Martingales II: Applications

Will Rosenbaum
Updated: December 18, 2012
Department of Mathematics
University of California, Los Angeles

1 General Setting

In this essay, we consider applications of Azuma's inequality to combinatorics and theoretical computer science. First, recall

See [1].

Theorem 1. (*Azuma's Inequality*) Suppose $X_0, X_1, \ldots, X_m$ is a Martingale that satisfies

$$|X_i - X_{i-1}| \leq c_i \quad \text{for} \quad i = 1, 2, \ldots, m.$$

Then for every $\lambda > 0$,

$$\mathbf{P}(X_m - X_0 > \lambda) < e^{-\lambda^2/2C} \quad \text{where} \quad C = \sum_{i=1}^{m} c_i^2.$$

In particular, if we have $c_i = 1$ for all i,

$$\mathbf{P}(X_m - X_0 > \lambda\sqrt{m}) < e^{-\lambda^2/2}.$$

We will apply Azuma's inequality primarily in the following setting. Let A and B be finite sets and let $\Omega = A^B$, the set of maps $g : B \longrightarrow A$. We define a measure on Ω by assigning probabilities

$$p_{ab} = \mathbf{P}(g(b) = a)$$

where the values of $g(b)$ are mutually independent. Consider an increasing sequence of subsets of B:

$$\varnothing = B_0 \subset B_1 \subset \cdots \subset B_m = B.$$

Given any random variable $L : \Omega \longrightarrow \mathbf{R}$, we define a Martingale $X_0, X_1, \ldots, X_m$ by

$$X_i(h) = \mathbf{E}\left(L(G) \,|\, g(b) = h(b) \quad \text{for} \quad b \in B_i\right).$$

We say that L satisfies the ***Lipschitz condition*** if there exist constants c_i, $i = 1, 2, \ldots, m$ such that if h and h' only differ only on $B_i \setminus B_{i-1}$ then

$$|L(h') - L(h)| \leq c_i.$$

Proposition 2. Suppose L satisfies the Lipschitz condition for $c_1, c_2, \ldots, c_m$. Then the corresponding Martingale satisfies

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

for all $1 = 1, 2, \ldots, m$ and $h \in A^B$.

Proof. For any $h \in \Omega$, let $H = \{h' \in \Omega \mid h'(b) = h(b) \text{ for all } b \in B_i\}$. Then

$$X_i(h) = \sum_{h' \in H} L(h') w_{h'} \quad \text{where} \quad w_{h'} = \mathbf{P}\left(g = h' \mid g = h \text{ on } B_i\right).$$

For $h' \in H$, let $H(h')$ be the family of h^* that agree on h' on all points except possibly $B_i \setminus B_{i-1}$. Notice that the sets $H(h')$ form a partition of the set of all h^*. Therefore,

$$X_{i-1}(h) = \sum_{h' \in H} \sum_{h^* \in H[h']} (L(h^*) q_{h^*}) w_{h'},$$

where $q_{h^*} = \mathbf{P}(g = h^* \text{ on } B_i | g = h \text{ on } B_{i-1})$. Hence, we compute

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

□

Now we can restate Azuma's inequality in this setting:

Theorem 3. Suppose L satisfies the Lipschitz condition for $c_1, c_2, \ldots, c_m$ and let $\mu = \mathbf{E}(L(g))$. Then for all $\lambda > 0$

$$\mathbf{P}(L(g) \geq \mu + \lambda) < e^{-\lambda^2/2C}$$
$$\mathbf{P}(L(g) \leq \mu - \lambda) < e^{-\lambda^2/2C}$$

where $C = c_1^2 + c_2^2 + \cdots + c_m^2$.

2 Random Graphs

Definition 4. We define $G(n,p) = (\Omega, \mathbf{P})$ to be the probability space where

See [1].

$$\Omega = \{G = (V, E) \mid |V| = n\}$$

with $\mathbf{P}$ determined by

$$\mathbf{P}(\{i, j\} \in E) = p$$

for all $i, j \in 1, 2, \ldots, n$ and $i \neq j$, where these events are mutually independent. In general, p is a function of n.

