Let $X_0, X_1, \ldots, X_n = X$ be the martingale given by exposing one coordinate of $\{0,1\}^n$ at a time. The Lipschitz condition holds for X: If y, y' differ in just one coordinate then $X(y) - X(y') \leq 1$. Thus, with $\mu = E[X]$

$$\Pr[X < \mu - \lambda\sqrt{n}] < e^{-\lambda^2/2} = \epsilon$$

$$\Pr[X > \mu + \lambda\sqrt{n}] < e^{-\lambda^2/2} = \epsilon$$

But

$$\Pr[X = 0] = |A|2^{-n} \geq \epsilon$$

so $\mu \leq \lambda\sqrt{n}$. Thus

$$\Pr[X > 2\lambda\sqrt{n}] < \epsilon$$

and

$$|B(A, 2\lambda\sqrt{n})| = 2^n \Pr[X \leq 2\lambda\sqrt{n}] \geq 2^n(1-\epsilon) \qquad \Box$$

Actually, a much stronger result is known. Let $B(s)$ denote the ball of radius s about $(0, \ldots, 0)$. The Isoperimetric Inequality proved by Harper in 1966 states that

$$|A| \geq |B(r)| \Rightarrow |B(A, s)| \geq |B(r+s)|$$

One may actually use this inequality as a beginning to give an alternate proof that $\chi(G) \sim n/2\log_2 n$ and to prove a number of the other results we have shown using martingales.

Deriving these asymptotic bounds from first principles is quite cumbersome.

As a second illustration let B be any normed space and let $v_1, \ldots, v_n \in B$ with all $|v_i| \leq 1$. Let $\epsilon_1, \ldots, \epsilon_n$ be independent with

$$\Pr[\epsilon_i = +1] = \Pr[\epsilon_i = -1] = \frac{1}{2}$$

and set

$$X = |\epsilon_1 v_1 + \ldots + \epsilon_n v_n|$$

Theorem 5.2.

$$\Pr[X - E[X] > \lambda\sqrt{n}] < e^{-\lambda^2/2}$$

$$\Pr[X - E[X] < -\lambda\sqrt{n}] < e^{-\lambda^2/2}$$

Proof. Consider $\{-1, +1\}^n$ as the underlying probability space with all $(\epsilon_1, \ldots, \epsilon_n)$ equally likely. Then X is a random variable and we define a martingale $X_0, \ldots, X_n = X$ by exposing one ϵ_i at a time. The value of ϵ_i can only change X by two so direct application of Theorem 4.1 gives $|X_{i+1} - X_i| \leq 2$. But let ϵ, ϵ' be two n-tuples differing only in the i-th coordinate.

$$X_i(\epsilon) = \frac{1}{2}\left[X_{i+1}(\epsilon) + X_{i+1}(\epsilon')\right]$$

so that

$$|X_i(\epsilon) - X_{i+1}(\epsilon)| = \frac{1}{2}\left|X_{i+1}(\epsilon') - X_{i+1}(\epsilon)\right| \leq 1$$

Now apply Azuma's Inequality. □

For a third illustration let ρ be the Hamming metric on $\{0,1\}^n$. For $A \subseteq \{0,1\}^n$ let $B(A, s)$ denote the set of $y \in \{0,1\}^n$ so that $\rho(x, y) \leq s$ for some $x \in A$. ($A \subseteq B(A, s)$ as we may take $x = y$.)

Theorem 5.3. Let $\epsilon, \lambda > 0$ satisfy $e^{-\lambda^2/2} = \epsilon$. Then

$$|A| \geq \epsilon 2^n \Rightarrow |B(A, 2\lambda\sqrt{n})| \geq (1 - \epsilon)2^n$$

Proof. Consider $\{0,1\}^n$ as the underlying probability space, all points equally likely. For $y \in \{0,1\}^n$ set

$$X(y) = \min_{x \in A} \rho(x, y)$$

where $w_{h'}$ is the conditional probability that $g = h'$ given that $g = h$ on B_{i+1}. For each $h' \in H$ let $H[h']$ denote the family of h^* which agree with h' on all points except (possibly) $B_{i+1} - B_i$. The $H[h']$ partition the family of h^* agreeing with h on B_i. Thus we may express

