Concentration inequalities: martingales

Next we study concentration of functions other than sums. Let $X_1, \dots, X_n$ be independent random variables taking values in some space $\mathcal{X}$ and let $f : \mathcal{X}^n \to \mathbb{R}$. Let $Z = f(X_1, \dots, X_n)$. We are interested in bounds for $Z - \mathbb{E}Z$. We start with the variance. Define

$$\mathbb{E}_i(\cdot) = \mathbb{E}(\cdot \mid X_1, \dots, X_i)$$

and let

$$\Delta_i = \mathbb{E}_i Z - \mathbb{E}_{i-1} Z$$

Then $$Z - \mathbb{E}Z = \sum_{i=1}^n \Delta_i$$ (Doob martingale decomposition)

Since $$\mathbb{E}[\Delta_j \Delta_i] = \mathbb{E}\, \mathbb{E}_i[\Delta_j \Delta_i] = \mathbb{E}[\Delta_i \underbrace{\mathbb{E}_i \Delta_j}_{=0}] = 0,$$ ($j > i$)

$$\mathrm{var}(Z) = \mathbb{E}\left(\sum_{i=1}^n \Delta_i\right)^2 = \sum_{i=1}^n \mathbb{E}[\Delta_i^2]$$

By independence, $$\mathbb{E}_i\left[\mathbb{E}^{(i)} Z\right] = \mathbb{E}_{i-1} Z$$
($\mathbb{E}^{(i)}$: expectation w.r.t. X_i only)

and therefore

$$\Delta_i^2 = \left(\mathbb{E}_i\left[Z - \mathbb{E}^{(i)} Z\right]\right)^2 \le \mathbb{E}_i\left(Z - \mathbb{E}^{(i)} Z\right)^2$$
(Jensen)

so $$\mathbb{E}[\Delta_i^2] \le \mathbb{E}\,\mathrm{var}^{(i)}(Z)$$

$\mathrm{var}^{(i)}$: variance w.r.t. X_i, conditionally on $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$

We have obtained

Efron-Stein inequality If $X_1, \dots, X_n$ are independent and $Z = f(X_1, \dots, X_n)$, then

$$\operatorname{var}(Z) \le \mathbb{E} \sum_{i=1}^n \operatorname{var}^{(i)}(Z) = \frac{1}{2} \mathbb{E} \sum_{i=1}^n (Z - Z_i')^2$$

$Z_i' = f(X_1, \dots, X_i', \dots, X_n)$ where X_i' is an independent copy of X_i

Note that the inequality is tight for $f(x_1, \dots, x_n) = \sum_{i=1}^n x_i$.

Corollary: If $f : \mathcal{X}^n \to \mathbb{R}$ is such that

$$\left| f(x_1, \dots, x_n) - f(x_1, \dots, x_{i-1}, x_i', x_{i+1}, \dots, x_n) \right| \le c_i$$

("bounded differences" property)

and $X_1, \dots, X_n$ are independent and $Z = f(X)$, then

$$\operatorname{var}(Z) \le \frac{1}{2} \sum_{i=1}^n c_i^2$$

Example longest common subsequence: let $X_1, \dots, X_n, Y_1, \dots, Y_n$ be i.i.d. Bernoulli 1/2 and let Z be the longest common subsequence. One can show that

$\frac{\mathbb{E} Z}{n} \to c$ to some unknown constant c $\left(= \frac{2}{1+\sqrt{2}} ??\right)$

By Efron-Stein, $\operatorname{var}(Z) \le n/2$

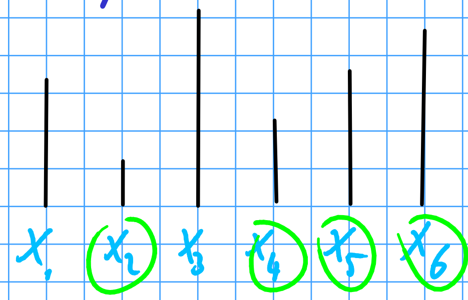
A variant: Since $\operatorname{var}(X) = \min_c \mathbb{E}(X-c)^2$, Efron-Stein implies that

$$\operatorname{var}(Z) \le \mathbb{E} \sum_{i=1}^n (Z - Z_i)^2$$

where $Z_i = f_i(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$

for any functions $f_i : \mathcal{X}^{n-1} \to \mathbb{R}$.

