

# The Gini Index in the Representation Theory of the General Linear Group

Grant Kopitzke

*Department of Mathematics, University of Wisconsin - Stevens Point, Stevens Point, WI 54481*  
*Email: [gpkopitzk@uwsp.edu](mailto:gpkopitzk@uwsp.edu)*

**Received:** February 27, 2023, **Accepted:** May 15, 2023, **Published:** May 26, 2023  
 The authors: Released under the CC BY-ND license (International 4.0)

**ABSTRACT:** The Gini index is a function that attempts to measure the amount of inequality in the distribution of a finite resource throughout a population. It is commonly used in economics as a measure of inequality of income or wealth. We define a discrete Gini index on the set of integer partitions with at most  $n$  parts and show how this function emerges in the representation theory of the complex general linear group.

**Keywords:** Earth mover's distance; General linear group; Gini index; Integer partition; Kostka-Foulkes polynomials; Lorenz curve; Representation theory; Wasserstein metric

**2020 Mathematics Subject Classification:** 05E10; 20G05; 91B82

## 1. Introduction

The Gini index, originally defined in 1912 by the Italian statistician Corrado Gini [8], is a function that measures statistical dispersion. Traditionally used in economics to measure how equitably a resource is distributed throughout a population, the Gini index has found a wide range of modern applications in fields including biology [2], the medical sciences [20], and climate science [21].

We will review the discretization of the Gini index to the set of partitions of a fixed positive integer  $n$  (defined in [13]), and will extend this discretization to the set of partitions of a positive multiple of  $n$  with  $n$  parts. In section 3, we review the representation theory of the complex general linear group,  $GL_n(\mathbb{C})$ , of invertible  $n \times n$  complex matrices – with an emphasis on representations of  $GL_n(\mathbb{C})$ -harmonic polynomials. We will then show how the discrete Gini index appears in this setting as the top homogeneous degree in which an irreducible rational representation of the general linear group occurs as a factor in the  $GL_n(\mathbb{C})$ -harmonic polynomials (Theorem 3.1).

## 2. The discrete Gini index

### 2.1 The Gini index of an integer partition

The classical Gini index is usually defined in terms of Lorenz curves. The Lorenz curve of a distribution is the graph of a function  $L$  for which  $L(x)$  is the percentage of total wealth possessed by the poorest  $x$  percent of the population [17]. From this definition it follows that, regardless of the distribution,  $L$  will be an increasing, convex function on the interval from 0 to 1,  $L(0) = 0$ , and  $L(1) = 1$ .

The data sets we use to measure inequality are necessarily discrete. To make the corresponding Lorenz curve continuous, the curve is often constructed by approximating the discrete data set with a smooth curve [12], or by linear interpolation [6]. Bypassing this process, a discrete version of the Gini index was defined in [13] as follows. Let  $\lambda = (\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n)$  be a partition of an integer  $n$  – that is, a finite tuple of  $n$  non-negative decreasing integers whose sum is equal to  $n$ . Each such partition corresponds to a wealth distribution in which there are  $n$  dollars distributed among  $n$  people such that the most wealthy person has  $\lambda_1$  dollars, and the least wealthy has  $\lambda_n$  dollars. To measure the amount of inequality in the wealth distribution  $\lambda$ , the Gini index of  $\lambda$  is then defined as

$$g(\lambda) := \binom{n+1}{2} - \sum_{i=1}^n i\lambda_i. \quad (1)$$

This definition is best understood in terms of discrete Lorenz curves. The discrete Lorenz curve of an integer partition  $\lambda$  of  $n$  is defined on the interval from 0 to  $n$  by

$$L_\lambda(x) := \begin{cases} 0 & x = 0 \\ \sum_{i=0}^{j-1} \lambda_{n-i} & x \in (j-1, j] \end{cases} \quad (2)$$

for  $j = 1, \dots, n$ .

**Example 2.1.** Suppose there are 5 dollars distributed among a population of 5 people according to the partition  $(3 \geq 1 \geq 1 \geq 0 \geq 0)$ . The most equitable distribution would be the so-called flat partition  $(1 \geq 1 \geq 1 \geq 1 \geq 1)$  in which each person has the same amount of money; the Lorenz curve of this distribution is sometimes called the line of equity or the line of equality. The Lorenz curves of each of these distributions are pictured in Figure 1.

