Math 554

Hemanshu Kaul

kaul@iit.edu

1963

Theorem [Chernoff-Hoeffding]

Let $X = \sum_{i=1}^{n} X_i$, $X_i \in [0,1]$ $\forall i$, $\mathbb{E}[X_i] = p_i$ ($\&$ $q_i = 1 - p_i$)
independent r.v.s

Then,
$$\mathbb{P}[X - np > tn] \leq \left(\left(\frac{p}{p+t}\right)^{p+t} \left(\frac{q}{q-t}\right)^{q-t} \right)^n,$$
where $\mu = \mathbb{E}[X] = \sum p_i = np$, $q = 1 - p$

Proof Same steps as before (!), except

$$\mathbb{E}[e^{\lambda X}] = \ldots = \prod_i \mathbb{E}[e^{\lambda X_i}] \leq \prod_i (p_i e^{\lambda} + q_i)$$

How?!?!

$e^{\lambda x}$ is a convex function, hence in $[0,1]$, it always lies below the straight line joining $(0,1)$ & $(1, e^{\lambda})$, that is $y = \alpha x + \beta$ where $\alpha = e^{\lambda} - 1$ and $\beta = 1$.

$$\therefore \mathbb{E}(e^{\lambda X_i}) \leq \mathbb{E}[\alpha X_i + \beta] = p_i e^{\lambda} + q_i \quad (!!)$$

A more useful bound on $\left(\left(\frac{p}{p+t}\right)^{p+t}\left(\frac{q}{q-t}\right)^{q-t}\right)^n$?

<u>Lemma</u> If $p+q=1$, $p\geq 0, q\geq 0$, and $0\leq t<q$, then for $f(t)=\ln\left(\left(\frac{p}{p+t}\right)^{p+t}\left(\frac{q}{q-t}\right)^{q-t}\right)$, $f(t)\leq -2t^2$

That is, $\left(\left(\frac{p}{p+t}\right)^{p+t}\left(\frac{q}{q-t}\right)^{q-t}\right)^n \leq e^{-2t^2 n}$

<u>Proof</u> $f'(t)=\ln\frac{p(q-t)}{(p+t)q}$ and $f''(t)=\frac{-1}{(p+t)(q-t)}\leq -4$.

Since $f(0)=f'(0)=0$, $f(t)=\int_0^t\int_0^t f''(s)ds \leq -2t^2$ $\blacksquare$

Use Taylor series expansion to bound such upper bound functions under different conditions on t.

Theorem Let $X_1, X_2, \ldots, X_n$ be independent random variables with $a_i \leq X_i \leq b_i \; \forall i$. Let $S_n = \sum_{i=1}^{n} X_i$ & $\mu = \mathbb{E}[S_n]$.

Then, for all $t \geq 0$

$$\mathbb{P}[|S_n - \mu| \geq t] \leq 2e^{-2t^2/\sum(b_i - a_i)^2}$$

Theorem Let $X_1, X_2, \ldots, X_n$ be independent random variables with $a_i \leq X_i \leq b_i \; \forall i$. Let $S_n = \sum_{i=1}^{n} X_i$ & $\mu = \mathbb{E}[S_n]$.

Then, for all $t \geq 0$

$$\mathbb{P}[|S_n - \mu| \geq t] \leq 2e^{-2t^2/\sum(b_i - a_i)^2}$$

Proof Try it!

① Consider $a_i = a$ & $b_i = b \; \forall i$. Derive a bound on $\mathbb{P}[S_n - \mu \geq t]$ by rescaling the C-H bound.

② Repeat the general argument to derive a bound on $\mathbb{P}[S_n - \mu \geq t]$. (Rescaling doesn't work here directly)

③ Apply the bound from ② to an appropriate r.v. to get a bound on $\mathbb{P}[S_n - \mu < -t]$

④ Combine the bounds from ② & ③ to get a bound on $\mathbb{P}[|S_n - \mu| \geq t]$

Another form —

Theorem Let $X_1, \ldots, X_n$ be independent r.v.s with $0 \leq X_i \leq 1$ and $\mu_i = \mathbb{E}[X_i]$. Denote $S_n = \sum_{i=1}^{n} X_i$, $\mu = \mathbb{E}[S_n] = \sum \mu_i$, $\sigma^2 = \mathrm{Var}[S_n]$

