Hoeffding spaces and Specht modules

Giovanni PECCATI* and Jean-Renaud PYCKE†

April 20, 2009

Abstract

It is proved that each Hoeffding space associated with a random permutation (or, equivalently, with extractions without replacement from a finite population) carries an irreducible representation of the symmetric group, equivalent to a two-block Specht module.

Key words – Exchangeability; Finite Population Statistics; Hoeffding Decompositions; Irreducible Representations; Random Permutations; Specht Modules; Symmetric Group.

MSC Classification – 05E10; 60C05

1 Introduction

Let $X_{(m)} = (X_1, ..., X_m)$ ($m \geq 2$) be a sample of random observations. According e.g. to [10], we say that $X_{(m)}$ is *Hoeffding-decomposable* if every symmetric statistic of $X_{(m)}$ can be written as an orthogonal sum of symmetric U-statistics with degenerated kernels of increasing orders. In the case where $X_{(m)}$ is composed of i.i.d. random variables, Hoeffding decompositions are a classic and very powerful tool for obtaining limit theorems, as $m \to \infty$, for sequences of general symmetric statistics of the vectors $X_{(m)}$. See e.g. [13], or the references indicated in the introduction to [10], for further discussions in this direction.

In recent years, several efforts have been made in order to provide a characterization of Hoeffding decompositions associated with *exchangeable* (and not necessarily independent) vectors of observations. See El-Dakkak and Peccati [8] and Peccati [10] for some general statements; see Bloznelis [2], Bloznelis and Götze [3, 4] and Zhao and Chen [15] for a comprehensive analysis of Hoeffding decompositions associated with extractions without replacement from a finite population.

In the present note, we are interested in building a new explicit connection between the results of [3, 4, 15] and the irreducible representations of the symmetric groups $\mathfrak{S}_n$, $n \geq 2$. In particular, our main result is the following.

Theorem 1 *Let $1 \leq m \leq n/2$, and let $X_{(m)} = (X(1), ..., X(m))$ be a random vector obtained as the first m extractions without replacement from a population of n individuals. For $l = 1, ..., m$, let SH_l be the lth symmetric Hoeffding space associated with $X_{(m)}$ (that is, SH_l is the vector space of all symmetric U-statistics with a completely degenerated kernel of order l). Then, for every $l = 1, ..., m$, there exists an action of $\mathfrak{S}_n$ on SH_l, such that SH_l is an irreducible representation of $\mathfrak{S}_n$. This representation is equivalent to a Specht module of shape $(n-l, l)$.*

*Équipe *Modal'X*. Université Paris Ouest - Nanterre La Défense. 200, Avenue de la République, 92000 Nanterre France. E-mail: giovanni.peccati@gmail.com

†Department of mathematics, University of Évry, France. E-mail: jrpycke@maths.univ-evry.fr

We refer the reader to the forthcoming Section 2 for some basic results on the representations of the symmetric group and two-block Specht modules. We will see that Theorem 1 provides *de facto* a new probabilistic characterization of two-block Specht modules, as well as some original insights into the combinatorial structure of Hoeffding spaces. Observe that the case where $n/2 < m \leq n$ can be reduced to the framework of present paper by standard arguments (see for instance [3, Proposition 1]). One should note that a connection between decompositions of symmetric statistics and representations of $\mathfrak{S}_n$ is already sketched in Diaconis' celebrated monograph [5]: in particular, the results of the present paper can be regarded as a probabilistic counterpart to the *spectral analysis on homogeneous spaces* developed in Chapters 7 and 8 of [5].

The rest of this note is organized as follows. In Section 2 we provide some background on the representations of the symmetric group. Sections 3 and 4 focus, respectively, on uniform random permutations and Hoeffding spaces. Section 5 contains the statements and proofs of our main results.

2 Background

For future reference, we recall that a *k-block partition* of the integer $n \geq 2$ is a k-dimensional vector of the type $\lambda = (\lambda_1, ..., \lambda_k)$, such that: (i) each λ_i is a strictly positive integer, (ii) $\lambda_i \geq \lambda_{i+1}$, and (iii) $\lambda_1 + \cdots + \lambda_k = n$. One sometimes writes $\lambda \vdash n$ to indicate that λ is a partition of n.

We also write $[n] = \{1, ..., n\}$ to indicate the set of the first n positive integers. Finally, given a finite set A, we denote by $\mathfrak{S}_A$ the group of all permutations of A, and we use the shorthand notation $\mathfrak{S}_{[n]} = \mathfrak{S}_n$, $n \geq 1$. In other words, when writing $x \in \mathfrak{S}_A$, we mean that

$$x : A \to A : a \mapsto x(a)$$

is a bijection from A to itself.

2.1 Some structures associated with two-block partitions

We now introduce some classic definitions and notation related to tableaux and tabloids; see Sagan [12, Chapter 2] (from which we borrow most of our terminology and notational conventions) for any unexplained concept or result. For the rest of the section, we fix two integers n and m, such that $1 \leq m \leq n/2$. Observe that $n - m \geq m$, and therefore the vector $(n - m, m)$ is a two-block partition of the integer n.

Remark. It is sometimes useful to adopt a graphical representation of tableaux and tabloids by means of *Ferrer diagrams*. Since we uniquely deal with two-block tableaux and tabloids, and for the sake of brevity, in what follows we shall not make use of this representation. See e.g. [12, Section 2.1] for a complete discussion of this point.

The following objects will be needed in the sequel.

– A (Young) *tableau* t of shape $(n - m, m)$ is a pair $t = \left(i_{(n-m)}; j_{(m)}\right)$ of *ordered* vectors of the type $i_{(n-m)} = (i_1, ..., i_{n-m})$, $j_{(m)} = (j_{n-m+1}, ..., j_n)$ such that $\{i_1, ..., i_{n-m}, j_{n-m+1}, ..., j_n\} = [n]$, that is, the union of the entries of $i_{(n-m)}$ and $j_{(m)}$ coincides with the first n integers (with no repetitions).

