MATH 262A CLASS NOTES, 4-13-04

DANIEL S. FELIX

We begin with an easy definition.

Definition 1. *A* martingale *is a sequence* $X_0, \ldots, X_t$ *of random variables satisfying*

$$E[X_{i+1}|X_0, \ldots, X_i] = X_i$$

for all $0 \leq i \leq t-1$.

The classic example of such a sequence is the standard random walk on the integers. More specifically, we let the X_i be given by the sum

$$X_i = \sum_{j=1}^{i} Y_j$$

where the Y_j are independent random variables taking the values 1 and -1, each with probability $\frac{1}{2}$. The variable Y_j indicates whether we move left or right at time j, while the sum X_i represents our position. It should come as no surprise that our expected position at time $i+1$ is always X_i:

$$\begin{aligned} E[X_{i+1}|X_0, \ldots, X_i] &= E[X_{i+1}|X_i] \\ &= \frac{1}{2}(X_i + 1) + \frac{1}{2}(X_i - 1) \\ &= X_i \end{aligned}$$

so that this process is a martingale.

More often that not, martingales are used in combinatorics to show that a particular random variable is concentrated around its mean. We can show (via the second moment method, for example) that for the above sequence $X_0, \ldots, X_t$ describing a t-step random walk, the outcome of X_t is concentrated in an interval of length $c\sqrt{t}$. This concentration property is not unique to the random walk, and in fact it is shared by all martingales satisfying the *Lipschitz condition*

$$|X_{i+1} - X_i| \leq 1 \quad \text{for all } 0 \leq i \leq t-1.$$

Of course, we may always scale a martingale (check that linear combinations of martingales are martingales) so that any martingale in which the differences $|X_{i+1} - X_i|$ are uniformly bounded by a constant c can be said to satisfy a version of the Lipschitz condition.

Theorem 1 (Azuma's Inequality)**.** *Let* $0 = X_0, \ldots, X_t$ *be a martingale satisfying the Lipschitz condition. Then for every* $\lambda > 0$,

$$P[X_t > \lambda] < e^{-\lambda^2/2t}.$$

Since the sequence $-X_0, \ldots, -X_t$ is also a martingale, an immediate corollary to this theorem is the fact that the random variable X_t is concentrated. That is, with $\lambda \mapsto \omega(t)\sqrt{t}$ for some $\omega(t)$ tending to infinity arbitrarily slowly, X_t lies in an interval of width $2\omega(t)\sqrt{t}$ with probability $1 - 2exp(-\omega(t)^2/2)$, which tends to 1. Also, notice that for martingales for which $X_0 \neq 0$ the concentration will occur around $X_0 = E[X_t]$. This result was first proven by Hoeffding in 1963 (see [5]) and only later by Azuma [2].

Proof. Our approach follows [1], though here we treat only the case in which the random variables $Y_i := X_i - X_{i-1}$ take on finitely many values. Notice that in general a martingale need not be discrete.

Given knowledge of X_{i-1}, let Y_i take the values a_{ij} with probabilities p_{ij}, respectively. Let $\alpha > 0$ (to be optimized later) and consider the expectation

$$E[e^{\alpha Y_i} | X_0, \ldots, X_{i-1}] = \sum_j p_{ij} e^{\alpha a_{ij}}.$$

Since $e^{\alpha x}$ is convex, we know that it lies below the straight line passing through the points $(-1, e^{-\alpha})$ and $(1, e^{\alpha})$ whenever $x \in [-1, 1]$. Points on this line are given by

$$\begin{aligned} y &= \left(\frac{e^{\alpha} - e^{-\alpha}}{2}\right) x + \left(\frac{e^{\alpha} - e^{-\alpha}}{2}\right) \\ &= \cosh(\alpha) x + \cosh(\alpha) \end{aligned}$$

so that

$$\begin{aligned} \sum_j p_{ij} e^{\alpha a_{ij}} &\leq \sum_j p_{ij} \left(\cosh(\alpha) a_{ij} + \cosh(\alpha)\right) \\ &= \cosh(\alpha) \sum_j (p_{ij} a_{ij} + p_{ij}) \\ &= \cosh(\alpha) \left(E[Y_i | X_0, \ldots, X_{i-1}] + 1\right) \\ &= \cosh(\alpha) \end{aligned}$$

since $X_0, \ldots, X_t$ is a martingale. This is at most $e^{\alpha^2/2}$ in the interval $[-1, 1]$, as a check of the taylor series of these functions will verify. Therefore,

$$\begin{aligned} E[e^{\alpha X_t}] &= E[e^{\alpha \sum_i Y_i}] \\ &= E[\prod_i e^{\alpha Y_i}] \\ &= E[(\prod_{i=1}^{t-1} e^{\alpha Y_i}) E[e^{\alpha Y_t} | X_0, \ldots, X_{t-1}]] \\ &\leq e^{\alpha^2/2} E[\prod_{i=1}^{t-1} e^{\alpha Y_i}] \\ &\leq e^{\alpha^2 t/2} \quad \text{(by induction)} . \end{aligned}$$

