Math 554

Hemanshu Kaul

kaul@iit.edu

Defn Let H be a graph with v vertices and e edges
we define density of H as $\rho(H) = e/v$ (half of average degree)
we call H balanced if no subgraph of H has density strictly greater than $\rho(H)$.

e.g. Trees, cliques, k-trees, k-reg connected, hypercubes, etc.

Theorem Let H be a balanced graph with density ρ.
Then, $t(n) = n^{-1/\rho}$ is a threshold function for the event that H is a subgraph of $G(n,p)$.

What does this thm tell us about $H = K_3$?

$\rho(K_3) = \frac{3}{3} = 1$, so $t(n) = n^{-1/1} = \frac{1}{n}$

Theorem Let H be a balanced graph with density ϱ.
Then $t(n) = n^{-1/\varrho}$ is a threshold function for "H is a subgraph of $G(n,p)$".

Proof Let H have v vertices and e edges, so $\varrho = e/v$
Denote $V(H)$ as $\{a_1, a_2, \ldots, a_v\}$
For any ordered v-tuple, $\beta = (b_1, b_2, \ldots, b_v)$ of distinct vertices $b_1, b_2, \ldots, b_v$ in $V(G(n,p))$,

let A_β be the event that "$G(n,p)$ contains an appropriately ordered copy of H on $(b_1, b_2, \ldots, b_v)$"

that is, A_β occurs if $b_i b_j \in E(G(n,p))$ whenever $a_i a_j \in E(H)$.

Theorem Let H be a balanced graph with density ρ.
Then $t(n) = n^{-1/\rho}$ is a threshold function for "H is a subgraph of $G(n,p)$".

Proof Let H have v vertices and e edges, so $\rho = e/v$

Denote $V(H)$ as $\{a_1, a_2, \dots, a_v\}$

For any ordered v-tuple, $\beta = (b_1, b_2, \dots, b_v)$ of distinct vertices $b_1, b_2, \dots, b_v$ in $V(G(n,p))$,

let A_β be the event that "$G(n,p)$ contains an appropriately ordered copy of H on $(b_1, b_2, \dots, b_v)$"

that is, A_β occurs if $b_i b_j \in E(G(n,p))$ whenever $a_i a_j \in E(H)$.

Let X_β = indicator r.v. for event A_β

Let $X = \sum_\beta X_\beta$

# of ordered copies of H, which could be larger than # copies of H

However, $X = 0$ & $X > 0$ are equivalent to absence & appearance of H in $G(n,p)$

$$
\begin{aligned}
\mathbb{E}[X] &= \sum \mathbb{E}[X_\beta] \\
&= \sum \mathbb{P}[A_\beta] \\
&= \sum p^e = \Theta(n^v p^e)
\end{aligned}
$$

So, $\mathbb{E}[X] = \Theta(n^{v} p^{e})$

If $p(n) << n^{-1/g} = n^{-v/e}$ then $p^{e} n^{v} << 1$, ie., $p^{e} n^{v} \to 0$

so, $\lim_{n\to\infty} \mathbb{E}[X] = 0$

& by Markov, $\mathbb{P}[X=0] \to 1$ as $n \to \infty$.

So, $\mathbb{E}[X] = \Theta(n^v p^e)$

If $\underline{p(n) << n^{-1/g}} = n^{-v/e}$ then $p^e n^v << 1$, ie., $p^e n^v \to 0$

so, $\lim_{n\to\infty} \mathbb{E}[X] = 0$

& by Markov, $\mathbb{P}[X=0] \to 1$ as $n \to \infty$.

Now, we assume $\underline{p(n) >> n^{-1/g}}$

$$\mathrm{Var}[X] = \sum_{\beta} \mathrm{Var}[X_\beta] + \sum_{\alpha \neq \beta} \mathrm{Cov}[X_\alpha, X_\beta]$$

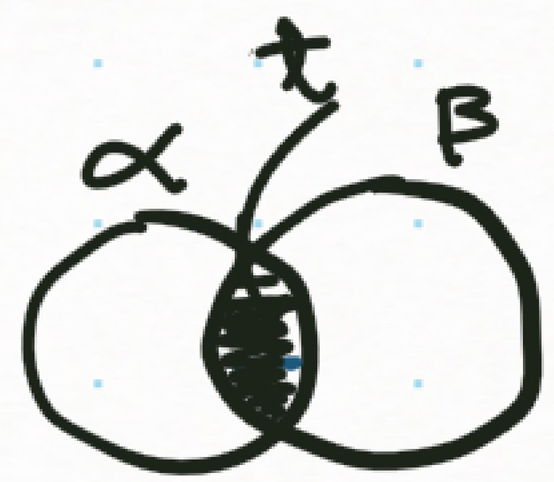
So, $\mathbb{E}[X] = \Theta(n^v p^e)$

If $\underline{p(n) << n^{-1/\rho}} = n^{-v/e}$ then $p^e n^v << 1$, ie., $p^e n^v \to 0$

so, $\lim_{n\to\infty} \mathbb{E}[X] = 0$

& by Markov, $\mathbb{P}[X=0] \to 1$ as $n\to\infty$.