– The set of the *columns* of the tableau $t = \left(i_{(n-m)}; j_{(m)}\right)$, noted $\{C_1, ..., C_{n-m}\}$, is the collection of (i) the ordered pairs

$$C_1 = (i_1, j_{n-m+1}), ..., C_m = (i_m, j_n) \tag{1}$$

(that is, the pairs composed of the first m entries of $i_{(n-m)}$ and the entries of $j_{(m)}$), and (ii) the remaining singletons of $i_{(n-m)}$, that is,

$$C_{m+1} = i_{m+1}, ..., C_{n-m} = i_{n-m}. \tag{2}$$

– For $l = 1, ..., n$, we write $V^{(n-l,l)}$ to indicate the class of the $\binom{n}{l}$ subsets of $[n]$ of size equal to l. This slightly unusual notation has been chosen in order to stress the connection between the set $V^{(n-l,l)}$ and the $\mathfrak{S}_n$-modules $M^{(n-l,l)}$ ($l \leq m$) to be defined below. The elements of $V^{(n-l,l)}$ are denoted by $\mathbf{a}_{(l)}$, $\mathbf{b}_{(l)}$, $\mathbf{i}_{(l)}$, $\mathbf{j}_{(l)}$,..., and so on.

– A *tabloid* of shape $(n-m, m)$ is a two-block partition of the set $[n]$, of the type

$$\gamma = \{\mathbf{a}_{(n-m)}; \mathbf{b}_{(m)}\} = \{\{a_1, ..., a_{n-m}\}; \{b_{n-m+1}, ..., b_n\}\}. \tag{3}$$

Of course, a tabloid γ of shape $(n-m, m)$ as in (3) is completely determined by the specification of set $\mathbf{b}_{(m)} = \{b_{n-m+1}, ..., b_n\} \in V^{(n-m,m)}$; to emphasize this dependence, we shall sometimes write $\gamma = \gamma(\mathbf{b}_{(m)})$. Note that the mapping $\mathbf{b}_{(m)} \mapsto \gamma(\mathbf{b}_{(m)})$ is a bijection between $V^{(n-m,m)}$ and the class of all tabloids of shape $(n-m, m)$.

– Given a tableau $t = \left(i_{(n-m)}; j_{(m)}\right)$ of shape $(n-m, m)$, we write $\{t\} = \{\mathbf{i}_{(n-m)}; \mathbf{j}_{(m)}\}$ (observe the boldface!) to indicate the tabloid defined by $\mathbf{i}_{(n-m)} = \{i_1, ..., i_{n-m}\}$ and $\mathbf{j}_{(m)} = \{j_{n-m+1}, ..., j_n\}$. In other words, $\{t\}$ is obtained as the two-block partition composed of the collection of the entries of $i_{(n-m)}$ and the collection of the entries of $j_{(m)}$. With the notation introduced at the previous point, one has that $\{t\} = \gamma(\mathbf{j}_{(m)})$.

Example. Let $n = 5$ and $m = 2$. Then, a tableau of shape $(3, 2)$ is $t = \left(i_{(3)}; j_{(2)}\right)$, where $i_{(3)} = (2, 1, 3)$ and $j_{(2)} = (5, 4)$. The columns of t are $C_1 = (2, 5)$, $C_2 = (1, 4)$ and $C_3 = 3$. The associated tabloid is $\{t\} = \{\mathbf{i}_{(3)}; \mathbf{j}_{(2)}\}$, where $\mathbf{i}_{(3)} = \{1, 2, 3\} \in V^{(2,3)}$ and $\mathbf{j}_{(2)} = \{4, 5\} \in V^{(3,2)}$.

2.2 Actions of $\mathfrak{S}_n$

Fix as before $n \geq 2$ and $1 \leq m \leq n/2$.

<u>*Actions on tableaux*</u>. For every $x \in \mathfrak{S}_n$ and every tableaux $t = \left(i_{(n-m)}; j_{(m)}\right)$, the action of x on t is defined as follows:

$$xt = \left(xi_{(n-m)}; xj_{(m)}\right), \tag{4}$$

where $xi_{(n-m)} = (x(i_1), ..., x(i_{n-m}))$ and $xj_{(m)} = (x(j_{n-m+1}), ..., x(j_n))$.

<u>*Actions on tabloids*</u>. For every $x \in \mathfrak{S}_n$ and every tabloid $\gamma(\mathbf{b}_{(m)}) = \{\mathbf{a}_{(n-m)}; \mathbf{b}_{(m)}\}$, we set

$$\begin{aligned} x\gamma(\mathbf{b}_{(m)}) &= x\{\{a_1, ..., a_{n-m}\}; \{b_{n-m+1}, ..., b_n\}\} \\ &= \{\{x(a_1), ..., x(a_{n-m})\}; \{x(b_{n-m+1}), ..., x(b_n)\}\} \end{aligned} \tag{5}$$

In particular, for every tableau t, one has $x\{t\} = \{xt\}$.