$$X_i(h) = \sum_{h' \in H} \sum_{h^* \in H[h']} [L(h^*)q_{h^*}]w_{h'}$$

where q_{h^*} is the conditional probability that g agrees with h^* on B_{i+1} given that it agrees with h on B_i. (This is because for $h^* \in H[h']$ $w_{h'}$ is also the conditional probability that $g = h^*$ given that $g = h^*$ on B_{i+1}.) Thus

$$\begin{aligned} |X_{i+1}(h) - X_i(h)| &= \left| \textstyle\sum_{h' \in H} w_{h'}[L(h') - \sum_{h^* \in H[h']} L(h^*)q_{h^*}] \right| \\ &\leq \textstyle\sum_{h' \in H} w_{h'} \sum_{h^* \in H[h']} |q_{h^*}[L(h') - L(h^*)]| \end{aligned}$$

The Lipschitz condition gives $|L(h') - L(h^*)| \leq 1$ so

$$|X_{i+1}(h) - X_i(h)| \leq \sum_{h' \in H} w_{h'} \sum_{h^* \in H[h']} q_{h^*} = \sum_{h' \in H} w_{h'} = 1 \qquad \Box$$

Now we can express Azuma's Inequality in a general form.
Theorem 4.2. Let L satisfy the Lipschitz condition relative to a gradation of length m and let $\mu = E[L(g)]$. Then for all $\lambda > 0$

$$\Pr[L(g) > \mu + \lambda\sqrt{m}] < e^{-\lambda^2/2}$$

$$\Pr[L(g) < \mu - \lambda\sqrt{m}] < e^{-\lambda^2/2}$$

5 Three Illustrations

Let g be the random function from $\{1, \ldots, n\}$ to itself, all n^n possible function equally likely. Let $L(g)$ be the number of values not hit, i.e., the number of y for which $g(x) = y$ has no solution. By Linearity of Expectation

$$E[L(g)] = n\left(1 - \frac{1}{n}\right)^n \sim \frac{n}{e}$$

Set $B_i = \{1, \ldots, i\}$. L satisfies the Lipschitz condition relative to this gradation since changing the value of $g(i)$ can change $L(g)$ by at most one. Thus
Theorem 5.1.

$$\Pr[|L(g) - \frac{n}{e}| > \lambda\sqrt{n}] < 2e^{-\lambda^2/2}$$

and ϵ was arbitrarily small. □

Using the same technique similar results can be achieved for other values of α. For any fixed $\alpha > \frac{1}{2}$ one finds that $\chi(G)$ is concentrated on some fixed number of values.

4 A General Setting

The martingales useful in studying Random Graphs generally can be placed in the following general setting which is essentially the one considered in Maurey [1979] and in Milman and Schechtman [1986]. Let $\Omega = A^B$ denote the set of functions $g : B \to A$. (With B the set of pairs of vertices on n vertices and $A = \{0, 1\}$ we may identify $g \in A^B$ with a graph on n vertices.) We define a measure by giving values p_{ab} and setting

$$\Pr[g(b) = a] = p_{ab}$$

with the values $g(b)$ assumed mutually independent. (In $G(n, p)$ all $p_{1b} = p, p_{0b} = 1 - p$.) Now fix a gradation

$$\emptyset = B_0 \subset B_1 \subset \ldots \subset B_m = B$$

Let $L : A^B \to R$ be a functional. (E.g., clique number.) We define a martingale $X_0, X_1, \ldots, X_m$ by setting

$$X_i(h) = E[L(g)|g(b) = h(b) \text{ for all } b \in B_i]$$

X_0 is a constant, the expected value of L of the random g. X_m is L itself. The values $X_i(g)$ approach $L(g)$ as the values of $g(b)$ are "exposed". We say the functional L satisfies the Lipschitz condition relative to the gradation if for all $0 \leq i < m$

$$h, h' \text{ differ only on } B_{i+1} - B_i \Rightarrow |L(h') - L(h)| \leq 1$$

Theorem 4.1. Let L satisfy the Lipschitz condition. Then the corresponding martingale satisfies

$$|X_{i+1}(h) - X_i(h)| \leq 1$$

for all $0 \leq i < m$, $h \in A^B$.