Remark 5. We can place $G(n,p)$ setting of the previous section. Let $B = \left\{1, 2, \ldots, \binom{n}{2}\right\}$ and $A = \{0, 1\}$. We can identify B with potential edges in a graph G sampled from $G(n,p)$ and we have $h(j) = 1$ if and only if the j-th edge (in some arbitrary ordering of edges) appears in G. That is, $p_{1j} = p$ and $p_{0j} = 1 - p$.

The following processes give rise to Martingales on $G(n,p)$. Let f be a graph theoretic function.

Example 6. (*vertex exposure*). Identify the vertices of G with $[n] = \{1, 2, \ldots, n\}$. Then define the Martingale $X_0, X_2, \ldots, X_n$ by

$$X_i(H) = \mathbf{E}\left(f(G) \,|\, \text{ for } x, y \leq i, \ \{x,y\} \in G \iff \{x,y\} \in H\right).$$

That is, we look at the first i vertices in H and see compute the expectation of $f(G)$ conditioned on the induced subgraph of H consisting of the first i vertices. Notice that $H_0(H) = \mathbf{E}(f(G))$ and $H_n(H) = f(H)$.

Example 7. (*edge exposure*). Label the pairs $\{i,j\} \subset [n]$ by $e_1, e_2, \ldots, e_m$ where $m = \binom{n}{2}$. Then we define the Martingale

$$X_i(H) = \mathbf{E}\left(f(G) \,|\, e_j \in G \iff e_j \in H, \ 1 \leq j \leq i\right).$$

In this case, we inspect each pair $\{i,j\} \subset [n]$ and see if H contains $e = \{i,j\}$ as an edge. As before $X_0(H) = \mathbf{E}(f)$ and $X_m(H) = f(H)$.

Definition 8. Let $G = (V, E)$ be a graph. Then a ***proper*** k***-coloring*** of G is a map $\psi : V \to [k]$ such that $\{u, v\} \in E$ implies $\psi(u) \neq \psi(v)$. That is, a proper k-coloring is an assignment of "colors" $\{1, 2, \ldots, k\}$ to the vertices of G such that adjacent vertices are assigned different colors. The ***chromatic number*** of G denoted $\chi(G)$ is the smallest number k such that G admits a proper k-coloring.

Theorem 9. (*Shamir and Spencer, 1987*). Let n and p be arbitrary and let $c = \mathbf{E}(\chi(G))$, where $G \sim G(n,p)$. Then

$$\mathbf{P}(|\chi(G) - c| > \lambda\sqrt{n-1}) < 2e^{-\lambda^2/2}.$$

Proof. Consider the vertex exposure Martingale on $G(n,p)$ with $f(G) = \chi(G)$. Adding a single vertex to a graph can increase the chromatic number by at most 1, so this Martingale satisfies the Lipschitz condition with $c_i = 1$ for all i. Thus the theorem follows from theorem 3. □

Remark 10. Computing $\chi(G)$ for random graphs is tricky business. However, it is known that $\chi(G) \sim n/2\log_2 n$ with high probability when $G \sim G(n, 1/2)$. In fact, this was proven by Bollobás using the Martingale method, but the proof is too technical to present here. See, for example, Alon and Spencer's book for details.

In the case of the previous theorem, the vertex exposure Martingale gives fairly tight bounds on concentration of chromatic number around its expected value. In the case of relatively sparse graphs, however, we can do much better.

Theorem 11. Let $p = n^{-\alpha}$ for $\alpha > \frac{5}{6}$ fixed. Let $G \sim G(n,p)$. Then there exists $u = u(n,p)$ such that with high probability

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

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

Lemma 12. Let α, c be fixed, $\alpha > 5/6$. Let $p = n^{-\alpha}$. Then with high probability, every $c\sqrt{n}$ vertices of $G \sim G(n,p)$ may be 3-colored.