$\mathfrak{S}_n$*-modules*. The symmetric group $\mathfrak{S}_n$ acts on $V^{(n-m,m)}$ in the standard way, namely: for every $x \in \mathfrak{S}_n$ and for every $\mathbf{j}_{(m)} = \{j_1, ..., j_m\} \in V^{(n-m,m)}$,

$$x\mathbf{j}_{(m)} = \{x(j_1), ..., x(j_m)\}. \tag{6}$$

Remark. By combining the above introduced notational conventions, one sees that, for every $x \in \mathfrak{S}_n$ and for every $\mathbf{j}_{(m)} = V^{(n-m,m)}$,

$$x\gamma(\mathbf{j}_{(m)}) = \gamma(x\mathbf{j}_{(m)}),$$

that is, x transforms the tabloid generated by $\mathbf{j}_{(m)}$ into the tabloid generated by $x\mathbf{j}_{(m)}$. Also, if $t = \left(i_{(n-m)}; j_{(m)}\right)$, then, for every $x \in \mathfrak{S}_n$,

$$x\{t\} = \{xt\} = x\gamma(\mathbf{j}_{(m)}) = \gamma(x\mathbf{j}_{(m)}).$$

The complex vector space of all complex-valued functions on $V^{(n-m,m)}$ is written $L\left(V^{(n-m,m)}\right)$. Plainly, the space $L\left(V^{(n-m,m)}\right)$ has dimension $\binom{n}{m}$, and a basis of $L\left(V^{(n-m,m)}\right)$ is given by the collection $\{\mathbf{1}_{\mathbf{j}_{(m)}} : \mathbf{j}_{(m)} \in V^{(n-m,m)}\}$, where $\mathbf{1}_{\mathbf{j}_{(m)}}\left(\mathbf{k}_{(m)}\right) = 1$ if $\mathbf{k}_{(m)} = \mathbf{j}_{(m)}$ and $\mathbf{1}_{\mathbf{j}_{(m)}}\left(\mathbf{k}_{(m)}\right) = 0$ otherwise. The group $\mathfrak{S}_n$ acts on $L\left(V^{(n-m,m)}\right)$ as follows: for $x \in \mathfrak{S}_n$, $\mathbf{k}_{(m)} \in V^{(n-m,m)}$ and $f \in L\left(V^{(n-m,m)}\right)$,

$$\begin{aligned} xf\left(\mathbf{k}_{(m)}\right) &= f\left(x^{-1}\mathbf{k}_{(m)}\right), \text{ so that, in particular,} \\ x\mathbf{1}_{\mathbf{j}_{(m)}} &= \mathbf{1}_{x\mathbf{j}_{(m)}}, \quad \mathbf{j}_{(m)} \in V^{(n-m,m)}. \end{aligned} \tag{7}$$

When endowed with the action (7), the set $L\left(V^{(n-m,m)}\right)$ carries a representation of $\mathfrak{S}_n$. In this case, we say that $L\left(V^{(n-m,m)}\right)$ is the *permutation module* associated with $(n-m, m)$, and we use the customary notation $L\left(V^{(n-m,m)}\right) = M^{(n-m,m)}$ (see [12, Section 2.1]).

Remark. Our definition of the permutation modules $M^{(n-m,m)}$ slightly differs from the one given e.g. in [12, Definition 2.1.5]. Indeed, we define $M^{(n-m,m)}$ as the vector space spanned by all indicators of the type $\mathbf{1}_{\mathbf{j}_{(m)}}$, $\mathbf{j}_{(m)} \in V^{(n-m,m)}$, endowed with the action (7), whereas in the above quoted reference $M^{(n-m,m)}$ is the space of all formal linear combinations of tabloids of shape $(n-m, m)$, endowed with the canonical extension of the action (5). The two definitions are equivalent, in the sense that they give rise to two isomorphic $\mathfrak{S}_n$-modules. We will see that the definition of $M^{(n-m,m)}$ chosen in this paper allows a more transparent connection with the theory of U-statistics based on random permutations.

2.3 A decomposition of $M^{(n-m,m)}$

We recall that the dual of $\mathfrak{S}_n$ coincides with the set $\left\{\left[S^\lambda\right] : \lambda \vdash n\right\}$, where $\left[S^\lambda\right]$ is the equivalence class of all irreducible representations of $\mathfrak{S}_n$ that are equivalent to a Specht module of index λ (see again [12, Section 2.1]). For every $\lambda \vdash n$, we will denote by χ^λ the character associated with the class $\left[S^\lambda\right]$, whereas D_λ is the associate dimension. Observe that $\chi^\lambda \in \mathbb{Z}$ for every λ

(see e.g. [14, Section 13.1]), and D_λ equals the number of standard tableaux (that is, tableaux with increasing rows and columns) of shape λ. In particular $D_{(n-1,1)} = n-1$ (see [12, Section 2.5]).

The next result ensures that the module $M^{(n-m,m)}$ is reducible. This fact is well-known (see e.g. [9, Example 14.4, p. 52] or [5, pp. 134-139]), and a proof is added here for the sake of completeness.

Proposition 2 *There exists a unique decomposition of $M^{(n-m,m)}$ of the type*

$$M^{(n-m,m)} = K_0^{(n-m,m)} \oplus K_1^{(n-m,m)} \oplus \cdots \oplus K_m^{(n-m,m)}. \tag{8}$$

Where the vector spaces (endowed with the action of $\mathfrak{S}_n$ described in (7)) $K_l^{(n-m,m)}$ are such that $K_0^{(n-m,m)} \in [S^{(n)}]$, and $K_l^{(n-m,m)} \in [S^{(n-l,l)}]$, $l = 1, ..., m$.