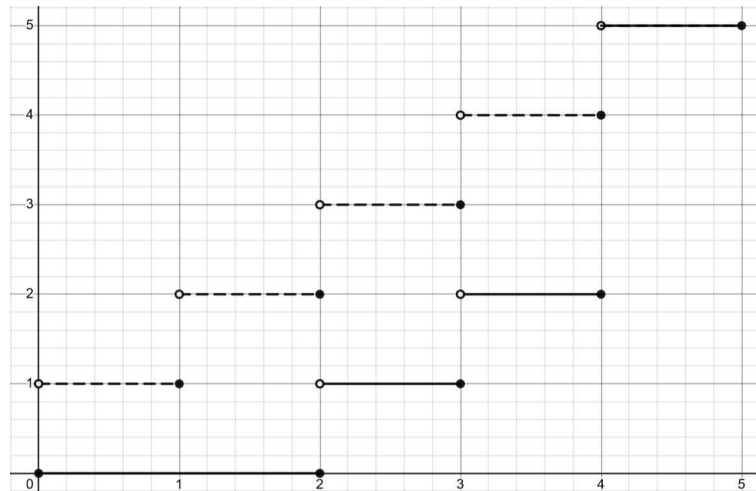


Figure 1: The line of equality (dashed) and the Lorenz curve of the partition  $(3 \geq 1 \geq 1 \geq 0 \geq 0)$  of 5 (solid).

The Gini index of a partition  $\lambda$  can easily be seen as the difference in the area between the line of equality and the curve  $L_\lambda(x)$ . Indeed, one computes that the area under the line of equality is  $\binom{n+1}{2}$ , and the area under  $L_\lambda(x)$  is  $\sum_{i=1}^n i\lambda_i$ . In Example 2.1 we see that the Gini index of  $(3 \geq 1 \geq 1 \geq 0 \geq 0)$  is 7. The Lorenz curve of  $\lambda$  will always coincide with the line of equality on the interval  $(n-1, n]$ , so the formula in Equation 1 simplifies to:

$$g(\lambda) = \binom{n}{2} - \sum_{i=1}^n (i-1)\lambda_i. \quad (3)$$

The function  $\sum_{i=1}^n (i-1)\lambda_i$  appears frequently throughout algebraic combinatorics and is sometimes referred to as the weighted total of the partition  $\lambda$  [18]; thus we will adopt the function notation

$$b(\lambda) := \sum_{i=1}^n (i-1)\lambda_i. \quad (4)$$

## 2.2 A generalization of the discrete Gini index

The definition of the Gini index in Equation 3 is unnecessarily restrictive, and can easily be generalized to allow for dollar amounts that are multiples of  $n$ . To that end, let  $n$  and  $k$  be positive integers, and consider a wealth distribution in which  $nk$  dollars are distributed among  $n$  people. The possible distributions are in one-to-one correspondence with the partitions of  $nk$  with at most  $n$  parts.

The discrete Gini index can be generalized to this setting by extending the notion of the discrete Lorenz curve and defining the Gini index of a distribution to be the area between the line of equality and the discrete Lorenz curve of the distribution. The Gini index in this setting is best understood via the following example.

**Example 2.2.** Suppose 15 dollars are distributed among 5 people according to the partition  $\lambda = (6 \geq 4 \geq 3 \geq 1 \geq 1)$ . The line of equality is the discrete Lorenz curve of the most equitable distribution, which in this setting is the flat partition  $(3 \geq 3 \geq 3 \geq 3 \geq 3)$ , or  $(3^5)$  in simplified form. The discrete Lorenz curves for these distributions are pictured in Figure 2.

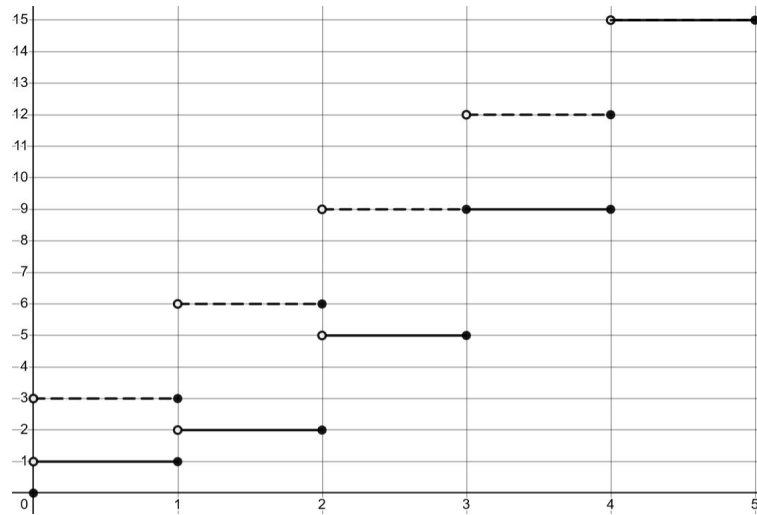


Figure 2: The line of equality (dashed) and the discrete Lorenz curve (solid) of the partition  $(6 \geq 4 \geq 3 \geq 1 \geq 1)$  of 15 with 5 parts.