Proof. Suppose T is a minimal set that is not 3-colorable. By the minimality of T, for each $x \in T$, $T \setminus \{x\}$ is 3-colorable. Therefore, x must have internal degree at least 3 (i.e., x has 3 neighbors in T). Thus, if T has t vertices, then it must have at least $3t/2$ edges. We can bound the probability G containing a set of t vertices ($t \leq c\sqrt{n}$) and $3t/2$ edges from above by

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

Standard bounds for binomial coefficients give

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

Therefore, each term in the sum is at most

$$\begin{aligned} \left(\frac{ne}{t} \frac{(te)^{3/2}}{3^{3/2}} n^{-3\alpha/2}\right)^t &\leq (c_1 n^{1-3\alpha/2} t^{1/2})^t \\ &\leq (c_2 n^{1-3\alpha/2} n^{1/4})^t \\ &= (c_2 n^{-\varepsilon})^t \end{aligned}$$

where $\varepsilon = \frac{3}{2}\alpha - \frac{5}{4} > 0$. Thus, the sum is $o(1)$, giving the desired result. □

Proof. (*of theorem 11*). Given ε arbitrarily small, let $u = u(n, p, \varepsilon)$ be the least integer such that

$$\mathbf{P}(\chi(G) \leq u) > \varepsilon.$$

Define $f(G)$ to be the minimal size of a set of S vertices such that $G \setminus S$ can be u-colored. Notice that f satisfies the vertex Lipschitz condition with $c_i = 1$ for all i, as adding a vertex to the graph could increase the size of S by at most 1. Applying the vertex exposure Martingale on $G(n, p)$ and f, we obtain

$$\begin{aligned} \mathbf{P}(f \leq \mu - \lambda\sqrt{n-1}) &< e^{-\lambda^2/2}, \\ \mathbf{P}(f \geq \mu + \lambda\sqrt{n-1}) &< e^{-\lambda^2/2}. \end{aligned}$$

Choose λ so that $e^{-\lambda^2/2} < \varepsilon$, so that all of the events have probability less than ε. By the definition of u, G is u colorable with probability at least ε, that is $\mathbf{P}(f = 0) > \varepsilon$. Thus, $\mu \leq \lambda\sqrt{n-1}$. By the second inequality,

$$\mathbf{P}(f \geq 2\lambda\sqrt{n-1}) \leq \mathbf{P}(f \geq \mu + \lambda\sqrt{n-1}) \leq \varepsilon.$$

With probability at least $1 - \varepsilon$, there is a u-coloring of all but at most $c'\sqrt{n}$ vertices. By the lemma (taking n sufficiently large) at least $1 - \varepsilon$, the remaining points may be colored with 3 further colors, giving a $(u + 3)$-coloring of G. The minimality of u implies that at least u colors are needed to color G. Putting everything together, we have

$$\mathbf{P}(u \leq \chi(G) \leq u + 3) \geq 1 - 3\varepsilon$$

where ε was arbitrarily small. □

Remark 13. The random graph $G(n, c/n)$ has been of particular historical interest, with some of the first results obtained by Erdös and Rényi. In particular, they demonstrated a phase transition around $c = 1$. For $c < 1$, the connected components of $G(n, c/n)$ typically have size $\mathcal{O}(\log n)$, while for $c > 1$ graphs typically have a unique "large" connected component of size linear in n. A recent paper of Krivelevich and Sudakov gives a particularly elegant proof of this phenomenon. In the critical case $c = 1$, typical graphs have a largest connected component of size $\theta(n^{2/3})$. A recent paper of Nachmias and Peres proves bounds for the critical case using Martingale methods.