Proof. It is sufficient to prove that

$$M^{(n-m,m)} \cong S^{(n)} \oplus \bigoplus_{l=1}^{m} S^{(n-l,l)},$$

where " $\cong$ " indicates equivalence between representations of $\mathfrak{S}_n$. According Young's Rule (see e.g. [12, Th. 2.11.2]), we know that

$$M^{(n-m,m)} \cong S^{(n)} \oplus \bigoplus_{l=1}^{m} K_{n,l,m} S^{(n-l,l)},$$

where the integers $K_{n,l,m}$ (known as *Kostka numbers*) count the number of generalized semistandard tableaux of shape $(n-l,l)$ and type $(n-m,m)$. This is equivalent to saying $K_{n,l,m}$ counts the ways of arranging $n-m$ copies of 1 and m copies of 2 in a Ferrer diagram of shape $(n-l,l)$, in such a way that the rows of the diagram are weakly increasing and the columns are strictly increasing. Since there is just one way of doing this, one infers that $K_{n,l,m} = 1$, and the proof is concluded. ■

Remarks. (i) (*Definition of two-block Specht modules*) For the sake of completeness, we recall here the definition of the modules $S^{(n)}$ and $S^{(n-m,m)}$, $1 \leq m \leq n/2$. First of all, one has that $S^{(n)} = \mathbb{C}$, and therefore $[S^{(n)}]$ is the class of representations of $\mathfrak{S}_n$ that are equivalent to the trivial representation. Now fix $1 \leq m \leq n/2$. For every tableau $t = (i_{(n-m)}; j_{(m)})$, define the columns $C_1, ..., C_{n-m}$ according to (1) and (2). Then, (a) for every $l = 1, ..., m$, write κ_{C_l} for the formal operator

$$\kappa_{C_l} = \mathrm{Id.} - (i_l \to j_l),$$

where $(i_l \to j_l)$ indicates the element of $\mathfrak{S}_n$ given by the translation sending i_l to j_l, and (b) define the composed operator $\kappa_t = \kappa_{C_1}\kappa_{C_2}\cdots\kappa_{C_m}$. Then, the Specht module of shape $(n-m,m)$ is the $\mathfrak{S}_n$-invariant subspace of $M^{(n-m,m)}$ spanned by the elements of the type

$$\kappa_t \mathbf{1}_{\mathbf{j}_{(m)}}, \quad \text{where } t = \left(i_{(n-m)}; j_{(m)}\right) \text{ is a tableau;} \tag{9}$$

note that, in the formula (9), t and $\mathbf{j}_{(m)}$ are related by the fact that $t = \left(i_{(n-m)}; j_{(m)}\right)$, and $\{t\} = \{\mathbf{i}_{(n-m)}; \mathbf{j}_{(m)}\}$.

(ii) Consider for instance the case $n=6$ and $m=2$, and select the tableau $t=\{(1,2,3,4)\,;(5,6)\}$. One has that $\mathbf{j}_{(2)}=\{5,6\}$,

$$\kappa_t = (\mathrm{Id.} - (1 \to 5))\,(\mathrm{Id.} - (2 \to 6))\,,$$

and one deduces that an element of $S^{(4,2)}$ is given by

$$\kappa_t \mathbf{1}_{\mathbf{j}_{(2)}} = \mathbf{1}_{\{5,6\}} - \mathbf{1}_{\{1,6\}} - \mathbf{1}_{\{5,2\}} + \mathbf{1}_{\{1,2\}}.$$

(iii) By recurrence, one deduces from Proposition 2 that the dimension of $K_l^{(n-m,m)}$, and therefore of $S^{(n-l,l)}$, is $D_{(n-l,l)} = \binom{n}{l} - \binom{n}{l-1}$, $l \leq n/2$.
(iv) From the previous discussion, we infer that $K_0^{(n-m,m)} = S^{(n)} = \mathbb{C}$.

3 Uniform random permutations

Fix $n \geq 2$. We consider a uniform *random permutation* X of $[n]$. This means that $X = X(\omega)$ is a random element with values in $\mathfrak{S}_n$, defined on some finite probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and such that, $\forall x \in \mathfrak{S}_n$, $\mathbf{P}(X = x) = (n!)^{-1}$. For $1 \leq m \leq n/2$ as before, we will write $X_{(m)}(\omega) = (X(1), ..., X(m))(\omega)$, and also, for every $y \in \mathfrak{S}_n$, $(Xy)_{(m)} = \{Xy(1), ..., Xy(m)\}$. Observe that Xy indicates the product of the deterministic permutation y with the random permutation X. It is clear that $X_{(m)}$ is an *exchangeable* vector, having the law of the first m extractions without replacement from the set $[n]$ (see e.g. Aldous [1] for any unexplained notion about exchangeability). A random variable T is called a (complex-valued) *symmetric statistic* of $X_{(m)}$ if T has the form

$$T = f(\{X(1), ..., X(m)\}), \quad \text{for some } f \in L\left(V^{(n-m,m)}\right).$$

In other words, a symmetric statistic is a random variable deterministically depending on the realization of $X_{(m)}$ as a non-ordered set. Note that, by a slight abuse of notation, in what follows we will write $f(\{X(1), ..., X(m)\}) = f\left(X_{(m)}\right)$ (other analogous conventions will be tacitly adopted).

We also write $L_s^2\left(X_{(m)}\right)$ to indicate the Hilbert space of symmetric statistics of $X_{(m)}$, endowed with the inner product

$$\begin{aligned}
\left\langle f_1\left(X_{(m)}\right), f_2\left(X_{(m)}\right)\right\rangle_{\mathbf{P}} &= \mathbf{E}\left[f_1\left(X_{(m)}\right)\overline{f_2\left(X_{(m)}\right)}\right] && (10)\\
&= \frac{1}{n!}\sum_{x\in\mathfrak{S}_n} f_1(x\{1,...,m\})\,\overline{f_2(x\{1,...,m\})} && (11)\\
&= \binom{n}{m}^{-1}\sum_{\mathbf{k}_{(m)}\in V^{(n-m,m)}} f_1\left(\mathbf{k}_{(m)}\right)\overline{f_2\left(\mathbf{k}_{(m)}\right)}.
\end{aligned}$$

Since the sum in (11) runs over the whole set $\mathfrak{S}_n$, it is clear that $\langle\cdot,\cdot\rangle_{\mathbf{P}}$ induces a $\mathfrak{S}_n$-invariant inner product on $M^{(n-m,m)}$ given by

$$\langle f_1, f_2\rangle_{(n-m,m)} = \left\langle f_1\left(X_{(m)}\right), f_2\left(X_{(m)}\right)\right\rangle_{\mathbf{P}}, \quad f_1, f_2 \in M^{(n-m,m)}; \tag{12}$$

in particular, the $\mathfrak{S}_n$-invariance of $\langle\cdot,\cdot\rangle_{(n-m,m)}$ yields that the spaces $K_i^{(n-m,m)}$ and $K_j^{(n-m,m)}$ are orthogonal with respect to $\langle\cdot,\cdot\rangle_{(n-m,m)}$ for every $0 \leq i \neq j \leq m$.