Example Longest increasing subsequence in a random permutation: let $X_1, \dots, X_n \sim \text{Unif}[0,1]$ and let Z denote the longest increasing subsequence. If Z_i' is the longest increasing subsequence with X_i deleted, then



$$0 \le Z - Z_i' \le 1$$

and
$$\sum_{i=1}^{n} (Z - Z_i') \le Z$$

so by Efron-Stein, $\text{var}(Z) \le \mathbb{E} Z$ ($\leftarrow \sim 2\sqrt{n}$)

Max-degree in $G(n,p)$

Largest eigenvalue of a random symmetric matrix

Let $\underline{A} = (X_{i,j})_{n \times n}$ be symmetric with $X_{i,j}$ independent $(i \le j)$, bounded: $|X_{i,j}| \le 1$. Let $Z = \sup_{u: \|u\|=1} u^T \underline{A} u$ be the largest eigenvalue. Let $\underline{A}_{i,j}$ be the same as $\underline{A}$ with $X_{i,j}$ replaced by an independent copy $X_{i,j}'$. Then

$$(Z - Z_{i,j}')_+ \le \left(v^T (\underline{A} - \underline{A}_{i,j}) v\right)_+ \le 2|v_i v_j| (X_{i,j} - X_{i,j}')_+ \le 4 |v_i v_j|$$

(v: eigenvector of A corresponding to the largest eigenvalue)

By Efron-Stein,

$$\text{var}(Z) \le \mathbb{E} \sum_{i \le j} (Z - Z_{i,j}')^2 \le 16 \sum_{i \le j} v_i^2 v_j^2 \le 16 \left(\sum_i v_i^2\right)^2 = 16,$$

independently on n !!!

An important application is for the adjacency matrix of $G(n,p)$ but it extends to the generalized random graph model as well.

Azuma's inequality

To obtain exponential concentration inequalities for functions of independent random variables, recall the Doob martingale decomposition

$$Z - \mathbb{E}Z = \sum_{i=1}^{n} \Delta_i \quad \text{with} \quad \Delta_i = \mathbb{E}_i Z - \mathbb{E}_{i-1} Z$$

Suppose that the martingale differences Δ_i are bounded: $|\Delta_i| \le c_i$ (such a condition holds if f has the bounded differences property)

functions of $X_1,\dots,X_{n-1}$

Then

$$\mathbb{E} e^{s(Z-\mathbb{E}Z)} = \mathbb{E}\left[e^{s\sum_{i=1}^{n}\Delta_i}\right]$$

$$= \mathbb{E}\,\mathbb{E}\left[e^{s\sum_{i=1}^{n}\Delta_i} \,\middle|\, X_1,\dots,X_{n-1}\right] = \mathbb{E}\left[e^{s\sum_{i=1}^{n-1}\Delta_i}\,\underbrace{\mathbb{E}\left[e^{s\Delta_n} \,\middle|\, X_1,\dots,X_{n-1}\right]}_{\le e^{s^2c_n^2/2} \text{ by Hoeffding's lemma}}\right]$$

$$\le e^{s^2c_n^2/2}\,\mathbb{E}\left[e^{s\sum_{i=1}^{n-1}\Delta_i}\right]$$

$$\le \dots \le e^{-s^2\sum_{i=1}^{n} c_n^2/2}$$

We obtain the Azuma-Hoeffding inequality for sums of bounded martingale differences. This implies the

Bounded differences inequality (or McDiarmid's inequality)

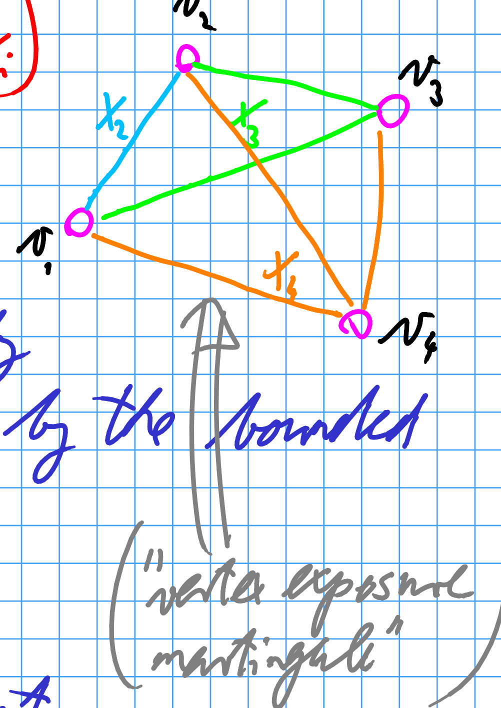
If $f : \mathcal{X}^n \to \mathbb{R}$ satisfies the bounded differences condition and $X_1,\dots,X_n$ are independent, then $Z = f(X_1,\dots,X_n)$ satisfies

$$\mathbb{P}(Z - \mathbb{E}Z > t) \le e^{-2t^2/\sum_{i=1}^{n} c_i^2}$$

As an example, consider the **chromatic number** of $\mathcal{G}(n,p)$. χ is the smallest number of colors needed to color the vertices of $\mathcal{G}(n,p)$ such that all pairs of adjacent vertices have different color.