Then, $\mathbb{P}[S_n - \mu > t] \leq \exp(-t^2/4\sigma^2)$ for $0 \leq t < 2\sigma^2$

Another form —

Theorem Let $X_1, \dots, X_n$ be independent r.v.s with $0 \le X_i \le 1$ and $\mu_i = \mathbb{E}[X_i]$. Denote $S_n = \sum_{i=1}^{n} X_i$, $\mu = \mathbb{E}[S_n] = \sum \mu_i$, $\sigma^2 = \mathrm{Var}[S_n]$

Then, $\mathbb{P}[S_n - \mu > t] \le \exp(-t^2/4\sigma^2)$ for $0 \le t < 2\sigma^2$

Proof. Use the moment gen. ftn. method and apply the inequalities $1+x \le e^x$ & $e^x \le 1+x+x^2$ for $|x|<1$, to $\mathbb{E}[e^{\lambda \sum (x_i - \mu_i)}] = \prod_i \mathbb{E}[e^{\lambda (x_i - \mu_i)}] \le \prod_i \mathbb{E}[\cdots\cdots] \le \cdots\cdots$ ∎

Another form —

Theorem Let $X_1, \dots, X_n$ be independent r.v.s with $0 \le X_i \le 1$ and $\mu_i = \mathbb{E}[X_i]$. Denote $S_n = \sum_{i=1}^{n} X_i$, $\mu = \mathbb{E}[S_n] = \sum \mu_i$, $\sigma^2 = \mathrm{Var}[S_n]$

Then, $\mathbb{P}[S_n - \mu > t] \le \exp(-t^2/4\sigma^2)$ for $0 \le t < 2\sigma^2$

Proof Use the moment gen. ftn. method and apply the inequalities $1+x \le e^x$ & $e^x \le 1+x+x^2$ for $|x|<1$, to $\mathbb{E}[e^{\lambda \Sigma(X_i - \mu_i)}] = \prod_i \mathbb{E}[e^{\lambda(X_i - \mu_i)}] \le \prod_i \mathbb{E}[\cdots\cdots] \le \cdots\cdot$ ∎

Cor Same conditions as above, but with $-1 \le X_i \le 1$ and $\mu_i = \mathbb{E}[X_i] = 0$, then

$\mathbb{P}[S_n > t] \le \exp(-t^2/4\sigma^2)$ for $0 \le t < 2\sigma^2$

Repeat the argument, or apply the Thm. to $X_i = 2Y_i - 1$, for $0 \le Y_i \le 1$.

Sampling Multidimensional Data (& a hint of VC-dimension)

Let P be a set of n points in $\mathbb{R}^d$, such as profiles of n people along d attributes

A query q specifies an interval in each dimension (geometrically, it forms a d-dim box).

The answer is the range (subset of P contained in the d-dim box)

The family R of all possible ranges from P defines the range space (P, R).

Sampling Multidimensional Data (& a hint of VC-dimension)

Let P be a set of n points in $\mathbb{R}^d$, such as profiles of
n people along d attributes

A query q specifies an interval in each dimension
(geometrically, it forms a d-dim box).

The answer is the range (subset of P contained in the d-dim box)

The family R of all possible ranges from P defines the
range space (P, R).

How big a random sample of the population P do we need to
accurately describe the fraction of population in any range?

Sampling Multidimensional Data (& a hint of VC-dimension)

Let P be a set of n points in $\mathbb{R}^d$, such as profiles of n people along d attributes

A query q specifies an interval in each dimension (geometrically, it forms a d-dim box).

The answer is the range (subset of P contained in the d-dim box)

The family R of all possible ranges from P defines the range space (P, R).

How big a random sample of the population P do we need to accurately describe the fraction of population in any range?

Theorem Given a d-dimensional range space (P, R), and positive constants ϵ & δ, let S randomly sample $(d/\epsilon^2)\ln(n/\delta)$ points from P. With probability at least $1-2\delta^{2d}$, we have

points in P returned by query q

$$\left|\frac{q(P)}{|P|} - \frac{q(S)}{|S|}\right| \le \epsilon \quad \text{for all } q \in R.$$

Proof Let $k = \lceil (d/\epsilon^2) \ln(n/\delta) \rceil$. For a fixed query q:

Let X_i be indicator r.v. for the membership of $s_i \in S$ in q,
for each random choice s_i of a point in S.

