Homework 8. Due Tue 10/16.

1. Consider Poisson trials with \( n = 5 \) and \( p_i \) equal to \( \frac{1}{2}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{3}{8} \). What is the mode for number of successes? What is the mean?

2. Let \( a = 3 \), and evaluate to five or so decimals the two quantities appearing below in the identity marked (SumIdentity). Compare.

3. (a) What is the probability that parents have two daughters, given that they have two children and the older one is a girl? What is the probability that parents have two daughters, given that they have two children and the younger one is a girl? What is the probability that parents have two daughters, given that they have two children and one of them is a girl?
   (b) Is the following identity always true:
   \[
   \Pr(A|B_1) \leq \Pr(A|B_1 \cup B_2).
   \]

4. Let \( X \) be the number of \( K_5 \)'s in \( G_{n,p} \). Determine exponents \( a, b \) such that \( \mathbf{E} X \approx n^a p^b \). Let \( X_i \) be the indicator of the \( i \)-th \( K_5 \). We know
   \[
   \text{Var}(X) \leq \sum_i \mathbf{E} X_i + \sum_{i \neq j} \text{Cov}(X_i, X_j).
   \]
   Determine exponents \( a, b \) such that \( \text{Cov}(X_i, X_j) \approx n^a p^b \) for each way in which the \( i \)-th and \( j \)-th \( K_5 \)'s might overlap: 2 vertices, 3 vertices, or 4 vertices. Determine the threshold \( p = p(n) \) for the appearance of a \( K_5 \), and prove both sides of your answer correct.

5. We started off the Probabilistic Method chapter with a theorem about Ramsey numbers: if
   \[
   \binom{n}{k} 2^{1-(k/2)} < 1 \quad (*)
   \]
then $R(k, k) > n$. As in the proof of that theorem, again let $\Omega$ be the space of all 2-colorings of the edges of $K_n$, with uniform probability. For each subsets $S$ of size $k$, let $A_S$ be the event “$S$ is monochromatic.”

Define a graph $G$ on the events $A_S$ by declaring $(S, T)$ to be an edge whenever $|S \cap T| \geq 2$; show that $G$ is a dependency graph for the events $A_S$. Use the LLL to prove a new theorem: if

$$4 \binom{k}{2} \binom{n}{k-2} 2^{1-\binom{k}{2}} < 1 \quad (**)$$

then $R(k, k) > n$.

What is the best lower bound on $R(10, 10)$ that can be proven using (*)? What is the best lower bound on $R(10, 10)$ that can be proven using (**) ? What is the smallest $k$ such that $R(k, k) > 5000$ can be proven using (*)? What is the smallest $k$ such that $R(k, k) > 5000$ can be proven using (**) ?

**Eighth week recap**

Tue 10/2. We talked about problem 5.12, p 121. The errata indicate that parts (a) and (b) are fine, but that (c) is wrong. Further, it is indicated that, even though the final conclusion of (c)-(f), that $\Pr(Z \geq \mu) \geq 1/2$, is true, proving it is too difficult for students of the course. (Whether this discussion matches up with what is in your book depends on which version of the book you have.) The latest version asks for just a computational confirmation through $\mu = 10$, but the errata itself appears to state the problem wrong; i.e., an error in the errata, as far as we can tell.

Proving inequalities analytically can be challenging. As a cultural addendum, we went over another famous and difficult proof that resembles a little bit Problem 5.12. Namely, consider Poisson trials. We have $n$ independent 0-1 variables $X_i$, with $p_i$ the probability that $X_i$ is 1. Let $p(n, k)$ be the probability that $\sum_{i=1}^n X_i$ equals $k$. We have the recursion

$$p(n, k) = p(n - 1, k - 1) \times p_n + p(n - 1, k - 1) \times (1 - p_n).$$
From here one may prove strict log-concavity

\[ p(n, k)^2 > p(n, k - 1)p(n, k + 1) \]

and from there that there is either a unique value \( k_* \) or at most two consecutive such values which satisfy

\[ p(n, k_*) \geq p(n, k) \text{ for all } k. \]

The value \( k_* \) is called the \textit{mode}. It is the most likely value of \( k \). The expected value, \( \mu \), of \( k \) is given by

\[
\mu = \sum_k kp(n, k) = \sum_{i=1}^n p_i.
\]


\[ |k_* - \mu| \leq 1. \]

This is a difficult proof with many cases, details, etc.

We finally got back to course. We proved that the choice of probability \( p \) in the randomized algorithm for finding a large independent set is optimal. We then stated and proved the theorem

\[
\Pr(X = 0) \leq \frac{\text{Var}(X)}{\text{E}X^2},
\]

which is referred to as the Second Moment Method (2MM). The hypothesis on random variable \( X \) is \( X \geq 0 \). The book also assumes \( X \) is integer-valued, but this does not seem to be used in the proof.

One place where \( X \geq 0 \) AND \( X \) is integer-valued seem to be needed is in the useful inequality

\[
\Pr(X \geq 1) = p_1 + p_2 + \cdots \leq p_1 + 2p_2 + \cdots = \text{E}X. \quad (*)
\]
We considered the following random variable $X$: take a random graph $G_{n,p}$ and let $X$ be the number of $K_4$’s contained therein. Then

$$E X = \binom{n}{4} p^6.$$

Using (⋆) above, we proved:

**Proposition 1.** Suppose that $p = p(n)$ depends on $n$ in such a way that $pn^{2/3} \to 0$. Then, the probability that $X = 0$ goes to 1 as $n \to \infty$.

Wed 10/3. We talked a bit about the bygone problem 5.12, and showed how it might be finished. Use would be made of a remarkable identity

$$\sqrt{a} \sum_{h=\infty}^{+\infty} e^{-\pi h^2 a} = \sum_{h=-\infty}^{+\infty} e^{-\pi h^2 / a}. \quad \text{(SumIdentity)}$$

We then used 2MM to prove (with $X$ still the number of $K_4$’s in $G_{n,p}$)

**Proposition 2.** Suppose that $p = p(n)$ depends on $n$ in such a way that $pn^{2/3} \to \infty$. Then, the probability that $X = 0$ goes to 0 as $n \to \infty$.

To apply 2MM you need an upper bound on $\text{Var}(X)$. Here we write $X = \sum_{i=1}^{n(i)} X_i$, a sum of indicators, take the formula for variance of a sum, bound $\text{Var}(X_i)$ by $E X_i$ and deal with the covariance terms.

Thu 10/4. The agenda said we would discuss LLL briefly, and then begin Chapter 7. However, the brief discussion became long when we found that we could not understand the notion of the dependency graph for a set of events $E_1, E_2, \ldots, E_n$. Towards the end, we began to suspect that the key was to say a dependency graph for a set of events $E_1, E_2, \ldots, E_n$.

**Ninth week topics**

Tue 10/9. More on LLL, using an unofficial handout. The book has a nice example with 2SAT, and we put another example in the next homework for the Probabilistic Method swan song.
Wed 10/10. And now, at last, intro to Chapter 7.

Thu 10/11. The random 2SAT algorithm based on Markov chain. Time permitting, move on to 3SAT.