Proof. Let H be the family of h' which agree with h on B_{i+1}. Then

$$X_{i+1}(h) = \sum_{h' \in H} L(h') w_{h'}$$

if T has t vertices it must have at least $\frac{3t}{2}$ edges. The probability of this occuring for some T with at most $c\sqrt{n}$ vertices is bounded from above by

$$\sum_{t=4}^{c\sqrt{n}} \binom{n}{t} \binom{\binom{t}{2}}{\frac{3t}{2}} p^{3t/2}$$

We bound

$$\binom{n}{t} \leq (\frac{ne}{t})^t \text{ and } \binom{\binom{t}{2}}{\frac{3t}{2}} \leq (\frac{te}{3})^{3t/2}$$

so each term is at most

$$\left[\frac{ne}{t}\frac{t^{3/2}e^{3/2}}{3^{3/2}}n^{-3\alpha/2}\right]^t \leq \left[c_1 n^{1-\frac{3\alpha}{2}} t^{1/2}\right]^t \leq \left[c_2 n^{1-\frac{3\alpha}{2}} n^{1/4}\right]^t = \left[c_2 n^{-\epsilon}\right]^t$$

with $\epsilon = \frac{3\alpha}{2} - \frac{5}{4} > 0$ and the sum is therefore $o(1)$.

Proof of Theorem 3.3. Let $\epsilon > 0$ be arbitrarily small and let $u = u(n, p, \epsilon)$ be the least integer so that

$$\Pr[\chi(G) \leq u] > \epsilon$$

Now define $Y(G)$ to be the minimal size of a set of vertices S for which $G-S$ may be u-colored. This Y satisfies the vertex Lipschitz condition since at worst one could add a vertex to S. Apply the vertex exposure martingale on $G(n,p)$ to Y. Letting $\mu = E[Y]$

$$\Pr[Y \leq \mu - \lambda\sqrt{n-1}] < e^{-\lambda^2/2}$$

$$\Pr[Y \leq \mu + \lambda\sqrt{n-1}] < e^{-\lambda^2/2}$$

Let λ satisfy $e^{-\lambda^2/2} = \epsilon$ so that these tail events each have probability less than ϵ. We defined u so that with probability at least ϵ G would be u-colorable and hence $Y = 0$. That is, $\Pr[Y = 0] > \epsilon$. The first inequality therefore forces $c \leq \lambda\sqrt{n-1}$. Now employing the second inequality

$$\Pr[Y \geq 2\lambda\sqrt{n-1}] \leq \Pr[Y \geq \mu + \lambda\sqrt{n-1}] \leq \epsilon$$

With probability at least $1-\epsilon$ there is a u-coloring of all but at most $c'\sqrt{n}$ vertices. By the Lemma almost always, and so with probability at least $1-\epsilon$, these points may be colored with 3 further colors, giving a $u+3$-coloring of G. The minimality of u guarantees that with probability at least $1-\epsilon$ at least u colors are needed for G. Altogether

$$\Pr[u \leq \chi(G) \leq u+3] \geq 1 - 3\epsilon$$

Delete from $\mathcal{C}$ one set from each such pair $\{A, B\}$. This yields a set $\mathcal{C}^*$ of edge disjoint k-cliques of G and

$$E[Y] \geq E[|\mathcal{C}^*|] \geq E[|\mathcal{C}|] - E[W'] = \mu q - \Delta q^2/2 = \mu^2/2\Delta \sim n^2/2k^4$$

where we choose $q = \mu/\Delta$ (noting that it is less than one!) to minimize the quadratic. □

We conjecture that Lemma 3.1 may be improved to $E[Y] > cn^2/k^2$. That is, with positive probability there is a family of k-cliques which are edge disjoint and cover a positive proportion of the edges.
Theorem 3.2.

$$\Pr[\omega(G) < k] < e^{-(c+o(1))\frac{n^2}{\ln^8 n}}$$

with c a positive constant.
Proof. Let $Y_0, \ldots, Y_m$, $m = \binom{n}{2}$, be the edge exposure martingale on $G(n, 1/2)$ with the function Y just defined. The function Y satisfies the edge Lipschitz condition as adding a single edge can only add at most one clique to a family of edge disjoint cliques. (Note that the Lipschitz condition would not be satisfied for the number of k-cliques as a single edge might yield many new cliques.) G has no k-clique if and only if $Y = 0$. Apply Azuma's Inequality with $m = \binom{n}{2} \sim n^2/2$ and $E[Y] \geq \frac{n^2}{2k^4}(1 + o(1))$. Then

