4. The First and Second Moment Methods

- Theorem 4.1 (Markov's Inequality). Let $X \geq 0$ be a nonnegative random variable with $E[X] < \infty$. Then for any $a > 0$,

$$\Pr(X \geq a) \leq E[X]/a.$$

Proof. In general terms:

$$E[X] = \int_0^\infty x \, dP = \int_0^a x\, dP + \int_a^\infty x\, dP \geq 0 + a \int_a^\infty dP$$
$$= a \Pr(X \geq a).$$

$$\Rightarrow \Pr(X \geq a) \leq E[X]/a.$$

[For clarity, let us review the proof in case X is integer-valued and $a \in \mathbb{N}_+$:

$$E[X] = \sum_{k=0}^{\infty} k \Pr(X=k) = \sum_{k=0}^{a-1} k \Pr(X=k) + \sum_{k=a}^{\infty} k \Pr(X=k)$$
$$\geq 0 + a \sum_{k=a}^{\infty} \Pr(X=k) = a \Pr(X \geq a).\ ]$$

$\square$

- Corollary 1. If $X \geq 0$ is an integer-valued random variable with $E[X] = \mu$, then

$$\Pr(X \geq 1) \leq \mu, \quad \text{i.e.} \quad \Pr(X=0) \geq 1-\mu. \quad \square$$

- Let then X_n, $n \geq 0$, be a sequence of random variables counting the number of occurrences of some feature in a random structure of size n. If $\mu_n = E[X_n] \to 0$ as $n \to \infty$, then the feature almost surely doesn't occur at all in sufficiently large structures.
- This the "first moment method".

- Let's say that we to the contrary want to show that some interesting feature does occur almost surely in sufficiently large random structures, i.e. that

$$\Pr(X_n > 0) \to 1 \text{ as } n \to \infty.$$

Now even $E[X_n] \to \infty$ isn't sufficient. (Consider e.g. a case where there are n equally probable structures of size n, one of which has 2^n occurrences of the feature and others 0.)

- However even this works out if the variance $\sigma_n^2 = \mathrm{Var}[X_n]$ stays small (more specifically if $\sigma_n^2/\mu_n^2 \to 0$).

- This is the "second moment method".

- Theorem 4.2 (Chebyshev's Inequality) Let X be a random variable with $E[X] = \mu$, $\mathrm{Var}[X] = \sigma^2 > 0$. Then for any $t > 0$,

$$\Pr(|X - \mu| \geq t\sigma) \leq \frac{1}{t^2}.$$

Proof. By definition, $\sigma^2 = E[(X-\mu)^2]$. Now the random variable $(X-\mu)^2$ is nonnegative, so by Markov's inequality:

$$\begin{aligned}\Pr(|X-\mu| \geq t\sigma) &= \Pr((X-\mu)^2 \geq t^2\sigma^2) \\ &\leq E[(X-\mu)^2] / t^2\sigma^2 \\ &= \frac{1}{t^2}. \quad \square\end{aligned}$$

- Corollary 2. If $\mu_n = E[X_n] > 0$ for n large, and $\frac{\sigma_n^2}{\mu_n^2} \to 0$ as $n \to \infty$, then $\Pr(X_n > 0) \to 1$ as $n \to \infty$.

Proof. If $X_n = 0$, then $|X_n - \mu_n| = \mu_n$. Choose $t_n = \frac{\mu_n}{\sigma_n}$ above; then:
$\Pr(X_n = 0) \leq \Pr(|X_n - \mu_n| \geq \mu_n) \leq \sigma_n^2/t_n^2 \to 0$ as $n \to \infty$. $\square$

- Example 1. Random graphs and threshold functions.

- Consider the family $\mathcal{G}_n$ of all (labelled, undirected) graphs on the vertex set $[n] = \{1, \dots, n\}$. Denote $N = \binom{n}{2}$. Then $|\mathcal{G}_n| = 2^N$.

  For a given $p \in [0,1]$, make $\mathcal{G}_n$ into a probability space $\mathcal{G}(n,p)$, by picking each edge uniformly and independently at random with prob. p.

  Thus, for any specific graph $H \in \mathcal{G}_n$ with $M \le N$ edges,
  $$\Pr(G_p = H) = p^M \underbrace{(1-p)}_{q}{}^{N-M}.$$

- This is called the Erdös – Rényi ensemble of random graphs. (Unfairly, because the ensemble was considered already before Erdös & Renyi's 1960 paper by E. Gilbert in 1959.)

- One of Erdös & Rényi's discoveries was that many structural properties of $\mathcal{G}(n,p)$ random graphs "emerge" at sharply defined threshold densities p.

- Definition Function $t = t(n)$ is a threshold for graph property Q if

  (i) $p(n) \ll t(n) \Rightarrow \Pr(G \in \mathcal{G}(n,p) \text{ has } Q) \longrightarrow 0$ as $n \to \infty$,

  (ii) $p(n) \gg t(n) \Rightarrow \Pr(G \in \mathcal{G}(n,p) \text{ has } Q) \longrightarrow 1$ as $n \to \infty$.

- Here "$f(n) \ll g(n)$" denotes $f(n) = o(g(n))$, i.e. $\frac{f(n)}{g(n)} \to 0$ as $n \to \infty$.

- Terminology: "random $G \in \mathcal{G}(n,p)$ has/doesn't have property Q asymptotically almost surely (a.a.s.)"

- Consider e.g. the property $Q \equiv$ "$\omega(G) \geq 4$", i.e. "G contains a 4-clique".

