3 Balls into Bins, Markov and Chebyshev Inequalities

The value of a random variable may be distant to its expectation. We will introduce a toy model "balls-into-bins" and then introduce the *concentration of probability measures*. We will see that the analysis of many randomized algorithms will eventually reduce to some basic questions about balls and bins.

3.1 Balls-into-bins model

Ball-into-bins is the following random process:

Throw m balls into n bins uniformly at random.

Many interesting questions can be asked about the process and today we mainly investigate some of them.

Example 3.1.
Empty bins / coupon collector
How many balls do we have to throw so that there are no empty bins? Clearly it is precisely the model of coupon collector. So the expected number of balls is $nH_n \approx n\ln n$. If we have thrown m balls, the probability that there is still at least one empty bin is $\Pr[\exists\text{empty bin}] \le n\left(1-\frac{1}{n}\right)^m \approx n\mathrm{e}^{-m/n}$. Thus if $m = n\ln n + cn$. The probability is upper bounded by e^{-c}.

Example 3.2.
Birthday paradox / collision
Consider the probability that some bin has more than one ball. The problem can also be described as the probability that two persons in the class have the same birthday, hence is called **birthday paradox**. Since each ball is thrown independently, the probability that no collision occurs at the k-th ball is thrown is $\frac{n-k+1}{n}$. Thus

$$\Pr[\text{no shared birthday}] = \prod_{k=1}^{m} \frac{n-k+1}{n} = \prod_{k=1}^{m-1}\left(1-\frac{k}{n}\right)$$
$$\le \exp\left(-\frac{\sum_{k=1}^{m-1} k}{n}\right) = \exp\left(-\frac{m(m-1)}{2n}\right),$$

where we use the inequality $1+x \le \mathrm{e}^x$. For $m = O(\sqrt{n})$, the probability can be arbitrarily close to 0.

It is called a *paradox* since it may seem surprising that only 23 individuals are required to reach a 50% probability of a shared birthday. One may believe that there have to be more individuals if they hasn't seen the result before.

Example 3.3.

Max load

Max load is the maximum number of balls among n bins. Let X_i be the number of balls in the i-th bin. We need to compute $X = \max_{i\in[n]} X_i$. Today we assume that $m = n$, and we only try to find a number k such that $Pr[X > k] = O(1)$.By the union bound, we have

$$\Pr[\max_i X_i > k] = \Pr[\exists i \colon X_k > k] \le n \Pr[X_i > k] \,.$$

So it suffices to determine the k such that $\Pr[X_i > k] = O(1/n)$.
Again by the union bound we have

$$\Pr[X_i > k] \le \binom{n}{k} \frac{1}{n^k} \le \left(\frac{k}{\mathrm{e}}\right)^{-k} \,.$$

It is easy to verify that $\left(\frac{k}{\mathrm{e}}\right)^{-k} = O\left(\frac{1}{n}\right)$ by setting $k = O\left(\frac{\log n}{\log\log n}\right)$.

Stirling's formula gives $k! \sim \sqrt{2\pi k}\left(\frac{k}{\mathrm{e}}\right)^k$.

3.2 Markov's inequality

For nonnegative random variables, the Markov's inequality is the first concentration inequality.

Theorem 3.1. Markov's inequality

Let X be a nonnegative random variable. Then

$$\Pr[X \ge t] \le \frac{\mathbb{E}[X]}{t} \,.$$

Proof. $\mathbb{E}X = \mathbb{E}[X \mid X \ge t] \Pr[X \ge t] + \mathbb{E}[X \mid X < t] \Pr[X < t] \ge t \Pr[X \ge t]$. □

There are lots of applications of Markov's inequality. For example, we can apply Markov's inequality directly to the Max Load problem mentioned above. Since $\mathbb{E}[X_i] = 1$, we derive that

$$\Pr\left[X_i > \frac{\log n}{\log\log n}\right] \le \frac{\log\log n}{\log n} \,,$$

but the upper bound is weaker than the previous result $O(1/n)$, mainly due to the fact that we only use $\mathbb{E}[X_i]$ to estimate the upper bound while lacking a lot of other information. Similarly, if we apply Markov's inequality to coupon collector, we can only get

$$\Pr[X > 100 n \ln n] < \frac{1}{100} \,,$$

which is much weaker than $\Pr[X > n\ln n + cn] < \mathrm{e}^{-c}$.