$\therefore$ $q(S)$ is the sum of k identical Bernoulli trials with success probability $\frac{q(P)}{|P|}$,

and expectation $k\,q(P)/|P|$.

By Chernoff, $\mathbb{P}\left[\left|q(S) - |S|\frac{q(P)}{|P|}\right| > k\epsilon\right] \leq 2e^{-2k\epsilon^2} = 2\delta^{2d} n^{-2d}$ — ⊛

Claim at most n^{2d} distinct subsets of P can be returned by q.

Consider the minimal d-dim box that contains the answer to the query, we can shrink the intervals in each dimension till we hit points from the answer, giving us at most $2d$ points in n that generate this box.
There are at most n^{2d} ways to specify that minimal box for q, so there are at most n^{2d} sets in the range.

$\therefore$ by claim & ⊛, probability $\geq 1 - 2\delta^{2d}$ for all ranges to have sizes approx. within ϵ. ∎
with union bound

Theorem Given d-dimensional range space (P, R) and positive constants ϵ & δ, let S randomly sample $(d/\epsilon^2)\ln(n/\delta)$ points from P. With probability $1-2\delta^{2d}$, $\left|\frac{q(P)}{|P|} - \frac{q(S)}{|S|}\right| \leq \epsilon$ holds for all $q \in R$.

Vapnik-Chervonenkis (1971) proved that amazingly only $O\left(\frac{d}{\epsilon^2}\ln\left(\frac{1}{\epsilon\delta}\right)\right)$ points are needed in the random sample to represent all the ranges, independent of n [# samples needed for 1 query, also work for all queries]
(Talagrand improved it to $O(\epsilon^{-2}(d + \ln(1/\delta)))$)

They introduced & used the notion of VC-dimension ν which is $\nu = 2d$ for axis aligned boxes, $\nu = d+1$ for balls and half-spaces, ...
And proved $O(\epsilon^{-2}(\nu + \ln(1/\delta)))$ works for all types of queries (not just boxes).

VC-dimension is a fundamental concept in Computational & Discrete geometry, and in Statistical Learning.

Discrepancy of a set system [Combinatorial Discrepancy, see book by Chazelle]

Given a finite family of finite sets, we want to color the underlying elements Red or Blue so that all sets have nearly the same number of Red and Blue elements.

Hypergraph $\mathcal{H} = (V, E)$ $\qquad \chi : V \to \{-1, +1\}$

let $d(e) = \sum_{v \in e} \chi(v)$, for each edge e.

$$\operatorname{disc}(\mathcal{H}, \chi) = \max_{e \in E} |d(e)|$$

<u>discrepancy of $\mathcal{H}$</u>, $\operatorname{disc}(\mathcal{H}) = \min_{\chi : V \to \{-1,+1\}} \operatorname{disc}(\mathcal{H}, \chi)$.

- $\operatorname{disc}(\mathcal{H}) \leq k \iff \exists$ a coloring χ of $\mathcal{H}$ s.t. $d(e) \leq k \;\; \forall e \in E$.
- If $\mathcal{H}$ is k-uniform then $\operatorname{disc}(\mathcal{H}) < k$ iff $\mathcal{H}$ is 2-colorable

Theorem Let $\mathcal{H}$ be a hypergraph on n vertices with m edges,
Then $\operatorname{disc}(\mathcal{H}) \leq \sqrt{2n \ln(2m)}$

Proof Let $\mathcal{H} = (V, E)$ with $|V| = n$, $|E| = m$.
Take a random 2-coloring of vertices, $f: V \to \{-1, +1\}$
Let $\alpha = \sqrt{2n \ln(2m)}$
For each $e \in E$, $\mathbb{P}[d(e) > \alpha] = \mathbb{P}\left[\sum_{v \in e} f(v) > \alpha\right]$

HW Bound this using (i) Chebyshev
(ii) Chernoff-Hoeffding

& compare.

Most well studied case is when $m = n$,
with a famous result of Spencer that $\operatorname{disc}(\mathcal{H}) \leq 6\sqrt{n}$.

Two vectors picked at random from S^{n-1}, unit sphere in $\mathbb{R}^n$, are almost orthogonal w.h.p.

Theorem Let $\vec{x}, \vec{y}$ be two vectors picked independently from S^{n-1}, then with probability $> 1 - 1/n$,
$$|\cos(\theta_{x,y})| = O\left(\sqrt{\frac{\ln 2n}{n}}\right),$$ where $\theta_{x,y}$ is the angle between x & y.