Now, we assume $\underline{p(n) >> n^{-1/\rho}}$



$$Var[X] = \sum_{\beta} Var[X_\beta] + \sum_{\alpha\neq\beta} Cov[X_\alpha, X_\beta]$$

Let α and β share $t \geqslant 2$ vertices, then the two copies of H on α & β have at most ? edges in common

$\therefore Cov[X_\alpha, X_\beta] \leq \mathbb{E}[X_\alpha X_\beta] \leq p^{?}$, since their union has at least ? edges

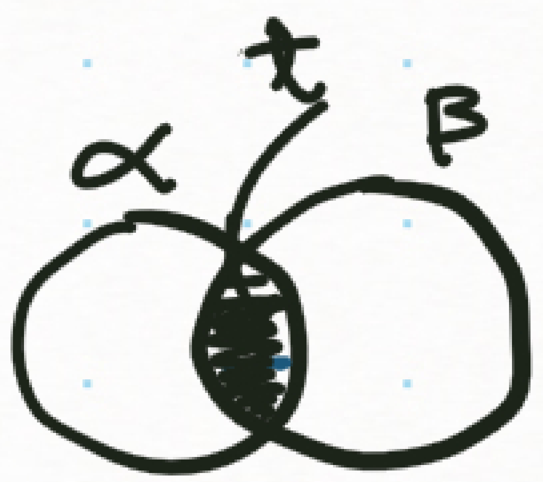
So, $\mathbb{E}[X] = \Theta(n^v p^e)$

If $p(n) << n^{-1/\rho} = n^{-v/e}$ then $p^e n^v << 1$, ie., $p^e n^v \to 0$

so, $\lim_{n\to\infty} \mathbb{E}[X] = 0$

& by Markov, $\mathbb{P}[X=0] \to 1$ as $n \to \infty$.

Now, we assume $p(n) >> n^{-1/\rho}$



$$\mathrm{Var}[X] = \sum_\beta \mathrm{Var}[X_\beta] + \sum_{\alpha \neq \beta} \mathrm{Cov}[X_\alpha, X_\beta]$$

Let α and β share $t \geq 2$ vertices, then the two copies of H on α & β have at most $t\rho$ edges in common

$\therefore \mathrm{Cov}[X_\alpha, X_\beta] \leq \mathbb{E}[X_\alpha X_\beta] \leq p^{2e-t\rho}$, since their union has at least $2e-t\rho$ edges

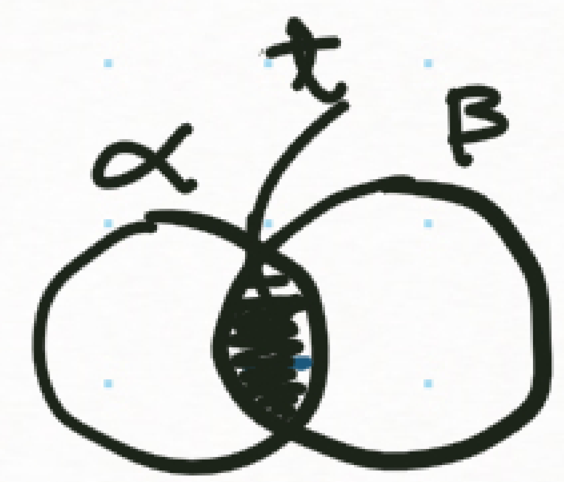
So, $\mathbb{E}[X] = \Theta(n^{v} p^{e})$

If $p(n) << n^{-1/\varrho} = n^{-v/e}$ then $p^e n^v << 1$, ie., $p^e n^v \to 0$

so, $\lim_{n\to\infty} \mathbb{E}[X] = 0$

& by Markov, $\mathbb{P}[X=0] \to 1$ as $n \to \infty$.

Now, we assume $p(n) >> n^{-1/\varrho}$



$$\mathrm{Var}[X] = \sum_{\beta} \mathrm{Var}[X_\beta] + \sum_{\alpha \neq \beta} \mathrm{Cov}[X_\alpha, X_\beta]$$

Let α and β share $t \geq 2$ vertices, then the two copies of H on α & β have at most $t\varrho$ edges in common

$\therefore \mathrm{Cov}[X_\alpha, X_\beta] \leq \mathbb{E}[X_\alpha X_\beta] \leq p^{2e - t\varrho}$, since their union has at least $2e - t\varrho$ edges

The number of pairs α, β sharing t vertices is $\Theta(n^{2v-t})$

(we can choose $2v-t$ vertices in $\binom{n}{2v-t}$ ways and there are only constant many ways to choose α & β from them since H is fixed so $v = |\alpha| = |\beta|$ is fixed)

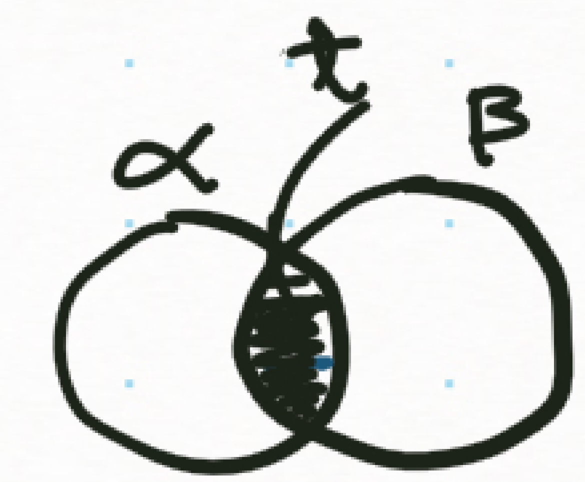
So, $\mathbb{E}[X] = \Theta(n^v p^e)$

If $\underline{p(n) << n^{-1/\varrho}} = n^{-v/e}$ then $p^e n^v << 1$, ie., $p^e n^v \to 0$

so, $\lim_{n\to\infty} \mathbb{E}[X] = 0$

& by Markov, $\mathbb{P}[X=0] \to 1$ as $n \to \infty$.