We define the Gini index  $g_{15,5}(\lambda)$  of this distribution to be the area between the line of equality and the discrete Lorenz curve of  $\lambda = (6 \geq 4 \geq 3 \geq 1 \geq 1)$ :

$$b((3^5)) - b(\lambda) = 13.$$

In general, if  $n$  and  $k$  are positive integers and  $\lambda$  is a partition of  $nk$  with at most  $n$  parts, then the Gini index of  $\lambda$  is the area between the line of equality (the discrete Lorenz curve of the flat partition  $(k^n)$ ) and the discrete Lorenz curve of  $\lambda$ :

$$g_{nk,n}(\lambda) := b((k^n)) - b(\lambda). \quad (5)$$

Notice that in the case  $k = 1$ , the function in Equation 5 specializes to the Gini index defined in Equation 3.

### 3. Representation theory

This section provides a brief overview of the basic representation theory of the complex general linear group and the structures relevant to our main result. Even so, some familiarity with the representation theory of classical groups is assumed.

#### 3.1 Representation theory of $GL_n(\mathbb{C})$

The complex general linear group,  $GL_n(\mathbb{C})$ , is the set of all  $n \times n$  complex invertible matrices, equipped with the operation of matrix multiplication. A finite dimensional\* representation of  $GL_n(\mathbb{C})$  is a homomorphism  $\rho : GL_n(\mathbb{C}) \rightarrow GL(V)$ , where  $GL(V)$  is the group of invertible linear transformations of a finite-dimensional complex vector space  $V$ . The representation is *rational* if, after choosing a basis for  $V$ ,  $\rho(g)$  is a matrix whose entries are fixed rational functions in the matrix entries of  $g$ , where  $g \in GL_n(\mathbb{C})$ . When the action of  $\rho$  is understood, the vector space  $V$  itself is often referred to as the representation.

An irreducible representation (irrep) is a nonzero representation with no proper nontrivial subrepresentations. In other words,  $V$  is irreducible if there are no non-trivial subspaces  $W \subset V$  that are closed under the action of  $GL_n(\mathbb{C})$  restricted to  $W$ . The general linear group is *reductive*, which implies that all rational representations can be written as a direct sum of irreducible rational representations. Irreps of  $GL_n(\mathbb{C})$  are classified by the Theorem of the Highest Weight, which we now review.

Let  $H$  denote the subgroup of  $GL_n(\mathbb{C})$  consisting of diagonal matrices. Let  $V$  be a rational representation of  $GL_n(\mathbb{C})$ . A vector  $v$  in  $V$  has *weight*  $\alpha = (\alpha_1, \dots, \alpha_n)$  in  $\mathbb{Z}^n$  if  $\rho(x)v = x_1^{\alpha_1} \cdots x_n^{\alpha_n} v$  for all  $x = \text{diag}(x_1, \dots, x_n)$  in  $H$ .

Let  $B$  denote the Borel subgroup of upper triangular matrices in  $GL_n(\mathbb{C})$ . The weight  $\alpha$  of a vector  $v$  in  $V$  is called a *highest weight* of  $V$  if  $\rho(B)v = \rho(\mathbb{C}^*)v$ , where  $\mathbb{C}^*$  is the set of all non-zero complex numbers. Two irreducible rational representations are isomorphic if and only if they have the same highest weight. These concepts are summarized in the Theorem of the Highest Weight.

\*Unless otherwise stated, all representations in this paper are finite-dimensional.

**Theorem 3.1** (Theorem of the Highest Weight). *For every tuple  $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{Z}^n$  with  $\alpha_1 \geq \dots \geq \alpha_n$ , there is a unique irreducible rational representation  $V^\alpha$  of  $GL_n(\mathbb{C})$  with highest weight  $\alpha$ . Moreover, all such irreducible rational representations of  $GL_n(\mathbb{C})$  are of this form.*

For more on the representation theory of  $GL_n(\mathbb{C})$  we refer the reader to [7], [19], and [9].