Proof

Lemma Let $\vec{a}$ be a unit vector in $\mathbb{R}^n$. Let $\vec{x} = \frac{1}{\sqrt{n}}(x_1, \ldots, x_n) \in \mathbb{R}^n$ be chosen from S^{n-1} by choosing each x_i independently at random from $\{-1, +1\}$.
Let $X = \vec{a} \cdot \vec{x} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_i x_i = \cos(\theta_{\vec{a},\vec{x}})$, then $\mathbb{P}[|X| > t] < 2e^{-nt^2/4}$

Pf. HW! Apply Variance version of Chernoff.
Why can't we apply usual Chernoff?

Two vectors picked at random from S^{n-1}, unit sphere in $\mathbb{R}^n$, are almost orthogonal w.h.p.

Theorem Let $\vec{x}, \vec{y}$ be two vectors picked independently from S^{n-1}, then with probability $> 1 - 1/n$,

$$|\cos(\theta_{x,y})| = O\left(\sqrt{\frac{\ln 2n}{n}}\right),$$

where $\theta_{x,y}$ is the angle between x & y.

Proof

Lemma Let $\vec{a}$ be a unit vector in $\mathbb{R}^n$. Let $\vec{x} = \frac{1}{\sqrt{n}}(x_1, \dots, x_n) \in \mathbb{R}^n$ be chosen from S^{n-1} by choosing each x_i independently at random from $\{-1, +1\}$.

Let $X = \vec{a} \cdot \vec{x} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_i x_i = \cos(\theta_{\vec{a},\vec{x}})$, then $\mathbb{P}[|X| > t] < 2e^{-nt^2/4}$

If two unit vectors $\vec{x}, \vec{y}$ are chosen at random in $\mathbb{R}^n$ then

$$\mathbb{P}\left[|\cos(\theta_{x,y})| > \sqrt{\frac{-4\ln \epsilon/2}{n}}\right] < \epsilon \qquad (\text{put } \epsilon = 2e^{-nt^2/4} \text{ in lemma})$$

Now take $\epsilon = 1/n$, to get $\mathbb{P}\left[|\cos(\theta_{x,y})| > \sqrt{\frac{4\ln 2n}{n}}\right] < 1/n$.

Johnson-Lindenstrauss Flattening Lemma

"low-distortion embedding of data in low dimension"

Given n points $z_1, z_2, \ldots, z_n$ in $\mathbb{R}^d$, we would like to find n points $u_1, \cdots, u_n$ in $\mathbb{R}^m$, where m is of low dimension (compared to n), and the distance metric restricted to these points is almost preserved: for given $0<\epsilon<1$

$$\circledast \quad \|z_i - z_j\|_2 \leq \|u_i - u_j\|_2 \leq (1+\epsilon)\|z_i - z_j\|_2 \quad \forall i,j$$

Johnson-Lindenstrauss Flattening Lemma

"low-distortion embedding of data in low dimension"

Given n points $z_1, z_2, \ldots, z_n$ in $\mathbb{R}^d$, we would like to find n points $u_1, \cdots, u_n$ in $\mathbb{R}^m$, where m is of low dimension (compared to n), and the distance metric restricted to these points is almost preserved: for given $0<\epsilon<1$

$$\circledast \quad \|z_i - z_j\|_2 \leq \|u_i - u_j\|_2 \leq (1+\epsilon) \|z_i - z_j\|_2 \quad \forall i,j$$

Thm [J-L 1984] $m = O(\epsilon^{-2} \log n)$ dimension suffices.

To represent n points in $\mathbb{R}^n$, we need to store n^2 numbers and store their distances in about n^2 numbers as well. By J-L, we can store only $n \log n$ numbers and still reconstruct any n^2 distances

By providing a low-dim representation of data, these embeddings dramatically speed up certain algorithms (often running time depends exponentially on the dimension of the working space). Applications in dimension reduction in databases, embeddings of networks into normed spaces, ...

Outline of the proof

Step1 Choose m vectors $x_1, \ldots, x_m \in \mathbb{R}^d$ at random by choosing each coordinate at random from $\left\{\sqrt{\frac{1+\epsilon}{m}}, -\sqrt{\frac{1+\epsilon}{m}}\right\}$.

