Symmetric polynomials and symmetric mean inequalities

Karl Mahlburg *

Department of Mathematics
Louisiana State University
Baton Rouge, LA 70803, U.S.A.

mahlburg@math.lsu.edu

Clifford Smyth

Department of Mathematics and Statistics
University of North Carolina Greensboro
Greensboro, NC 27402, U.S.A.

cdsmyth@uncg.edu

Submitted: Nov 22, 2012; Accepted: Aug 27, 2013; Published: Sep 6, 2013
Mathematics Subject Classifications: 05E05, 26E60, 60C05

Abstract

We prove generalized arithmetic-geometric mean inequalities for quasi-means arising from symmetric polynomials. The inequalities are satisfied by all positive, homogeneous symmetric polynomials, as well as a certain family of non-homogeneous polynomials; this family allows us to prove the following combinatorial result for marked square grids.

Suppose that the cells of a $n \times n$ checkerboard are each independently filled or empty, where the probability that a cell is filled depends only on its column. We prove that for any $0 \leqslant \ell \leqslant n$, the probability that each column has at most ℓ filled sites is less than or equal to the probability that each row has at most ℓ filled sites.

Keywords: symmetric means; symmetric polynomials; arithmetic-geometric mean inequality

1 Introduction

Let n be a positive integer. We define an n-variable *orthant function* to be a continuous function $F : \mathbb{R}^n_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ such that $F(\boldsymbol{x}) = F(x_1, \ldots, x_n)$ is monotonically increasing in each x_i and that also has a strictly increasing *diagonal restriction*, $f_F(y) = F(y, \ldots, y)$. Given an n-variable orthant function F, we define the following n-variable orthant functions associated with F: the *quasi-arithmetic mean*, the *quasi-geometric mean*, and the

*Partially supported by an NSF Postdoctoral Fellowship administered by the Mathematical Sciences Research Institute through its core grant DMS-0441170 and by NSF Grant DMS-1201435..

quasi-mean; respectively, they are

$$\mathcal{A}_F(\boldsymbol{x}) := f_F^{-1}\left(\frac{f_F(x_1)+\cdots+f_F(x_n)}{n}\right), \tag{1.1}$$
$$\mathcal{G}_F(\boldsymbol{x}) := f_F^{-1}\left(\prod_{j=1}^{n} f_F(x_j)^{\frac{1}{n}}\right),$$
$$\mathcal{M}_F(\boldsymbol{x}) := f_F^{-1}\left(F(x)\right).$$

Note that these means have been studied classically (see [2], Chapter III).

Some care is needed to verify that these definitions are well-defined. One must note that since $f_F(y)$ is strictly increasing and continuous, its range $R := f_F(\mathbb{R}_{\geqslant 0})$ is of the form $R = [f_F(0), M)$ or $[f_F(0), +\infty)$, according to the value $M = \lim_{y\to+\infty} f_F(y)$. Furthermore, f_F is a bijection and f_F^{-1} is strictly increasing and continuous. Since R is closed under taking arithmetic and geometric means of its elements, $\mathcal{A}_F(x)$ and $\mathcal{G}_F(x)$ are well-defined. Since F is monotonically increasing in each variable, it also satisfies

$$f_F(\underline{\boldsymbol{x}}) \leqslant F(\boldsymbol{x}) \leqslant f_F(\overline{\boldsymbol{x}})$$

where $\underline{\boldsymbol{x}} := \min\{x_i : 1 \leqslant i \leqslant n\}$ and $\overline{\boldsymbol{x}} := \max\{x_i : 1 \leqslant i \leqslant n\}$. This implies that $F(\mathbb{R}^n_{\geqslant 0}) = R$ and so $\mathcal{M}_F(x)$ is well-defined also.

If $M = \mathcal{A}_F, \mathcal{G}_F$, or $\mathcal{M}_F$, we therefore have that M is a *quasi-mean*. In particular, M satisfies the following usual properties of a mean: it is continuous and monotonically increasing in each variable, $\underline{\boldsymbol{x}} \leqslant M(\boldsymbol{x}) \leqslant \overline{\boldsymbol{x}}$ for all $\boldsymbol{x} \in \mathbb{R}_{\geqslant 0}$, and $M(y,\ldots,y) = y$ for all $y \geqslant 0$. Note that the f_F^{-1} in the definition of M ensures the final "identity" property, $M(y,\ldots,y) = y$. However, M is not necessarily linearly homogeneous, i.e. we do not necessarily have $M(\lambda\boldsymbol{x}) = \lambda M(\boldsymbol{x})$ for all $\lambda \geqslant 0$ and $\boldsymbol{x} \in \mathbb{R}^n_{\geqslant 0}$. The function $F(\boldsymbol{x}) = (1+x_1)(1+x_2)$ provides a simple counterexample for each type of M.

Since f_F and hence f_F^{-1} are strictly increasing, $\mathcal{A}_F$ and $\mathcal{G}_F$ are strictly increasing in each variable. An arithmetic-geometric mean inequality between $\mathcal{G}_F$ and $\mathcal{A}_F$ also easily follows: for all $\boldsymbol{x} \in \mathbb{R}^n_{\geqslant 0}$,

$$\mathcal{G}_F(\boldsymbol{x}) \leqslant \mathcal{A}_F(\boldsymbol{x}),$$

with equality if and only if all x_i are equal. This is also Theorem 85 of [2], with $\psi = \log f_F$, $\chi = f_F$, and $q \equiv 1/n$. Note that the classical arithmetic-geometric mean inequality can be recovered from this by setting $F(\boldsymbol{x}) = (x_1 \cdots x_n)^{1/n}$.