With every $f \in M^{(n-m,m)}$, we associate the $\mathfrak{S}_n$-indexed stochastic process

$$Z_f(x,\omega) = Z_f(x) := f\left(xX_{(m)}\right), \quad x \in \mathfrak{S}_n,$$

and, for every $\lambda \vdash n$, we define

$$\begin{aligned} Z_f^\lambda(x,\omega) &= Z_f^\lambda(x) := \frac{D_\lambda}{n!}\sum_{g\in\mathfrak{S}_n}\chi^\lambda(g)\, f\left((g^{-1}x)X_{(m)}\right) \\ f^\lambda\left(\mathbf{l}_{(m)}\right) &= \frac{D_\lambda}{n!}\sum_{x\in\mathfrak{S}_n}\chi^\lambda(x)\, f\left(x^{-1}\mathbf{l}_{(m)}\right), \quad \mathbf{l}_{(m)} \in V^{(n-m,m)}, \end{aligned} \tag{13}$$

so that $f^\lambda\left(X_{(m)}\right) = Z_f^\lambda(e)$, where e is the identity element in $\mathfrak{S}_n$.

The following facts will be used in the subsequent analysis. The proofs are standard and omitted – see e.g. the results from [11] and [14] evoked below for further details.

(a) Since (8) holds, $f^\lambda = 0$ for every $f \in M^{(n-m,m)}$ if and only if λ is different from $(n-l,l)$, $l = 0, ..., m$ (see e.g. [14, Theorem 8, Section 2.6]) and moreover: $f^{(n)} \in K_0^{(n-m,m)}$ and, for every $l = 1, ..., m$, $f^{(n-l,l)} \in K_l^{(n-m,m)}$ (as defined in (8)).

(b) Thanks to exchangeability, for every $f \in M^{(n-m,m)}$ the class

$$\left\{Z_f, Z_f^{(n-l,l)} : l = 0, ..., m\right\},$$

has a $\mathfrak{S}_n$-invariant law, with respect to the canonical action of $\mathfrak{S}_n$ on itself (i.e., $x \cdot y = xy$, $x, y \in \mathfrak{S}_n$).

(c) Due to the orthogonality of isotypical spaces (see e.g. (see [7, Theorem 4.4.5], and also [11, Theorem 4-3]), for every $x, y \in \mathfrak{S}_n$, $f, h \in M^{(n-m,m)}$ and $0 \leq i \neq j \leq m$,

$$\mathbf{E}\left[Z_f^{(n-i,i)}(x)\,\overline{Z_h^{(n-j,j)}(y)}\right] = \mathbf{E}\left[f^{(n-i,i)}\left(xX_{(m)}\right)\overline{h^{(n-j,j)}\left(yX_{(m)}\right)}\right] \tag{14}$$

$$\mathbf{E}\left[f^{(n-i,i)}\left((Xx)_{(m)}\right)\overline{h^{(n-j,j)}\left((Xy)_{(m)}\right)}\right] = 0, \tag{15}$$

where, here and in the sequel (by a slight abuse of notation) we use the convention $(n-0,0) = (n)$.

(d) Due to [11, Theorem 4-4] and point (a) above, for every $x \in \mathfrak{S}_n$ and every $f \in M^{(n-m,m)}$,

$$Z_f(x) = Z_f^{(n)}(x) + \sum_{l=1}^m Z_f^{(n-l,l)}(x), \tag{16}$$

where $Z_f^{(n)}(x) = \mathbf{E}[Z_f(x)] = \mathbf{E}\left[f\left(X_{(m)}\right)\right]$. In particular,

$$f\left(X_{(m)}\right) = \mathbf{E}\left[f\left(X_{(m)}\right)\right] + \sum_{l=1}^m f^{(n-l,l)}\left(X_{(m)}\right) \tag{17}$$

and therefore, for every $f,h\in M^{(n-m,m)}$,

$$\mathbf{E}\left[f\left(X_{(m)}\right)\overline{h\left(X_{(m)}\right)}\right]=\mathbf{E}\left[f\left(X_{(m)}\right)\right]\overline{\mathbf{E}\left[h\left(X_{(m)}\right)\right]}+\sum_{l=1}^{m}\mathbf{E}\left[f^{(n-l,l)}\left(X_{(m)}\right)\overline{h^{(n-l,l)}\left(X_{(m)}\right)}\right] \tag{18}$$

(e) Due to [11, Theorem 5-1], for every $0\leq i\neq j\leq m$ and $f,h\in M^{(n-m,m)}$,

$$\sum_{x\in\mathfrak{S}_n}Z_f^{(n-i,i)}(x,\omega)\overline{Z_h^{(n-j,j)}(x,\omega)}=\sum_{x\in\mathfrak{S}_n}f^{(n-i,i)}\left(xX_{(m)}\right)\overline{h^{(n-j,j)}\left(xX_{(m)}\right)}=0. \tag{19}$$