Let $u_i = (z_i \cdot x_1, z_i \cdot x_2, \ldots\ldots, z_i \cdot x_m) \in \mathbb{R}^m$, $i = 1, \ldots, n$.

To show: $u_1, \ldots, u_m$ satisfy $\circledast$ with positive probability.

Outline of the proof

Step 1 Choose m vectors $x_1, \dots, x_m \in \mathbb{R}^d$ at random by choosing each coordinate at random from $\left\{\sqrt{\frac{1+\varepsilon}{m}}, -\sqrt{\frac{1+\varepsilon}{m}}\right\}$

Let $u_i = (z_i \cdot x_1, z_i \cdot x_2, \dots, z_i \cdot x_m) \in \mathbb{R}^m$, $i = 1, \dots, n$.

To show: $u_1, \dots, u_m$ satisfy ⊛ with positive probability.

Step 2 Express $\|u_i - u_j\|^2$ in terms of z_k's & x_i's

$$\|u_i - u_j\|^2 = \sum_{k=1}^{m} \left(\sum_{l=1}^{d} (z_{il} - z_{jl}) x_{kl} \right)^2$$

For a fixed pair i, j, denote $z := z_i - z_j$ & $u := u_i - u_j$, so

$$\|u\|^2 = \|u_i - u_j\|^2 = \sum_{k=1}^{m} \left(\sum_{l=1}^{d} z_l x_{kl} \right)^2$$

Outline of the proof

Step1 Choose m vectors $x_1, \ldots, x_m \in \mathbb{R}^d$ at random by choosing each coordinate at random from $\left\{\sqrt{\frac{1+\epsilon}{m}}, -\sqrt{\frac{1+\epsilon}{m}}\right\}$.

Let $u_i = (z_i \cdot x_1, z_i \cdot x_2, \ldots\ldots, z_i \cdot x_m) \in \mathbb{R}^m$, $i = 1, \ldots, n$.

To show: $u_1, \ldots, u_m$ satisfy ⊛ with positive probability.

Step2 Express $\|u_i - u_j\|^2$ in terms of z_k's & x_i's

$$\|u_i - u_j\|^2 = \sum_{k=1}^{m}\left(\sum_{\ell=1}^{d} (z_{i\ell} - z_{j\ell})\, x_{k\ell}\right)^2$$

For a fixed pair i, j, denote $z := z_i - z_j$ & $u := u_i - u_j$, so

$$\|u\|^2 = \|u_i - u_j\|^2 = \sum_{k=1}^{m}\left(\sum_{\ell=1}^{d} z_\ell x_{k\ell}\right)^2$$

Step3 $\mathbb{E}[\|u\|^2] = (1+\epsilon)\,\|z\|^2 =: \mu$

let X_k be the r.v. $\left(\sum_{\ell=1}^{d} z_\ell x_{k\ell}\right)^2$ then show $\mathbb{E}[X_k] = \frac{1+\epsilon}{m}\|z\|^2$

(expand out the expression for X_k)

Step 4 Show $\|u\|^2$ is concentrated around its mean.

Lemma $\exists$ constants $c_1 > 0$, $c_2 > 0$ s.t.

(i) $\mathbb{P}[\|u\|^2 > (1+t)\mu] < e^{-c_1 t^2 m}$

(ii) $\mathbb{P}[\|u\|^2 < (1-t)\mu] < e^{-c_2 t^2 m}$

By moment gen. function & taylor series estimate $\frac{1}{2}\exp(t)+\frac{1}{2}\exp(-t) < \exp(t^2)$

Step 4 Show $\|u\|^2$ is concentrated around its mean.

Lemma $\exists$ constants $c_1 > 0$, $c_2 > 0$ s.t.

(i) $\mathbb{P}[\|u\|^2 > (1+t)\mu] < e^{-c_1 t^2 m}$

(ii) $\mathbb{P}[\|u\|^2 < (1-t)\mu] < e^{-c_2 t^2 m}$

By moment gen. function & taylor series estimate $\frac{1}{2}\exp(t)+\frac{1}{2}\exp(-t) < \exp(t^2)$

Step 5 By the lemma, $\exists\, c > 0$ s.t.

$$\mathbb{P}[(\|u\|^2 > (1+t)\mu) \text{ or } (\|u\|^2 < (1-t)\mu)] < e^{-ct^2 m}$$

There are $\binom{n}{2}$ r.v.s of the type $\|u\|^2 = \|u_i - u_j\|^2$

with $t = \epsilon/2$, by the union bound, the probability that any of these r.v.s is not within $(1 \pm \epsilon/2)$ of their expectation is bdd. by $\binom{n}{2} \exp(-c \frac{\epsilon^2}{4} m)$.