Now, we assume $\underline{p(n) >> n^{-1/\varrho}}$



$$Var[X] = \sum_{\beta} Var[X_\beta] + \sum_{\alpha \neq \beta} Cov[X_\alpha, X_\beta]$$

Let α and β share $t \geq 2$ vertices, then the two copies of H on α & β have at most $t\varrho$ edges in common

$\therefore Cov[X_\alpha, X_\beta] \leq \mathbb{E}[X_\alpha X_\beta] \leq p^{2e-t\varrho}$, since their union has at least $2e-t\varrho$ edges

The number of pairs α, β sharing t vertices is $\Theta(n^{2v-t})$

$$\therefore \sum_{|\alpha \cap \beta| = t} Cov[X_\alpha, X_\beta] \leq \Theta(n^{2v-t} p^{2e-t\varrho}) = \Theta((n^v p^e)^{2-t/v})$$

So $Var[X] \leq O\left(\sum_{t=2}^{v} (n^v p^e)^{2-t/v}\right)$

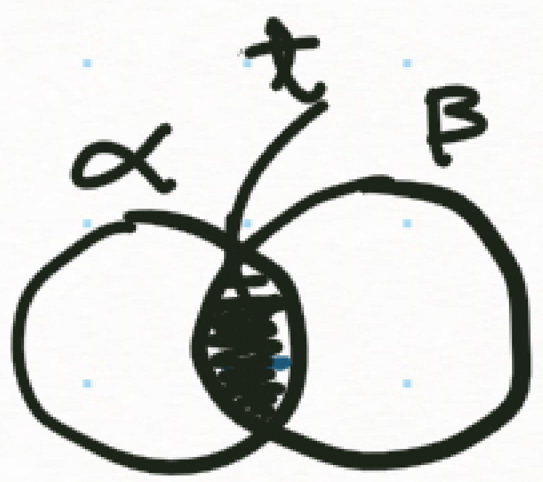
So, $\mathbb{E}[X] = \Theta(n^{v} p^{e})$

If $p(n) << n^{-1/\varrho} = n^{-v/e}$ then $p^{e} n^{v} << 1$, ie., $p^{e} n^{v} \to 0$

so, $\lim_{n\to\infty} \mathbb{E}[X] = 0$

& by Markov, $\mathbb{P}[X=0] \to 1$ as $n \to \infty$.

Now, we assume $p(n) >> n^{-1/\varrho}$



$$\mathrm{Var}[X] = \sum_{\beta} \mathrm{Var}[X_\beta] + \sum_{\alpha \neq \beta} \mathrm{Cov}[X_\alpha, X_\beta]$$

Let α and β share $t \geq 2$ vertices, then the two copies of H on α & β have at most $t\varrho$ edges in common

$\therefore \mathrm{Cov}[X_\alpha, X_\beta] \leq \mathbb{E}[X_\alpha X_\beta] \leq p^{2e - t\varrho}$, since their union has at least $2e - t\varrho$ edges

The number of pairs α, β sharing t vertices is $\Theta(n^{2v-t})$

$$\therefore \sum_{|\alpha \cap \beta| = t} \mathrm{Cov}[X_\alpha, X_\beta] \leq \Theta(n^{2v-t} p^{2e - t\varrho}) = \Theta((n^{v} p^{e})^{2 - t/v})$$

So $\mathrm{Var}[X] \leq \Theta\left(\sum_{t=2}^{v} (n^{v} p^{e})^{2 - t/v}\right)$

$$\therefore \lim_{n\to\infty} \frac{\mathrm{Var}[X]}{(\mathbb{E}[X])^2} = \lim_{n\to\infty} \Theta\left(\sum_{t=2}^{v} (n^{v} p^{e})^{-t/v}\right) = 0, \text{ since } \lim_{n\to\infty} n^{v} p^{e} = \infty \quad \blacksquare$$

<u>Proposition</u> If H is not balanced then $t(n) = n^{-v/e}$ is not a threshold function for appearance of H.

Proposition If H is not balanced then $t(n) = n^{-v/e}$ is not a threshold function for appearance of H.

Proof Pick a subgraph H' of H s.t. $\frac{e(H')}{v(H')} > \frac{e}{v}$

Let $v/e > \gamma > v'/e'$ & $p(n) = n^{-\gamma}$

Since $p(n) = n^{-\gamma} >> t(n) = n^{-v/e}$, there must be a copy of H w.h.p.

Since $p(n) = n^{-\gamma} << n^{-v'/e'}$, the expected number of copies of H' must tend to 0 as $n \to \infty$

which implies w.h.p. there are no copies of H'

which means there are no copies of H.

contradiction $\blacksquare$

So, what should be the threshold function for the appearance of any graph H?

Proposition If H is not balanced then $t(n) = n^{-v/e}$ is not a threshold function for appearance of H.

Proof Pick a subgraph H' of H s.t. $\frac{e(H')}{v(H')} > \frac{e}{v}$ [Let H' be a subgraph of max density]

Let $v/e > \gamma > v'/e'$ & $p(n) = n^{-\gamma}$

Since $p(n) = n^{-\gamma} >> t(n) = n^{-v/e}$, there must be a copy of H w.h.p.

Since $p(n) = n^{-\gamma} << n^{-v'/e'}$, the expected number of copies of H' must tend to 0 as $n \to \infty$

which implies w.h.p. there are no copies of H'

which means there are no copies of H.

contradiction $\blacksquare$

Theorem For any graph H, let max density of H be $\rho_{max} = \max_{H' \subseteq H} \left(\frac{e(H')}{v(H')}\right)$

Then, $t(n) = n^{-1/\rho_{max}}$ is a threshold function for appearance of H.

Much more is known about Threshold functions in $G(n,p)$.

Theorem [Bollobás & Thomason 1987] Every non-trivial monotone graph property has a threshold function.

Sharp Thresholds

e.g. ① $\mathbb{P}[G(n, \frac{c_n}{n}) \text{ contains a } K_3] \longrightarrow \begin{cases} 0 & \text{if } c_n \to -\infty \\ 1-e^{-c^3/6} & \text{if } c_n \to c, \text{constant} \\ 1 & \text{if } c_n \to \infty \end{cases}$

② $\mathbb{P}[G(n, \frac{\log n + c_n}{n}) \text{ is connected}] \longrightarrow \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-e^{-c}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty \end{cases}$

③ $\mathbb{P}\left[\begin{array}{l} G(n, \frac{\log n + \log\log n + c_n}{n}) \\ \text{has a Hamiltonian cycle} \end{array}\right] \rightarrow \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-e^{-c}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty \end{cases}$

Conjecture [k-colorability Threshold]

For all $k \geq 3$, $\exists$ constant $d_k > 0$ s.t. for any constant $d > 0$

$$\mathbb{P}[G(n, d/n) \text{ is } k\text{-colorable}] \to \begin{cases} 1 & \text{if } d < d_k \\ 0 & \text{if } d > d_k \end{cases}$$

Whats known?

