8. Bounding Large Deviations

8.1 Chernoff's Inequality

- Many alternative forms of "Chernoff bounds" exist; we only consider a simple and transparent special case. (For more, see Alon & Spencer, Appendix A.)

- Generalisations: Hoeffding's Inequality, Bernstein's Inequality, McDiarmid's Inequality, Azuma's Inequality.

- **Theorem** 8.1 (Chernoff 1952). Let Y_i, $i = 1, \dots, n$, be mutually independent random variables with

$$\Pr(Y_i = +1) = \Pr(Y_i = -1) = \frac{1}{2}.$$

Then for any $a > 0$:

$$\Pr(S_n > a) < e^{-a^2/2n}.$$

Proof Let $\lambda > 0$ be an arbitrary parameter (to be optimised later). Then for $i = 1, \dots, n$:

$$E[e^{\lambda Y_i}] = \frac{1}{2}(e^{\lambda} + e^{-\lambda}) = \cosh(\lambda).$$

Observe that for all $\lambda > 0$, $\cosh(\lambda) < e^{\lambda^2/2}$. (Compare e.g. the Taylor series termwise.)

Consider then the random variable $e^{\lambda X_n} = \prod_i e^{\lambda Y_i}$. Since the Y_i are mutually independent:

$$E[e^{\lambda X_n}] = \prod_{i=1}^{n} E[e^{\lambda Y_i}] = (\cosh(\lambda))^n < e^{\lambda^2 n/2}.$$

Note that $X_n > a$ iff $e^{\lambda X_n} > e^{\lambda a}$. Thus by Markov's inequality:

$$\Pr(X_n > a) = \Pr(e^{\lambda X_n} > e^{\lambda a}) \leq E[e^{\lambda X_n}]/e^{\lambda a} < e^{\lambda^2 n/2 - \lambda a}.$$

Optimising the bound by choosing $\lambda = a/n$ yields
$\Pr(X_n > a) < e^{-a^2/2n}$, as desired. □

8.2 Azuma's Inequality

- A sequence of random variables $X_0, X_1, \dots, X_n$ is a martingale if for all $i = 0, \dots, n-1$:

$$E[X_{i+1} \mid X_i, X_{i-1}, \dots, X_0] = X_i.$$

- Example 1: X_i = a player's fortune after i games at a fair casino.

- Example 2: X_i as in Thm 8.1 (the "coin-flip martingale"), i.e.

$$\begin{cases} X_0 = 0, \\ X_i = Y_1 + \dots + Y_i, \end{cases}$$ where the Y_i are mutually indep. and $\Pr(Y_i = +1) = \Pr(Y_i = -1) = \frac{1}{2}$.

Then:

$$E[X_{i+1} \mid X_i, X_{i-1}, \dots, X_0] = E[X_{i+1} \mid X_i]$$

$$= E[X_i + Y_{i+1} \mid X_i] = X_i + \underbrace{E[Y_{i+1} \mid X_i]}_{0 \text{ by indep.}} = X_i.$$

- Theorem 8.2 (Azuma 1967) Let $0 = X_0, \dots, X_n$ be a martingale with $|X_{i+1} - X_i| \le 1$ for all $i = 0, \dots, n-1$. Then for any $a > 0$:

$$\Pr(X_n > a) < e^{-a^2/2n}.$$

Proof. Simple generalisation of the proof of Thm 8.1.

Set $Y_i = X_i - X_{i-1}$, so that $|Y_i| \le 1$, and

$$E[Y_i \mid X_{i-1}, \dots, X_0] = E[X_i \mid X_{i-1}, \dots, X_0] - E[X_{i-1} \mid X_{i-1}, \dots, X_0]$$

$$= X_{i-1} - X_{i-1} = 0.$$

Then, as in Thm 8.1, for any $\lambda > 0$:

$$E[e^{\lambda Y_i} \mid X_{i-1}, \dots, X_0] \overset{(*)}{\leq} \cosh(\lambda) < e^{\lambda^2/2}.$$

(*) By the fact that $E[Y_i]=0$, $|Y_i| \leq 1$, and the convexity of $f(y) = e^{\lambda y}$.