Step 4 Show $\|u\|^2$ is concentrated around its mean.

Lemma $\exists$ constants $c_1 > 0$, $c_2 > 0$ s.t.

(i) $\mathbb{P}[\|u\|^2 > (1+t)\mu] < e^{-c_1 t^2 m}$

(ii) $\mathbb{P}[\|u\|^2 < (1-t)\mu] < e^{-c_2 t^2 m}$

By moment gen. function & taylor series estimate $\frac{1}{2}\exp(t) + \frac{1}{2}\exp(-t) < \exp(t^2)$

Step 5 By the lemma, $\exists\, c > 0$ s.t.

$$\mathbb{P}[(\|u\|^2 > (1+t)\mu) \text{ or } (\|u\|^2 < (1-t)\mu)] < e^{-ct^2 m}$$

There are $\binom{n}{2}$ r.v.s of the type $\|u\|^2 = \|u_i - u_j\|^2$

With $t = \epsilon/2$, by the union bound, the probability that any of these r.v.s is not within $(1 \pm \epsilon/2)$ of their expectation is bdd. by $\binom{n}{2} \exp(-c\frac{\epsilon^2}{4} m)$.

Step 6 Choose $m > (8 \log n + \log c)/\epsilon^2$, then with positive probability, all r.v.s are close to their expectation within factor $(1 \pm \epsilon/2)$, i.e.

$$(1 - \epsilon/2)(1+\epsilon)\|z_i - z_j\|^2 \leq \|u_i - u_j\|^2 \leq (1+\epsilon/2)(1+\epsilon)\|z_i - z_j\|^2 \quad \forall i, j$$

Step 4 Show $\|u\|^2$ is concentrated around its mean.

Lemma $\exists$ constants $c_1>0$, $c_2>0$ s.t.

(i) $\mathbb{P}[\|u\|^2 > (1+t)\mu] < e^{-c_1 t^2 m}$

(ii) $\mathbb{P}[\|u\|^2 < (1-t)\mu] < e^{-c_2 t^2 m}$

By moment gen. function & taylor series estimate $\frac{1}{2}\exp(t)+\frac{1}{2}\exp(-t)<\exp(t^2)$

Step 5 By the lemma, $\exists\, c>0$ s.t.

$$\mathbb{P}[(\|u\|^2>(1+t)\mu) \text{ or } (\|u\|^2<(1-t)\mu)] < e^{-ct^2 m}$$

There are $\binom{n}{2}$ r.v.s of the type $\|u\|^2 = \|u_i-u_j\|^2$

With $t=\epsilon/2$, by the union bound, the probability that any of these r.v.s is not within $(1\pm\epsilon/2)$ of their expectation is bdd. by $\binom{n}{2}\exp(-c\frac{\epsilon^2}{4}m)$.

Step 6 Choose $m>(8\log n+\log c)/\epsilon^2$, then with positive probability, all r.v.s are close to their expectation within factor $(1\pm\epsilon/2)$, i.e.

$$(1-\epsilon/2)(1+\epsilon)\|z_i-z_j\|^2 \le \|u_i-u_j\|^2 \le (1+\epsilon/2)(1+\epsilon)\|z_i-z_j\|^2 \quad \forall i,j$$

Since $1\le(1-\epsilon/2)(1+\epsilon)\le 9/8$ & $(1+\frac{\epsilon}{2})\le(1+\epsilon)$, $\|z_i-z_j\|^2\le\|u_i-u_j\|^2\le(1+\epsilon)^2\|z_i-z_j\|^2$.

Hajós & Hadwiger's conjectures

Recall from Math 553,

H is a minor of G if H can be obtained from G by deleting edges and vertices, & by contracting edges.
(Take a subgraph of G & contract some of its edges to get H).

H is a subdivision of G if H can be obtained from a subgraph of G by contracting induced paths to edges.
(Subdividing edges of H into induced paths gives us a subgraph of G).

Kuratowski (1930) G without $K_{3,3}$ and K_5 as subdivisions is planar.

