Combinatorics at the interface of probability, analysis and experimental mathematics

AEC Summer School, Hagenberg, Austria
August 2016
Dan Romik

Homework Problem Set

A. Experimental math problems

1. Download and install on your computer a symbolic math software application such as `SageMath` or `Mathematica`. Play with it for a while to get familiar with using the software for doing some simple computations (e.g., manipulation of polynomials, symbolic differentiation and integration).

2. Think of, and execute, a cool programming exercise to illustrate a mathematical concept or theorem. (For example: the $3x + 1$ map; some concept in basic calculus; an exploration of the parity of binomial coefficients; etc. If you are still short of inspiration, the `Wolfram Demonstrations` project has many beautiful `Mathematica`-based demos.)

3. If you are using `Mathematica`, download from the course web page `https://www.math.ucdavis.edu/~romik/aec-2016/` the experimental math notebook `experimental-math.nb` that I prepared. (If you don't use `Mathematica`, you can download a PDF printout of the notebook.) This notebook suggests several possible experimental math projects you could work on. One of the assignments is to come up with your own project involving an open problem you are interested in.

B. Theory problems

1. Let $m \geq n \geq 1$ be integers. Assume that two candidates, Alice and Bob, are competing in an election with $m + n$ voters, of which m are voting "A" for Alice and n are voting "B"for Bob. An (m, n)**-ballot sequence** is an ordering $v_1 \dots v_{m+n} \in \{\mathrm{A}, \mathrm{B}\}^{m+n}$ of the list of votes

such that at any time during the vote-counting process, Alice is ahead of, or tied with, Bob. For example, if $m = 5$ and $n = 3$ then

AABBAABA

is a $(5, 3)$-ballot sequence, but

ABABBAAA

is not a $(5, 3)$-ballot sequence.

(a) Prove that the number $B_{m,n}$ of (m, n)-ballot sequences is given by
$$B_{m,n} = \frac{m - n + 1}{m + 1} \times \binom{m+n}{m}.$$
This can be interpreted probabilistically as the statement that if the $m + n$ votes are read in a random order, the probability that Alice maintains a lead or tie over Bob throughout the counting process is precisely $\frac{m-n+1}{m+1}$.

(b) In particular, in the case $m = n$, deduce that $B_{n,n}$ is equal the nth Catalan number, $C_n = \frac{1}{n+1}\binom{2n}{n}$. (This is one of literally hundreds of interpretations of the combinatorial meaning of the Catalan numbers, surveyed in the recently published book "Catalan Numbers" by Richard Stanley.)

(c) Show that (m, n)-ballot sequences are in bijection with standard Young tableaux of shape (m, n) (=a Young diagram with one row of length m and another row of length n).

(d) Show that the result of (a) can be derived using the bijection of part (c) together with the hook-length formula.

(e) Show that standard Young tableaux of shape $\lambda = (\lambda_1, \dots, \lambda_k)$ are in bijection with "$(\lambda_1, \dots, \lambda_k)$-ballot sequences", which are generalizations of ordinary (m, n)-ballot sequences corresponding to an election of k candidates in which candidate j received λ_j votes for each $j = 1, \dots, k$, and where the votes are counted in an order for which at each point during the count candidate 1 has the most votes, candidate 2 has the second-highest number of votes, etc. (Thus, the hook length formula can be thought of as an elegant generalization of the solution to the 2-candidate ballot problem.)

2. The Erdős-Szekeres theorem cited in the lecture has many proofs. Find three different ones (by thinking about the question yourself and/or looking for proofs online, which is also a useful exercise since it is likely to lead you to other interesting and related topics).

3. (a) Show using the Erdős-Szekeres theorem that for a permutation $\sigma \in S_n$, the inequality

$$\mathrm{LDS}(\sigma)\,\mathrm{LIS}(\sigma) \geq n$$

   holds.

   (b) Explain how this implies the lower bound $L_n \geq \frac{1}{2}\sqrt{n}$ (discussed in the lecture) on the expected maximal of a longest increasing subsequence of a random permutation $\sigma_n \in S_n$.

   (c) Show that in fact the bound can be improved to $L_n \geq \sqrt{n}$.

4. Show that for any $D > e$ there exists a constant $c > 0$ such that if $k \geq D\sqrt{n}$ then

$$\frac{1}{k!}\binom{n}{k} < \frac{n^k}{(k!)^2} < e^{-c\sqrt{n}}.$$

   (This inequality was used in the proof of the upper bound $L_n \leq 3\sqrt{n}$.)

5. Consider the following alternative "proof" of the hook-length formula: Given a Young diagram λ with n boxes, choose a uniformly random filling of the boxes of λ with the numbers $1, \ldots, n$. There are $n!$ possibilities, and the probability of the event E that the resulting filling is a standard Young tableau is precisely $d_\lambda/n!$. On the other hand, we can represent the event E as an intersection of events

$$E = \bigcap_{(i,j)} F_{i,j}$$

   over all boxes (i,j) of λ, where $F_{i,j}$ is the event that the number written in box (i,j) is the smallest from among the numbers appearing in the hook of (i,j).

   Clearly the probability of each event $F_{i,j}$ is $1/h_\lambda(i,j)$, the reciprocal of the hook length of (i,j). The events $F_{i,j}$ are independent, so the probability of their intersection is the product of the probabilities, and we get precisely the hook-length formula.

   (a) Find the flaw in the proof. Why doesn't it work?

(b) Moreover, prove that this argument is not just incomplete, but actually incorrect.

(c) Given a rooted tree T with n vertices, we can define an analogue of a standard Young tableau to be a labeling of the vertices of the tree with the numbers $1, \ldots, n$ such that the labels are strictly increasing as one moves further away from the root of the tree. Prove the following analogue of the hook-length formula for trees, which states that the number d_T of "tree standard Young tableaux," as defined above, is equal to $n!$ divided by the product of hook lengths of the vertices of the tree, where the hook length of a vertex v counts the number of vertices that are ... . Hint: adapt the incorrect argument given above to the setting of trees and show that in this case it is actually correct.

C. Recommended exercises from my book

- 1.6(a), 1.7, 1.8, 1.9, 1.10, 1.11, 1.13, 1.14, 1.15, 1.16 (pp. 72–76)
- 4.9, 4.10(a), 4.10(b), 4.10(e) (pp. 265-267)
- 5.20, 5.21 (p. 328; see also p. 312)