In this paper we study functions whose quasi-means provide a refinement of the preceding arithmetic-geometric mean inequality. Namely, we are interested in $\mathcal{S}$, which we define to be the set of all orthant functions F for which

$$\mathcal{G}_F(\boldsymbol{x}) \leqslant \mathcal{M}_F(x) \leqslant \mathcal{A}_F(\boldsymbol{x}), \qquad \text{for all } \boldsymbol{x} \in \mathbb{R}^n_{\geqslant 0}. \tag{1.2}$$

Proposition 1.1. *The set $\mathcal{S}$ satisfies the following properties.*

1. *If $F(x_1,\ldots,x_n)$ is a homogeneous symmetric polynomial with positive coefficients, then $F \in \mathcal{S}$.*

2. If $F_i(x_1, \ldots, x_n) \in \mathcal{S}$ for $i = 1, 2$, then $F_1 \cdot F_2 \in \mathcal{S}$.

Remark 1.2. The set $\mathcal{S}$ is not closed under addition. For example, $F(\boldsymbol{x}) = 1 + x_1 x_2$ is the sum of two homogeneous functions but $\mathcal{G}_F \leqslant \mathcal{M}_F$ fails to hold when $x_1 \neq x_2$.

For integers ℓ and n with $0 \leqslant \ell \leqslant n$ we define

$$\eta_\ell(\boldsymbol{x}) := \sum_{j=0}^{\ell} \sum_{\substack{I \subseteq [n] \\ |I| = j}} x_I, \tag{1.3}$$

where we have adopted the common notation $x_I := \prod_{i \in I} x_i$, along with the convention $x_\varnothing = 1$. Note that η_ℓ is an orthant function if and only if $\ell \geqslant 1$. We denote the diagonal restriction of η_ℓ by

$$\mu_\ell(y) := \eta_\ell(y, \ldots, y) = \sum_{j=0}^{\ell} \binom{n}{j} y^j. \tag{1.4}$$

In Section 3 we will prove the following result, which states that $\eta_\ell \in \mathcal{S}$ for $1 \leqslant \ell \leqslant n$.

Theorem 1.3. *If ℓ and n are integers with $0 \leqslant \ell \leqslant n$ and $n \geqslant 1$, then*

$$\left(\prod_{i=1}^{n} \mu_\ell(x_i) \right)^{1/n} \leqslant \eta_\ell(\boldsymbol{x}) \leqslant \frac{1}{n} \sum_{i=1}^{n} \mu_\ell(x_i), \text{ for all } \boldsymbol{x} \in \mathbb{R}^n_{\geqslant 0}.$$

Equality holds throughout if and only if $\ell = 0$, $\ell = n$, or $x_1 = \cdots = x_n$.

These quasi-mean inequalities have an appealing application to combinatorial probability. Let $\{X_{ij}, 1 \leqslant i, j \leqslant n\}$ be a collection of independent Bernoulli random variables with probability p_j; in other words, $\mathbf{P}(X_{ij} = 1) = p_j$ and $\mathbf{P}(X_{ij} = 0) = 1 - p_j$. Using these, we further define the random variables

$$C_j := \sum_{i=1}^{n} X_{ij} \quad \text{for } 1 \leqslant j \leqslant n,$$
$$R_i := \sum_{j=1}^{n} X_{ij} \quad \text{for } 1 \leqslant i \leqslant n.$$

Using Theorem 1.3, we prove bounds relating the distributions of the C_js and R_is.

Theorem 1.4. *Suppose that $p_1, \ldots, p_n \in [0, 1]$. If ℓ is an integer with $0 \leqslant \ell \leqslant n$, then*

$$\mathbf{P}\left(\max_{1 \leqslant j \leqslant n} \{C_j\} \leqslant \ell \right) \leqslant \mathbf{P}\left(\max_{1 \leqslant i \leqslant n} \{R_i\} \leqslant \ell \right)$$

Theorem 1.4 has two immediate combinatorial reformulations: one to marked square grids and another to random bipartite graphs. First, suppose that markers are independently placed in the squares of an n by n grid such that the probability that a marker is placed in a square depends only on that square's column. Then for all integers ℓ with $0 \leqslant \ell \leqslant n$, the probability that every column has at most ℓ markers is less than or equal to the probability that each row has at most ℓ markers.

Alternatively, suppose that G is a finite complete bipartite graph with bipartition (A, B), where $|A| = |B| = n$. Let H be a random subgraph of G where each edge e of G is independently selected to belong to H with a probability that depends only on the left vertex $e \cap A$. If $\ell \geqslant 0$, then

$$\mathbf{P}\left(\max_{a\in A}\{d_H(a)\} \leqslant \ell\right) \leqslant \mathbf{P}\left(\max_{b\in B}\{d_H(b)\} \leqslant \ell\right).$$

Remark 1.5. In the grid formulation, both events in the inequality require that at most $n\ell$ squares be occupied. Similarly, both events in the bipartite graph formulation require that there be at most $n\ell$ edges.

The remainder of the paper is as follows. In Section 2 we study the basic properties of homogeneous symmetric polynomials and the functions in $\mathcal{S}$, and also prove Proposition 1.1. In Section 3 we use Lagrange multipliers and polynomial inequalities to prove Theorem 1.3. We conclude in Section 4, where we describe the relationship between the quasi-mean inequalities and combinatorial probability, and prove Theorem 1.4.

2 Properties of symmetric polynomials and $\mathcal{S}$

As found in Chapter 7 of [4], one can define various homogeneous (graded) bases for the ring of symmetric polynomials on n variables; we will make explicit use of the *elementary symmetric polynomials* $\{e_j(\boldsymbol{x}) \mid 0 \leqslant j \leqslant n\}$, where the degree j polynomial is defined as $e_j(\boldsymbol{x}) := \sum\{x_I : I \subseteq [n], |I| = j\}$. We will also use the *monomial symmetric polynomials*, which are defined as follows.