Theorem 4.3 The property $Q \equiv$ "$\omega(G) \geq 4$" has threshold $t(n) = n^{-2/3}$.

Proof.

(i) We apply the first-moment method to show that $t(n)$ is an upper threshold for Q.

Let $X = X(G)$ be a random variable that counts the number of 4-cliques in G. Then X can be represented as a sum of indicator variables

$$X = \sum_S X_S,$$

where $X_S(G) \sim$ "a given set S of 4 vertices forms a clique in G".

Clearly $\Pr(X_S = 1) = p^{\binom{4}{2}} = p^6$, and by linearity of expectation:

$$E[X] = \sum_S E[X_S] = \binom{n}{4} p^6 \sim n^4 p^6 \longrightarrow 0 \text{ as } n \to \infty,$$

if $p(n) \ll n^{-2/3}$.

By Corollary 1, thus, if $p(n) \ll n^{-2/3}$, then

$$\Pr(X \geq 1) \leq E[X] \to 0 \text{ as } n \to \infty.$$

(ii) If $p(n) \gg n^{-2/3}$, then $E[X] \to \infty$ as $n \to \infty$.
To derive from this, by the second-moment method, that also $\Pr(X \geq 1) \to 1$, we need to bound $\mathrm{Var}[X]$. (More precisely show that $\mathrm{Var}[X] = o(E[X]^2)$.)

- Lemma 4.4 For $X = X_1 + \cdots + X_m$,

$$\mathrm{Var}[X] = \sum_{i=1}^{m} \mathrm{Var}[X_i] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j],$$

where $\mathrm{Cov}[Y, Z] = E[YZ] - E[Y]E[Z]$ is the covariance of variables Y, Z. (Note that if $Y \perp Z$, then $E[YZ] = E[Y]E[Z]$ and $\mathrm{Cov}[Y, Z] = 0$.) $\square$

- Lemma 4.5 If X is an indicator variable, then

$$\mathrm{Var}[X] \leq E[X].$$

Proof. Denote $\Pr(X = 1) = p$. Then

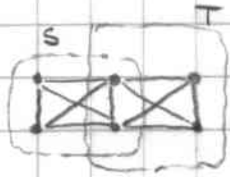
$$\begin{aligned} \mathrm{Var}[X] &= E[(X-p)^2] = p(1-p)^2 + (1-p)(-p)^2 \\ &= p(1-p) \\ &\leq p = E[X]. \quad \square \end{aligned}$$

- By lemmas 4.4 & 4.5, in this case thus:

$$\begin{aligned} \mathrm{Var}[X] &= \sum_{S} \mathrm{Var}[X_S] + \sum_{S \neq T} \mathrm{Cov}[X_S, X_T] \\ &\leq \sum_{S} E[X_S] + \sum_{S \neq T} \mathrm{Cov}[X_S, X_T] \\ &= O(n^4 p^6) + \sum_{S \neq T} \mathrm{Cov}[X_S, X_T]. \end{aligned}$$

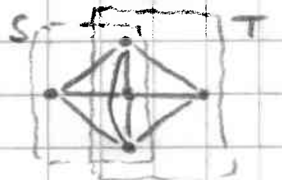
- Now if $X_S \perp X_T$, then $\mathrm{Cov}[X_S, X_T] = 0$, so we only need to worry about cases where cliques on S and T have common edges, i.e. $|S \cap T| = 2$ or $|S \cap T| = 3$.

  - Case $|S \cap T| = 2$: $\binom{n}{6}$ pairs, each with $\mathrm{Cov}[X_S, X_T] \leq E[X_S X_T] = p^{11}$.



6 vertices, 11 edges

  - Case $|S \cap T| = 3$: $\binom{n}{5}$ pairs, each with $\mathrm{Cov}[X_S, X_T] \leq E[X_S X_T] = p^9$.



5 vertices, 9 edges

- Thus altogether:

$$\mathrm{Var}[X] = O(n^4 p^6) + O(n^6 p^{11}) + O(n^5 p^9)$$

$$= o(n^8 p^{12}) \quad , \text{ since } p \gg n^{-2/3}.$$

On the other hand,

$$E[X]^2 \sim (n^4 p^6)^2 = n^8 p^{12}.$$

- Hence by Corollary 2, if $p(n) \gg n^{-2/3}$

$$\Pr(X > 0) \to 1 \quad \text{as } n \to \infty. \quad \square$$

- This overlap-order counting technique can be extended to yield a very general result: define the <u>density</u> of a graph $G = (V, E)$ as $\rho(G) = |E|/|V|$, and say that G is <u>balanced</u> if $\rho(G') \leq \rho(G)$ for all subgraphs G' of G.

- <u>Theorem</u> 4.6 (Erdős & Rényi 1960). Let H be a balanced graph. Then the graph property "G has a subgraph isomorphic to H" has threshold $n^{-1/\rho(H)}$. $\square$

- Technical note: Recall that Cor. 2 claims
  $Var[X] = o(E[X]^2) \Rightarrow X > 0$ a.a.s.

This can be strengthened to:

Corollary 2'. Assume that $E[X] > 0$ and $Var[X] = o(E[X]^2)$.
Then $X \sim E[X]$ a.a.s.

Proof. By Chebyshev's inequality, for every $\varepsilon > 0$:

$$\Pr(|X - E[X]| \geq \varepsilon E[X]) \leq \frac{Var[X]}{\varepsilon^2 E[X]^2}.$$

$$\Rightarrow \Pr((1-\varepsilon)E[X] \leq X \leq (1+\varepsilon)E[X]) = 1 - o\left(\frac{1}{\varepsilon^2}\right).$$

$\square$