Representations and the symmetric group - Diaconis data

George McNinch

2024-02-07

character table of S_5

Let's use `GAP` to find the character table of S_5.

```
gap> G:=SymmetricGroup(5);
Sym( [ 1 .. 5 ] )
gap> tab:=CharacterTable(G);
CharacterTable( Sym( [ 1 .. 5 ] ) )
gap> Display(tab);
CT1

     2  3  2  3  1  1  2  .
     3  1  1  .  1  1  .  .
     5  1  .  .  .  .  .  1

       1a 2a 2b 3a 6a 4a 5a
    2P 1a 1a 1a 3a 3a 2b 5a
    3P 1a 2a 2b 1a 2a 4a 5a
    5P 1a 2a 2b 3a 6a 4a 1a

X.1     1 -1  1  1 -1 -1  1
X.2     4 -2  .  1  1  . -1
X.3     5 -1  1 -1 -1  1  .
X.4     6  . -2  .  .  .  1
X.5     5  1  1 -1  1 -1  .
X.6     4  2  .  1 -1  . -1
X.7     1  1  1  1  1  1  1
gap>
```

Diaconis example – survey data

This data is taken from the paper (Diaconis 1989)

It describes 5,738 completed ballots rank-ordering 5 candidates.

View a rank-ordered ballot as an element of the symmetric group S_5; we want to study the *frequency function* f.

first ranking table

TABLE 2
Percentage of voters ranking candidate i in position j

Candidate	Rank 1	2	3	4	5
1	18	26	23	17	15
2	14	19	25	24	18
3	28	17	14	18	23
4	20	17	19	20	23
5	20	21	20	19	20

the regular representation

TABLE 3
Decomposition of the regular representation

M =	V_1	⊕	V_2	⊕	V_3	⊕	V_4	⊕	V_5	⊕	V_6	⊕	V_7
Dim 120	1		16		25		36		25		16		1
SS/120	2286		298		459		78		27		7		0

This diagram shows the decomposition of the regular representation into *isotypic components*.

Be careful: the notation Diaconis is using here does not match that used by `GAP` above. For example, the representation Diaconis writes as V_3 is the isotypic component determined by the irreducible representation labeled `X.5` by `GAP`.

The second row reflects the decomposition of the frequency function f. Namely, write

$$f = \sum_{i=1}^{7} f_i \quad \text{with } f_i \in V_i.$$

The second row entries are the "sums of squares" $\langle f_i, f_i \rangle$.

Remember that we can compute the f_i using the *idempotents* in $\mathbb{C}[G]$.

For example,

$$f_1 = \frac{1}{5!} \sum_{\sigma \in S_5} \sigma.f$$

More generally, if χ_i denotes the character of the irreducible representation L_i with $V_i = \mathbb{C}[G]_{(L_i)}$ then

$$f_i = \frac{1}{5!} \sum_{\sigma \in S_5} \chi_i(\sigma^{-1})\sigma.f$$

Note that $\langle f_3, f_3 \rangle = 459$ is relatively large (ignoring $\langle f_1, f_1 \rangle$ since f_1 is trivial).

normalizing the first-order data

TABLE 4
First order effects

Candidate	Rank 1	2	3	4	5
1	−94	371	165	−145	−296
2	−372	−70	267	268	−92
3	461	−187	−354	−97	178
4	24	−175	−58	16	193
5	−18	62	−19	−41	17

THe i, j entry in this table is the number of votes ranking candidate i in the j-th position, minus the sample size over 5.

In particular, rows and columns sum to 0.

This normalization can also be achieved as follows:

Let f_2 be the projection on V_2, and consider the functions

$$\sigma \mapsto \delta_{i,\sigma(j)}.$$

The i, j entry of the preceding table is $\langle f_2, \delta_{i,\sigma(j)} \rangle$

TABLE 5
Second order, unordered effects

Candidate	Rank 1, 2	1, 3	1, 4	1, 5	2, 3	2, 4	2, 5	3, 4	3, 5	4, 5
1, 2	−137	−20	18	140	111	22	4	6	−97	−46
1, 3	476	−88	−179	−209	−147	−169	−160	107	128	241
1, 4	−189	51	113	24	−9	98	99	−65	23	−146
1, 5	−150	57	47	45	43	49	56	−48	−53	−48
2, 3	−42	84	19	−61	30	−16	82	−76	−39	72
2, 4	157	−20	−43	−25	−93	−76	−56	8	38	112
2, 5	22	−44	7	15	−117	69	25	62	99	−138
3, 4	−265	−7	72	199	39	140	85	19	−52	−233
3, 5	−169	10	88	70	78	44	47	−51	−36	−80
4, 5	296	−24	−142	−130	−5	−163	−128	38	−9	267

Interpretation in this last table:

Compute the projection f_3 of f into the component V_3 of $M = \mathbb{C}[S_5]$.

Now, consider the easily understood functions

$$\sigma \mapsto \delta_{\{i,i'\},\{\sigma(j),\sigma(j')\}}$$

in $\mathbb{C}[S_5]$ for distinct i, i' and distinct j, j'.

The space of these functions is a 100 dimensional subspace of W $\mathbb{C}[G]$.

The entries in the table are the *inner products*

$$\langle f_3, \delta_{\{i,i'\},\{\sigma(j),\sigma(j')\}} \rangle$$

Summary observations

The data were to elect a president for the American Psychological Association. Candidates 1 and 3 were `clinicians` while candidates 4 and 5 were `academicians`, two groups within the association with somewhat divergent perspectives.

In the second-order table, we see a preference for candidates 1 & 3 witnessed by the entry 376 corresponding to the entry for candidates $\{1, 3\}$ and ranks $\{1, 2\}$.

And we see a (slightly smaller) preference for candidates 4 and 5 witnessed by the entry 296 corresponding to the entry for candidates $\{4, 5\}$ and ranks $\{1, 2\}$.

Bibliography

Bibliography

Diaconis, Persi. 1989. "A Generalization of Spectral Analysis with Application to Ranked Data." *The Annals of Statistics* 17 (3): 949–79. https://www.jstor.org/stable/2241705.