An important application of the Markov's inequality is *Monte Carlo vs. Las Vegas*.

We have seen many randomized algorithms, which appear to differ in their behavior: some always run in a fixed amount of time, but may produce the wrong answer (e.g., polynomial identity testing, Karger's min-cut); while others always succeed, but may run for longer than you expect (e.g., coupon collector, QuickSelect).
We generally divide randomized algorithms into two classes: **Monte Carlo** algorithms and **Las Vegas** algorithms, depending on of which events we are talking about the probability, $\Pr[\text{output is correct}]$ or $\Pr[\text{algorithm halts}]$.

Originally named by László Babai.

- **Monte Carlo algorithm**: always halt in finite time but may output the wrong answer, i.e., $\Pr[\text{error}]$ is larger than 0. These algorithms are further divided into those with *one-sided* and *two-sided error*.
- **Las Vegas algorithm**: the output is always correct but the running time may be unbounded. However, the expected running time is required to be bounded.

The class of decision problems solvable by a polynomial time Monte Carlo algorithm with one-sided error is known as RP ("Randomized Polynomial time"). One-sided error algorithms only make errors in one direction. For example, if the true answer is "yes", then $\Pr[\text{yes}] \geq \varepsilon > 0$ while if the true answer is "no" then $\Pr[\text{no}] = 1$. In other words, a "yes" answer is guaranteed to be accurate, while a "no" answer is uncertain. We can increase our confidence in a "no" answer by running the algorithm many times (using independent random bits each time); we can then output "yes" if we see any "yes" answer after t repetitions, and output "no" otherwise. The probability of error in this scheme is at most $(1-\varepsilon)^t$, so if we want to make this smaller than any desired $\delta > 0$ it suffices to take the number of trials t to be $O(\log 1/\delta)$ (where the constant in the O depends on ε).

The class of decision problems solvable by a polynomial time Monte Carlo algorithm with two-sided error is known as BPP ("Bounded-error Probabilistic Polynomial time"). Two-sided error algorithms make errors in both directions; thus, e.g., if the true answer is "yes" then $\Pr[\text{yes}] \geq \frac{1}{2}+\varepsilon$, and if the true answer is "no" then $\Pr[\text{no}] \geq \frac{1}{2}+\varepsilon$. It should be clear that the containments P $\subseteq$ RP $\subseteq$ BPP and RP $\subseteq$ NP hold. The relationship between BPP and NP is unknown. For a two-sided error algorithm, we can also increase our confidence by running the algorithm many times and then taking the *majority vote*. The following standard calculation shows that the number of trials t required to ensure an error of at most δ is again $O(\log 1/\delta)$.

Proof. The probability that the majority vote algorithm yields an error is equal to the probability that we obtain at most $t/2$ correct outputs in t trials, which is bounded above by

$$\begin{aligned}
\sum_{k=0}^{t/2} \binom{t}{k} \left(\frac{1}{2}+\varepsilon\right)^k \left(\frac{1}{2}-\varepsilon\right)^{t-k} &\leq \sum_{k=0}^{t/2} \binom{t}{k} \left(\frac{1}{2}+\varepsilon\right)^{t/2} \left(\frac{1}{2}-\varepsilon\right)^{t/2} \\
&= \left(\frac{1}{4}-\varepsilon^2\right)^{t/2} \sum_{k=0}^{t/2} \binom{t}{k} \\
&\leq \left(\frac{1}{4}-\varepsilon^2\right)^{t/2} 2^{t-1} \\
&\leq (1-4\varepsilon^2)^{t/2}.
\end{aligned}$$

□

The class of decision problems solvable by a polynomial time Las Vegas algorithm is known as ZPP ("Zero- error Probabilistic Polynomial time"). We can use Markov inequality to show that ZPP $\subseteq$ RP. Suppose we have a Las Vegas algorithm $\mathcal{A}$ who terminates in X steps on some input with $\mathbb{E}[X] = T$. We can turn it into a Monte Carlo algorithm by running $\mathcal{A}$ for $3T$ steps. If the $\mathcal{A}$ terminates before $3T$, we just output the answer. Otherwise, we output an arbitrary symbol (e.g., output "no"). Then the probability that $\mathcal{A}$ makes a mistake is bounded by $\Pr[X > 3T] \leq 1/3$, and this algorithm is always correct if it outputs "yes". Whether RP or BPP $\subseteq$ ZPP is unknown. However, if we have a Monte Carlo algorithm that fails with some probability, but we can tell when it fails, then it can be converted into a Las Vegas algorithm. In fact, we can run it again and again until it succeeds, which means that we can eventually succeed with probability 1 (but with a potentially unbounded running time).

