Sketch of Solutions to Final Exam

Probabilistic Methods in CS

19 February 2003, 10:00-13:00

1 Question 1

The question. Consider an n by n integer grid on which we have to place points. We assume for simplicity that n is prime. A configuration of three points (p_1, p_2, p_3) is *forbidden* if p_1 and p_2 are in the same row, p_1 and p_3 are in the same column, and the distance between p_1 and p_2 (which we denote by d_2) and the distance between p_1 and p_3 (which we denote by d_3) satisfy the equation $7d_2 + 19d_3 = 0$ (modulo n). (n being prime implies that for every $1 \leq d_2 \leq n-1$ there is a unique value of $1 \leq d_3 \leq n-1$ that satisfies the equation.) Show that $\Omega(n^{3/2})$ points can be placed without forming any forbidden configuration.

This question was similar to a previous homework assignment, and essentailly all students got it right. Hence I'll not bother to write an answer. However, let me mention that one student correctly observed that for the particular linear equation given in the question, $\Omega(n^2)$ points arranged in a $n/27$ by $n/27$ subsquare do not contain a forbidden configuration.

2 Question 2

The question. Consider n^2 bins arranged in an n by n matrix. We throw independently uniformly at random n^2 balls into the n^2 bins. Let m denote the number of balls that fall in the most loaded bin. Show that with probability at least $1/2$, it simultaneously holds that every row contains at least one bin with at least $m/3$ balls. (You may assume here that n is "large enough").

No student really gave a full answer. This question was related to homework assignment 8, question 2. Recall that the homework question was: For

n balls, n bins and arbitrary k, what is the expected number of bins with k balls. Let E_k denote this value. Then

$$E_k = n\binom{n}{k}(\frac{1}{n})^k(\frac{n-1}{n})^{n-k}.$$

Apparently, most people did not bother to simplify this expression and see what it gives for various values of k. $\binom{n}{k}$ can be approximated by $n^k/k!$. $(1-1/n)^{n-k}$ is roughly $1/e$ for values of k that are small relative to n. Hence

$$E_k \simeq \frac{n}{e \cdot k!}.$$

Thus the expected number of bins with k balls for values of $k = 0, 1, 2, 3\ldots$ is roughly $n/e, n/e, n/2e, n/6e\ldots$. The expected number of bins with k balls drops off like $k!$, which is more than exponential in k. Of specific interest to the exam question are the following observations:

1. Let k_0 be the value of k for which the expected number of bins with k or more balls is smaller than $1/2$. Then $k_0 \simeq \frac{\log n}{\log\log n}$.

2. For $k_1 = \frac{k_0}{3}$, the expected number of bins with k_1 balls is $\Omega(n^{2/3})$. This is an immediate consequence of the superexponential decrease in E_k, that implies for $k_2 = 2k_0/3$ that

$$\frac{E_0}{E_{k_1}} \leq \frac{E_{k_1}}{E_{k_2}} \leq \frac{E_{k_2}}{E_{k_0}}.$$

   Via simple manipulations we get that $(E_{k_1})^3 \geq (E_0)^2 E_{k_0}$. In fact, we did not need an estimate on the value of k_0 in order to conclude that $E_{k_1} = \Omega(n^{2/3})$.

3. The probability that the actual number of bins with k_1 balls is below $E_{k_1} - c\sqrt{2n}$ is at most e^{-c^2}. This is a straightforward application of Azuma's inequality, as each ball can change the number of bins with k_1 balls by at most 1. Hence with high probability, the number of bins with k_1 balls is at least $\Omega(n^{2/3})$.

Equipped with these observations, the answer to the exam question is not difficult. (Unfortunately, I guess that most students did not make these observations in the context of the homework assignment.) Here we throw $N = n^2$ balls in $N = n^2$ bins. Now the corresponding $k_0 \simeq \frac{\log N}{\log\log N}$, and

$k_1 = k_0/3$. With high probability m (the number of balls in the most loaded bin) is at most k_0 (by Markov's inequality), and there are $\Omega(N^{2/3}) = \Omega(n^{4/3})$ bins with k_1 (and hence $m/3$) balls. Calls these bins *heavy*.

To finish of the answer, think of each heavy bin as a v-ball, and each row as a v-bin ("v" stands for virtual). Under this interpretation, deciding how to distribute the $n^{4/3}$ heavy bins in the n rows is the same as throwing $n^{4/3}$ v-balls in v-bins, under the extra provision that each v-bin gets at most n v-balls. Without the extra provision, we already saw in question 3 of homework assignment 8 that with high probability all v-bins have a v-ball. With the extra provision, things only get better, as if we ever throw into a v-bin more than n v-balls, we can get the extra v-balls out of the v-bin and rethrow them again, making it even less likely that there is a v-bin left without a v-ball.

