Fall 2013

COMP 691D Probabilistic Methods in Computer Science
Midterm exam
Solutions

1. Let G be a finite undirected graph. A *bicoloured graph based on* G is G with each edge coloured green or red. A bicoloured graph is called *even* if its vertex set can be partitioned onto two parts in such a way that the two endpoints of every green edge belong to two distinct parts and every red edge has both of its endpoints in one of the two parts. The *oddness* of a bicoloured graph is the smallest number of its edges such that switching the colours of these edges (green $\leftrightarrow$ red) produces an even graph.

Let $f(n,m)$ denote the the largest k with the following property:

> For every graph G with n vertices and m edges
> there is a bicoloured graph based on G with oddness at least k.

1.1 [5 points]. Write down an inequality in n, m, k which implies $f(n,m) > k$.

One answer:

$$2^n \sum_{i\le k} \binom{m}{i} < 2^m. \tag{1}$$

Justification: Let G be a graph with n vertices and m edges. Call a bicoloured graph based on G *bad* if its oddness is at most k. All bad bicoloured graphs based on G can be manufactured by choosing (i) a partition of the vertex set of G into parts and (ii) a set S of at most k edges of G: then an edge with both endpoints in one of the two parts is coloured green if and only if it belongs to S and an edge whose two endpoints edge belong to two distinct parts is coloured red if and only if it belongs to S Since there are 2^n choices in (i) and $\sum_{i\le k}\binom{m}{i}$ choices in (ii), the total number of bad colourings is at most

$$2^n \sum_{i\le k} \binom{m}{i}.$$

If this number is less than 2^m, then there must be a good bicoloured graph based on G, and so $f(n,m) > k$.

1.2. [5 points] Find a function g (the larger the better) for which the preceding inequality implies $f(n,m) > g(n,m)$.

One answer:

$$g(n,m) \;=\; \frac{m}{2} - \sqrt{(0.5\ln 2)nm}.$$

Justification: When $t = \sqrt{(0.5\ln 2)n/m}$, the Bernstein-Okamoto bound guarantees that

$$2^n \sum_{i\le(0.5-t)m} \binom{m}{i} \;<\; 2^m \exp(n\ln 2 - 2t^2 n) \;\le\; 2^m,$$

and so inequality (1) is satisfied whenever $k \le g(n,m)$.

2. Given a positive integer n, let us write $P_n = \{1,2,\ldots,n\}$ and $N_n = \{-1,-2,\ldots,-n\}$. By an (m,n) sequence, we will mean a sequence of (not necessarily distinct) m integers coming from $P_n \cup N_n$. Such a sequence will be called *safe* if it does not include any integer x along with its opposite $-x$; otherwise it will be called *unsafe.*

2.1. [5 points] Find a function s (the faster growing the better) such that $\lim_{n\to\infty} m(n)/s(n) = 0$ implies that almost all (m,n) sequences are safe. Justify your answer.

The best answer: $s(n) = n^{1/2}$.
One justification: The following algorithm will produce exclusively safe (m,n) sequences: (i) choose a set S of m elements of P_n, (ii) convert S into a subset T of $P_n \cup N_n$ by deciding, for each element x of S, whether to replace it by $-x$ or not, (iii) order the m elements of T into a sequence. Since there are $\binom{n}{m}$ choices in (i) and 2^m choices in (ii) and $m!$ choices in (iii), it follows that there are at least $\binom{n}{m}\cdot 2^m \cdot m!$ safe (m,n) sequences; the ratio of this number to the number $(2n)^m$ of all (m,n) sequences comes to $\binom{n}{m}\cdot m!/n^m$, which is $1-o(1)$ when $m = o(n^{1/2})$.

2.2. [5 points] Find a function u (the slower growing the better) such that $\lim_{n\to\infty} m(n)/u(n) = +\infty$ implies that almost all (m,n) sequences are unsafe. Justify your answer.

The best answer: $u(n) = n^{1/2}$.
One justification: Let m be a function of n such that $\lim_{n\to\infty} m(n)/n^{1/2} = +\infty$; let k be any integer-valued function such that $k(n) = \Theta(n^{1/2})$; let $A(n)$ denote the number of all (m,n) sequences where fewer than $k(n)$ dis-

tinct integers appear in the first $2k(n)$ positions; let $B(n)$ denote the number of safe (m, n) sequences where at least $k(n)$ distinct integers appear in the first $2k(n)$ positions. We will prove that $A(n) = o((2n)^{m(n)})$ and that $B(n) = o((2n)^{m(n)})$.

If fewer than $k(n)$ distinct integers appear in the first $2k(n)$ positions, then some integer appears at least three times in these $2k(n)$ positions. All (m, n) sequences where some integer appears at least three times in the first $2k(n)$ positions can be manufactured by (i) choosing the integer that will appear at least three times, (ii) choosing three of the first $2k(n)$ positions, where this integer will appear, and (iii) filling the remaining $m(n) - 3$ positions with integers from $P_n \cup N_n$. Since there are $2n$ choices in (i) and $\binom{2k(n)}{3}$ choices in (ii) and $(2n)^{m(n)-3}$ choices in (iii), there are at most $\binom{2k(n)}{3}(2n)^{m(n)-2)}$ such instances; the ratio of this number to the number $(2n)^{m(n)}$ of all (m, n) sequences comes to $\binom{2k(n)}{3}/(2n)^2$, which is $o(1)$.

If a safe (m, n) sequence has at least $k(n)$ distinct integers appearing in the first $2k(n)$ positions, then opposites of these $k(n)$ integers cannot appear in any of the subsequent $m(n) - 2k(n)$ positions. It folllows that the number of such sequences is at most

$$(2n)^{2k(n)} \cdot (2n - k(n))^{m(n)-2k(n)};$$

the ratio of this number to the number $(2n)^{m(n)}$ of all (m, n) sequences comes to $(1 - k(n)/2n)^{m(n)-2k(n)}$, which is $o(1)$ since $(1 - k(n)/2n)^{m(n)-2k(n)} \leq \exp(-k(n)(m(n) - 2k(n))/2n)$.