Hence

$$\begin{aligned} E[e^{\lambda X_n}] &= E\Big[\prod_{i=1}^{n} e^{\lambda Y_i}\Big] \\ &= E\Big[\Big(\prod_{i=1}^{n-1} e^{\lambda Y_i}\Big) \cdot E[e^{\lambda Y_n} \mid X_{n-1}, \dots, X_0]\Big] \\ &< E\Big[\prod_{i=1}^{n-1} e^{\lambda Y_i}\Big] \cdot e^{\lambda^2/2} \leq \dots \leq \\ &\leq e^{\lambda^2 n/2}. \end{aligned}$$

Therefore:

$$\begin{aligned} \Pr(X_n > a) &= \Pr(e^{\lambda X_n} > e^{\lambda a}) \\ &< e^{\lambda^2 n/2 - \lambda a}. \end{aligned}$$

Choosing $\lambda = a/n$ yields the desired result □

- Corollary 8.3 Let $0 = X_0, \dots, X_n$ be a martingale with $|X_{i+1} - X_i| \leq 1$ for all $i = 0, \dots, n-1$. Then for any $\lambda > 0$:

$$\Pr(X_n > \lambda\sqrt{n}) < e^{-\lambda^2/2}. \quad \square$$

- Corollary 8.4 Let $c = X_0, \dots, X_n$ be a martingale with $|X_{i+1} - X_i| \leq 1$ for all $i = 0, \dots, n-1$. Then for any $\lambda > 0$:

$$\Pr(|X_n - c| > \lambda\sqrt{n}) < 2e^{-\lambda^2/2}. \quad \square$$

- Example 3. The edge exposure martingale.

Let $f(G)$ be any graph-theoretic function (e.g. chromatic number, clique number, ...) Let's say we are interested in bounding the deviation of $f(G)$ from $E[f(G)]$ in the space of random graphs $\mathcal{G}(n,p)$.

We define a martingale describing this deviation as follows. Let the $m = \binom{n}{2}$ "potential" edges of a $\mathcal{G}(n,p)$ random graph be indexed in some order as $1,\dots,m$. For a given subgraph H on the first i vertices, we define

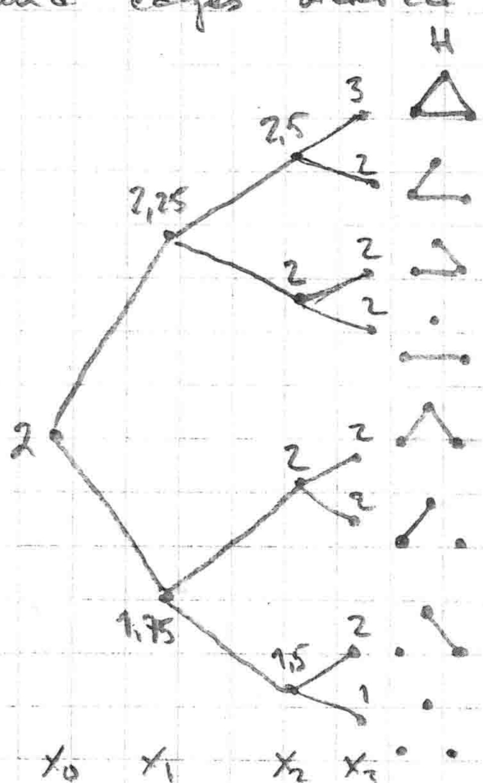
$$X_i(H) = E[f(G) \mid \text{edges } e_1,\dots,e_i \text{ in } G \text{ are fixed as in } H].$$

Thus, as special cases:

$$X_0(H) = E[f(G)], \quad G \sim \mathcal{G}(n,p), \quad \text{independent of } H;$$

$$X_m(H) = f(H), \quad \text{for any "fully exposed" graph } H.$$

The following figure illustrates this "edge exposure martingale" for $f \sim$ chromatic number, $n = 3$, $p = \frac{1}{2}$, and edges ordered "bottom, left, right":



[Check that the sequence $X_0,\dots,X_m$ indeed satisfies the martingale condition.]

- Example 4. The vertex exposure martingale.

  This is defined similarly as the edge exposure martingale, but based on "revealing" the vertices $1, \dots, n$ of a $G(n,p)$ graph and their internal edges in some given order.

- A graph-theoretic function f satisfies the edge (vertex) Lipschitz condition if whenever graphs H and H' differ in only one edge (vertex), then $|f(H) - f(H')| \leq 1$

- Lemma 8.5 If f satisfies the edge (vertex) Lipschitz condition, then the corresponding edge (vertex) exposure martingale satisfies $|X_{i+1} - X_i| \leq 1$.

  Proof. Exercise. $\square$

- As an illustration of the technique, consider the concentration of the chromatic number of $G(n,p)$ random graphs around its expected value $c = E[\chi(G)]$ (which we do not know).

- Theorem 8.6 (Shamir & Spencer 1987). Let n, p be arbitrary and $c = E[\chi(G)]$, where $G \sim G(n,p)$. Then

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

  Proof. Consider the vertex exposure martingale $X_1, \dots, X_n$ on $G(n,p)$ with $f(H) = \chi(H)$. A single vertex can always be assigned a new colour, so the vertex Lipschitz condition holds. Apply Azuma's Inequality in the form of Corollary 8.4. $\square$

  Note that for the v.-e. martingale, $E[\chi(G)] = X_1(H)$ for any H.