3 Balls and Bins

Imagine you have n buckets labelled $1, 2, \ldots, n$. You throw m balls into the buckets, choosing a bucket uniformly at random for each each ball. Let $f : [m] \to [n]$ be the function assigning each ball to its bin. Let F be the number of empty buckets after throwing all m balls. We can easily compute the expected value of F: See [4].

$$\mathbf{E}(F) = n\left(1 - \frac{1}{n}\right)^m \approx ne^{-m/n}$$

Let $B_i = \{1, 2, \ldots, i\}$ for $i = 0, 1, \ldots, m$ and define the Martingale

$$X_i = \mathbf{E}\left(F \mid f(j) \text{ for all } j \in B_i\right).$$

Notice that F satisfies the Lipschitz condition with $c_i = 1$ for all i. Indeed, changing the placement of the i-th ball can change the number of unoccupied bins by at most 1. Therefore, we may apply theorem 3 to obtain

$$\mathbf{P}(|F - \mathbf{E}(F)| \geq \lambda\sqrt{m}) \leq 2e^{-\lambda^2/2}.$$

4 Sums of Vectors in a Normed Space

Let B be a normed space with $v_1, v_2, \ldots, v_n \in B$ with $|v_i| \leq 1$ for all i. Let $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ be i.i.d. with $\mathbf{P}(\varepsilon_i = 1) = \mathbf{P}(\varepsilon_i = -1) = 1/2$ for all i. Set See [1].

$$X = |\varepsilon_1 v_1 + \varepsilon_2 v_2 + \cdots + e_n v_n|.$$

Consider the underlying probability space $\{-1, 1\}^n$ with all $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ equally likely. Define the Martingale $\mathbf{E}(X) = X_0, X_1, \ldots, X_n = X$ by exposing one ε_i at a time. The value of each ε_i can only change X by at most 2, so $|X_i - X_{i-1}| \leq 2$ by proposition 2. However, if ε and ε' are n-tuples differing only only on the i-th coordinate,

$$X_{i-1}(\varepsilon) = (X_i(\varepsilon) + X_i(\varepsilon'))/2$$

hence,

$$|X_{i-1}(\varepsilon) - X_i(\varepsilon)| = |X_i(\varepsilon') - X_i(\varepsilon)|/2 \leq 1.$$

So we can apply theorem 3 to obtain

$$\mathbf{P}(X - \mathbf{E}(X) > \lambda\sqrt{n}) < e^{-\lambda^2/2}$$
$$\mathbf{P}(X - \mathbf{E}(X) < -\lambda\sqrt{n}) < e^{-\lambda^2/2}$$

5 Traveling Salesman

For n points $x_1, x_2, \ldots, x_n \in [0,1]^2$, let $T(x_1, x_2, \ldots, x_n)$ denote the shortest length of a circuit visiting all n points. That is

See [3].

$$T(x_1, x_2, \ldots, x_n) = \min_{\pi \in S_n} \sum_{i=1}^{n} \left| x_{\pi(i)} - x_{\pi(i+1)} \right|$$

where by convention $\pi(n+1) = \pi(1)$. We consider the case where the points x_i are chosen independently and uniformly at random from $[0,1]^2$. Let $X_1, X_2, \ldots, X_n$ be iid, distributed uniformly in $[0,1]^2$ and denote the random variable $T = T(X_1, X_2, \ldots, X_n)$. We will show that T is tightly concentrated around its mean.

For $y \in [0,1]^2$, define

$$Y_i = \min \left\{ |y - X_j| \mid j = i+1, i+2, \ldots, n \right\}.$$

That is, Y is the distance from y to $\{X_{i+1}, X_{i+2}, \ldots, X_n\}$. We claim that

$$\mathbf{P}(Y_i > t) \leq (1 - C_1 t^2)^{n-k-1} \leq e^{-C_1 t^2 (n-k-1)} \quad \text{for some} \quad C_1 > 0. \qquad (*)$$

To see this holds for $C_1 = \frac{1}{2}$, note that

$$\mu\left([0,1]^2 \cap \mathcal{B}_t(y)\right) \geq \frac{1}{2} t^2 \quad \text{for any} \quad y \in [0,1]^2$$

where μ is Lebesgue measure on $[0,1]^2$ and $\mathcal{B}_t(y)$ is the ball of radius t centered at y. From $(*)$, we can bound $\mathbf{E}(Y)$:

$$\mathbf{E}(Y_i) = \int_0^\infty \mathbf{P}(Y_i \geq t)\, dt \leq \int_0^\infty e^{-C_1 t^2 (n-k-1)}\, dt \leq \frac{C_2}{\sqrt{n-i}} \qquad (**)$$

for some $C_2 > 0$.