### 3.2 $GL_n(\mathbb{C})$ –Harmonic polynomials

The discrete Gini index arises as the top degree in which an irrep appears as a direct summand in the space of harmonic polynomials on the Lie algebra of  $GL_n(\mathbb{C})$ . In this section, we provide a cursory overview of the  $GL_n(\mathbb{C})$ –harmonic polynomials.

It is well-known that the Lie algebra  $\mathfrak{g}$  of  $G = GL_n(\mathbb{C})$  is the Lie algebra of  $n \times n$  complex matrices,  $M_n(\mathbb{C})$ . The general linear group acts on its Lie algebra by the adjoint representation,

$$\text{Ad}(g)X := gXg^{-1},$$

for  $g \in G$  and  $X \in \mathfrak{g}$ . Choose a basis  $X_1, \dots, X_{n^2}$  for  $\mathfrak{g}$ , and define the algebra of polynomial functions on  $\mathfrak{g}$  by identifying

$$\mathbb{C}[\mathfrak{g}] = \mathbb{C}[X_1, \dots, X_{n^2}].$$

The group  $G$  acts on  $\mathfrak{g}$  via the adjoint representation, and therefore also acts on  $\mathbb{C}[\mathfrak{g}]$  by conjugation. The algebra  $\mathbb{C}[\mathfrak{g}]$  is thus an infinite dimensional graded representation of  $G$  with gradation

$$\mathbb{C}[\mathfrak{g}] = \bigoplus_{d \geq 0} \mathbb{C}[\mathfrak{g}]_d,$$

where  $\mathbb{C}[\mathfrak{g}]_d$  is the vector space of homogeneous degree  $d$  polynomials in  $\mathbb{C}[\mathfrak{g}]$ . The ring of  $G$ –invariant polynomials in  $\mathbb{C}[\mathfrak{g}]$  is the set

$$\mathbb{C}[\mathfrak{g}]^G = \{f \in \mathbb{C}[\mathfrak{g}] : g \cdot f = f \text{ for all } g \in G\}.$$

The coinvariant ring of  $G$  is the quotient

$$\mathbb{C}[\mathfrak{g}]_G := \mathbb{C}[\mathfrak{g}] / \mathbb{C}[\mathfrak{g}]_+^G$$

of the full polynomial ring by the ideal  $\mathbb{C}[\mathfrak{g}]_+^G$  of  $G$ –invariant polynomials without constant term.

Let  $\partial_i = \frac{\partial}{\partial x_i}$ , and for  $f \in \mathbb{C}[\mathfrak{g}]$ , define  $f(\partial) = f(\partial_1, \dots, \partial_{n^2})$ . The  $G$ –invariant differential operators with no constant term is the set

$$\mathcal{D}(\mathfrak{g})_+^G = \{f(\partial) | f \in \mathbb{C}[\mathfrak{g}]_+^G\}.$$

The module of  $G$ –harmonic polynomials on  $\mathfrak{g}$  is then defined by

$$\mathcal{H}(\mathfrak{g}) = \{f \in \mathbb{C}[\mathfrak{g}] | \Delta(f) = 0 \text{ for all } \Delta \in \mathcal{D}(\mathfrak{g})_+^G\}.$$

As a (infinite dimensional) representation of  $G$ ,  $\mathcal{H}(\mathfrak{g})$  is isomorphic to the coinvariant ring  $\mathbb{C}[\mathfrak{g}]_G$ .

The  $G$ –harmonic polynomials also form a graded representation of  $GL_n(\mathbb{C})$ ;

$$\mathcal{H}(\mathfrak{g}) = \bigoplus_{d=0}^{\infty} \mathcal{H}(\mathfrak{g})^d,$$

where  $\mathcal{H}^d(\mathfrak{g}) = \mathcal{H}(\mathfrak{g}) \cap \mathbb{C}[\mathfrak{g}]_d$ . This fact is non-trivial since it says that if a function is harmonic, then so are its homogeneous components. In general, every polynomial function can be expressed as a sum of  $G$ –invariant functions multiplied by  $G$ –harmonic polynomials. In other words, there is a surjection

$$\mathbb{C}[\mathfrak{g}]^G \otimes \mathcal{H}(\mathfrak{g}) \longrightarrow \mathbb{C}[\mathfrak{g}] \longrightarrow 0$$

obtained by linearly extending multiplication. Kostant showed in [15] that the product of invariants and harmonics is unique, and thus

$$\mathbb{C}[\mathfrak{g}] \cong \mathbb{C}[\mathfrak{g}]^G \otimes \mathcal{H}(\mathfrak{g}).$$