For a positive integer n, let $\Delta(n) := \{\boldsymbol{\lambda} \in \mathbb{R}^n_{\geqslant 0} : \lambda_1 \geqslant \cdots \geqslant \lambda_n\}$, and define the *weight* of such a vector $\boldsymbol{\lambda}$ as $|\boldsymbol{\lambda}| := \lambda(1) + \cdots + \lambda(n)$. If $\boldsymbol{\lambda} \in \Delta(n)$, the monomial symmetric polynomial associated to $\boldsymbol{\lambda}$ is defined as

$$M_{\boldsymbol{\lambda}}(\boldsymbol{x}) := \frac{1}{n!}\sum_{\sigma\in S_n} x_{\sigma(1)}^{\lambda(1)}\cdots x_{\sigma(n)}^{\lambda(n)}. \tag{2.1}$$

Note that this is homogeneous of degree $|\boldsymbol{\lambda}|$.

We now recall various inequalities between symmetric polynomials. Suppose that $\boldsymbol{\lambda_i} = (\lambda_i(1), \ldots, \lambda_i(n)) \in \Delta(n)$ for $i \in \{1, 2\}$. We say that $\boldsymbol{\lambda_1}$ *majorizes* $\boldsymbol{\lambda_2}$ if and only if $|\boldsymbol{\lambda_1}| = |\boldsymbol{\lambda_2}|$, and

$$\lambda_1(1) + \cdots + \lambda_1(j) \geqslant \lambda_2(1) + \cdots + \lambda_2(j)$$

for all $1 \leqslant j < n$. In this case, we write $\boldsymbol{\lambda_1} \succeq \boldsymbol{\lambda_2}$. Muirhead's inequalities can be concisely stated as follows.

Theorem (Muirhead [3])**.** *Suppose that* $\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2} \in \Delta(n)$*. The inequality*

$$M_{\boldsymbol{\lambda_1}}(\boldsymbol{x}) \geqslant M_{\boldsymbol{\lambda_2}}(\boldsymbol{x}) \tag{2.2}$$

is satisfied for all $\boldsymbol{x} \in \mathbb{R}^n_{\geqslant 0}$ *if and only if* $\boldsymbol{\lambda_1} \succeq \boldsymbol{\lambda_2}$*.*

Note that [3] only contains the case where both $\boldsymbol{\lambda_i}$ have integral parts, while [2] (Theorem 45, p. 45) contains the general result above. We also recall the following elementary result from the general theory of series inequalities; see [2] (Theorem 368, p. 261).

Theorem (Rearrangement Inequality)**.** *Suppose that* $a_1 \geqslant \ldots \geqslant a_n \geqslant 0$ *and* $b_1 \geqslant \ldots \geqslant b_n \geqslant 0$*. If* $\sigma \in S_n$ *is a permutation, then*

$$\sum_{i=1}^{n} a_i b_i \geqslant \sum_{i=1}^{n} a_i b_{\sigma(i)}. \tag{2.3}$$

Our final preliminary observations address the inequalities in (1.2) individually. Let $\mathcal{L}$ denote the set of orthant functions F that satisfy the left inequality:

$$\mathcal{G}_F(\boldsymbol{x}) \leqslant \mathcal{M}_F(\boldsymbol{x}), \text{ for all } \boldsymbol{x} \in \mathbb{R}^n_{\geqslant 0},$$

and let $\mathcal{R}$ denote the set of orthant functions that satisfy the right inequality:

$$\mathcal{M}_F(\boldsymbol{x}) \leqslant \mathcal{A}_F(\boldsymbol{x}), \text{ for all } \boldsymbol{x} \in \mathbb{R}^n_{\geqslant 0}.$$

Clearly we have $\mathcal{S} = \mathcal{L} \cap \mathcal{R}$. The following properties follow immediately from the definitions of the quasi-means.

Proposition 2.1. *The classes* $\mathcal{L}$ *and* $\mathcal{R}$ *have the following closure properties:*

1. *If* $F_i \in \mathcal{L}$ *for* $i = 1, 2$*, then* $F_1^a \cdot F_2^b \in \mathcal{L}$ *for all* $a, b \geqslant 0$*.*
2. *If* $F_i \in \mathcal{R}$ *for* $i = 1, 2$*, then* $aF_1 + bF_2 \in \mathcal{R}$ *for all* $a, b \geqslant 0$*.*

The preceding facts now allow us to prove our first result about $\mathcal{S}$.

Proof of Proposition 1.1. We first prove statement 1. Suppose F is a homogeneous symmetric polynomial with positive coefficients and total degree w. It can be written as

$$F(\boldsymbol{x}) = \sum_{i=1}^{k} a_i M_{\boldsymbol{\lambda_i}}(\boldsymbol{x}), \tag{2.4}$$

where $a_i > 0$, $\boldsymbol{\lambda_i} \in \Delta(n)$ and $|\boldsymbol{\lambda_i}| = w$ for all i. Writing $A := \sum_{i=1}^k a_i > 0$, Muirhead's theorem (2.2) then implies that

$$A \cdot M_{(w/n,\ldots,w/n)}(\boldsymbol{x}) \leqslant F(\boldsymbol{x}) \leqslant A \cdot M_{(w,0,\ldots,0)}(\boldsymbol{x}), \text{ for all } \boldsymbol{x} \in \mathbb{R}^n_{\geqslant 0}. \tag{2.5}$$

We finish the proof by showing that this is equivalent to the statement $F \in \mathcal{S}$. Since $f_F(y) = F(y, \ldots, y) = Ay^w$, we find that the leftmost expression in (2.5) is the same as

$$A \cdot \frac{1}{n!} \cdot (x_1 \cdots x_n)^{w/n} \cdot n! = \left(\prod_{i=1}^{n} f_F(x_i) \right)^{1/n}.$$

Similarly, the rightmost expression in (2.5) equals

$$A \cdot \frac{1}{n!} (x_1^w + \cdots + x_n^w) \cdot (n-1)! = \frac{1}{n} \sum_{i=1}^{n} f_F(x_i).$$

This completes the first part of the proof.

