A re-examination of the Diaconis-Graham inequality on the symmetric group

Petros Hadjicostas
(joint work with Chris Monico)

Right-invariant metrics on the set of permutations S_n of the first n positive integers were introduced by Diaconis and Graham in 1977 in a paper published in the *Journal of the Royal Statistical Society, Series B.* Each such metric provides a distance between two permutations of the first n positive integers in such a way that, if one changes the order of the numbers in the two permutations in exactly the same way (i.e., if the number in position i of the first permutation goes to position j, then the number in position i of the second permutation also goes to position j), then the distance between the two permutations stays the same.

Estivill-Castro (1991) introduced the idea of a right-invariant pseudo-metric by modifying the triangle inequality. In Statistics, if a right-invariant metric or pseudo-metric is properly standardized, then it becomes a non-parametric correlation coefficient between two numerical variables (where the original data are replaced by ranks). In Computer Science, the distance between a permutation and the identity element of the symmetric group can be considered as a *measure of disorder* of the permutation.

If a is a permutation in S_n (and a_i is the number in position i of a), one can use the following measures of disorder: $I(a)$, the number of inversions in a; $SQ(a)$, the sum of squares of the differences $a_i - i$; $D(a)$, the sum of the absolute values of the differences $a_i - i$; and $EX(a)$, the smallest number of successive exchanges of elements in a needed to sort a. In 1849, Cayley proved that $EX(a)$ equals n minus the number of *cycles* in the permutation a.

Diaconis and Graham proved that $I(a)+EX(a) \leq D(a) \leq 2\, I(a)$. (A weaker version of these inequalities, $I(a) \leq D(a) \leq 2\, I(a)$, has been used, for example, by Computer scientists in the analysis of ranking of the top search results by Internet search engines.) Their proof of the second inequality is easy, and equality holds if and only if a has no 3-inversions (a 3-inversion is a triplet (a_i, a_j, a_k) such that $i < j < k$ but $a_i > a_j > a_k$). Their proof of the first inequality is quite long, and not intuitive. In addition, they do not characterize when equality holds.

In this talk, we give a more intuitive proof of the first inequality, and we characterize all the cases where equality holds for the first inequality. Our proof is based on the idea that, if the order of two numbers in a that form an inversion is switched, then one can easily characterize how each one of the above measures of disorder changes. As observed by Diaconis and Graham, the number of permutations a in S_n that satisfy the second inequality equals the n-th Catalan number, but we do not know how often the first inequality holds – we present some numerical calculations for this case. For small n, the number of a in S_n that satisfy equality in the first inequality (say c_n) seems to be close to the number of permutations in S_n that avoid the pattern 1234; or the number of permutations that avoid the pattern 1324; or the number of permutations that avoid the pattern 1342. Unfortunately, for large n, the number c_n starts to diverge from these three numbers (that are related to pattern avoidance in permutations).