Order the vertices $v_1, v_2, \ldots, v_n$ and let X_i be the collection

$$X_i = \left(\mathbb{1}_{x_1 \sim x_i}, \ldots, \mathbb{1}_{x_{i-1} \sim x_i}\right)$$



These are independent random variables that determine χ.
Also, changing X_i can change the value of χ by at most 1, so by Efron-Stein $\mathrm{var}(\chi) \le \frac{n-1}{2}$ and by the bounded differences inequality

$$\mathbb{P}\left(|\chi - \mathbb{E}\chi| > t\right) \le e^{-2t^2/n-1}$$

so random fluctuations are at most of the order of $\sqrt{n}$.

When p is small, in fact much more can be said.

Theorem Let $p = n^{-\alpha}$ for some $\alpha > 5/6$. Then there exists a value $u = u(n,p)$ such that w.h.p.,

$$u \le \chi \le u + 3.$$

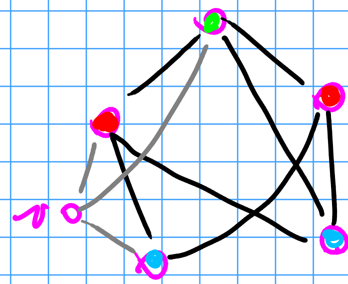
First we need a simple lemma.

Lemma Let c be fixed. Then w.h.p., every subgraph of $\mathcal{G}(n, n^{-\alpha})$ with at most $c\sqrt{n}$ vertices is 3-colorable.

Proof Suppose not and let T be a subgraph of minimal size which is not 3-colorable.

For any vertex v of T, $T - v$ is 3-colorable but T is not, so v must have degree at least 3 in T.

But then T has at least $3t/2$ edges where $t = |T|$.



$$\mathbb{P}\left(\exists \text{ subgraph } T \text{ with } \le c\sqrt{n} \text{ vertices and } \ge \tfrac{3t}{2} \text{ edges}\right)$$

$$\le \sum_{t=4}^{c\sqrt{n}} \binom{n}{t} \binom{\binom{t}{2}}{\frac{3t}{2}} p^{3t/2} \le \sum_{t=4}^{c\sqrt{n}} \left(\frac{ne}{t}\right)^t \left(\frac{te}{3}\right)^{3t/2} p^{3t/2}$$

$$\le \sum_{t=4}^{c\sqrt{n}} \left(c' \frac{n}{t} t^{3/2} \cdot p^{3/2}\right)^t = \sum_{t=4}^{c\sqrt{t}} \left(c' n^{5/4} n^{-3\alpha/2}\right)^t$$

$$\binom{t^2/2}{3t/2} \le \left(\frac{t^2 e}{3t}\right)^{3t/2}$$

$$\to 0$$

Proof of the theorem

Fix $\varepsilon > 0$ small and define u as the smallest positive integer such that $\mathbb{P}(\chi \le u) > \varepsilon$

Define Y as the size of the smallest vertex set S such that $G - S$ is u-colorable. By definition, $\mathbb{P}(Y = 0) > 0$. Y satisfies the bounded differences condition w.r.t. the "vertex exposure" variables X_i defined earlier. Thus,

$$\mathbb{P}\left(Y < \mathbb{E}Y - \sqrt{2(n-1)\log \tfrac{1}{\varepsilon}}\right) < \varepsilon$$

and

$$\mathbb{P}\left(Y > \mathbb{E}Y + \sqrt{2(n-1)\log \tfrac{1}{\varepsilon}}\right) < \varepsilon$$

From the first inequality and $\mathbb{P}(Y=0) > \varepsilon$, we get $\mathbb{E}Y \le \sqrt{2(n-1)\log\frac{1}{\varepsilon}}$. Plugging this in the second,

$$\mathbb{P}\left(Y > 2\sqrt{2(n-1)\log\frac{1}{\varepsilon}}\right) < \varepsilon.$$

Thus, with probability $\ge 1-\varepsilon$ we can color all but $\sqrt{8\left(\log\frac{1}{\varepsilon}\right)n}$ vertices. By the lemma, we can color the rest with 3 additional colors w.p. $1-\varepsilon$, so w.p. $1-2\varepsilon$, $u+3$ colors suffice. The minimality of u guarantees that w.p. $1-\varepsilon$ at least u colors are needed, so $\mathbb{P}(u \le \chi \le u+3) \ge 1-3\varepsilon$ ∎

Janson's inequality

In the second moment method we use Chebyshev's inequality to deduce that if $Z \geq 0$, then

$$\mathbb{P}(Z=0) \leq \frac{\mathrm{var}(Z)}{(\mathbb{E}Z)^2}.$$

Janson's inequality gives a significantly sharper bound in many important cases.