Each degree  $d$  homogeneous component,  $\mathcal{H}(\mathfrak{g})^d$ , is a finite-dimensional rational representation of  $G$ , and therefore decomposes into a finite direct sum of irreducible representations of  $G$ . We then pose the question, “given an irreducible rational representation  $V^\alpha$  of  $GL_n(\mathbb{C})$  with highest weight  $\alpha$ , in what degrees does  $V^\alpha$  occur in the direct sum decomposition of  $\mathcal{H}(\mathfrak{g})^d$ ?” This question can be further formalized by considering the *graded multiplicity* of the representation  $V^\alpha$ , which is defined as follows. If  $V^\alpha$  is a finite dimensional irreducible rational representation of  $G$  with highest weight  $\alpha$ , we denote by  $[V^\alpha, \mathcal{H}(\mathfrak{g})^d]$  the multiplicity with which  $V^\alpha$

occurs in the direct sum decomposition of the homogeneous degree  $d$  harmonic polynomials,  $\mathcal{H}(\mathfrak{g})^d$ . The degrees in which and multiplicities with which the representation  $V^\alpha$  occurs within the harmonic polynomials is summarized in what is called the *graded multiplicity polynomial* of  $V^\alpha$  in  $\mathcal{H}(\mathfrak{g})$ , and is defined by

$$m_\alpha(q) := \sum_{d=0}^{\infty} [V^\alpha, \mathcal{H}(\mathfrak{g})^d] q^d, \quad (6)$$

where  $q$  is a dummy variable. It follows from [15] that  $m_\alpha(q) \neq 0$  if and only if  $\alpha \neq 0$  and  $\alpha_1 + \cdots + \alpha_n = 0$ . We now state our main result.

**Theorem 3.1.** *Let  $V^\alpha$  be a  $GL_n(\mathbb{C})$  irrep of highest weight  $\alpha = (\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n)$  with  $\alpha_1 + \alpha_2 + \cdots + \alpha_n = 0$ . Let  $k \in \mathbb{N}$  be such that  $k \geq |\alpha_n|$ , and let  $\lambda := \alpha + (k^n)$ . Then  $\deg m_\alpha(q) = g_{nk,n}(\lambda)$ , where  $g_{nk,n}$  is the generalized Gini index defined in Equation 5.*

It is worth noting that the Gini index  $g_{nk,n}(\lambda)$  is independent of the choice of integer  $k$ , so long as  $k \geq |\alpha_n|$ . In simple terms, Theorem 3.1 states that for a given irreducible rational representation  $V^\alpha$  of  $GL_n(\mathbb{C})$ , the graded multiplicity polynomial  $m_\alpha(q)$  is either zero, or of degree  $g_{nk,n}(\lambda)$ , where  $\lambda = \alpha + (k^n)$ , and  $k$  is any integer greater than  $|\alpha_n|$ . In other words, the maximal homogeneous degree in which an irreducible representation of  $GL_n(\mathbb{C})$  appears in the  $GL_n(\mathbb{C})$ -harmonics is completely determined by the Gini index. The proof of this result is provided in Section 3.3.

### 3.3 Kostka-Foulkes polynomials, and proof of main result

If  $\lambda = (\lambda_1 \geq \cdots \geq \lambda_n)$  is a partition of  $nk$  with at most  $n$  parts, let  $\alpha \in \mathbb{Z}^n$  be given by  $\alpha := \lambda - (k^n)$ . Then  $\alpha_1 + \cdots + \alpha_n = 0$  and  $k \geq |\alpha_n|$ . We will show that, in this setting, the degree of  $m_\alpha(q)$  is  $g_{nk,n}(\lambda)$  by relating the graded multiplicities to *Kostka-Foulkes polynomials*.

Given two partitions  $\lambda$  and  $\mu$  of a positive integer  $n$ , the Kostka number  $K_{\lambda\mu}$  is the number of semistandard Young tableaux of shape  $\lambda$  and weight  $\mu$ . The Kostka numbers can also be defined as the coefficients obtained by expressing the Schur polynomial  $s_\lambda$  as a linear combination of monomial symmetric functions  $m_\mu$ :