We now turn to statement 2. Throughout this part of the proof we write f_i as shorthand for f_{F_i}. By part 1 of Proposition 2.1 we need only show that if $F_i \in \mathcal{S}$ for $i = 1, 2$, then $F_1 \cdot F_2 \in \mathcal{R}$. This is equivalent to showing that

$$(F_1F_2)(\boldsymbol{x}) \leqslant \frac{1}{n} \sum_{i=1}^{n} f_{F_1F_2}(\boldsymbol{x}), \tag{2.6}$$

and we can immediately rewrite $f_{F_1F_2} = f_1 \cdot f_2$. Using the assumption that $F_i \in \mathcal{R}$, we find that the left side of (2.6) satisfies

$$(F_1F_2)(\boldsymbol{x}) \leqslant \left(\frac{1}{n} \sum_{i=1}^{n} f_1(x_i) \right) \left(\frac{1}{n} \sum_{i=1}^{n} f_2(x_i) \right) = \frac{1}{n^2} \sum_{i_1, i_2 = 1}^{n} f_1(x_{i_1}) f_2(x_{i_2}). \tag{2.7}$$

Define the one-step shift cyclical permutation by $\sigma(i) := i + 1$ for $1 \leqslant i \leqslant n - 1$, and $\sigma(n) := 1$. Reordering the x_i if necessary so that $x_1 \geqslant \ldots \geqslant x_n \geqslant 0$, we then further rewrite the sum from (2.7) as

$$\frac{1}{n^2} \sum_{i_1, i_2 = 1}^{n} f_1(x_{i_1}) f_2(x_{i_2}) = \frac{1}{n^2} \sum_{j=0}^{n-1} \sum_{i=1}^{n} f_1(x_i) f_2\left(x_{\sigma^j(i)}\right).$$

The Rearrangement Inequality, (2.3), now implies that the largest term in the outer summation occurs when $j = 0$, so

$$(F_1F_2)(\boldsymbol{x}) \leqslant \frac{1}{n} \sum_{i=1}^{n} (f_1 f_2)(x_i), \tag{2.8}$$

which verifies (2.6). $\square$

3 Proof of Theorem 1.3

3.1 Overview

In this section we prove Theorem 1.3, devoting most of our effort to the left inequality. In particular, we consider the level sets of $\eta_\ell(\boldsymbol{x})$ and then apply the technique of Lagrange multipliers in order to determine the extremal behavior of

$$U(\boldsymbol{x}) := \prod_{j=1}^{n} \mu_\ell(x_j)^{\frac{1}{n}}.$$

For each R in the range of η_ℓ on $\mathbb{R}^n_{\geqslant 0}$, define the surface $\Omega(R) = \{\boldsymbol{x} \in \mathbb{R}^n_{\geqslant 0} : \eta_\ell(\boldsymbol{x}) = R\}$. Given $d > 0$ and an integer k with $1 \leqslant k \leqslant n$, we define $\boldsymbol{c}_k(d) = d(\boldsymbol{e_1} + \cdots + \boldsymbol{e_k}) \in \mathbb{R}^n$, where $\boldsymbol{e_i}$ is the i-th standard basis vector. We refer to a point in $\mathbb{R}^n_{\geqslant 0}$ as a *k-diagonal* point if it is any coordinate permutation of a point of the form $\boldsymbol{c}_k(d)$ for some $d > 0$.

Lemma 3.1. *If $1 \leqslant \ell \leqslant n-1$, $1 < R < \infty$, and $\boldsymbol{z}$ is a maxima of $U(\boldsymbol{x})$ on $\Omega(R)$ then $\boldsymbol{z}$ is a k-diagonal point for some k with $1 \leqslant k \leqslant n$.*

We will prove Lemma 3.1 in Section 3.2 using the method of Lagrange multipliers (see Theorem 3.3).

By the symmetry of f and η_ℓ, if we restrict our attention to a single point $\boldsymbol{x} \in \mathbb{R}^n_{\geqslant 0}$, we may assume that there is some $0 \leqslant k \leqslant n$ such that $x_1, \ldots, x_k > 0$ and $x_{k+1} = \cdots = x_n = 0$. With this in mind, we generalize the functions defined in (1.3) and (1.4) by setting

$$\begin{aligned} \eta_{\ell,k}(\boldsymbol{x}) &:= \eta_\ell(x_1, \ldots, x_k, 0, \ldots, 0), \\ \mu_{\ell,k}(y) &:= \eta_\ell(\underbrace{y, \ldots, y}_{k \text{ times}}, 0, \ldots, 0). \end{aligned} \tag{3.1}$$

These are related to our earlier definitions by $\eta_\ell = \eta_{\ell,n}$ and $\mu_\ell = \mu_{\ell,n}$, and we also have the further relations

$$\eta_{\ell,k}(\boldsymbol{x}) = \sum_{\substack{I \subseteq [k] \\ |I| \leqslant \ell}} x_I, \qquad \mu_{\ell,k}(y) = \eta_{\ell,k}(y, \ldots, y) = \sum_{j=0}^{\ell} \binom{k}{j} x^j, \tag{3.2}$$

and finally, $\eta_{\ell,k}(\boldsymbol{x}) = \eta_{k,k}(\boldsymbol{x}) = \prod_{i=1}^{k}(1 + x_i)$ if $\ell \geqslant k$.

Lemma 3.2. *If $0 \leqslant \ell \leqslant n$, $1 \leqslant k \leqslant n$, and $y \geqslant 0$, then*

$$\mu^k_{\ell,n}(y) \leqslant \mu^n_{\ell,k}(y).$$

The inequality is tight if and only if $\ell = 0$, $\ell = n$, $y = 0$ or $k = n$.

We will prove Lemmas 3.1 and 3.2 in Sections 3.2 and 3.3, respectively. We now show how they imply Theorem 1.3.

Proof of Theorem 1.3. If $\ell = 0$ or $\boldsymbol{x} = 0$, then all terms in the two inequalities are equal to 1. If $x_1 = \cdots = x_n = y$ then all terms are identically equal to $\mu_\ell(y)$. If $\ell = n$, the left inequality is an identity and the right inequality is an application of the arithmetic-geometric mean inequality. In general, the right inequality follows from Proposition 1.1 part 1 and Proposition 2.1 part 2. So it suffices to prove the left inequality in the case where $\boldsymbol{x} \in \mathbb{R}^n_{\geqslant 0}$, $\boldsymbol{x} \neq 0$ and $1 \leqslant \ell \leqslant n-1$.

