**(2) (2012), 233–256; special issue to celebrate Charles Pearce's 70th birthday.
[14] R. Tyrrell Rockafellar, *Convex Analysis*, Princeton Landmarks in Mathematics and Physics (reprint of the 1970 Princeton Mathematical Series edn) (Princeton University Press, Princeton, NJ, 1997).
[15] R. D. Stern and R. Coe, 'A model fitting analysis of daily rainfall', *J. Roy. Statist. Soc. Ser. A* **147**(1) (1984), 1–34.
[16] D. S. Wilks and R. L. Wilby, 'The weather generation game: a review of stochastic weather models', *Prog. Phys. Geog.* **23**(3) (1999), 329–357.

JONATHAN BORWEIN, (deceased), previously at
School of Mathematical and Physical Sciences,
CARMA, The University of Newcastle,
Callaghan, NSW 2308, Australia

PHIL HOWLETT, Scheduling and Control Group (SCG),
Centre for Industrial and Applied Mathematics (CIAM),
School of Information Technology and Mathematical Sciences,
University of South Australia, Australia
e-mail: phil.howlett@unisa.edu.au