$$s_\lambda = \sum_{\mu} K_{\lambda\mu} m_\mu.$$

These numbers are significant in representation theory, as  $K_{\lambda\mu}$  counts the dimension of the weight space corresponding to  $\mu$  in the irreducible representation  $V^\lambda$  with highest weight  $\lambda$ . This result is often expressed in terms of Kostant's partition function  $\mathcal{P} : \mathbb{Z}^n \rightarrow \mathbb{Z}$ . Let  $\epsilon_i \in \mathbb{Z}^n$  denote the tuple with a 1 in the  $i$ -th position, and 0's elsewhere. Define  $\epsilon_{ij} = \epsilon_i - \epsilon_j$ . The tuples  $\epsilon_{ij}$  with  $i \neq j$  form the root system of the Lie algebra  $\mathfrak{sl}_n(\mathbb{C})$ , and the positive roots are generally defined as the  $\epsilon_{ij}$  with  $i < j$ . Kostant's partition function  $\mathcal{P}(\alpha)$  counts the number of ways in which a weight  $\alpha$  of  $\mathfrak{sl}_n(\mathbb{C})$  can be expressed as a sum of positive roots of  $\mathfrak{sl}_n(\mathbb{C})$ . This function can also be defined as the coefficients in the series

$$\sum_{\alpha \in \mathbb{Z}^n} \mathcal{P}(\alpha) x_1^{\alpha_1} \cdots x_n^{\alpha_n} = \prod_{1 \leq i < j \leq n} \left(1 - \frac{x_i}{x_j}\right)^{-1}$$

where  $x_1, \dots, x_n$  are indeterminants. The Kostka numbers can then be written in terms of  $\mathcal{P}$ , yielding Kostant's weight multiplicity formula

$$K_{\lambda\mu} = \sum_{w \in S_n} \epsilon(w) \mathcal{P}(w(\lambda + \rho) - (\mu + \rho)),$$

where  $\rho = (n-1, n-2, \dots, 1, 0)$ . The  $q$ -analogue of Kostant's partition function can be defined as

$$\sum_{\alpha \in \mathbb{Z}^n} \mathcal{P}_q(\alpha) x_1^{\alpha_1} \cdots x_n^{\alpha_n} = \prod_{1 \leq i < j \leq n} \left(1 - q \frac{x_i}{x_j}\right)^{-1}$$

Substituting  $\mathcal{P}$  for its  $q$ -analogue, we obtain the Kostka-Foulkes polynomial

$$K_{\lambda\mu}(q) := \sum_{w \in S_n} \epsilon(w) \mathcal{P}_q(w(\lambda + \rho) - (\mu + \rho)).$$

Hesselink connected  $K_{\lambda\mu}(q)$  to the graded multiplicity of an irrep  $V^\alpha$  of  $GL_n(\mathbb{C})$  [11]. If  $\alpha \in \mathbb{Z}^n$  is decreasing and sums to zero, and if  $\lambda = \alpha + (k^n)$ , with  $k \geq |\alpha_n|$ , then the graded multiplicity polynomial  $m_\alpha(q)$  defined in Equation 6 is given by

$$m_\alpha(q) = \sum_{w \in S_n} \epsilon(w) \mathcal{P}_q(w(\alpha + \rho) - \rho).$$

Since  $\lambda = \alpha + (k^n)$ , it follows that

$$m_\alpha(q) = K_{\lambda(k^n)}(q).$$

For more on this result see [3] and [10].

The Kostka-Foulkes polynomials are understood combinatorially in [16] and [18]. From this interpretation, the degree of the polynomial  $K_{\lambda\mu}(q)$  can be expressed in terms of the weighted totals of  $\lambda$  and  $\mu$ , defined in Equation 4, and is given by  $b(\mu) - b(\lambda)$ .

In summary, the only irreducible rational representations  $V^\alpha$  with highest weight  $\alpha \in \mathbb{Z}^n$  for which  $m_\alpha(q) \neq 0$  are those whose highest weight satisfies  $\alpha_1 + \dots + \alpha_n = 0$ . If  $m_\alpha(q)$  is nonzero, then we have

$$m_\alpha(q) = K_{\lambda(k^n)}(q),$$

where  $k$  is any integer satisfying  $k \geq |\alpha_n|$  and  $\lambda = \alpha + (k^n)$ . From [16] and [18] It follows that

$$\deg m_\alpha(q) = b((k^n)) - b(\lambda) = g_{nk,n}(\lambda).$$

This proves Theorem 3.1.

## 4. The Gini index and the Earth Mover’s Distance

This formulation of a discrete Gini index was motivated by recent research on the discrete Earth Mover’s Distance (EMD) – also known as the Wasserstein distance. Consider two different distributions of  $s$  pebbles into  $n$  distinct piles. Simply put, the one-dimensional EMD counts the least number of moves needed to convert the first distribution into the second, where each *move* involves moving one pebble to a neighboring pile. An example of this procedure is provided in Figure 3.