We may then apply the Markov inequality as follows:

$$\begin{aligned} P[X_t > \lambda] &= P[e^{\alpha X_t} > e^{\alpha \lambda}] \\ &< e^{-\alpha \lambda} E[e^{\alpha X_t}] \\ &\leq e^{-\alpha \lambda} e^{\alpha^2 t/2} \\ &= e^{-\lambda^2/2t} \end{aligned}$$

after letting $\alpha = \lambda/t$. □

It would be quite a coincidence if some sequence of random variables "happened" to be a martingale. It is then fitting that the two most common of applications of Azuma's inequality in random graph theory are to sequences of random variables which we *force* to be martingales. For some concreteness we will let $G(n,p)$ be the probability space in what follows.

The first process is the *edge exposure martingale*. We begin with a function whose domain is the set of all graphs (for instance, the maximum cardinality of a set of vertex disjoint triangles). We label the possible edges of our random graph $e_1, \dots, e_t$, where $t = \binom{n}{2}$. We then define the sequence of random variables $X_0, \dots, X_t$ by defining their evaluations on an arbitrary (fixed) graph H:

$$X_i(H) = E[f(G) | \, e_j \in G \iff e_j \in H \quad \text{for all } 1 \leq j \leq i].$$

It is helpful to think of $X_0(H), \dots, X_t(H)$ as a sequence of approximations to the true value of $f(H)$. In the beginning we have no information about the structure of H, so our best guess is the average value $E[f(G)] = X_0(H)$. We are slowly given more information; at time i we know which of the edges $e_1, \dots, e_i$ actually lie in H. Using this information, we slowly update our best guess as to the value of $f(H)$. This process ends at time t, when we have seen the entire graph, whereupon we conclude with the best guess possible, $X_t(H) = f(H)$.

It is worthy to note that the use of expectation (and not the use of edges) in the above definition forces our sequence of random variables to be a martingale (exercise: check it). In fact, speaking very roughly, any discrete information revealing technique can be substituted to give a new type of martingale. Take the *vertex exposure martingale*, for instance. In this process the value of $f(H)$ is approximated in $|V(G)| = n$ steps. At time i the vertex v_i is revealed, so that our guess $X_i(H)$ is made with the knowledge of the induced subgraph $G|_{[i]}$ having vertex set $\{v_1, \dots, v_i\}$. That is,

$$X_i(H) = E[f(G) \,|\, G|_{[i]} = H|_{[i]}].$$

The vertex exposure martingale always lends itself to a tighter concentration than the edge exposure, as $n \ll \binom{n}{2}$, though in general it may be harder to ensure that the corresponding sequence (X_i) is Lipschitz.

The first application of these martingales to $G(n,p)$ was the following result of Shamir and Spencer [4].

Theorem 2. *A random graph G in $G(n,p)$ has chromatic number $\chi(G)$ satisfying*

$$P[|\chi(G) - E[\chi(G)]| > \lambda\sqrt{n-1}] \leq 2e^{-\lambda^2/2}.$$

Proof. Simply apply Azuma's inequality to the vertex exposure martingale. The sequence is Lipschitz since the introduction of a new vertex can increase the chromatic number of a graph by at most 1. □

This method can be greatly refined in special cases. One of the most well known is due to Bollobás, which we state without proof.

Theorem 3. *Let $p(n) = n^{-\alpha}$ for $\frac{1}{2} < \alpha < 1$. Then a random graph G has chromatic number $\chi(G)$ satisfying*

$$P[u \leq \chi(G) \leq u + \left\lceil \frac{2\alpha+1}{2\alpha-1} \right\rceil] \to 1 \quad \text{as } n \to \infty.$$

This result is strongest when $\alpha > 5/6$, for which $\chi(G)$ concentrates on five values! See [3] for details.

References

[1] N. Alon and J. Spencer, *The Probabilistic Method*, Second Edition, John Wiley and Sons, New York (2000).
[2] K. Azuma, Weighted Sums of Certain Dependent Variables, *Tôhoku Math. J.* **3** (1967), 357-367.
[3] B. Bollobás, *Random Graphs*, Second Edition, Cambridge University Press, Cambridge (2001).
[4] E. Shamir and J. Spencer, Sharp Concentration of the chromatic number on random graphs $G_{n,p}$, *Combinatorica* **7** (1981).
[5] W. Hoeffding, Probability Inequalities for Sums of Bounded Random Variables, *J. Amer. Statist. Ass.* **58** (1963), 13-30.