Since η_ℓ is strictly increasing in each variable, a non-zero $\boldsymbol{x}$ is contained in the surface $\Omega(R)$ with $R = R(\boldsymbol{x}) = \eta_\ell(\boldsymbol{x}) > \eta_\ell(0) = 1$. Since $\eta_\ell(\boldsymbol{x}) \geqslant 1 + x_i$ for all $1 \leqslant i \leqslant n$, $\Omega(R)$ is bounded. Since $\eta_\ell(\boldsymbol{x})$ is continuous, $\Omega(R)$ is also closed, and hence compact.

This means that there exists at least one point $\boldsymbol{z} \in \Omega(R)$ at which $f(\boldsymbol{x})$ takes its maximum value on $\Omega(R)$. Since $R > 1$, $\boldsymbol{z} \neq 0$ and, by Lemma 3.1, $\boldsymbol{z} = \boldsymbol{c}_k(d)$, for some $1 \leqslant k \leqslant n$ and $d > 0$. By Lemma 3.2, if $k < n$, then

$$f(\boldsymbol{z}) = (\mu_{\ell,n}(d))^{\frac{k}{n}} < \mu_{\ell,k}(d) = \eta_\ell(\boldsymbol{z}) = R.$$

However, if $k = n$, we have $f(\boldsymbol{z}) = \mu_\ell(d) = \eta_\ell(\boldsymbol{z}) = R$.

□

3.2 Lagrange multipliers and maxima

In this section we prove Lemma 3.1 using Proposition 3.3.1 of [1], p.284. We restate this result as Theorem 3.3, a form more suitable to our purposes.

Theorem 3.3 (The method of Lagrange multipliers with inequality constraints.)**.** *Let $f, h_1, \ldots, h_m, g_1, \ldots, g_r : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable functions. Suppose $\boldsymbol{z}$ is a point at which $f(\boldsymbol{x})$ has a local maximum over $\Omega = \{\boldsymbol{x} : h_1(\boldsymbol{x}) = \cdots = h_m(\boldsymbol{x}) = 0, g_1(\boldsymbol{x}), \ldots, g_r(\boldsymbol{x}) \geqslant 0\}$. Suppose also that $\boldsymbol{z}$ is regular, i.e. $\{\nabla h_1(\boldsymbol{z}), \ldots, \nabla h_m(\boldsymbol{z})\} \cup \{\nabla g_i(\boldsymbol{z}) : i \in A(\boldsymbol{z})\}$ is a linearly independent set where $A(\boldsymbol{x}) := \{1 \leqslant j \leqslant r : g_j(\boldsymbol{x}) = 0\}$. Then there exist unique Lagrange multiplier vectors $\boldsymbol{\lambda'} \in \mathbb{R}^m$ and $\boldsymbol{\rho'} \in \mathbb{R}^r_{\geqslant 0}$, such that*

$$\begin{aligned} \tfrac{\partial L}{\partial x_i}(\boldsymbol{z}, \boldsymbol{\lambda'}, \boldsymbol{\rho'}) &= 0, \quad 1 \leqslant i \leqslant n, \\ \rho'_j &= 0, \quad j \notin A(z) \end{aligned}$$

where $L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\rho}) := f(\boldsymbol{x}) + \sum_{i=1}^m \lambda_i h_i(\boldsymbol{x}) + \sum_{j=1}^r \rho_j g_j(\boldsymbol{x})$.

Proof of Lemma 3.1. Theorem 3.3 applies to the present setting with $\Omega(R)$ as the set Ω, constraint function $h(\boldsymbol{x}) := \eta_\ell(\boldsymbol{x}) - R$, and inequality constraints $g_i(\boldsymbol{x}) := x_i$ for $1 \leqslant i \leqslant n$. The Lagrangian function is then

$$L(\boldsymbol{x}, \lambda, \boldsymbol{\rho}) := U(\boldsymbol{x}) + \lambda(\eta_l(\boldsymbol{x}) - R) + \sum_{i=1}^n \rho_i x_i.$$

Let $\boldsymbol{z} \in \Omega(R)$ be a point at which $U(\boldsymbol{x})$ attains its maximum value over $\Omega(R)$. Since $R > 1$, $\boldsymbol{z}$ has at least one non-zero coordinate. By symmetry we may assume $z_1, z_2, \ldots, z_k > 0$ and $z_{k+1} = \ldots = z_n = 0$ for some $k \geqslant 1$. Since $\nabla h(\boldsymbol{z})$ trivially has positive coordinates, $\nabla h(\boldsymbol{z})$ together with $\nabla g_i(\boldsymbol{z}) = \boldsymbol{e_i}$, for $k+1 \leqslant i \leqslant n$, form a linearly independent set. Thus $\boldsymbol{z}$ is regular, and the conditions of Theorem 3.3 are met.

The theorem statement now implies that there is a constant λ' such that

$$\frac{1}{n}\frac{\mu'_{\ell,n}(z_i)}{\mu_{\ell,n}(z_i)}U(\boldsymbol{z}) + \lambda'\frac{\partial \eta_\ell}{\partial x_i}(\boldsymbol{z}) = 0, \qquad 1 \leqslant i \leqslant k.$$

It is trivial to check that $U(\boldsymbol{z}) \geqslant 1$ and $\frac{\partial \eta_\ell}{\partial x_i}(\boldsymbol{z}) \geqslant 1$ for all i with $1 \leqslant i \leqslant n$. Thus, if we define

$$\gamma_i := \frac{1}{n}\frac{\mu'_{\ell,n}(z_i)}{\mu_{\ell,n}(z_i)\frac{\partial \eta_\ell}{\partial x_i}(\boldsymbol{z})}, \qquad 1 \leqslant i \leqslant k,$$

then $\gamma_1 = \cdots = \gamma_k = -\lambda'/f(z)$. Note that

$$\gamma_i = \frac{\sum\limits_{j=0}^{\ell-1}\binom{n-1}{j}z_i^j}{\mu_{\ell,n}(z_i)\sum\limits_{j=0}^{\ell-1}\sum\limits_{\substack{I\subseteq[k]-i\\|I|=j}} z_I}, \qquad 1 \leqslant i \leqslant k.$$