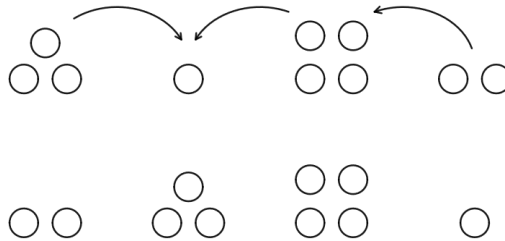


Figure 3: The minimum number of moves required to convert the top distribution,  $(3, 1, 4, 2)$ , into the bottom distribution,  $(2, 3, 4, 1)$ , is 3.

The Gini index and EMD are both measures of dissimilarity and are, in fact, measuring the same quantity.

A composition of a positive integer  $s$  into  $n$  parts is an  $n$ -tuple of non-negative integers whose sum is  $s$ . The EMD of such compositions was discussed in [1] (the one-dimensional case) and [4] (a higher dimensional generalization) – generating functions and connections to representation theory were found in both instances. In [5] it was shown that if  $\mu$  and  $\lambda$  are compositions of  $s$  into  $n$  parts, then  $EMD(\mu, \lambda)$  can be expressed as the symmetric difference of the Young diagrams of the words of  $\mu$  and  $\lambda$ . The word of a composition  $\mu$  is a tuple of integers constructed by writing “0”  $\mu_1$  times, then “1”  $\mu_2$  times, and so on, ending with “ $n - 1$ ” written  $\mu_n$  times. If  $\mu$  is a composition of  $s$  into  $n$  parts, then the word of  $\mu$  will be weakly increasing with length  $s$ . The Young diagram of a word is a collection of square boxes arranged in left-justified rows, whose ascending row lengths are given by the word. The symmetric difference of the diagrams is the number of boxes in their union minus their intersection.

**Example 4.1.** *The EMD of the compositions  $\mu = (3, 1, 4, 2)$  and  $\lambda = (2, 3, 4, 1)$  can be found as in Figure 3, or by using Young diagrams as shown in [5]. The word of  $\mu$  is  $(0, 0, 0, 1, 2, 2, 2, 2, 3, 3)$  and the word of  $\lambda$  is  $(0, 0, 1, 1, 1, 2, 2, 2, 2, 3)$ . Their Young diagrams and symmetric difference are shown in Figure 4.*

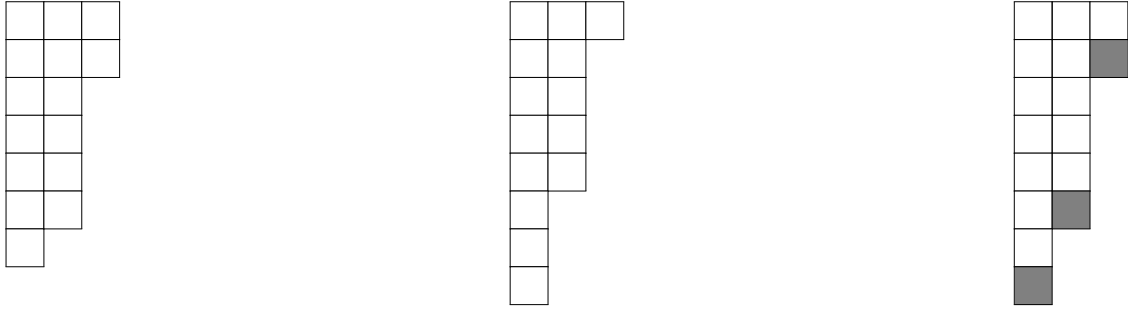


Figure 4: The Young diagrams of the words of  $\mu$  (left) and  $\lambda$  (middle) and their union (right), with symmetric difference shaded.

The number of shaded cells in the final diagram is the symmetric difference. Hence, for Example 4.1,  $EMD(\mu, \lambda) = 3$ .

In general,  $EMD(\mu, \lambda)$  can be expressed in terms of the weighted totals of  $\mu$  and  $\lambda$  (defined in Equation 4) as

$$EMD(\mu, \lambda) = b(\mu) - b(\lambda)$$

whenever  $\mu$  and  $\lambda$  are partitions of  $s$  into  $n$  parts, and  $\mu$  majorizes  $\lambda$  – that is, whenever  $\sum_{i=1}^k \mu_i \geq \sum_{i=1}^k \lambda_i$  for all  $1 \leq k \leq n$ . The discrete Gini index can also be defined in terms of Young diagrams as in [13] and [14]. From this interpretation, in the special case where  $s = nk$ , it follows that

$$EMD((k^n), \lambda) = b((k^n)) - b(\lambda) = g_{nk,n}(\lambda).$$

That is, the discrete Gini index is a restriction of the discrete one-dimensional Earth Mover’s Distance to the set of partitions of  $nk$  with at most  $n$  parts.