- **Achlioptas & Friedgut 2000**: For every $k \geq 3$, $\exists\, d_k(n)$ s.t. for all $\epsilon > 0$ and $d(n) > 0$,

$$\mathbb{P}[G(n, \tfrac{d(n)}{n}) \text{ is } k\text{-colorable}] \to \begin{cases} 1 & \text{if } d(n) < d_k(n) - \epsilon \\ 0 & \text{if } d(n) > d_k(n) + \epsilon \end{cases}$$

But its not known whether $\lim_{n \to \infty} d_k(n)$ exists.

- **Achlioptas & Naor 2006**: For each fixed $d > 0$, w.h.p. $\chi(G(n, d/n)) \in \{k_d, k_d + 1\}$ where $k_d = \min\{k \in \mathbb{N} : 2k \log k > d\}$
- **Bollobás 1987**: $\chi(G(n, 1/2)) \sim n / 2\log_2 n$ w.h.p.

To learn more about Random Graphs, Random Graph models, and their properties:

(*) ① Frieze & Karonski, Introduction to Random Graphs, 2016.

followed by

② Janson, Luczak, Rucinski, Random Graphs, 2000.

③ Penrose, Random Geometric Graphs, 2003.

(*) ④ van der Hofstad, Random Graphs & Complex Networks,
Vol 1 (2017)
Vol 2 (2021+)

(*) available on author's website

Addendum Probabilistic method in Number Theory & Analysis

Let $\nu(n)$ = # of distinct primes dividing n

Theorem [Hardy-Ramanujan 1917] "almost all n have approx. $\log\log n$ prime factors"

For every $\epsilon > 0$, $\exists\, C$ such that all but ϵ-fraction of $x \in [n]$ satisfy

$$|\nu(x) - \log\log n| \leq C\sqrt{\log\log n}$$

Addendum Probabilistic method in Number Theory & Analysis.

Let $\nu(n)$ = # of distinct primes dividing n

Theorem [Hardy-Ramanujan 1917] "almost all n have approx $\log\log n$ prime factors"

For every $\epsilon > 0$, $\exists\, C$ such that all but ϵ-fraction of $x \in [n]$ satisfy $|\nu(x) - \log\log n| \leq C\sqrt{\log\log n}$

Proof (Outline) [Turán 1934]

Choose $x \in [n]$ u.a. random.

For prime p, let $X_p = \begin{cases} 1 & \text{if } p \mid x \\ 0 & \text{otherwise.} \end{cases}$

Set $M = n^{1/10}$, and $X = \sum_{\substack{p \leq M \\ p \text{ prime}}} X_p$ (# prime factors of x at most $n^{1/10}$)

Addendum Probabilistic method in Number Theory & Analysis

Let $\nu(n)$ = # of distinct primes dividing n

Theorem [Hardy-Ramanujan 1917] "almost all n have $\log\log n$ prime factors"

For every $\epsilon > 0$, $\exists\, C$ such that all but ϵ-fraction of $x \in [n]$ satisfy $|\nu(x) - \log\log n| \le C\sqrt{\log\log n}$

Proof (Outline) [Turán 1934]

Choose $x \in [n]$ u.a. random.

For prime p, let $X_p = \begin{cases} 1 & \text{if } p \mid x \\ 0 & \text{otherwise.} \end{cases}$

Set $M = n^{1/10}$, and $X = \sum_{\substack{p \le M \\ p \text{ prime}}} X_p$ (# prime factors of x at most $n^{1/10}$)

Since x cannot have more than 10 prime factors $> n^{1/10}$, we have $\nu(x) - 10 \le X(x) \le \nu(x)$, i.e., $\boxed{|X(x) - \nu(x)| \le 10}$ — $\circledast$

Hence we can focus on analyzing X.

Since exactly $\lfloor n/p \rfloor$ positive integers $\leq n$ are divisible by p,

$$\mathbb{E}[X_p] = \mathbb{P}[X_p = 1] = \mathbb{P}[p \text{ divides } x] = \lfloor n/p \rfloor / n = \frac{1}{p} + O(1/n)$$

Hence,

$$\mathbb{E}[X] = \sum_{p \leq M} \left(1/p + O(1/n)\right) = \log\log n + O(1),$$ using Merten's Theorem

$$\sum_{p \leq t} 1/p = \log\log t + \underbrace{O(1)}_{\text{Meissel-Merten constant}}$$

Var[X] ?

Intuitively, distinct primes behave independently.

If $pq | n$ then X_p & X_q are independent

If $pq \nmid n$ then there is some covariance contribution when n is large

Since exactly $\lfloor n/p \rfloor$ positive integers $\leq n$ are divisible by p,

$$\mathbb{E}[X_p] = \mathbb{P}[X_p = 1] = \mathbb{P}[p \text{ divides } x] = \lfloor n/p \rfloor / n = \frac{1}{p} + O(1/n)$$

Hence,

$$\mathbb{E}[X] = \sum_{p \leq M} \left(1/p + O(1/n)\right) = \log\log n + O(1),$$ using Merten' Theorem

$$\sum_{p \leq t} 1/p = \log\log t + O(1)$$

Meissel-Merten constant

since exactly $\lfloor n/p \rfloor$ positive integers $\leq n$ are divisible by p,

$$\mathbb{E}[X_p] = \mathbb{P}[X_p = 1] = \mathbb{P}[p \text{ divides } x] = \lfloor n/p \rfloor / n = \frac{1}{p} + O(1/n)$$