A Las Vegas algorithm is actually a Monte Carlo algorithm that we know when we have succeeded. The heuristic for remembering which class is which is that the names were chosen to appeal to English speakers: in Las Vegas, the dealer can tell you whether you've won or lost, but in Monte Carlo, *le croupier ne parle que Français*, so you have no idea what he's saying.

3.3 CHEBYSHEV'S INEQUALITY

A common trick to improve concentration is to consider $\mathbb{E}[f(X)]$ instead $\mathbb{E}[X]$ for some nondecreasing $f : \mathbb{R} \to \mathbb{R}$, since

$$\Pr[X \geq a] = \Pr[f(X) \geq f(a)] .$$

Taking $f(x) = x^2$ we obtain another concentration inequality: Chebyshev's inequality.

Theorem 3.2. Chebyshev's inequality
For any random variable X and constant t,

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathbf{Var}[X]}{t^2} .$$

Variance $\mathbf{Var}[X] = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, which is usually denoted σ^2. $\mathbb{E}X$ is usually denoted μ.

Proof.

$$\begin{aligned}
\sigma^2 = \mathbb{E}[(X-\mu)^2] &= \mathbf{Pr}[|X-\mu| \geq t] \cdot \mathbb{E}[(X-\mu)^2 \mid |X-\mu| \geq t] \\
&\quad + \mathbf{Pr}[|X-\mu| \leq t] \cdot \mathbb{E}[(X-\mu)^2 \mid |X-\mu| \leq t] \\
&\geq \mathbf{Pr}[|X-\mu| \geq t] \cdot t^2.
\end{aligned}$$

□

The use of Chebyshev's inequality is often referred as the "second-moment method" as it uses the variance of the random variable. Here we see some examples of Chebyshev's inequality.

Recall the coupon collector's problem. Now we apply Chebyshev's inequality. The key ingredient is the variance of X. Since $X = X_1 + \cdots + X_n$ and $X_1, \ldots, X_n$ are independent, we have

$$\mathbf{Var}[X] = \mathbf{Var}[X_1] + \cdots + \mathbf{Var}[X_n] .$$

Recall that X_i follows geometric distribution $X_i \sim \text{Geom}\left(\frac{n-i+1}{n}\right)$ We have the following lemma for the geometric distribution.

Lemma 3.3.
Suppose $Y \sim \text{Geom}(p)$. Then $\mathbf{Var}[Y] = \frac{1-p}{p^2}$.

Proof. Clearly $\mathbb{E}[Y]^2 = \frac{1}{p^2}$. Now we compute $Z = \mathbb{E}[Y^2]$:

$$\begin{aligned}
Z &= \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p \\
(1-p)Z &= \sum_{k=1}^{\infty} (k-1)^2 (1-p)^{k-1} p \\
\implies pZ &= \sum_{k=1}^{\infty} (2k-1)(1-p)^{k-1} p \\
&= 2\mathbb{E}[Y] - \sum_{k=1}^{\infty} (1-p)^{k-1} p \\
&= 2\frac{1}{p} - 1 = \frac{2-p}{p} \\
\implies Z &= \frac{2-p}{p^2} .
\end{aligned}$$

□

Applying this lemma, we have

$$\mathbf{Var}[X] = \sum_{i=1}^{n} \mathbf{Var}[X_i] = \sum_{i=0}^{n-1} \frac{1 - \frac{n-i}{n}}{\frac{(n-i)^2}{n^2}} = \sum_{i=0}^{n-1} \frac{ni}{(n-i)^2} \le n^2 \sum_{i=1}^{\infty} i^{-2} = \frac{\pi^2 n^2}{6}.$$

Finally, the Chebyshev's inequality gives that

$$\Pr[X > nH_n + cn] \le \Pr[|X - \mathbb{E}[X]| > cn] \le \frac{\mathbf{Var}[X]}{(cn)^2} \le \frac{\pi^2}{6c^2}.$$

Another famous example is the Weierstrass approximation theorem in analysis.