Wagner (1937) G without $K_{3,3}$ and K_5 as minors is planar.

We know $\omega(G) \leq \chi(G) \not\leq f(\omega(G))$

What if we replaced clique by clique minor or subdivision?

Hadwiger's Conjecture (1936)

For each $t \geq 1$, if G does not contain K_{t+1} as a minor, then $\chi(G) \leq t$

$t=1$ trivial

$t=2$ trivial (G K_3-minor-free $\Rightarrow$ G tree)

$t=3$ short argument

$t=4$ equivalent to 4-color thm (Wagner 1937)

$t=5$ equivalent to 4-color thm (Robertson-Seymour-Thomas 1994)

$t \geq 6$ open

G K_{t+1}-minor-free $\Rightarrow \chi(G) \leq O(t\sqrt{\log t})$ (Kostochka / Thomason 1984)

$\leq O(t(\log\log t)^6)$ (Norin, Postle, Song 2020+)

Hadwiger's Conjecture (1936)

For each $t \geq 1$, if G does not contain K_{t+1} as a minor, then $\chi(G) \leq t$

$t=1$ trivial

$t=2$ trivial (G K_3-minor-free $\Rightarrow$ G tree)

$t=3$ short argument

$t=4$ equivalent to 4-color thm (Wagner 1937)

$t=5$ equivalent to 4-color thm (Robertson-Seymour-Thomas 1994)

$t \geq 6$ open

G K_{t+1}-minor-free $\Rightarrow \chi(G) \leq O(t\sqrt{\log t})$ (Kostochka / Thomason 1984)

$\leq O(t(\log\log t)^6)$ (Norin, Postle, Song 2020+)

Hajós Conjecture (1961)

For each $t \geq 1$, if G is K_{t+1}-subdivision-free then $\chi(G) \leq t$.

True for $t \leq 3$

False for all $t \geq 6$ (Catlin 1979)

$t = 4, 5$??

Theorem [Erdös-Fajtlowicz 1981]

In $G(n, 1/2)$, almost every graph has $\chi \geq n/3\log n$ and has no subdivision of $K_{3\sqrt{n}}$.

Almost every graph is a counterexample to Hajós's conjecture!

Theorem [Erdös-Fajtlowicz 1981]

In $G(n,1/2)$, almost every graph has $\chi \geq n/3\log n$ and has no subdivision of $K_{3\sqrt{n}}$.

Proof $\mathbb{P}[\alpha(G) \geq x] = \mathbb{P}[\exists$ set of x vertices that form an ind. set$]$

$$\leq \binom{n}{x} 2^{-\binom{x}{2}} \leq \left(n\, 2^{-\frac{x-1}{2}}\right)^x \leq 2^{-2\log n+3} = 1/8n^2 \to 0$$

as $n \to \infty$

let $x = 2\log_2 n + 3$, so $2^{x-1/2} = 2n$

$\therefore\ \alpha(G) \leq 2\log n + 3 < 3\log_2 n$ w.h.p., i.e., $\chi(G) \geq n/3\log n$ w.h.p.

Theorem [Erdös-Fajtlowicz 1981]

In $G(n, 1/2)$, almost every graph has $\chi \geq n/3\log n$ and has no subdivision of $K_{3\sqrt{n}}$.

Proof Suppose G contains $K_{c\sqrt{n}}$-subdivision, $c = 3$.

Since $K_{c\sqrt{n}}$ contains $(3\sqrt{n})(3\sqrt{n}-1)/2 > 4n$ edges, G must contain $4n$ disjoint paths.

And, every vertex of G must either be a vertex of the $K_{3\sqrt{n}}$-subdivision or be contained in at most one such path.

Since there are only n vertices of G, at least $3n$ of the paths must be edges

Theorem [Erdös-Fajtlowicz 1981]

In $G(n, 1/2)$, almost every graph has $\chi \geq n/3\log n$ and has no subdivision of $K_{3\sqrt{n}}$.

Proof Suppose G contains $K_{c\sqrt{n}}$-subdivision, $c = 3$.

Since $K_{c\sqrt{n}}$ contains $(3\sqrt{n})(3\sqrt{n}-1)/2 > 4n$ edges, G must contain $4n$ disjoint paths.

And, every vertex of G must either be a vertex of the $K_{3\sqrt{n}}$-subdivision or be contained in at most one such path.