Hence,

$$\mathbb{E}[X] = \sum_{p \leq M} \left(1/p + O(1/n)\right) = \log\log n + O(1),$$

using Merten' Theorem

$$\sum_{p \leq t} 1/p = \log\log t + O(1)$$

Meissel-Merten constant

If $p \neq q$, then $X_p X_q = 1 \iff pq \mid x$, so

$$|\mathrm{Cov}[X_p, X_q]| = |\mathbb{E}[X_p X_q] - \mathbb{E}[X_p]\,\mathbb{E}[X_q]| = \left| \frac{\lfloor n/pq \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n} \frac{\lfloor n/q \rfloor}{n} \right| = O(1/n)$$

$$\therefore \sum_{p \neq q} |\mathrm{Cov}[X_p, X_q]| \lesssim M^2/n \lesssim n^{2/10 - 1} = n^{-4/5} \quad \text{—} (**)$$

And, $\mathrm{Var}[X_p] = \mathbb{E}[X_p] - (\mathbb{E}[X_p])^2 = (1/p)(1 - 1/p) + O(1/n)$. —$(***)$

Combining $(**)$ & $(***)$,

$$\mathrm{Var}[X] = \sum_{p \leq M} \mathrm{Var}[X_p] + \sum_{p \neq q} \mathrm{Cov}[X_p, X_q]$$

$$= \sum_{p \leq M} 1/p + O(1) = \log\log n + O(1) \sim \mathbb{E}[X]$$

Since exactly $\lfloor n/p \rfloor$ positive integers $\le n$ are divisible by p,

$$\mathbb{E}[X_p] = \mathbb{P}[X_p = 1] = \mathbb{P}[p \text{ divides } x] = \lfloor n/p \rfloor / n = \frac{1}{p} + O(1/n)$$

Hence,

$$\mathbb{E}[X] = \sum_{p \le M} \left(1/p + O(1/n) \right) = \log\log n + O(1),$$ using Merten's Theorem

$$\sum_{p \le t} 1/p = \log\log t + \underset{\substack{\downarrow \\ \text{Meissel-Merten constant}}}{O(1)}$$

If $p \ne q$, then $X_p X_q = 1 \iff pq \mid x$, so

$$|\mathrm{Cov}[X_p, X_q]| = |\mathbb{E}[X_p X_q] - \mathbb{E}[X_p]\,\mathbb{E}[X_q]| = \left| \frac{\lfloor n/pq \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n} \frac{\lfloor n/q \rfloor}{n} \right| = O(1/n)$$

$$\therefore \sum_{p \ne q} |\mathrm{Cov}[X_p, X_q]| \lesssim M^2/n \lesssim n^{2/10 - 1} = n^{-4/5} \quad \text{—} \; (**)$$

And, $$\mathrm{Var}[X_p] = \mathbb{E}[X_p] - (\mathbb{E}[X_p])^2 = (1/p)(1 - 1/p) + O(1/n). \quad \text{—} \; (***)$$

Combining $(**)$ & $(***)$, $$\mathrm{Var}[X] = \sum_{p \le M} \mathrm{Var}[X_p] + \sum_{p \ne q} \mathrm{Cov}[X_p, X_q]$$
$$= \sum_{p \le M} 1/p + O(1) = \log\log n + O(1) \sim \mathbb{E}[X]$$

By Chebyshev, for every $\lambda > 0$,

$$\mathbb{P}\left[|X - \log\log n| \ge \lambda \sqrt{\log\log n} \right] \le \frac{(\mathrm{Var}[X])^2}{\lambda^2 (\log\log n)} = \frac{1}{\lambda^2} + o(1),$$ Now replace X by ν by $(*)$. ∎

Weierstrass Approximation Theorem 1885

"Every continuous function can be approximated by a polyn."

Let $f:[0,1]\to\mathbb{R}$ be a continuous function

For any $\epsilon>0$, $\exists$ polynomial $p(x)$ s.t. $|p(x)-f(x)|\leq\epsilon \quad \forall x\in[0,1]$.

Weierstrass Approximation Theorem 1885

"Every continuous function can be approximated by a polyn."

Let $f : [0,1] \to \mathbb{R}$ be a continuous function

For any $\epsilon > 0$, $\exists$ polynomial $p(x)$ s.t. $|p(x) - f(x)| \leq \epsilon \quad \forall x \in [0,1]$.

Proof [Bernstein 1912] Let $P_n(x) = \sum_{i=0}^{n} B_i(x) f(i/n)$

where $B_i(x) = \binom{n}{i} x^i (1-x)^{n-i} =$ for $0 \leq i \leq n$

Weierstrass Approximation Theorem 1885

"Every continuous function can be approximated by a polyn."

Let $f:[0,1]\to\mathbb{R}$ be a continuous function

For any $\epsilon>0$, $\exists$ polynomial $p(x)$ s.t. $|p(x)-f(x)|\leq\epsilon \quad \forall x\in[0,1]$.

Proof [Bernstein 1912] Let $P_n(x)=\sum_{i=0}^{n} B_i(x)\, f(i/n)$

where $B_i(x)=\binom{n}{i}x^i(1-x)^{n-i} = \mathbb{P}[\mathrm{Bin}(n,x)=i]$ for $0\leq i\leq n$

Binomial distribution

Claim $|P_n(x)-f(x)|\leq\epsilon \quad \forall x\in[0,1]$

Weierstrass Approximation Theorem 1885

"Every continuous function can be approximated by a polyn."

Let $f:[0,1]\to\mathbb{R}$ be a continuous function

For any $\epsilon>0$, $\exists$ polynomial $p(x)$ s.t. $|p(x)-f(x)|\leq\epsilon$ $\forall x\in[0,1]$.