Theorem 3.4. Weierstrass approximation theorem
Suppose $f : [0,1] \to \mathbb{R}$ is a continuous function. For every $\varepsilon > 0$, there exists a polynomial $p(x)$ such that

$$\forall x \in [0,1], \quad |p(x) - f(x)| \le \varepsilon.$$

Proof. Since $[0,1]$ is compact, f is uniformly continuous and bounded. Without loss of generality, assume $|f(x)| \le 1$. There exists $\delta > 0$ such that $|f(x) - f(y)| \le \frac{\varepsilon}{2}$ for all $|x - y| \le \delta$.

By Bernstein, 1912.

Now, we approximate f by

$$P_n(x) = \sum_{i=0}^{n} E_i(x) f(\frac{i}{n}),$$

where

$$E_i(x) = \mathbf{Pr}[\mathsf{Bin}(n,x) = i] = \binom{n}{i} x^i (1-x)^i.$$

Note that $E_i(x)$ peaks at $\frac{i}{n}$ and decays away from $\frac{i}{n}$. Since Bin(n,x) has expectation nx and variance $nx(1-x) \le \frac{n}{4}$, with Chebyshev's inequality we have

$$\sum_{i:|i-nx|>n^{2/3}} E_i(x) = \mathbf{Pr}[|\mathsf{Bin}(n,x) - nx| > n^{2/3}] \le n^{-1/3}.$$

Note that $\sum_{i=0}^{n} E_i(x) = 1$. Taking $n > \max\{64\varepsilon^{-3}, \delta^{-3}\}$, we have

$$\begin{aligned} |P_n(x) - f(x)| &\le \sum_{i=0}^{n} E_i(x) |f(\frac{i}{n}) - f(x)| \\ &\le \sum_{|i-nx| \le n^{2/3}} E_i(x) \cdot \frac{\varepsilon}{2} + 2n^{-1/3} < \varepsilon, \end{aligned}$$

which completes the proof. □

3.4 Threshold of Random Graphs

We now study the properties of random graphs $\mathcal{G}(n,p)$. The notation $\mathcal{G}(n,p)$ specify a distribution over all simple undirected graphs of n vertices. In this model, each of the $\binom{n}{2}$ possible edges exists with probability p independently. Therefore, the expected number of edges in the graph is $\binom{n}{2}p$ and each vertex has expected degree $(n-1)p$.

Definition 3.1.
A graph property $\mathcal{P}$ is a subset of all graphs.

We say a graph property $\mathcal{P}$ is monotone increasing/decreasing if for any $G \in \mathcal{P}$, any graph we obtain through adding/deleting edges in G always belongs to $\mathcal{P}$. For instance, for a fixed graph H, the graph property

$$\mathcal{P}_1 = \{G \mid H \text{ is an induced sub-graph of } G\}$$

is monotone increasing. The graph property

$$\mathcal{P}_2 = \{G \mid G \text{ is a planar graph}\}$$

is monotone decreasing. However,

$$\mathcal{P}_3 = \{G \mid G \text{ contains a vertex of degree } 1\}$$

is not monotone. A graph property $\mathcal{P}$ is non-trivial if for any sufficiently large n, there always exists a graph with n vertices in $\mathcal{P}$ and another graph not in $\mathcal{P}$.

What we want to discuss is the following natural problem.

> Given a graph property $\mathcal{P}$, for which $p = p_n$ is $\mathcal{P}$ true for $\mathcal{G}(n,p)$ with high probability?

Let's start from the easiest case. Suppose $\mathcal{P} = \{G : K_3 \subseteq G\}$. Now, consider $G \sim \mathcal{G}(n,p_n)$. Let X be the number of K_3 in graph G, which is a random variable. Clearly, $\mathbb{E}[X] = \binom{n}{3}p^3$.

We will use $f \ll g$ to denote $f = o(g)$, and use $f \gg g$ to denote $g = o(f)$.

If $p \ll \frac{1}{n}$, then $\mathbf{Pr}[X \geq 1] = o(1)$ by Markov's inequality. If $p \gg \frac{1}{n}$, let's first prove that $\mathbf{Var}[X] = o(\mathbf{E}[X]^2)$. Denote $\mathcal{S}$ as the set of all subsets of vertices in G of size 3, and denote X_T the indicator variable of the set T inducing a triangle in G. Obviously, $X = \sum_{T \in \mathcal{S}} X_T$. Notice that

$$\begin{aligned}
\mathbf{Cov}[X_{T_1}, X_{T_2}] &= \mathbb{E}[X_{T_1}X_{T_2}] - \mathbb{E}[X_{T_1}] \cdot \mathbb{E}[X_{T_2}] \\
&= p^{|E(T_1 \cup T_2)|} - p^{|E(T_1) + E(T_2)|} \\
&= \begin{cases} 0 & |T_1 \cap T_2| \leq 1 \\ p^5 - p^6 & |T_1 \cap T_2| = 2 \end{cases}.
\end{aligned}$$