Summarizing, with high probability there are $\Omega(n^{4/3})$ bins that have $m/3$ balls, and because $n^{4/3} \gg n \log n$, with high probability every row has at least one such bin.

The solution that I gave above was pretty simple (if one knows the background), but clever in the sense that it avoids complicated computations. A different legitimate approach was to say that every row is very likely to have roughly n balls. Then one could use the union bound to show that no row has more than $2k_0$ balls (for k_0 as defined above for n). Then one would need to show that every row has at least one bin with $k_2 = 2k_0/3$ balls. The expected number of such bins within a row is only roughly $n^{1/3}$. Now one runs into trouble if one wants to prove a concentration result using martingales, because a deviation of $\Omega(\sqrt{n})$ is larger than the expectation itself. One can get around this by using Talagrand's inequality, because a witness to having $n^{1/3}$ such bins requires only $k_2 n^{1/3}$ balls. (There are other ways that work as well, e.g., by breaking each row into $n^{1/3}$ segments, and proving that each segment has constant probability of having a bin with k_2 balls.) However, to use Talagrand's inequality, one needs the median number of bins with k_2 balls, rather than the expectation. This adds some complications to the proof, but I will not go into them (and anyway, nobody seems to have reached this point).

3 Question 3

The question. There are n players $P_1, \ldots, P_n$ in a room who play the following game. A hat is placed on the head of each person. This hat may be either red or blue. Each player can see the colors of the hats of all other players, but not of her own hat. Based on this information alone, each player needs to guess the color of her own hat. Let r denote the number of red hats, b the number of blue hats, $m = \max[r, b]$, and t the number of players who guessed correctly the true color of their hats. (Note that r, b, m, t are not known to the players.) The goal of this problem is to design a strategy for the players that ensures that t is not much smaller than m.

1. Randomized strategy. Consider the following strategy. Every player counts how many red hats she sees on the other player's heads. If this number is below $(n - \sqrt{n})/2$ then she guesses *blue* as the color of her own hat. If this number is above $(n + \sqrt{n})/2$ then she guesses *red*. If this number is $k + (n - \sqrt{n})/2$ for $1 \leq k \leq \sqrt{n}$, then she tosses a (biased) coin and with probability $k/\sqrt{n}$ guesses *red*, and with probability $1 - k/\sqrt{n}$ guesses *blue*. All players use this randomized strategy. Observe that whenever $m > 1 + (n + \sqrt{n})/2$, the strategy for every player becomes deterministic, and $t = m$. Show that for every placement of hats on the players, with high probability $t > m - O(\sqrt{n})$. More specifically, show that for $c > 0$, the probability of the event "$t < m - \sqrt{cn}$" behaves like $e^{-\Omega(c)}$.

2. Deterministic strategy. A deterministic strategy for a player is a table with 2^{n-1} entries that for each sequence of colors seen for the hats of the other players specifies a color as output. A deterministic strategy for all players is a collection of n such strategies, one for each player. (Different players may have different strategies.) Show the existence of a deterministic strategy that ensures that for every placement of hats on the players, $t > m - O(\sqrt{n \log n})$.

3. Open question. Is there a deterministic strategy that ensures $t > m - O(\sqrt{n})$? I do not know the answer to this question at the moment, so if you happen to figure out the answer, do let me know!

Part (1) of the question was done correctly by some. As noted in the question, for $m > 1 + (n + \sqrt{n})/2$, the strategy for every player becomes deterministic and $t = m$. So it remains to deal with $m \leq 1 + (n + \sqrt{n})/2$. Here, for every fixed assignment of hats, simple computations that I omit show that $E[t] = n/2 - O(\sqrt{n})$. Moreover, once the assignment is fixed, the probability of every player to guess her hat correctly is independent of all other guesses, and hence standard use of Chernoff bounds shows that the probability of the event "$t < m - \sqrt{cn}$" behaves like $e^{-\Omega(c)}$. (The Ω notation covers the fact that we need c to be large enough.)

Part (2) of the question can be solved using the uniform version of the local lemma. List all possible placement of hats, where we call the sequence of colors a *configuration*. There are 2^n possible configurations, and with each configuration i we associate a random experiment in which every player guesses a color for her hat according to the randomized strategy of part (1). The bad event b_i says that in the random experiment for configuration i, $t < m - \sqrt{cn \log n}$. For c a large enough constant, the probability of the bad event b_i is at most $p = 1/4n$. Each event (configuration i) depends on only n other bad events, namely, those configurations that differ from configuration i in the color of the hat of exactly one player (because then this player makes the same decision in both configurations). Hence the dependency graph has degree $d = n$, and the condition $ep(d+1) < 1$ is satisfied. By the local lemma, there is a way of assigning guesses to all players in all configurations such that no bad event happens. This is a deterministic strategy for the game. (Note that the question only asked to prove the existence of a strategy.)

For part (3), I still do not know the answer.