Proof [Bernstein 1912] Let $P_n(x)=\sum_{i=0}^{n} B_i(x)\, f(i/n)$

where $B_i(x)=\binom{n}{i}x^i(1-x)^{n-i} = \mathbb{P}[\mathrm{Bin}(n,x)=i]$ for $0\leq i\leq n$

$$|P_n(x)-f(x)| \leq \sum_{i:|i-nx|\leq n^{2/3}} B_i(x)\,|f(i/n)-f(x)| + \sum_{i:|i-nx|>n^{2/3}} B_i(x)\,[|f(i/n)|+|f(x)|]$$

Weierstrass Approximation Theorem 1885

"Every continuous function can be approximated by a polyn."

Let $f:[0,1]\to\mathbb{R}$ be a continuous function

For any $\epsilon>0$, $\exists$ polynomial $p(x)$ s.t. $|p(x)-f(x)|\leq\epsilon$ $\forall x\in[0,1]$.

Proof [Bernstein 1912] Let $P_n(x)=\sum_{i=0}^{n} B_i(x)\, f(i/n)$

where $B_i(x)=\binom{n}{i}x^i(1-x)^{n-i} = \mathbb{P}[\mathrm{Bin}(n,x)=i]$ for $0\leq i\leq n$

$$|P_n(x)-f(x)| \leq \sum_{i:|i-nx|\leq n^{2/3}} B_i(x)\,|f(i/n)-f(x)| + \sum_{i:|i-nx|>n^{2/3}} B_i(x)\,\big[|f(i/n)|+|f(x)|\big]$$

Since $\mathrm{Bin}(n,x)$ has mean nx and variance $nx(1-x)\leq n$,

by Chebyshev's ineq, $\sum_{i:|i-nx|>n^{2/3}} B_i(x) = \mathbb{P}\big[|\mathrm{Bin}(n,x)-nx|>n^{2/3}\big]\leq n^{-1/3}$

Weierstrass Approximation Theorem 1885

"Every continuous function can be approximated by a polyn."

Let $f:[0,1]\to\mathbb{R}$ be a continuous function

For any $\epsilon>0$, $\exists$ polynomial $p(x)$ s.t. $|p(x)-f(x)|\le\epsilon$ $\forall x\in[0,1]$.

Proof [Bernstein 1912] Let $P_n(x)=\sum_{i=0}^{n} B_i(x)\, f(i/n)$

where $B_i(x)=\binom{n}{i}x^i(1-x)^{n-i}=\mathbb{P}[\mathrm{Bin}(n,x)=i]$ for $0\le i\le n$

$$|P_n(x)-f(x)|\le \sum_{i:|i-nx|\le n^{2/3}} B_i(x)\,|f(i/n)-f(x)| + \sum_{i:|i-nx|>n^{2/3}} B_i(x)\,[|f(i/n)|+|f(x)|]$$

Since $\mathrm{Bin}(n,x)$ has mean nx and variance $nx(1-x)\le n$,

by Chebyshev's ineq, $\sum_{i:|i-nx|>n^{2/3}} B_i(x) = \mathbb{P}[|\mathrm{Bin}(n,x)-nx|>n^{2/3}]\le n^{-1/3}$

And, since f is continuous on a compact set $[0,1]$,

f is uniformly continuous, $\exists\delta>0$ s.t. $|f(x)-f(y)|\le\epsilon/2$ $\forall |x-y|<\delta$ in $[0,1]$.

f is bounded, $\exists M>0$ s.t. $|f(x)|\le M$ $\forall x\in[0,1]$.

$\therefore \exists n\in\mathbb{Z}^+$ s.t. $n^{-1/3}<\delta$ and $\sum_{i:|i-nx|>n^{2/3}} B_i(x)<n^{-1/3}\le \epsilon/4M$

Weierstrass Approximation Theorem 1885

"Every continuous function can be approximated by a polyn."

Let $f:[0,1]\to\mathbb{R}$ be a continuous function

For any $\epsilon>0$, $\exists$ polynomial $p(x)$ s.t. $|p(x)-f(x)|\leq\epsilon$ $\forall x\in[0,1]$.

Proof [Bernstein 1912] Let $P_n(x)=\sum_{i=0}^{n}B_i(x)\,f(i/n)$

where $B_i(x)=\binom{n}{i}x^i(1-x)^{n-i}=\mathbb{P}[\mathrm{Bin}(n,x)=i]$ for $0\leq i\leq n$

$$|P_n(x)-f(x)|\leq\sum_{i:|i-nx|\leq n^{2/3}}B_i(x)\,|f(i/n)-f(x)|+\sum_{i:|i-nx|>n^{2/3}}B_i(x)\,[|f(i/n)|+|f(x)|]$$

$$\leq\sum_{i:|i/n-x|\leq n^{-1/3}<\delta}B_i(x)\,|f(i/n)-f(x)|+\epsilon/4M\,(M+M)$$

$$\leq\frac{\epsilon}{2}\sum_{i:\ldots}B_i(x)+2M\,\epsilon/4M$$

$$\leq\frac{\epsilon}{2}\underbrace{\sum_{i=0}^{n}B_i(x)}_{=?}+\epsilon/2=\epsilon/2\,(1)+\epsilon/2=\epsilon$$

$\blacksquare$

Weierstrass Approximation Theorem 1885

"Every continuous function can be approximated by a polyn."

Let $f:[0,1]\to\mathbb{R}$ be a continuous function

For any $\epsilon>0$, $\exists$ polynomial $p(x)$ s.t. $|p(x)-f(x)|\le\epsilon \quad \forall x\in[0,1]$.