## Acknowledgements

We would like to thank the referee for their suggestions that significantly improved the content and exposition of this article.

## References

- [1] R. Bourn and J. F. Willenbring, *Expected value of the one-dimensional earth mover’s distance*, *Algebr. Stat.* 11(1) (2020), 53–78.
- [2] Y. Cai, D. S. Chatelet, R. P. Howlin, Z. Wang, and J. S. Webb, *A novel application of Gini coefficient for the quantitative measurement of bacterial aggregation*, *Sci. Reports* 9 (2019), 19002.
- [3] J. ésarménien, B. Leclerc, and J.-Y. Thibon, *Hall-Littlewood functions and Kostka-Foulkes polynomials in representation theory*, *Sém. Lothar. Combin.* 32 (1994), Article B32c.
- [4] W. Q. Erickson, *A generalization for the expected value of the earth mover’s distance*, *Algebr. Stat.* 12(2) (2021), 139–166.
- [5] W. Q. Erickson, *Comparing weighted difference and earth mover’s distance via young diagrams*, arXiv 2104.07273 (2023).
- [6] F. A. Farris, *The Gini index and measures of inequality*, *Amer. Math. Monthly* 117(10) (2010), 851–864.
- [7] W. Fulton, *Young tableaux*, Vol. 35 of London Mathematical Society Student Texts, Cambridge University Press, Cambridge, 1997. With applications to representation theory and geometry.
- [8] C. Gini, *Variabilità e mutabilità*, reprinted in *Memorie di Metodologica Statistica*, Libreria Eredi Virgilio Veschi, 1955.
- [9] R. Goodman and N. R. Wallach, *Symmetry, representations, and invariants*, Volume 255 of Graduate Texts in Mathematics, Springer, Dordrecht, 2009.
- [10] R. K. Gupta, *Generalized exponents via Hall-Littlewood symmetric functions*, *Bull. Amer. Math. Soc. (N.S.)* 16(2) (1987), 287–291.
- [11] W. H. Hesselink, *Characters of the nullcone*, *Math. Ann.* 252(3) (1980), 179–182.
- [12] R. T. Jantzen and K. Volpert, *On the mathematics of income inequality: splitting the Gini index in two*, *Amer. Math. Monthly* 119(10) (2012), 824–837.

- [13] G. Kopitzke, *The Gini index of an integer partition*, J. Integer Seq. 23(9) (2020), Article 20.9.7.
- [14] G. Kopitzke, *The Gini index in algebraic combinatorics and representation theory*, ProQuest LLC, Ann Arbor, MI, 2021. Thesis (Ph.D.)—The University of Wisconsin - Milwaukee.
- [15] B. Kostant, *Lie group representations on polynomial rings*, Bull. Amer. Math. Soc. 69 (1963), 518–526.
- [16] A. Lascoux and M. Schützenberger, *Sur une conjecture de h. o. foulkes*, C. R. Acad. Sci. Paris Sér. A-B 286(7) (1978), A323–A324.
- [17] M.O. Lorenz, *Methods of measuring the concentration of wealth*, Publ. Am. Stat. Assoc. 9(70) (1905), 209–219.
- [18] I. G. Macdonald, *Symmetric functions and Hall polynomials*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New York, 1979.
- [19] R. P. Stanley, *GL( $n, \mathbf{C}$ ) for combinatorialists*. In Surveys in combinatorics (Southampton, 1983), Volume 82 of London Math. Soc. Lecture Note Ser., 187–199. Cambridge Univ. Press, Cambridge, 1983.
- [20] M. E. Sánchez-Hechavarría, S. Ghiya, R. Carrazana-Escalona, S. Cortina-Reyna, A. Andreu-Heredia, C. Acosta-Batistia, and N. A. Saá-Muñoz, *Introduction of application of Gini coefficient to heart rate variability spectrum for mental stress evaluation*, Arquivos Brasileiros de Cardiol. 113(4) (2019), 725–733.
- [21] F. Teng, X. Pan, and C. Zhang, *Metric of carbon equity: Carbon Gini index based on historical cumulative emission per capita*, Adv. Clim. Chang. Research 2(3) (2011), 134–140.