We are now ready to construct our Martingale. Predictably, we choose the Martingale

$$Z_i = \mathbf{E}\left(T \mid X_1, X_2, \ldots, X_i\right) \quad \text{for} \quad i = 0, 1, \ldots, n.$$

That is, we reveal the points X_i one-by-one until we have $Z_n = T$. By $(**)$, for any $x_i, x_i' \in [0,1]^2$

$$\left|\mathbf{E}\left(T \mid (X_1, \ldots, X_i) = (x_1, \ldots, x_i)\right) - \mathbf{E}\left(T \mid (X_1, \ldots, X_i) = (x_1, \ldots, x_i')\right)\right| \leq \frac{2C_2}{\sqrt{n-i}}.$$

Therefore, the Lipschitz condition applies with $c_i = 2C_2/\sqrt{n-i}$. Since

$$\sum_{i=0}^{n-1} \frac{4C_2^2}{n-i} \leq \frac{1}{2} c \ln n \quad \text{for some} \quad c > 0.$$

Thus, by Azuma's inequality, we have

$$\mathbf{P}(|T - \mathbf{E}(T)| \geq \lambda) \leq 2e^{-\lambda^2 / c \ln n}.$$

6 The Symmetric Group

Let S_n be the symmetric group on n elements endowed with the metric

See [2].

$$d(\sigma, \pi) = \frac{1}{n} \left|\{i \,|\, \sigma(i) \neq \pi(i)\}\right|$$

and uniform measure. For each $i = 1, 2, \ldots, n$, consider the sets

$$A_{j_1, j_2, \ldots, j_i} = \{\sigma \in S_n \,|\, \sigma(1) = j_1, \ldots, \sigma(i) = j_i\}.$$

The sets of the form $A_{j_1, j_2, \ldots, j_i}$ form a partition $\mathcal{X}^i$ of S_n. Let $A = A_{j_1, j_2, \ldots, j_i}$, $B = A_{j_1, \ldots, j_i, p}$, and $C = A_{j_1, \ldots, j_i, q}$ so that $B, C \subset A$. Let τ be the transposition that interchanges p and q. Define the bijection $\phi : B \longrightarrow C$ by $\phi(\sigma) = \tau \circ \sigma$. Then $d(\sigma, \phi(\sigma)) \leq 2/n$, hence S_n has length

$$\ell = \sqrt{\sum_{i=1}^{n} \frac{4}{n^2}} = \frac{2}{\sqrt{n}}.$$

Therefore, we have

$$\mu\left(\left\{F \geq \int F \, d\mu + r\right\}\right) \leq e^{-nr^2/8}.$$

In particular, the concentration function satisfies

$$\alpha_{(S_n, d, \mu)}(r) \leq e^{-nr^2/32}.$$

7 Sources

[1] Noga Alon and Joel Spencer. *The Probabilistic Method*. John Wiley & Sons, Inc., Hoboken, New Jersey, third edition, 2008.

[2] Michel Ledoux. *The Concentration of Measure Phenomenon*. The American Mathematical Society, P. O. Box 6248 Providence, Rhode Island, 2001.

[3] Colin McDiarmid. On the method of bounded differences, 1989.

[4] Michael Mitzenmacher and Eli Upfal. *Probability and Computing*. Cambridge University Press, 32 Avenue of the Americas, New York, NY 10013, 2009.