Also, we have

$$\mathbf{Var}[X_T] = \mathbb{E}[X_T^2] - \mathbb{E}[X_T]^2 = p^3 - p^6.$$

Therefore,

$$\begin{aligned}
\mathbf{Var}[X] &= \sum_{T \in \mathcal{S}} \mathbf{Var}[X_T] + \sum_{\substack{T_1, T_2 \in \mathcal{S} \\ T_1 \neq T_2}} \mathbf{Cov}[X_{T_1}, X_{T_2}] \\
&= \binom{n}{3}(p^3 - p^6) + \sum_{\substack{T_1, T_2 \in \mathcal{S} \\ |T_1 \cap T_2| = 2}} (p^5 - p^6) \\
&= \binom{n}{3}(p^3 - p^6) + \binom{n}{2}(n-2)(n-3)(p^5 - p^6) \\
&\lesssim n^3p^3 + n^4p^5 \\
&= o(n^6p^6).
\end{aligned}$$

The last equality above holds as $p \gg \frac{1}{n}$. This implies that $\mathbf{Var}[X] = o(\mathbf{E}[X]^2)$. Based on Chebyshev's inequality, we can see that $\mathbf{Pr}[X = 0] = o(1)$.

The behavior of a random graph containing K_3 as a subgraph has a "threshold".

Definition 3.2. Thresholds
We say r_n is a threshold function for some graph property $\mathcal{P}$ if

$$\mathbf{Pr}[\mathcal{G}(n,p_n) \in \mathcal{P}] \to \begin{cases} 0 & \text{if } p_n/r_n \to 0 \\ 1 & \text{if } p_n/r_n \to \infty \end{cases}.$$

From above, we show the following theorem.

Theorem 3.5.
A threshold function for containing a K_3 is $\frac{1}{n}$.

Similarly, we can show that $p = n^{-2/3}$ is a threshold for containing a K_4.

We now consider some general cases. Suppose we have a random variable $X = X_1 + \ldots + X_m$, where X_i is the indicator of event E_i. By Markov's inequality, it is easy to show that $X = 0$ with high probability if $\mathbb{E}[X] = o(1)$. However, it is difficult to show $X > 0$ with high probability if $\mathbb{E}[X] \neq o(1)$. To apply Chebyshev's inequality, we need to bound the variance first.

We say $i \sim j$ if $i \neq j$ and E_i, E_j are not independent. If $i \neq j$ and $i \nsim j$, we clearly have $\mathbf{Cov}[X_i, X_j] = 0$. Otherwise,

$$\mathbf{Cov}[X_i, X_j] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] \leq \mathbb{E}[X_i X_j] = \mathbf{Pr}[E_i \wedge E_j].$$

Also note that $\mathbf{Var}[X_i] \leq \mathbb{E}[X_i^2] = \mathbb{E}[X_i]$, which implies that

$$\mathbf{Var}[X] \leq \mathbb{E}[X] + \sum_{i \sim j} \mathbf{Pr}[E_i \wedge E_j].$$

Define $\Delta := \sum_{i\sim j} \mathbf{Pr}[E_i \wedge E_j]$. We hope $\mathbf{Var}[X] = o(\mathbf{E}[X])^2$, so if $\mathbf{E}[X] \to \infty$, $\Delta = o(\mathbf{E}[X])^2$ suffices. Moreover,

$$\sum_{i \sim j} \mathbf{Pr}[E_i \wedge E_j] = \sum_i \mathbf{Pr}[E_i] \sum_{j \sim i} \mathbf{Pr}[E_j \mid E_i].$$

In many symmetric cases, $\sum_{j\sim i} \mathbf{Pr}[E_j \mid E_i]$ does not depend on i. Denote Δ^* this value. Therefore, $\Delta = \sum_i \mathbf{Pr}[E_i]\Delta^* = \mathbb{E}[X]\Delta^*$. This gives us the following lemma.

Or we may set

$$\Delta^* = \max_i \sum_{j \sim i} \mathbf{Pr}[E_j \mid E_i]$$

in asymmetric cases.