If j, k are non-negative integers, define

$$Z_{j,k} := \sum_{\substack{I\subseteq[k]\setminus\{1,2\}\\|I|=j}} z_I.$$

Suppose that $z_3, \ldots, z_k$ are fixed. We will show that the equality $\gamma_1 = \gamma_2$ holds if and only if $z_1 = z_2$. By symmetry, this will then prove that $z_1 = \cdots = z_k$, completing the proof that $\boldsymbol{z}$ is a k-diagonal point.

Observe that we can write

$$\gamma_1 = \frac{\sum\limits_{j=0}^{\ell-1}\binom{n-1}{j}z_1^j}{\mu_{\ell,n}(z_1)\left((1+z_2)\sum_{j=0}^{\ell-2}Z_{j,k} + Z_{\ell-1,k}\right)}, \tag{3.3}$$

and

$$\gamma_2 = \frac{\sum\limits_{j=0}^{\ell-1}\binom{n-1}{j}z_2^j}{\mu_{\ell,n}(z_2)\left((1+z_1)\sum_{j=0}^{\ell-2}Z_{j,k} + Z_{\ell-1,k}\right)}. \tag{3.4}$$

Comparing (3.3) and (3.4), it is now clear that $\gamma_1 = \gamma_2$ if and only if $\Gamma_{\ell,k}(z_1) = \Gamma_{\ell,k}(z_2)$, where

$$\Gamma_{\ell,k}(y) := \frac{\sum\limits_{j=0}^{\ell-1}\binom{n-1}{j}y^j \cdot \left((1+y)\sum_{j=0}^{\ell-2}Z_{j,k} + Z_{\ell-1,k}\right)}{\mu_{\ell,n}(y)}. \tag{3.5}$$

To conclude, we show that $\Gamma_{\ell,k}(y)$ is strictly decreasing on $[0,\infty)$ and, hence, one-to-one. Note that

$$\Gamma_{\ell,k}(y) = \widetilde{\Gamma}_{\ell,k}(y) \cdot \left(A + \frac{B}{1+y}\right),$$

where

$$\widetilde{\Gamma}_{\ell,k}(y) := \frac{\sum\limits_{j=0}^{\ell-1} \binom{n-1}{j} y^j \cdot (1+y)}{\mu_{\ell,n}(y)}.$$

and where $A, B > 0$ are constants. We show that $\widetilde{\Gamma}'_{\ell,k}(y) < 0$. This implies $\widetilde{\Gamma}_{\ell,k}(y)$ is strictly decreasing and hence $\Gamma_{\ell,k}(y)$ is strictly decreasing as well. Simple algebra shows that

$$\widetilde{\Gamma}_{\ell,k}(y) = 1 - \frac{\binom{n-1}{\ell} y^\ell}{\sum_{j=0}^{\ell} \binom{n}{j} y^j},$$

and thus

$$\widetilde{\Gamma}'_{\ell,k}(y) = \binom{n-1}{\ell} \cdot \frac{-\ell y^{\ell-1} \sum_{j=0}^{\ell} \binom{n}{j} y^j + y^\ell \sum_{j=0}^{\ell} j \binom{n}{j} y^{j-1}}{\mu_{\ell,n}(y)^2}.$$

The numerator simplifies to

$$-z^\ell \sum_{j=0}^{\ell} (\ell - j) \binom{n}{j} y^j < 0,$$

and the proof is complete. □

3.3 Majorization and polynomial inequalities

In this section we use a partial order on polynomials in order to prove Lemma 3.2. Let $f(y) = \sum_{n \geqslant 0} c(n) y^n$ and $g(y) = \sum_{n \geqslant 0} d(n) y^n$ be polynomials in y with real coefficients. We say that f is dominated by g in the *coefficient partial order*, denoted $f \sqsubseteq g$, if and only if $c(n) \leqslant d(n)$ for all $n \geqslant 0$. If $f \sqsubseteq g$, and $c(n) < d(n)$ for some $n \geqslant 0$, then we denote this by $f \sqsubset g$. We write $0 \sqsubseteq f$ if and only if f has all coefficients non-negative and $0 \sqsubset f$ if and only f has all coefficients non-negative and at least one coefficient positive.

Proposition 3.4. *If $a > b \geqslant 0$ are integers, then*

$$\mu_{\ell,a}(y)\mu_{\ell,b}(x) \sqsubseteq \mu_{\ell,a-1}(x)\mu_{\ell,b+1}(x).$$

We have $\mu_{\ell,a}(x)\mu_{\ell,b}(x) \sqsubset \mu_{\ell,a-1}(x)\mu_{\ell,b+1}(x)$ if and only if, additionally, $a \geqslant b+2$, $a-1 \geqslant \ell$, and $\ell \geqslant 1$.

The proof of Proposition 3.4 requires the two following lemmas.

Lemma 3.5. *If A, B, M, N are integers with $A \geqslant B \geqslant 0$ and $M \geqslant N \geqslant 0$, then*

$$\binom{A}{M}\binom{B}{N} \geqslant \binom{A}{N}\binom{B}{M}.$$

For these same ranges of parameters, the inequality is strict if and only if $A > B$, $M > N$, $A \geqslant M$, and $B \geqslant N$.