4 Hoeffding spaces

We now define a class of subspaces of $L_s^2\left(X_{(m)}\right)$ (the notation is the same as in [8, 10]): $SU_0=\mathbb{C}$, and, for $l=1,...,m$, SU_l is the vector subspace generated by the functionals of $X_{(m)}$ of the type

$$T_\phi\left(X_{(m)}\right)=\sum_{\{k_1,...,k_l\}\in V^{(m-l,l)}}\phi\left(X\left(k_1\right),...,X\left(k_l\right)\right), \tag{20}$$

for some $\phi\in L\left(V^{(n-l,l)}\right)$. A random variable such as (20) is called a *U-statistic* based on $X_{(m)}$, with a *symmetric kernel* ϕ of order l. One has that $SU_l\subset SU_{l+1}$ (see e.g. [10]) and $SU_m=L_s^2\left(X_{(m)}\right)$. The collection of the *symmetric Hoeffding spaces* associated to $X_{(m)}$, noted $\{SH_l:l=0,...,m\}$ is defined as follows: $SH_0=SU_0$, and

$$SH_l=SU_l\cap SU_{l-1}^{\perp},$$

where the symbol $\perp$ means orthogonality with respect to the inner product $\langle\cdot,\cdot\rangle_{\mathbf{P}}$ defined in (10), so that

$$L_s^2\left(X_{(m)}\right)=\bigoplus_{l=0}^{m}SH_l,$$

where the direct sum $\bigoplus$ is again in the sense of $\langle\cdot,\cdot\rangle_{\mathbf{P}}$.

Following [3, Section 2], we define the real coefficients

$$\begin{aligned}d_{l,j}&=\prod_{r=j}^{l-1}\frac{n-r}{n-r-j},\quad l=2,3,...,m,\ 1\leq j\leq l-1,\\ d_{l,l}&=N_{l,l}=1,\quad l=1,...,m,\\ N_{l,j}&=-\sum_{i=j}^{l-1}\binom{l-j}{i-j}d_{l,i}N_{i,j},\quad l=2,3,...,m,\quad 1\leq j\leq l-1.\end{aligned} \tag{21}$$

The following result can be proved by using the content of [3, Section 2], or as a special case of [10, Theorem 11].

Proposition 3 *Keep the assumptions and notation of this section. Then, for $l=1,...,m$, the following assertions are equivalent:*

(i) $f\left(X_{(m)}\right)\in SH_l$;

(ii) *there exists* $\phi \in L\left(V^{(n-l,l)}\right)$ *such that*

$$f\left(X_{(m)}\right) = \sum_{\{k_1,...,k_l\} \in V^{(m-l,l)}} \phi\left(X\left(k_1\right),...,X\left(k_l\right)\right), \tag{22}$$

and

$$\mathbf{E}\left[\phi\left(X\left(1\right),...,X\left(l\right)\right) \mid X\left(1\right),...,X\left(l-1\right)\right] = 0.$$

Moreover, for every $h\left(X_{(m)}\right) \in L_s^2\left(X_{(m)}\right)$*, the orthogonal projection of* $h\left(X_{(m)}\right)$ *on* SH_l*,* $l = 1,...,m$*, is given by*

$$\mathsf{proj}\left(h\left(X_{(m)}\right) \mid SH_l\right) = \sum_{\{k_1,...,k_l\} \in V^{(m-l,l)}} \phi_h^{(l)}\left(X\left(k_1\right),...,X\left(k_l\right)\right),$$

where, for every $\{j_1,...,j_l\} \in V^{(n-l,l)}$*,*

$$\begin{aligned} & \phi_h^{(l)}\left(j_1,..,j_l\right) \\ = \ & d_{m,l} \sum_{a=1}^{l} N_{l,a} \sum_{1 \leq i_1 < ... < i_a \leq l} \mathbf{E}\left[h\left(X_{(m)}\right) - \mathbf{E}\left(h\left(X_{(m)}\right)\right) \mid X\left(1\right) = j_{i_1},...,X\left(a\right) = j_{i_a}\right]. \end{aligned} \tag{23}$$

The kernel ϕ of the U-statistic $f\left(X_{(m)}\right)$ appearing in (22) is said to be *completely degenerated*. Completely degenerated kernels are related to the notion of *weak independence* in [10, Theorem 6]. Note that, in the above quoted references, the content of Proposition 3 is proved for real valued symmetric statistics (the extension of such results to complex random variables is immediate: just consider separately the real and the imaginary parts of each statistic). Formula (23) completely characterizes the symmetric Hoeffding spaces associated to $X_{(m)}$: it can be obtained by recursively applying an appropriate version of the Möbius inversion formula (see e.g. [12, Exercise 18, Section 5.6]), on the lattice of the subsets of $[n]$ (see also [10, Theorem 11], for a generalization of (23) to the case of Generalized Urn Sequences). In the next section we state and prove the main result of this note, that is, that the spaces SH_l, $l = 1,...,m$, admit a further algebraic characterization in terms of Specht modules.

5 Hoeffding spaces and two-blocks Specht modules

5.1 Main results and some consequences

The main achievement of this note is the following statement, which is a more precise reformulation of Theorem 1, as stated in the Introduction. The proof is deferred to Section 5.2.

Theorem 4 *Under the above notation and assumptions, for every* $f\left(X_{(m)}\right) \in L_s^2\left(X_{(m)}\right)$ *and every* $l = 0,1,...,m$*, the following assertions are equivalent:*

1. $f\left(X_{(m)}\right) \in SH_l$*;*

2. $f \in K_l^{(n-m,m)}$*, where the* $\mathfrak{S}_n$*-module* $K_l^{(n-m,m)}$ *is defined through formula (8) (in particular,* $K_l^{(n-m,m)} \in \left[S^{(n-l,l)}\right]$*).*

We now list some consequences of Theorem 4. They can be obtained by properly combining Proposition 3 with the five facts (**a**)–(**e**), as listed at the end of Section 3.