Since there are only n vertices of G, at least $3n$ of the paths must be edges.

Fix $U \subseteq V(G)$, $|U| = 3\sqrt{n}$, $e(U)$ = # edges induced by U.

If U forms a $K_{3\sqrt{n}}$-subdivision then $e(U) \geq 3n$.

By Chernoff, $\mathbb{P}\left[|e(U) - \mathbb{E}[e(U)]| \geq \frac{1}{4}\mathbb{E}[e(U)]\right] \leq 2e^{-\mathbb{E}[e(U)]/48}$

where $\mathbb{E}[e(U)] = \frac{1}{2}\frac{(3\sqrt{n})(3\sqrt{n}-1)}{2} = \frac{9n - 3\sqrt{n}}{4}$

Theorem [Erdös-Fajtlowicz 1981]

In $G(n, 1/2)$, almost every graph has $\chi \geq n/3\log n$ and has no subdivision of $K_{3\sqrt{n}}$.

Proof Suppose G contains $K_{c\sqrt{n}}$-subdivision, $c=3$.
Since $K_{c\sqrt{n}}$ contains $(3\sqrt{n})(3\sqrt{n}-1)/2 > 4n$ edges, G must contain $4n$ disjoint paths.

And, every vertex of G must either be a vertex of the $K_{3\sqrt{n}}$-subdivision or be contained in at most one such path.

Since there are only n vertices of G, at least $3n$ of the paths must be edges.
Fix $U \subseteq V(G)$, $|U| = 3\sqrt{n}$, $e(U) = $ # edges induced by U.
If U forms a $K_{3\sqrt{n}}$-subdivision then $e(U) \geq 3n$.
By Chernoff, $\mathbb{P}[|e(U) - \mathbb{E}[e(U)]| \geq \frac{1}{4}\mathbb{E}[e(U)]] \leq 2e^{-\mathbb{E}[e(U)]/48}$
where $\mathbb{E}[e(U)] = \frac{1}{2}\frac{(3\sqrt{n})(3\sqrt{n}-1)}{2} = \frac{9n-3\sqrt{n}}{4}$

$\therefore \mathbb{P}[e(U) \geq 3n] \leq 2e^{-(9n-3\sqrt{n})/192} < e^{-n/25}$
so, $\mathbb{P}[U$ forms a $K_{3\sqrt{n}}$-subdivision$] <$

Theorem [Erdös-Fajtlowicz 1981]

In $G(n, 1/2)$, almost every graph has $\chi \geq n/3\log n$ and has no subdivision of $K_{3\sqrt{n}}$.

Proof Suppose G contains $K_{c\sqrt{n}}$-subdivision, $c=3$.
Since $K_{c\sqrt{n}}$ contains $(3\sqrt{n})(3\sqrt{n}-1)/2 > 4n$ edges, G must contain $4n$ disjoint paths.

And, every vertex of G must either be a vertex of the $K_{3\sqrt{n}}$-subdivision or be contained in at most one such path.

Since there are only n vertices of G, at least $3n$ of the paths must be edges.
Fix $U \subseteq V(G)$, $|U| = 3\sqrt{n}$, $e(U) = \#$ edges induced by U.
If U forms a $K_{3\sqrt{n}}$-subdivision then $e(U) \geq 3n$.
By Chernoff, $\mathbb{P}[|e(U) - \mathbb{E}[e(U)]| \geq \frac{1}{4}\mathbb{E}[e(U)]] \leq 2e^{-\mathbb{E}[e(U)]/48}$
where $\mathbb{E}[e(U)] = \frac{1}{2}\frac{(3\sqrt{n})(3\sqrt{n}-1)}{2} = \frac{9n-3\sqrt{n}}{4}$

$\therefore \mathbb{P}[e(U) \geq 3n] \leq 2e^{-(9n-3\sqrt{n})/192} < e^{-n/25}$
so, $\mathbb{P}[U$ forms a $K_{3\sqrt{n}}$-subdivision$] < e^{-n/25}$. $\Rightarrow \mathbb{P}[\exists\, K_{3\sqrt{n}}$-subdivision$]$
$< \binom{n}{3\sqrt{n}} e^{-n/25} < \left(\frac{e\sqrt{n}}{3}\right)^{3\sqrt{n}} e^{-n/25}$
$\to 0$ as $n \to \infty$