

CMSC 858C, Spring 2009: Ungraded Homework 2

Note: We will have *ungraded* homework assignments such as this one, as well as ones that will be graded. I will post the solutions for all the assignments some time after they are handed out. You will get the most out of this course if you do your best to solve all the homework problems (whether they are graded or not) by yourself. The suggested time period to finish this assignment is 10 days.

0. Limited independence uses “polynomial” functions in tail bounds, while the Chernoff bounds with fully-independent random variables use “exponential” type functions. However, it is interesting to note that we can go back and forth between these two notions!

- Read and understand the tail inequalities from Section 2 of the Bellare-Rompel paper, as well as their proof from the Appendix of that paper. (See how bounds for fully-independent random variables turn out to be useful for the t -wise independent case!)
- Read the paper by Schmidt, Siegel & Srinivasan, available at <http://www.cs.umd.edu/~srin/PDF/ch-bounds.pdf>, up to the end of Section 2.2. Note how the exponential function of Chernoff-Hoeffding can essentially be replaced by polynomials – the elementary symmetric polynomials, denoted S_k in the paper. (Section 2.3 of this paper contains Bellare-Rompel type bounds, but the Bellare-Rompel approach and final results are cleaner.)
- Note that some of the questions below make use of (at least one of) the above two papers.

1. We show tail bounds for sums of “negatively correlated” random variables. Let X_1, X_2, \dots, X_n be Bernoulli random variables¹ that are not necessarily independent. Let $X = \sum_i X_i$.

- Suppose we have, for all subsets $S \subseteq \{1, 2, \dots, n\}$, that $\Pr[\bigwedge_{i \in S} (X_i = 1)] \leq \prod_{i \in S} \Pr[X_i = 1]$. Prove that the upper tail of X can be bounded by the upper-tail value given by the analogous Chernoff-Hoeffding bound (which would have held if the X_i were independent).
- Similarly suppose we have, for all subsets $S \subseteq \{1, 2, \dots, n\}$, that $\Pr[\bigwedge_{i \in S} (X_i = 0)] \leq \prod_{i \in S} \Pr[X_i = 0]$. Prove that the lower tail of X can be bounded by the lower-tail value given by the analogous Chernoff-Hoeffding bound.

2. Motivated in part by the Schmidt-Siegel-Srinivasan ([SSS]) paper, part(a) of this problem suggests a tail-bound approach that is worth keeping in mind, especially when working with upper-tail bounds with large relative-deviations δ .

(a) Suppose X_1, X_2, \dots, X_n are Bernoulli random variables that are not necessarily independent. Let $X = \sum_i X_i$. For any integer $a \geq 1$, prove, in the notation of [SSS], that

$$\Pr[X \geq a] \leq \mathbf{E}[S_a(X_1, X_2, \dots, X_n)].$$

(b) We have n bins, and n balls are thrown uniformly at random and independently into these bins. Use the approach of part (a) of this problem to re-prove the result that the maximum load on any bin is at most $O(\log n / \log \log n)$ with high probability.

¹That is, they take values in $\{0, 1\}$

3. We have n bins, and n balls are thrown uniformly at random and independently into these bins. Let X denote the number of empty bins. Prove that $\mathbf{E}[X] \sim n/e$. Also show that X has upper- and lower-tail bounds of magnitude given by the “analogous” Chernoff upper- and lower-tail bounds. (Consider what “analogous” could reasonably mean here.)

4. Prove the Chebyshev-Cantelli inequality by constructing an appropriate quadratic function.

5. We will now present an *explicit construction* of a multi-set S as in problem 3 from graded homework 2. Consider the field $GF[2^m]$, the unique field with 2^m elements. We will need the following result: each element x of this field can also be encoded in some bijective manner as an m -bit string $\text{enc}(x)$, which has the following property:

for any $x, y \in GF[2^m]$, if $z = x + y$ is their sum in the field, then $\text{enc}(z) = \text{enc}(x) \oplus \text{enc}(y)$, where the “ \oplus ” denotes the bit-wise XOR of the two m -bit vectors $\text{enc}(x)$ and $\text{enc}(y)$. (In particular, $\text{enc}(0)$ is the all-zeroes vector.)

For elements u and v of this field and any non-negative integer i , we define a bit

$$f_i(u, v) = (\text{enc}(u^i) \cdot \text{enc}(v)) \bmod 2,$$

where u^i is simply the field operation of multiplying u by itself i times, and “ $\text{enc}(u^i) \cdot \text{enc}(v) \bmod 2$ ” is just the dot-product (inner product) of the two m -bit vectors $\text{enc}(u^i)$ and $\text{enc}(v)$, taken mod 2.

Now suppose we choose elements U and V of $GF[2^m]$ uniformly at random and independently, and define random bits X_1, X_2, \dots, X_n as follows: $X_i = f_i(U, V)$. Prove that for all nonempty $A \subseteq \{1, 2, \dots, n\}$,

$$1/2 - \frac{n}{2^{m+1}} \leq \Pr \left[\left(\bigoplus_{i \in A} X_i \right) = 1 \right] \leq 1/2.$$

(Hint: Over any field \mathbf{F} , any polynomial of any degree t can have at most t roots.)