Corollary 5 *Under the above notation and assumptions,*

1. *for every $l = 1, ..., m$, $f \in M^{(n-m,m)}$ and $\mathbf{i}_{(m)} = \{i_1, ..., i_m\} \in V^{(n-m,m)}$,*

$$f^{(n-l,l)}\left(\mathbf{i}_{(m)}\right) \tag{24}$$

$$= \frac{D_{(n-l,l)}}{n!} \sum_{x \in \mathfrak{S}_n} \chi^{(n-l,l)}(x) f\left(x^{-1}\mathbf{i}_{(m)}\right) \tag{25}$$

$$= \sum_{\{i_1,...,i_l\} \subseteq \mathbf{i}_{(m)}} d_{m,l} \sum_{a=1}^{l} N_{l,a} \times$$

$$\sum_{1 \leq s_1 < ... < s_a \leq l} \mathbf{E}\left[f\left(X_{(m)}\right) - \mathbf{E}\left(f\left(X_{(m)}\right)\right) \mid X(1) = i_{s_1}, ..., X(a) = i_{s_a}\right],$$

where $D_{(n-l,l)} = \binom{n}{l} - \binom{n}{l-1}$.

2. *for every $l = 1, ..., m$, every symmetric U-statistic, based on $X_{(m)}$ and with a completely degenerated kernel of order l, has the form (24) for some $f \in M^{(n-m,m)}$. It follows that SH_l is an irreducible $\mathfrak{S}_n$-module, carrying a representation in $\left[S^{(n-l,l)}\right]$.*

For instance, by using [12, Exercice 5.d, p. 87], we deduce from (24) that for every $\mathbf{i}_{(m)} = \{i_1, ..., i_m\} \in V^{(n-m,m)}$ and $f \in M^{(n-m,m)}$,

$$\frac{n-1}{n!} \sum_{x \in \mathfrak{S}_n} \{(\text{number of fixed points of } x) - 1\} \times f\left(x\mathbf{i}_{(m)}\right)$$

$$= \prod_{r=1}^{m-1} \frac{n-r}{n-r-1} \sum_{s=1}^{m} \mathbf{E}\left[f\left(X_{(m)}\right) - \mathbf{E}\left(f\left(X_{(m)}\right)\right) \mid X(1) = i_s\right].$$

The next result gives an algebraic explanation of a property of degenerated U-statistics, already pointed out – in the more general framework of Generalized Urn Sequences – in [10, Corollary 9]. Basically, it states that the orthogonality, between two completely degenerated U-statistics of different orders, is preserved after shifting one of the two arguments. It can be useful when determining the covariance between two U-statistics based on two urn sequences of different lenghts.

Corollary 6 *Let $f, h \in M^{(n-m,m)}$ be such that $f\left(X_{(m)}\right) \in SH_j$ and $h\left(X_{(m)}\right) \in SH_l$ for some $1 \leq j \neq l \leq m$. Consider moreover an element $\mathbf{k}_{(m)} = \{k_1, ..., k_m\} \in V^{(n-m,m)}$ such that, for some $r = 0, ..., m$, $\mathsf{Card}\left(\mathbf{k}_{(m)} \cap \{1, ..., m\}\right) = r$, and note $X'_{(m)} = (X(k_1), ..., X(k_m))$. Then,*

$$\mathbf{E}\left(f\left(X_{(m)}\right)\overline{h\left(X'_{(m)}\right)}\right) = 0.$$

Proof. Due to the exchangeability of the vector $(X(1),...,X(n))$, we can assume, without loss of generality, that

$$\mathbf{k}_{(m)} = \{1,...,r,m+1,...,2m-r\}.$$

Now introduce the permutation (written as a product of translations)

$$y = (r+1 \to m+1)(r+2 \to m+2)\cdots(m \to 2m-r), \tag{26}$$

and note that

$$\mathbf{E}\left(f\left(X_{(m)}\right)\overline{h\left(X'_{(m)}\right)}\right) = \mathbf{E}\left(f\left(X_{(m)}\right)\overline{h\left((Xy)_{(m)}\right)}\right),$$

so that the conclusion derives immediately from formula (15), by setting $x = e$ and y as in (26). ■

5.2 Remaining proofs

The key of the proof of Theorem 4 is nested in the following Lemma.

Lemma 7 *Let the previous notation prevail. Then,*

1. *for each $l = 1,...,m$, a basis of SU_l is given by the set of random variables*

$$\left\{\eta_{\mathbf{i}_{(l)}}\left(X_{(m)}\right) : \mathbf{i}_{(l)} \in V^{(n-l,l)}\right\},$$

where, for each $\mathbf{k}_{(m)} \in V^{(n-m,m)}$,

$$\eta_{\mathbf{i}_{(l)}}\left(\mathbf{k}_{(m)}\right) = \begin{cases} 1 & \text{if } \mathbf{i}_{(l)} \subseteq \mathbf{k}_{(m)} \\ 0 & \text{otherwise;} \end{cases} \tag{27}$$

2. *for each $l = 1,...,m$, the restriction of the action (7) of $\mathfrak{S}_n$ to the vector subspace of $M^{(n-m,m)}$ generated by the set $\{\eta_{\mathbf{i}_{(l)}} : \mathbf{i}_{(l)} \in V^{(n-l,l)}\}$, defined in (27), is equivalent to the action carried by the $\mathfrak{S}_n$-module $M^{(n-l,l)}$.*