$$\begin{aligned} \Pr[\omega(G) < k] &= \Pr[Y = 0] & &\leq \Pr[Y - E[Y] \leq -E[Y]] \\ &\leq e^{-E[Y]^2/2\binom{n}{2}} & &\leq e^{-(c'+o(1))n^2/k^8} \\ &= e^{-(c+o(1))n^2/\ln^8 n} & & \end{aligned}$$

as desired. □

Here is another example where the martingale approach requires an inventive choice of graphtheoretic function.
Theorem 3.3. Let $p = n^{-\alpha}$ where α is fixed, $\alpha > \frac{5}{6}$. Let $G = G(n, p)$. Then there exists $u = u(n, p)$ so that almost always

$$u \leq \chi(G) \leq u + 3$$

That is, $\chi(G)$ is concentrated in four values.

We first require a technical lemma that had been well known.
Lemma 3.4. Let α, c be fixed $\alpha > \frac{5}{6}$. Let $p = n^{-\alpha}$. Then almost always every $c\sqrt{n}$ vertices of $G = G(n, p)$ may be 3-colored.
Proof. If not, let T be a minimal set which is not 3-colorable. As $T - \{x\}$ is 3-colorable, x must have internal degree at least 3 in T for all $x \in T$. Thus

Theorem 2.4 (Shamir, Spencer[1987]) Let n, p be arbitrary and let $c = E[\chi(G)]$ where $G \sim G(n,p)$. Then

$$\Pr[|\chi(G) - c| > \lambda\sqrt{n-1}] < 2e^{-\lambda^2/2}$$

Proof. Consider the vertex exposure martingale $X_1, \ldots, X_n$ on $G(n,p)$ with $f(G) = \chi(G)$. A single vertex can always be given a new color so the vertex Lipschitz condition applies. Now apply Azuma's Inequality. □

Letting $\lambda \to \infty$ arbitrarily slowly this result shows that the distribution of $\chi(G)$ is "tightly concentrated" around its mean. The proof gives no clue as to where the mean is.

3 Chromatic Number

We have previously shown that $\chi(G) \sim n/2\log_2 n$ almost surely, where $G \sim G(n, 1/2)$. Here we give the original proof of Béla Bollobás using martingales. We follow the earlier notations setting $f(k) = \binom{n}{k}2^{-\binom{k}{2}}$, k_0 so that $f(k_0 - 1) > 1 > f(k_0)$, $k = k_0 - 4$ so that $k \sim 2\log_2 n$ and $f(k) > n^{3+o(1)}$. Our goal is to show

$$\Pr[\omega(G) < k] = e^{-n^{2+o(1)}},$$

where $\omega(G)$ is the size of the maximum clique of G. We shall actually show in Theorem 3.2 a more precise bound. The remainder of the argument is as given earlier.

Let $Y = Y(H)$ be the maximal size of a family of edge disjoint cliques of size k in H. This ingenious and unusual choice of function is key to the martingale proof.

Lemma 3.1. $E[Y] \geq \frac{n^2}{2k^4}(1 + o(1))$

Proof. Let $\mathcal{K}$ denote the family of k-cliques of G so that $f(k) = \mu = E[|\mathcal{K}|]$. Let W denote the number of unordered pairs $\{A, B\}$ of k-cliques of G with $2 \leq |A \cap B| < k$. Then $E[W] = \Delta/2$, with Δ as described earlier, $\Delta \sim \mu^2 k^4 n^{-2}$. Let $\mathcal{C}$ be a random subfamily of $\mathcal{K}$ defined by setting, for each $A \in \mathcal{K}$,

$$\Pr[A \in \mathcal{C}] = q,$$

q to be determined. Let W' be the number of unordered pairs $\{A, B\}$, $A, B \in \mathcal{C}$ with $2 \leq |A \cap B| < k$. Then

$$E[W'] = E[W]q^2 = \Delta q^2/2$$

Note that $X_1(H) = E[f(G)]$ is constant as no edges have been exposed and $X_n(H) = f(H)$ as all edges have been exposed.

