max planck institut
informatik

SAARLAND UNIVERSITY

COMPUTER SCIENCE

Prof. Dr. Benjamin Doerr **Winter 2012/13**

Exercises for Randomized Methods in Computer Science

http://www.mpi-inf.mpg.de/departments/d1/teaching/ws12/rmcs/

Assignment 3 Due: Wednesday, November 7, 2012

Exercise 0 (*Reminder Variances*) Read the two pages linked from the course webpage on Chebyshev's inequality. Check that you know all the small facts not made explicit (e.g., that the two different definitions of variance and covariance as the same, or Claim 1 to 6) and that you are able to prove them. If not, ask a colleague, look them up in any textbook etc.

Exercise 1

a) [elementary] For the coupon collector problem, use the Markov, Chebyshev and Chernoff inequality to bound deviations from the mean. Note: For each part, you will have to regard the right random variables...

b) [elementary] We still collect coupons of n different types, but now we want each coupon r times. r may be a constant or may be a function of n, e.g., $r = log(n)$ or $r = n$. How many rounds to we need (an asymptotic statement, that is, $\Theta(\dots)$ is enough)? Prove your results. After having done this, have a look at `http://epubs.siam.org/doi/abs/10.1137/1116027`. In particular, compare your result to this for $r = \log n$.

c) [difficult] In this exercise, we grow (rooted) binary trees of height at most h. We start with the tree consisting of a single node, which hence is the root of this tree and the only leaf at the same time. Each round, we choose a leaf of our current tree uniformly at random from all leaves. If this leaf is at height h, we do nothing (but count this as a round, of course). If this leaf is at height less than h, we append two new nodes as children to the leaf (which then is not a leaf anymore; the two new nodes naturally are leaves). This process ends when we have a complete binary tree of height h. How many rounds does this take? Give a good upper bound! Hint: No complicated maths is necessary. The problem is easy if you look at it from the right angle and then are generous in the estimates.

Exercise 2 (*elementary to moderate*) Another useful Chernoff bound is the following: Let $X_1, \ldots, X_n$ be independent random variables taking values in $[0,1]$. Let $X = \sum_{i \in [n]} X_i$. Then

$$\Pr(X \geq E(X) + \lambda) \leq \exp(-2\lambda^2/n)$$

for all $\lambda \geq 0$. Naturally, the same bound holds for the event "$X \leq E(X) - \lambda$". These bounds shall be used in this exercise.

Let $H = (V,E)$ a hypergraph, that is, V is a finite set ("vertices") and E is a set of subsets of V ("edges"). We usually abbreviate $n := |V|$ and $m := |E|$. The discrepancy problem asks for a 2-coloring of V such that, ideally, each edge $e \in E$ contains the same number of vertices in each color. For convenience, we use -1 and $+1$ as colors, that is, we regard colorings $f : V \to \{-1,+1\}$. This allows to measure the (signed) imbalance in an edge simply by $f(e) := \sum_{v \in e} f(v)$. Now the discrepancy of a coloring f is simply the maximum absolute imbalance: $\operatorname{disc}(H,f) = \max_{e \in E} |f(e)|$. The discrepancy of the hypergraph H is the discrepancy of the best possible coloring, that is, $\operatorname{disc}(H) = \min_f \operatorname{disc}(H,f)$.

a) [elementary] Prove that for any hypergraph $H = (V,E)$, we have $\operatorname{disc}(H) \leq \sqrt{2n\log(2m)}$. To this aim, reformulate the above Chernoff bound to deal with random variables taking values in $[-1,1]$.

b) [elementary] What discrepancy bound do to you get in (a) if you want a random coloring to fulfill your bound with probability at least $1/2$?

c) [elementary] A very simple hypergraph is the one of two-dimensional combinatorial boxes. It has as vertex set a grid $[k] \times [k]$. Each set of the form $S \times T$ with $S,T \subseteq [k]$ is an edge of this hypergraph. What discrepancy bound does (a) give? [challenging/frustrating] What is the best coloring you can find constructively?

d) [moderate; needs a small but clever idea] Improve the bound in (a) by a factor of $\sqrt{2}$. Improve the bound in (c) by at least a factor of 2, better by a little more...

Exercise 3 (*moderate to difficult*) Imagine that there are n people, each knowing each other (and their phone numbers). One of them has a great news. This news is disseminated among this group according the following simple mechanism. In each round of the protocol, everyone already knowing the news calls a person chosen independently and uniformly at random and tells him/her the news. How many rounds are necessary and sufficient to spread the news? Again, a $\Theta(...)$ statement is sufficient. Hint on how to proceed: This exercise only needs arguments that we discussed in the lecture, and none of them in a particularly ingenious manner. The difficulty stems from the fact that several arguments have to be used in the right way. Consequently, try to first understand what is going on, then try to transform this intuitive understanding into a proof.]