Proofread 5-9-2005

CMSC 27400-1/37200-1 Combinatorics and Probability **Spring 2005**

Lecture 18: May 9, 2005

Instructor: László Babai *Scribe: Raghav Kulkarni*

TA SCHEDULE: TA sessions are held in Ryerson-255, Monday, Tuesday and Thursday 5:30–6:30pm.
INSTRUCTOR'S EMAIL: laci@cs.uchicago.edu
TA's EMAIL: hari@cs.uchicago.edu, raghav@cs.uchicago.edu

Concentration Inequalities

The random variables $X_1, \dots, X_n$ are independent if $(\forall a_1, \dots, a_n \in \mathbb{R})(P(X_1 = a_1 \text{ and} \dots \text{ and} X_n = a_n) = P(X_1 = a_1) \dots P(X = a_n))$.

Exercise 18.1 *If $X_1, \dots, X_n$ are independent random variables then for every subset $S \subseteq [n]$ $\{X_i \mid i \in S\}$ are independent.*

Exercise 18.2

1. *Give an example of 3 random variables which are pairwise independent but* not *independent.*
2. *For every integer n, construct n random variables which are $(n-1)$-wise independent but* not *independent.*

Exercise 18.3 *Suppose $1 \le k \le n$. (a) $\left(\frac{n}{k}\right)^k \le \frac{n}{k} < \left(\frac{en}{k}\right)^k$*
(b) $\sum_{i=0}^{k} \binom{n}{i} < \left(\frac{en}{k}\right)^k$
(c) $\frac{2^{2n}}{2n+1} < \binom{2n}{n} < 2^{2n}$.

Let $0 < p < 1$. *Definition:* The *binary entropy function* $H(p) := -p \log_2 p - (1-p)\log_2(1-p)$.

Exercise 18.4 $\frac{2^{H(p)n}}{n+1} < \binom{n}{k} < 2^{H(p)n}$ *where $p = \frac{k}{n}$.*

Notice that the Exercise 18.3 is a special case for $p = 1/2$ since $H(1/2) = 1$.

Exercise 18.5

1. *Show that for $p \ne 1/2$,* $\quad 0 < H(p) < 1$.

2. *$H(p) = H(1-p)$.*
3. *Show that $H(0) = H(1) = 0$. (Hint: Show that $\lim_{x\to 0} H(x) = 0$.)*

Exercise 18.6 *If $n_i \to \infty$, $\frac{k_i}{n_i} \to p$ then*

$\binom{n_i}{k_i} \sim \frac{c_p}{\sqrt{n}} 2^{H(p)n}$ *where $0 < p < 1$. (Hint: Stirling's formula.)*

1. *Find the value of $c = c(p)$.*

Chernoff's bound (tail estimate) (concentration inequality):
Suppose $Y_1, \ldots, Y_n$ are independent random variables such that $E(Y_i) = 0$ and $|Y_i| \le 1$. Let $X = \sum_{i=1}^n Y_i$ then $P(X \ge a) \le 2\mathrm{e}^{\frac{-a^2}{2n}}$.
HOMEWORK: Read proof from handouts.
Let $H :=$ # heads in a sequence of n coin flips. By Chebyshev's inequality,

$$P(|H - \frac{n}{2}| \ge \epsilon n) \le \frac{Var(H)}{(\epsilon n)^2} = \frac{1}{4\epsilon^2 n}.$$

$H = \sum_{i=1}^n T_i$ where $T_i = 1$ with probability 1/2 and 0 with the same probability. $\mathrm{var}(H) = \sum_{i=1}^n \mathrm{var}(T_i) = \frac{n}{4}$. ($T_i$ are pairwise independent.)

$$Y_i = T_i - 1/2.$$

$$E(Y_i) = 0.$$

$$X = \sum_{i=1}^n T_i.$$

$$E(X) = 0.$$

$$\begin{aligned} H &= \sum_{i=1}^n T_i \\ &= \sum_{i=1}^n (Y_i + \frac{1}{2}) \\ &= X + \frac{n}{2}. \end{aligned}$$

$X = H - \frac{n}{2}$. $P(|X| \ge \epsilon n) \le 2^{\frac{-\epsilon^2 n^2}{2n}} = 2^{\frac{-\epsilon^2 n}{2}} \to 0$ at an exponential rate.

$|Y_i| < 1$. $T_i = 1$ or 0 with probability 1/2. $Y_i = 1/2$ or -1/2 with probability 1/2. Infact we can apply Chernoff to $2X = \sum_{i=1}^n 2Y_i$. $P(|X| \ge \epsilon n) = P(|2X| \ge 2\epsilon n) \le 2e^{\frac{-(-2\epsilon n)^2}{2n}} = 2e^{-2\epsilon^2 n}$.

Exercise 18.7 *Use Chernoff's bound to prove that for almost all graphs* $\deg_{\max} \leq \frac{n}{2} + \sqrt{n \ln n(1+\epsilon)}$.

Exercise 18.8 $P((\#\ \mathit{edges} - \frac{1}{2}\binom{n}{2}) > \epsilon n^2) < e^{-c\epsilon^2 n^2}$.

Exercise 18.9 *$P(|(\#\ \mathit{triangles} - \frac{\binom{n}{3}}{8})| > \epsilon n^3) < \mathrm{e}^{-c(\epsilon)n^3}$. *(Hint:* $\#\ \mathit{triangles} = \sum_{i=1} \binom{n}{3} Y_i$*.)*