2 Large Deviations

Maurey [1979] applied a large deviation inequality for martingales to prove an isoperimetric inequality for the symmetric group S_n. This inequality was useful in the study of normed spaces; see Milman and Schechtman [1986] for many related results. The applications of martingales in Graph Theory also all involve the same underlying martingale results used by Maurey, which are the following.
Theorem 2.1 (Azuma's Ineqality) Let $0 = X_0, \ldots, X_m$ be a martingale with

$$|X_{i+1} - X_i| \leq 1$$

for all $0 \leq i < m$. Let $\lambda > 0$ be arbitrary. Then

$$\Pr[X_m > \lambda\sqrt{m}] < e^{-\lambda^2/2}$$

Corollary 2.2 Let $c = X_0, \ldots, X_m$ be a martingale with

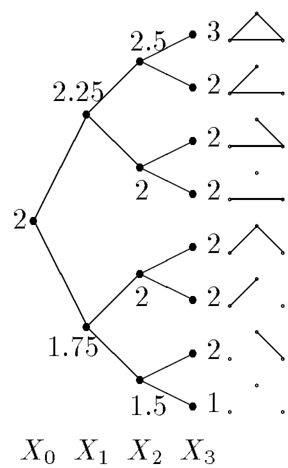
$$|X_{i+1} - X_i| \leq 1$$

for all $0 \leq i < m$. Then

$$\Pr[|X_m - c| > \lambda\sqrt{m}] < 2e^{-\lambda^2/2}.$$

A graph theoretic function f is said to satisfy the *edge Lipschitz condition* if whenever H and H' differ in only one edge then $|f(H) - f(H')| \leq 1$. It satisfies the *vertex Lipschitz condition* if whenever H and H' differ at only one vertex $|f(H) - f(H')| \leq 1$.
Theorem 2.3 When f satisfies the edge Lipschitz condition the corresponding edge exposure martingale satisfies $|X_{i+1} - X_i| \leq 1$. When f satisfies the vertex Lipschitz condition the corresponding vertex exposure martingale satisfies $|X_{i+1} - X_i| \leq 1$.

We prove these results in a more general context later. They have the intuitive sense that if knowledge of a particular vertex or edge cannot change f by more than one then exposing a vertex or edge should not change the expectation of f by more than one. Now we give a simple application of these results.



The edge exposure martingale with $n = m = 3$, f the chromatic number, and the edges exposed in the order "bottom,left,right". The values $X_i(H)$ are given by tracing from the central node to the leaf labelled H.

The figure shows why this is a martingale. The conditional expectation of $f(H)$ knowing the first $i-1$ edges is the weighted average of the conditional expectations of $f(H)$ where the i-th edge has been exposed. More generally - in what is sometimes referred to as a Doob martingale process - X_i may be the conditional expectation of $f(H)$ after certain information is revealed as long as the information known at time i includes the information known at time $i-1$.

The Vertex Exposure Martingale. Again let $G(n,p)$ be the underlying probability space and f any graphtheoretic function. Define $X_1,\ldots,X_n$ by

$$X_i(H) = E[f(G)|\text{for } x,y \leq i, \{x,y\} \in G \longleftrightarrow \{x,y\} \in H]$$

In words, to find $X_i(H)$ we expose the first i vertices and all their internal edges and take the conditional expectation of $f(G)$ with that partial information. By ordering the edges appropriately the vertex exposure martingale may be considered a subsequence of the edge exposure martingale.

Lecture 8: Martingales

1 Definitions

A martingale is a sequence $X_0, \ldots, X_m$ of random variables so that for $0 \leq i < m$,

$$E[X_{i+1}|X_i] = X_i$$

The Edge Exposure Martingale Let the random graph $G(n,p)$ be the underlying probability space. Label the potential edges $\{i,j\} \subseteq [n]$ by $e_1, \ldots, e_m$, setting $m = \binom{n}{2}$ for convenience, in any specific manner. Let f be any graphtheoretic function. We define a martingale $X_0, \ldots, X_m$ by giving the values $X_i(H)$. $X_m(H)$ is simply $f(H)$. $X_0(H)$ is the expected value of $f(G)$ with $G \sim G(n,p)$. Note that X_0 is a constant. In general (including the cases $i = 0$ and $i = m$)

$$X_i(H) = E[f(G)|e_j \in G \longleftrightarrow e_j \in H, 1 \leq j \leq i]$$

In words, to find $X_i(H)$ we first expose the first i pairs $e_1, \ldots, e_i$ and see if they are in H. The remaining edges are not seen and considered to be random. $X_i(H)$ is then the conditional expectation of $f(G)$ with this partial information. When $i = 0$ nothing is exposed and X_0 is a constant. When $i = m$ all is exposed and X_m is the function f. The martingale moves from no information to full information in small steps.