Proof [Bernstein 1912] Let $P_n(x)=\sum_{i=0}^{n} B_i(x)\, f(i/n)$

where $B_i(x)=\binom{n}{i}x^i(1-x)^{n-i} = \mathbb{P}[\mathrm{Bin}(n,x)=i]$ for $0\le i\le n$

$$|P_n(x)-f(x)| \le \sum_{i:|i-nx|\le n^{2/3}} B_i(x)\,|f(i/n)-f(x)| + \sum_{i:|i-nx|>n^{2/3}} B_i(x)\,[|f(i/n)|+|f(x)|]$$

$$\le \sum_{i:|i/n-x|\le n^{-1/3}<\delta} B_i(x)\,|f(i/n)-f(x)| + \epsilon/4M\,(M+M)$$

$$\le \frac{\epsilon}{2}\sum_{i:\dots} B_i(x) + 2M\,\epsilon/4M$$

$$\le \frac{\epsilon}{2}\underbrace{\sum_{i=0}^{n} B_i(x)}_{=(x+(1-x))^n=1} + \epsilon/2 = \epsilon/2\,(1)+\epsilon/2 = \epsilon$$

$\blacksquare$

So far we have seen that two elementary probabilistic ideas have many fruitful applications in Combinatorics, (and Number theory, Analysis):

First moment Method

For X non-neg. integer valued r.v,

$\mathbb{E}[X] < 1 \Rightarrow \mathbb{P}[X=0] > 0$, i.e. $\exists$ instance with $X=0$

$\mathbb{E}[X] \geq k \Rightarrow \mathbb{P}[X \geq k] > 0$, i.e. $\exists$ instance with $X \geq k$

Second moment Method

Chebyshev's Ineq – $\mathbb{P}[|X-\underbrace{\mu}_{\mathbb{E}[X]}| \geq t] \leq \mathrm{Var}[X]/t^2$

For X non-neg. integer valued r.v., $\mathbb{P}[X=0] \leq \dfrac{\mathrm{Var}[X]}{(\mathbb{E}[X])^2}$

so, if $\mathrm{Var}[X] << \mu^2$ then $\mathbb{P}[X>0] \to 1$.

We are interested in inequalities of the form:

$\mathbb{P}[|X-\mu| \geq t] \leq f_X(t)$ where $f_X(t) \to 0$ rapidly when t grows.

$\mathbb{P}[\,|X-\mu| \geq t\,] \leq f_X(t)$ type of inequalities (often called "large deviation" or "concentration of measure" inequalities) have been studied (& discovered & applied) for the past 60 years using —

1950s
- Moment Generating Method leading to Chernoff-Hoeffding

1960s
- Martingales & Bounded Differences approach leading to Bounded Diff. Ineq. & Azuma-Hoeffding

1990s
- Isoperimetric inequalities in Geometry and Product Spaces leading to Talagrand's ineq.

2000s
- Information-theoretic approaches through log-Sobolev inequalities and Entropy

⋮

What is "concentration of measure" phenomenon?

In a long sequence of tossing a fair coin,
it is highly likely that heads will come up nearly half the time.

Traditionally, this statement is made precise by the Law of Large numbers — a limit Theorem ($\bar{X}_n \to \mu$ as $n \to \infty$).

What is "concentration of measure" phenomenon?

In a long sequence of tossing a fair coin,
it is highly likely that heads will come up nearly half the time.

Traditionally, this statement is made precise by the Law of Large numbers — a limit Theorem ($\overline{X}_n \to \mu$ as $n \to \infty$).
But what about the "rate of convergence"?
Consider an iid sequence of Bernoulli r.v.s $\langle \epsilon_i \rangle_{i=1}^{N}$ with $\mathbb{P}[\epsilon_i = 1] = \mathbb{P}[\epsilon_i = -1] = 1/2$, ind. Then it can shown ("Chernoff-type")

$$\mathbb{P}\left[\left| B_N - \overset{\leftarrow \text{mean}}{\frac{N}{2}} \right| \geq t\right] \leq 2\exp\left(-\frac{2t^2}{N}\right),$$

where $B_N = \#1$s in the seq. $\langle \epsilon_i \rangle_{i=1}^{N}$

What is "concentration of measure" phenomenon?

In a long sequence of tossing a fair coin,
it is highly likely that heads will come up nearly half the time.

Traditionally, this statement is made precise by the Law of Large numbers — a limit Theorem ($\overline{X}_n \to \mu$ as $n \to \infty$).

But what about the "rate of convergence"?

Consider an iid sequence of Bernoulli r.v.s $\langle \epsilon_i \rangle_{i=1}^{N}$ with $\mathbb{P}[\epsilon_i = 1] = \mathbb{P}[\epsilon_i = -1] = 1/2$, ind. Then it can shown ("Chernoff-type")

$$\mathbb{P}\left[\left| B_N - \overset{\leftarrow \text{mean}}{\frac{N}{2}} \right| \geq t\right] \leq 2\exp\left(\frac{-2t^2}{N}\right),$$ where $B_N = \#1\text{s}$ in the seq. $\langle \epsilon_i \rangle_{i=1}^{N}$

Lets define $X = \sum_{i=1}^{N} \epsilon_i \; (= 2B_N - N)$, then above statement becomes

$$\mathbb{P}\left[\left|\frac{X}{N}\right| \geq t\right] \leq 2\exp\left(\frac{-t^2 N}{2}\right),$$ that is, $\frac{X}{N}$ is essentially zero

M. Talagrand states that "concentration of measure phenomenon" is essentially:

A r.v. that depends (in a "smooth" way) on the influence of many independent variables (but not too much on any one of them) is $\underbrace{\text{essentially constant.}}_{\text{in the sense we described previously}}$

How can we prove the sort of inequalities shown previously?

Here is Bernstein's idea (use moment gen. ftn.) —

Let $X = \sum_{i=1}^{n} X_i$, $X_i = \begin{cases} 1 & \text{with probab. } p \\ 0 & \text{with probab. } q = 1-p \end{cases}$, independent trials

For $\lambda > 0$, $\mathbb{P}[X > m] = \mathbb{P}[e^{\lambda X} > e^{\lambda m}] \leq \dfrac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda m}}$, by Markov's ineq.

Let us bound $\mathbb{E}[e^{\lambda X}]$ (moment generating function of r.v. X)

Here is Bernstein's idea (use moment gen. ftn.) —

Let $X = \sum_{i=1}^{n} X_i$, $X_i = \begin{cases} 1 & \text{with probab. } p \\ 0 & \text{with probab. } q = 1-p \end{cases}$, independent trials

For $\lambda > 0$, $\mathbb{P}[X > m] = \mathbb{P}[e^{\lambda X} > e^{\lambda m}] \leq \dfrac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda m}}$, by Markov's ineq.