Let $X_1, \dots, X_n$ be independent Bernoulli with parameter $\mathbb{P}(X_i = 1) = p_i$. Let I be a collection of subsets of $\{1, \dots, n\}$. For each $\alpha \in I$, define $Y_\alpha = \prod_{i \in \alpha} X_i$.

We are interested in $Z = \sum_{\alpha \in I} Y_\alpha$.

Example The number of triangles in a random graph of $\binom{n}{2}$ possible edges is of this form.

Note that $\mathrm{var}(Z) = \sum_{\alpha, \beta \in I} \mathbb{E} Y_\alpha Y_\beta - \sum_{\alpha, \beta} \mathbb{E} Y_\alpha \mathbb{E} Y_\beta$

$$= \sum_{\alpha, \beta : \alpha \cap \beta \neq \emptyset} \left(\mathbb{E} Y_\alpha Y_\beta - \mathbb{E} Y_\alpha \mathbb{E} Y_\beta \right) \leq \sum_{\alpha \cap \beta \neq \emptyset} \mathbb{E} Y_\alpha Y_\beta \overset{\text{def}}{=} \Delta$$

Theorem For all $0 \leq t \leq \mathbb{E}Z$,

$$\mathbb{P}(Z \leq \mathbb{E}Z - t) \leq e^{-t^2/2\Delta}$$

In particular, $\mathbb{P}(Z=0) \leq e^{-(\mathbb{E}Z)^2/2\Delta}$

Proof The proof is based on estimating the moment generating function of Z.

Define $G(\lambda) = \log \mathbb{E} e^{-\lambda(Z - \mathbb{E}Z)} = \log \mathbb{E}\left[e^{-\lambda Z}\right] - \lambda \mathbb{E}Z$

Then $G'(\lambda) = \frac{\mathbb{E}\left[Z e^{-\lambda Z}\right]}{\mathbb{E} e^{-\lambda Z}} - \mathbb{E}Z = \sum_{\alpha \in I} \frac{\mathbb{E}\left[Y_\alpha e^{-\lambda Z}\right]}{\mathbb{E} e^{-\lambda Z}} - \lambda \mathbb{E}Z.$

For any fixed α, we may write

$$Z = \underbrace{\sum_{\beta:\alpha\cap\beta\neq\emptyset}' Y_\beta}_{U_\alpha} + \underbrace{\sum_{\beta:\alpha\cap\beta=\emptyset}' Y_\beta}_{Z_\alpha}$$

Note that $\mathbb{E}\left[Y_\alpha e^{\lambda Z}\right] = \mathbb{E}\left[e^{\lambda Z} \mid Y_\alpha = 1\right]\cdot \mathbb{E}Y_\alpha$

Since we are after lower tails, we consider $\lambda < 0$.

$$\mathbb{E}\left[e^{\lambda Z} \mid Y_\alpha = 1\right] = \mathbb{E}\left[e^{\lambda U_\alpha} e^{\lambda Z_\alpha} \mid Y_\alpha = 1\right]$$

($e^{\lambda U_\alpha}$ and $e^{\lambda Z_\alpha}$: decreasing functions of $\{X_1,\dots,X_n\}\setminus\{X_i : i\in\alpha\}$)

$$\geq \mathbb{E}\left[e^{\lambda U_\alpha} \mid Y_\alpha = 1\right]\cdot \mathbb{E}\left[e^{\lambda Z_\alpha} \mid Y_\alpha = 1\right]$$

by Harris' inequality

$$= \mathbb{E}\left[e^{\lambda U_\alpha} \mid Y_\alpha = 1\right]\cdot \mathbb{E}\left[e^{\lambda Z_\alpha}\right] \quad (Y_\alpha \text{ and } Z_\alpha \text{ are independent})$$

$$\geq \mathbb{E}\left[e^{\lambda U_\alpha} \mid Y_\alpha = 1\right]\cdot \mathbb{E}\left[e^{\lambda Z}\right] \quad (\text{since } Z_\alpha \leq Z)$$