Lemma 3.6.
If $\mathbb{E}[X] \to \infty$ and $\Delta^* = o(\mathbb{E}[X])$, then $X > 0$ with high probability.

In fact, by Chebyshev's inequality, we have

$$\Pr[(1-\varepsilon)\mathbb{E}X \leq X \leq (1+\varepsilon)\mathbb{E}X] \geq 1 - \frac{\mathbf{Var}[X]}{\varepsilon^2 \mathbb{E}[X]^2} = 1 - o(1)$$

for any constant $0 < \varepsilon < 1$.

Now consider the property of containing K_4. For any set S consisting of exactly four vertices, let E_S be the event that S forms a K_4 in the random graph. For any S, T of size 4, $S \sim T$ if and only if $|S \cap T| \geq 2$. There are two cases:

- $|S \cap T| = 2$:

$$\sum_T \mathbf{Pr}[E_T \mid E_S] \leq 6\binom{n}{2}\mathbf{Pr}[E_T \mid E_S] = 6\binom{n}{2}p^5 \approx n^2 p^5;$$

- $|S\cap T|=3$:

$$\sum_T \mathbf{Pr}[E_T \mid E_S] = 4(n-4)\mathbf{Pr}[E_T \mid E_S] \le 4np^3 \approx np^3.$$

Therefore, $\Delta^* \approx n^2p^5 + np^3 = o(n^4p^6) = o(\mathbb{E}[X])$ if $n^2p \gg 1$ and $np \gg 1$.

One may ask letting X be the number of a general graph H, can we still say that $X>0$ with high probability if $\mathbb{E}[X] \to \infty$? This is actually not correct. Suppose H is the graph obtained by adding an edge to K_4. Then, $\mathbb{E}[X] \approx n^5p^7 \to \infty$ if $p \gg n^{-5/7}$. However, there is no K_4 in $\mathcal{G}(n,p)$ if $p \ll n^{-2/3}$.

The correct answer is to consider the maximum edge-vertex ratio.

Definition 3.3.
The edge-vertex ratio of $G=(V,E)$ is defined as $\rho(G)=|E|/|V|$. The maximum sub-graph ratio is given by $m(G)=\max_{H\subseteq G}\rho(H)$.

Theorem 3.7. (Bollobás, 1981)
Let $H=(V,E)$ be a fixed graph. Then $p=n^{-1/m(H)}$ is a threshold function for containing H as a subgraph.
Furthermore, if $p \gg n^{-1/m(H)}$, then with high probability $X_H = \#$ copies of H in $\mathcal{G}(n,p)$ satisfies

$$X_H \approx \mathbb{E}[X] = \binom{n}{|V|}\frac{|V|!}{|\mathrm{Aut}(H)|}p^{|E|} \approx \frac{n^{|V|}p^{|E|}}{|\mathrm{Aut}(H)|}.$$

Proof. Let H' be the sub-graph of H achieving the maximum edge-vertex ratio, i.e., $m(H)=\rho(H')$. If $p \ll n^{-1/m(H)}$, then $\mathbb{E}[X_{H'}]=o(1)$, which implies that $X_{H'}=0$ with high probability.

Now assume that $p \gg n^{-1/m(H)}$. Count the labelled copies of H in $\mathcal{G}(n,p)$. Let L be a labelled copy of H in K_n. A_L be the event of $L \subseteq \mathcal{G}(n,p)$. For fixed L, we have

$$\Delta^* = \sum_{L'\sim L}\mathbf{Pr}[A_{L'}|A_L] = \sum_{L'\sim L} p^{|E(L')\setminus E(L)|}.$$

Note that the number of L' such that $L' \sim L$ is approximately $n^{|V(L')\setminus V(L)|}$, and

$$p \gg n^{-1/m(H)} \gg n^{-1/\rho(L'\cap L)} = n^{-|V(L')\cap V(L)|/|E(L')\cap E(L)|}.$$

So, we have

$$\Delta^* \approx \sum n^{|V(L')\setminus V(L)|}p^{|E(L')\setminus E(L)|} \ll n^{|V(L)|}p^{|E(L)|},$$

which implies that $\Delta^* \ll \mathbb{E}[X_H]$. Therefore, $\mathbf{Var}[X] = \mathbb{E}[X_H] + o(\mathbb{E}[X_H])^2$, which completes the proof. □