Proof. Fix $l = 1,...,m$, and observe that, for every $\mathbf{i}_{(l)} \in V^{(n-l,l)}$,

$$\eta_{\mathbf{i}_{(l)}}\left(X_{(m)}\right) = \sum_{\{k_1,...,k_l\} \in V^{(m-l,l)}} \mathbf{1}_{\mathbf{i}_{(l)}}\left(\{X(k_1),...,X(k_l)\}\right),$$

so that the first part of the statement follows from the definition of SU_l, and the fact that every $\phi \in V^{(m-l,l)}$ is a linear combination of functions of the type $\mathbf{1}_{\mathbf{i}_{(l)}}(\cdot)$. To prove the second part, first recall that a basis of the $\mathfrak{S}_n$-module $M^{(n-l,l)}$ is given by the set $\left\{\mathbf{1}_{\mathbf{i}_{(l)}}(\cdot) : \mathbf{i}_{(l)} \in V^{(n-l,l)}\right\}$, and that the action of $\mathfrak{S}_n$ on $M^{(n-l,l)}$ is completely described by the action

$$x\mathbf{1}_{\mathbf{i}_{(l)}} = \mathbf{1}_{x\mathbf{i}_{(l)}}.$$

We can therefore construct a $\mathfrak{S}_n$-isomorphism between $\left\{\eta_{\mathbf{i}_{(l)}} : \mathbf{i}_{(l)} \in V^{(n-l,l)}\right\}$ and $M^{(n-l,l)}$ by linearly extending the mapping

$$\tau\left(\eta_{\mathbf{i}_{(l)}}\right) = \mathbf{1}_{\mathbf{i}_{(l)}}, \quad \mathbf{i}_{(l)} \in V^{(n-l,l)},$$

and by observing that, for every $\mathbf{k}_{(m)} \in V^{(n-m,m)}$, $\mathbf{i}_{(l)} \in V^{(n-l,l)}$ and $x \in \mathfrak{S}_n$,

$$x\eta_{\mathbf{i}_{(l)}}\left(\mathbf{k}_{(m)}\right) = \eta_{\mathbf{i}_{(l)}}\left(x^{-1}\mathbf{k}_{(m)}\right) = \eta_{x\mathbf{i}_{(l)}}\left(\mathbf{k}_{(m)}\right).$$

This concludes the proof. ■

End of the proof of Theorem 4. Since $SU_0 = SH_0 = K_0^{(n-m,m)} = \mathbb{C}$, the relation between representations

$$M^{(n-l,l)} \cong S^{(n)} \oplus S^{(n-1,1)} \oplus \cdots \oplus S^{(n-l,l)}, \quad \forall l = 1, ..., m,$$

along with Lemma 7, implies that the restriction of the action (7) of $\mathfrak{S}_n$ to those $f \in L\left(V^{(n-m,m)}\right)$ such that $f\left(X_{(m)}\right) \in SH_l$ is an element of $\left[S^{(n-l,l)}\right]$. This yields that each one of the $m+1$ summands in the decomposition

$$M^{(n-m,m)} = \mathbb{C} \oplus \bigoplus_{l=1}^{m} \left\{f : f\left(X_{(m)}\right) \in SH_l\right\}$$

is an irreducible $\mathfrak{S}_n$-submodule of $M^{(n-m,m)}$. Since the decomposition (8) of $M^{(n-m,m)}$ is unique, this gives

$$\left\{f : f\left(X_{(m)}\right) \in SH_l\right\} = K_l^{(n-m,m)},$$

as required. ■

Acknowledgement. We are grateful to Omar El-Dakkak for useful comments.

References

[1] D.J. Aldous (1983). Exchangeability and related topics. *École d'été de Probabilités de Saint-Flour XIII.* LNM **1117**. Springer, New York.

[2] M. Bloznelis (2005). Orthogonal decomposition of symmetric functions defined on random permutations. *Combinatorics, Probability and Computing*, **14**, 249-268.

[3] M. Bloznelis and F. Götze (2001). Orthogonal decomposition of finite population statistics and its applications to distributional asymptotics. *The Annals of Statistics* **29** (3), 353-365

[4] M. Bloznelis and F. Götze (2002). An Edgeworth expansion for finite population statistics. *The Annals of Probability*, **30**, 1238-1265

[5] P. Diaconis (1988). *Group Representations in Probability and Statistics.* IMS Lecture Notes – Monograph Series **11**. Hayward, California

[6] R. Dudley (2001). *Real analysis and probability* (*2nd edition*). Wadsworth and Brooks/Cole, Pacific Grove, CA.

[7] Duistermaat J.J. and Kolk J.A.C. (1997). *Lie groups.* Springer-Verlag, Berlin-Heidelberg-New York.

[8] O. El-Dakkak and G. Peccati (2008). Hoeffding decompositions and urn sequences. *The Annals of Probability* **36**(6), 2280-2310.

[9] G.D. James (1978) *The representation theory of the symmetric groups*, Lecture Notes in Mathematics **682**. Springer-Verlag. Berlin Heidelberg New York.

[10] G. Peccati (2004). Hoeffding-ANOVA decompositions for symmetric statistics of exchangeable observations. *The Annals of Probability* **32**(3A), 1796-1829.

[11] G. Peccati and J.-R. Pycke (2008). Decomposition of stochastic processes based on irreducible group representations. To appear in: *Theory of Probability and its Applications.*

[12] Sagan B.E. (2001). *The Symmetric Group. Representations, Combinatorial Algorithms and Symmetric Functions* (2nd edition). Springer, New York.

[13] R.J. Serfling (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

[14] Serre J.-P. (1977). *Linear representations of finite groups.* Graduate Texts in Mathematics **42**. Springer, New York.

[15] L. Zhao and X. Chen (1990). Normal approximation for finite-population U-statistics. *Acta Mathematicae Applicatae Sinica* **6** (3), 263-272