Thus,

$$\frac{\mathbb{E}\left[Z e^{\lambda Z}\right]}{\mathbb{E}Z} \geq \mathbb{E}\left[e^{\lambda Z}\right] \sum_{\alpha\in I}' \frac{\mathbb{E}Y_\alpha}{\mathbb{E}Z}\,\underbrace{\mathbb{E}\left[e^{\lambda U_\alpha} \mid Y_\alpha = 1\right]}_{\geq e^{\lambda \mathbb{E}[U_\alpha \mid Y_\alpha = 1]} \text{ by Jensen}}$$

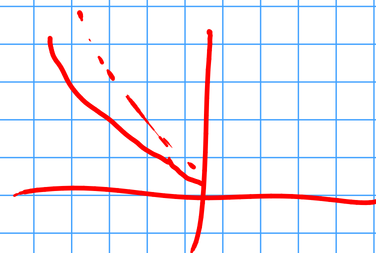
$$\geq \mathbb{E}\left[e^{\lambda Z}\right] \exp\left(\lambda \sum_{\alpha\in I}' \frac{\mathbb{E}Y_\alpha}{\mathbb{E}Z}\cdot \mathbb{E}\left[U_\alpha \mid Y_\alpha = 1\right]\right) \quad (\text{by Jensen})$$

$$= \mathbb{E}\left[e^{\lambda Z}\right] \exp\left(\lambda \frac{\sum_\alpha' \mathbb{E}\left[Y_\alpha U_\alpha\right]}{\mathbb{E}Z}\right)$$

$$= \mathbb{E}\left[e^{\lambda Z}\right] \exp\left(\lambda \frac{\Delta}{\mathbb{E}Z}\right).$$

we have obtained that for all $\lambda < 0$,

$$G'(\lambda) \ge EZ \cdot \left(\exp\left(\frac{\lambda \Delta}{EZ}\right) - 1 \right)$$



Since $G(0) = 0$, integrating between λ and 0,

$$G(\lambda) \le -EZ \int_\lambda^0 \left(\exp\left(x \frac{\Delta}{EZ}\right) - 1 \right) dx$$

$$= \frac{(EZ)^2}{\Delta} \varphi\left(\frac{\lambda \Delta}{EZ}\right) \quad \text{where } \varphi(x) = e^x - 1 - x$$

Since for $x > 0$, $\varphi(-x) \le \frac{x^2}{2}$,

$$G(\lambda) = \log E e^{-\lambda(Z - EZ)} \le \frac{(EZ)^2}{\Delta} \cdot \frac{\lambda^2 \Delta^2}{2(EZ)^2} = \lambda^2 \Delta / 2$$

This implies the announced sub-Gaussian estimate. ■

For the probability of non-existence one can also show

$$P(Z=0) \le \exp\left(-EZ + \frac{\tilde{\Delta}}{2}\right)$$

$$P(Z=0) \le \exp\left(-\frac{(EZ)^2}{EZ + \tilde{\Delta}}\right)$$

with $\tilde{\Delta} = \sum_{\substack{\alpha, \beta: \\ \alpha \ne \beta, \alpha \cap \beta \ne \emptyset}} E Y_\alpha Y_\beta$

Note that

$$P(Z=0) = P(Y_{\alpha_1} = 0, \dots, Y_{\alpha_m} = 0) = \prod_{i=1}^m P(Y_{\alpha_i} = 0 \mid Y_{\alpha_1} = 0, \dots, Y_{\alpha_{i-1}} = 0)$$

$$\ge \prod_{i=1}^m P(Y_{\alpha_i} = 0).$$

Triangles in $G(n, \frac{c}{n})$ Here we have $\binom{n}{3}$ terms. The bound gives

$$P(Z=0) \ge \left(1 - \left(\frac{c}{n}\right)^3\right)^{\binom{n}{3}} \sim e^{-c^3/6}$$

← we got this from Poisson approximation as well.

We also have

$$EZ = \binom{n}{3}\left(\frac{c}{n}\right)^3 \sim \frac{c^3}{6}, \quad \Delta = 2\binom{n}{4}\binom{4}{2}\left(\frac{c}{n}\right)^5 \sim (EZ) \cdot \frac{c^5}{2n}$$

so by Janson's inequality,

$$P(Z=0) \le \exp\left(-EZ\left(1 + O\left(\tfrac{1}{n}\right)\right)\right) \sim e^{-c^3/6}.$$