Proof. If $A = B$ or $M = N$ then both sides of the inequality are identically equal. If $M > A$, then $M > B$ and both sides are 0. If $N > B$, then $M > B$ and, again, both sides are 0. We may now assume $A > B, M > N, A \geqslant M$, and $B \geqslant N$. Canceling factorials, the desired inequality becomes

$$(A - N)_{M-N} > (B - N)_{M-N},$$

where, for non-negative integers n and real numbers α, $(\alpha)_n := \prod_{i=0}^{n-1}(x - i)$ is the falling factorial. Since $(\alpha)_n > (\beta)_n$ if $\alpha > \beta \geqslant n - 1 \geqslant 0$, we are done. $\square$

Lemma 3.6. *Suppose that f_1, f_2, g_1, g_2 are polynomials. If $0 \sqsubseteq f_1 \sqsubseteq f_2$ and $0 \sqsubseteq g_1 \sqsubseteq g_2$, then $f_1 g_1 \sqsubseteq f_2 g_2$. If, additionally, $f_1 \sqsubset f_2$ and $0 \sqsubset g_2$, then $f_1 g_1 \sqsubset f_2 g_2$.*

Proof. Denoting the coefficients by $f_i = \sum_{j\geqslant 0} f_{i,j} y^j$ and $g_i = \sum_{j\geqslant 0} g_{i,j} y^j$ for $i = 1, 2$, the conditions of the lemma state that $0 \leqslant f_{1,j} \leqslant f_{2,j}$ and $0 \leqslant g_{1,j} \leqslant g_{2,j}$ for all $j \geqslant 0$. Writing their products as $f_1 g_1 = \sum_{l\geqslant 0} a_l y^l$ and $f_2 g_2 = \sum_{l\geqslant 0} b_l y^l$, the coefficients then satisfy

$$a_l = \sum_{j,k\geqslant 0, j+k=l} f_{1,j} g_{1,k} \leqslant \sum_{j,k\geqslant 0, j+k=l} f_{2,j} g_{2,k} = b_l,$$

since $f_{1,j} g_{1,k} \leqslant f_{2,j} g_{2,k}$ for all $j, k \geqslant 0$.

If there is additionally some pair j, k such that $0 \leqslant f_{1,j} < f_{2,j}$ and $0 < g_{2,k}$, then $f_{1,j} g_{1,k} < f_{2,j} g_{2,k}$, and therefore the stronger conclusion $a_{j+k} < b_{j+k}$ holds. $\square$

Proof of Proposition 3.4. By definition, proving that $\mu_{\ell,a}(y)\mu_{\ell,b}(y) \sqsubseteq \mu_{\ell,a-1}(y)\mu_{\ell,b+1}(y)$ is the same as proving that for each $0 \leqslant m \leqslant 2\ell$ we have

$$\sum_{d=M'}^{M} \binom{a}{d}\binom{b}{m-d} \leqslant \sum_{d=M'}^{M} \binom{a-1}{d}\binom{b+1}{m-d}, \tag{3.6}$$

where $M' := \max\{0, m - \ell\}$ and $M := \min\{\ell, m\}$. Likewise, the stronger condition that $\mu_{\ell,a}(y)\mu_{\ell,b}(y) \sqsubset \mu_{\ell,a-1}(y)\mu_{\ell,b+1}(y)$ is equivalent to additionally proving that there is an m with $0 \leqslant m \leqslant 2\ell$ for which (3.6) is strict. Applying Pascal's identity $\binom{a}{d} = \binom{a-1}{d} + \binom{a-1}{d-1}$ to the left-side and $\binom{b+1}{m-d} = \binom{b}{m-d} + \binom{b}{m-d-1}$ to the right, and then cancelling like terms, (3.6) becomes

$$\sum_{d=M'}^{M} \binom{a-1}{d-1}\binom{b}{m-d} \leqslant \sum_{d=M'}^{M} \binom{a-1}{d}\binom{b}{m-d-1}. \tag{3.7}$$

Furthermore, after a summation index shift all of the terms but one in (3.7) cancel, leaving only $d = M'$ on the left-side and $d = M$ on the right:

$$\binom{a-1}{M'-1}\binom{b}{m-M'} \leqslant \binom{a-1}{M}\binom{b}{m-M-1}.$$

If $M' = 0$, then this inequality is trivially satisfied, and thus so is (3.6). We therefore need only consider the case that $M' = m - \ell$. This means that $m \geqslant \ell$, so in this case $M = \ell$, and (3.6) is finally equivalent to the inequality

$$\binom{a-1}{m-\ell-1}\binom{b}{\ell} \leqslant \binom{a-1}{\ell}\binom{b}{m-\ell-1}.$$

We now apply Lemma 3.5 with $A = a-1$, $B = b$, $M = \ell$, $N = m-\ell-1$ to complete the proof. The inequality easily follows. It is also easy to see that (3.6) is strict in the cases where $a \geqslant b+2$, $a-1 \geqslant \ell$, and $2\ell \geqslant m \geqslant \ell+1$ (and hence $\ell \geqslant 1$).

□

Remark 3.7. Proposition 3.4 (and its proof) can also be interpreted combinatorially. In particular, consider two rows consisting of a and b square cells, respectively. The y^m coefficient in

$$\sum_{i=0}^{\ell}\binom{a}{i}y^i \cdot \sum_{j=0}^{\ell}\binom{b}{j}y^j$$

is the number of ways of marking exactly m of the cells subject to the restriction that there are at most ℓ marked cells in each row, and the result then states that if $a > b$, then there are at least as many ways to mark two rows of length $a-1$ and $b+1$ subject to the same restriction.

Corollary 3.8. *If $y \geqslant 0$, $\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2} \in \Delta(m)$ each have integer coordinates, and $\boldsymbol{\lambda_1} \succeq \boldsymbol{\lambda_2}$ then*

$$\prod_{i=1}^{m}\mu_{\ell,\lambda_1(i)}(y) \sqsubseteq \prod_{i=1}^{m}\mu_{\ell,\lambda_2(i)}(y).$$

.