Let us bound $\mathbb{E}[e^{\lambda X}]$ (moment generating function of r.v. X)

$$\mathbb{E}[e^{\lambda X}] = \mathbb{E}[e^{\lambda \sum_i X_i}] = \mathbb{E}[\prod_i e^{\lambda X_i}] = \prod_i \mathbb{E}[e^{\lambda X_i}] \text{ , by independence}$$
$$= (pe^{\lambda} + q)^n$$

Here is Bernstein's idea (use moment gen. ftn.) —

let $X=\sum_{i=1}^{n} X_i$, $X_i=\begin{cases}1 & \text{with probab. } p\\ 0 & \text{with probab. } q=1-p\end{cases}$, independent trials

For $\lambda>0$, $\mathbb{P}[X>m] = \mathbb{P}[e^{\lambda X} > e^{\lambda m}] \leq \dfrac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda m}}$, by Markov's ineq.

Let us bound $\mathbb{E}[e^{\lambda X}]$ (moment generating function of r.v. X)

$\mathbb{E}[e^{\lambda X}] = \mathbb{E}[e^{\lambda \sum_i X_i}] = \mathbb{E}[\prod_i e^{\lambda X_i}] = \prod_i \mathbb{E}[e^{\lambda X_i}]$, by independence

$= (pe^{\lambda}+q)^n$

Now, let $m=(p+t)n$
& note $\mu = \mathbb{E}[X] = np$,

$$\mathbb{P}[X-\mu > tn] = \mathbb{P}[X-pn>tn] = \mathbb{P}[X>m] \leq \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda m}} = \frac{(pe^{\lambda}+q)^n}{e^{\lambda(p+t)n}} = \left(\frac{pe^{\lambda}+q}{e^{\lambda(p+t)}}\right)^n$$

Here is Bernstein's idea (use moment gen. ftn.) —

Let $X = \sum_{i=1}^{n} X_i$, $X_i = \begin{cases} 1 & \text{with probab. } p \\ 0 & \text{with probab. } q = 1-p \end{cases}$, independent trials

For $\lambda > 0$, $\mathbb{P}[X > m] = \mathbb{P}[e^{\lambda X} > e^{\lambda m}] \leq \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda m}}$, by Markov's ineq.

Let us bound $\mathbb{E}[e^{\lambda X}]$ (moment generating function of r.v. X)

$$\mathbb{E}[e^{\lambda X}] = \mathbb{E}[e^{\lambda \sum_i X_i}] = \mathbb{E}[\prod_i e^{\lambda X_i}] = \prod_i \mathbb{E}[e^{\lambda X_i}] \text{ , by independence}$$
$$= (pe^{\lambda} + q)^n$$

Now, let $m = (p+t)n$
& note $\mu = \mathbb{E}[X] = np$,

$$\mathbb{P}[X - \mu > tn] = \mathbb{P}[X - pn > tn] = \mathbb{P}[X > m] \leq \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda m}} = \frac{(pe^{\lambda} + q)^n}{e^{\lambda (p+t) n}} = \left(\frac{pe^{\lambda} + q}{e^{\lambda (p+t)}}\right)^n$$

To minimize this bound, set $e^{\lambda} = (p+t)q / p(q-t)$, assuming $0 \leq t < q$,

$$\mathbb{P}[X - pn > tn] \leq \left(\left(\frac{p}{p+t}\right)^{p+t} \left(\frac{q}{q-t}\right)^{q-t} \right)^n$$

1952

Thm [Chernoff Bound] Let $X=\sum_{i=1}^{n} X_i$ where $X_i=\begin{cases}1, & \text{with prob. } p_i \\ 0, & \text{with prob. } q_i=1-p_i\end{cases}$ independent trials

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

(note, $\mu=\mathbb{E}[X]=np$)

1952

Thm [Chernoff Bound] Let $X = \sum_{i=1}^{n} X_i$ where $X_i = \begin{cases} 1, & \text{with prob. } p_i \\ 0, & \text{with prob. } q_i = 1 - p_i \end{cases}$ independent trials

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

(note, $\mu = \mathbb{E}[X] = np$)

Proof Same steps as before, except

$$\mathbb{E}[e^{\lambda X}] = \ldots\ldots = \prod_i \mathbb{E}[e^{\lambda X_i}] = \prod_{i=1}^{n} (p_i e^{\lambda} + q_i)$$

1952

Thm [Chernoff Bound] Let $X = \sum_{i=1}^{n} X_i$ where $X_i = \begin{cases} 1, & \text{with prob. } p_i \\ 0, & \text{with prob. } q_i = 1 - p_i \end{cases}$ independent trials

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

(note, $\mu = \mathbb{E}[X] = np$)

Proof. Same steps as before, except

$$\mathbb{E}[e^{\lambda X}] = \ldots\ldots = \prod_i \mathbb{E}[e^{\lambda X_i}] = \prod_{i=1}^{n} (p_i e^{\lambda} + q_i)$$

$$\leq \left(\frac{\sum (p_i e^{\lambda} + q_i)}{n} \right)^n, \text{ by A.M-G.M inequality}$$

$$= (p e^{\lambda} + q)^n$$

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 $\mu = \mathbb{E}[X] = \sum p_i = np$, $q = 1 - p$

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

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

How?!?!

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

$$\mathbb{E}[e^{\lambda X}] = \dots = \prod_i \mathbb{E}[e^{\lambda X_i}] \leq \prod_i (p_i e^{\lambda} + q_i)$$
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 $\left(\left(\frac{p}{p+t}\right)^{p+t}\left(\frac{q}{q-t}\right)^{q-t}\right)^n$?

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

Proof $f'(t) = \ln \frac{p(q-t)}{(p+t)q}$ 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.