Proof. Suppose $\boldsymbol{\lambda_1} \neq \boldsymbol{\lambda_2}$. By definition, there must then be two indices $1 \leqslant \alpha < \beta \leqslant m$ such that $\lambda_1(\alpha) > \lambda_2(\alpha)$ and $\lambda_1(\beta) < \lambda_2(\beta)$. Define $\boldsymbol{\lambda_1'}$ by setting

$$\lambda_1'(\alpha) := \lambda_1(\alpha) - 1, \qquad \lambda_1'(\beta) := \lambda_1(\beta) + 1,$$

and $\lambda_1'(i) := \lambda_1(i)$ for all $i \neq \alpha, \beta$. Importantly, it is still true that $\boldsymbol{\lambda_1'}$ majorizes $\boldsymbol{\lambda_2}$.

Noting that $\lambda_1(\alpha) > \lambda_2(\alpha) \geqslant \lambda_2(\beta) > \lambda_1(\beta)$, Proposition 3.4 now states that

$$\mu_{\ell,\lambda_1(\alpha)}(y)\mu_{\ell,\lambda_1(\beta)}(y) \sqsubseteq \mu_{\ell,\lambda_1'(\alpha)}(y)\mu_{\ell,\lambda_1'(\beta)}(y)$$

which, combined with Lemma 3.6, implies that

$$\prod_{i=1}^{m}\mu_{\ell,\lambda_1(i)}(x) \sqsubseteq \prod_{i=1}^{m}\mu_{\ell,\lambda_1'(i)}(x). \tag{3.8}$$

If $\boldsymbol{\lambda_1'} = \boldsymbol{\lambda_2}$, then (3.8) gives the statement of the corollary. Otherwise, the above procedure is repeated (a finite number of steps) until this is the case. $\square$

Applying this result with the partitions $\boldsymbol{\lambda_1} = n^k$ and $\boldsymbol{\lambda_2} = k^n$ will finally complete the proof of Lemma 3.2.

Proof of Lemma 3.2. Corollary 3.8 implies $\mu_{\ell,n}(y)^k \sqsubseteq \mu_{\ell,k}(y)^n$. Since this partial order requires that all coefficients be dominated, this immediately implies that $\mu_{\ell,n}(y)^k \leqslant \mu_{\ell,k}(y)^n$ for all $y \geqslant 0$. Clearly, if $k = n$, $\ell = 0$, $\ell = n$, or $x = 0$, then $\mu_{\ell,n}(x)^k = \mu_{\ell,k}(x)^n$. It therefore remains to be shown that the inequality is strict if $1 \leqslant \ell \leqslant n-1$, $k \leqslant n-1$ and $x > 0$.

Proposition 3.4 implies that $\mu_{\ell,n}\mu_{\ell,0} \sqsubset \mu_{\ell,n-1}\mu_{\ell,1}$. Following the proof method of Proposition 3.4, we introduce the dummy term $\mu_{\ell,0} = 1$ and find

$$\mu_{\ell,n}^k = \mu_{\ell,n}^{k-1}(\mu_{\ell,n}\mu_{\ell,0}) \sqsubset \mu_{\ell,n}^{k-1}(\mu_{\ell,n-1}\mu_{\ell,1}) \sqsubseteq \mu_{\ell,k}^n.$$

The second relation follows from Lemma 3.6, and the third follows from Corollary 3.8. Since $\mu_{\ell,n}^k \sqsubset \mu_{\ell,k}^n$ and $x > 0$, we conclude that $\mu_{\ell,n}(x)^k < \mu_{\ell,k}(x)^n$ $\square$

4 Inequalities for sums of Bernoulli random variables

In this brief section we describe the relationship between our quasi-mean inequalities in Theorem 1.3 and the distributions of sums of Bernoulli random variables.

Proof of Theorem 1.4. The inequality is trivial if $\ell = n$, so we henceforth assume that $\ell < n$. Furthermore, if $p_i = 1$ for some i with $1 \leqslant i \leqslant n$, then $\mathbf{P}(C_i \leqslant \ell) = 0$ and the inequality is again trivially true. We therefore also assume that $p_i \in [0,1)$ for each i.

All of the events $\{C_j \leqslant \ell\}$ are independent, and their individual probabilities are given by

$$\mathbf{P}(C_j \leqslant \ell) = \sum_{0 \leqslant m \leqslant \ell} \binom{n}{m} p_j^m (1-p_j)^{n-m}.$$

Thus

$$\mathbf{P}\Big(\max_{1 \leqslant j \leqslant n}\{C_j\} \leqslant \ell\Big) = \prod_{j=1}^{n} \sum_{0 \leqslant m \leqslant \ell} \binom{n}{m} p_j^m (1-p_j)^{n-m}. \tag{4.1}$$

Similarly, the events $\{R_i \leqslant \ell\}$ are also independent, and their probabilities are given by

$$\mathbf{P}(R_i \leqslant \ell) = \sum_{0 \leqslant m \leqslant \ell} \sum_{\substack{I \subseteq [n] \\ |I| = m}} \prod_{j \in I} p_j \prod_{j \notin I} (1-p_j),$$

so

$$\mathbf{P}\Big(\max_{1 \leqslant i \leqslant n}\{R_i\} \leqslant \ell\Big) = \left(\sum_{0 \leqslant m \leqslant \ell} \sum_{\substack{I \subseteq [n] \\ |I| = m}} \prod_{j \in I} p_j \prod_{j \notin I} (1-p_j) \right)^n. \tag{4.2}$$

Dividing (4.1) and (4.2) by $\prod_{j=1}^{n}(1-p_j)$, we see that the desired inequality is equivalent to the left inequality from Theorem 1.3 with $x_i = p_i/(1-p_i)$ for $1 \leqslant i \leqslant n$. $\square$

References

[1] D. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1995.

[2] G. Hardy, J. Littlewood, and G. Pólya, *Inequalities, 2nd. ed.*, Cambridge University Press, Cambridge, 1952.

[3] R. Muirhead, *Some methods applicable to identities of symmetric algebraic functions of n letters*, Proc. Edinburgh Math. Soc. **21** (1902/03), 144–157.

[4] R. Stanley, *Enumerative Combinatorics, vol. 2*, Cambridge University Press, Cambridge, 1999.