Chapter 4: The Second Moment

1 Basics

After the expectation the most vital statistic for a random variable X is the *variance*. We denote it $Var[X]$. It is defined by

$$Var[X] = E[(X - E[X])^2]$$

and measures how spread out X is from its expectation. We shall generally, following standard practice, let μ denote expectation and σ^2 denote variance. The positive square root σ of the variance is called the *standard deviation*. With this notation here is our basic tool.
Chebyschev's Inequality. For any positive λ

$$\Pr[|X - \mu| \geq \lambda\sigma] \leq \frac{1}{\lambda^2}$$

Proof.

$$\sigma^2 = Var[X] = E[(X - \mu)^2] \geq \lambda^2\sigma^2 \Pr[|X - \mu| \geq \lambda\sigma] \qquad \square$$

The use of Chebyschev's Inequality is called the *Second Moment Method*. Chebyschev's Inequality is best possible when no additional restrictions are placed on X as X may by $\mu + \lambda\sigma$ and $\mu - \lambda\sigma$ with probability $1/2\lambda^2$ and otherwise μ. Note, however, that when X is a normal distribution with mean μ and standard deviation σ then

$$\Pr[|X - \mu| \geq \lambda\sigma] = 2\int_\lambda^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

and for λ large this quantity is asymptotically $\sqrt{2/\pi}e^{-\lambda^2/2}/\lambda$ which is significantly smaller than $1/\lambda^2$. In Chapters 7,8 we shall see examples where X is the sum of "nearly independent" random variables and these better bounds can apply.

Suppose we have a decomposition

$$X = X_1 + \ldots + X_m$$

Then $Var[X]$ may be computed by the formula

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

Here the second sum is over ordered pairs and the *covariance* $Cov[Y, Z]$ is defined by

$$Cov[Y, Z] = E[YZ] - E[Y]E[Z]$$

In general, if Y, Z are independent then $Cov[Y, Z] = 0$. This often simplifies considerably variance calculations. Now suppose further, as will generally be the case in our applications, that the X_i are indicator random variables - i.e., that $X_i = 1$ if a certain event A_i holds and otherwise $X_i = 0$. If X_i is one with probability $p_i = \Pr[A_i]$ then

$$Var[X_i] = p_i(1 - p_i) \leq p_i = E[X_i]$$

and so

$$Var[X] \leq E[X] + \sum_{i \neq j} Cov[X_i, X_j]$$

2 Number Theory

The second moment method is an effective tool in number theory. Let $v(n)$ denote the number of primes p dividing n. (We do not count multiplicity though it would make little difference.) The folllowing result says, roughly, that "almost all" n have "very close to" $\ln \ln n$ prime factors. This was first shown by Hardy and Ramanujan in 1920 by a quite complicated argument. We give the three page proof of Paul Turan [1934] a proof that played a key role in the development of probabilistic methods in number theory.
Theorem 2.1 Let $\omega(n) \to \infty$ arbitrarily slowly. Then number of x in $\{1, \ldots, n\}$ such that

$$|v(n) - \ln \ln n| > \omega(n)\sqrt{\ln \ln n}$$

is $o(n)$.
Proof. Let x be randomly chosen from $\{1, \ldots, n\}$. For p prime set

$$X_p = \begin{cases} 1 & \text{if } p|x \\ 0 & \text{otherwise} \end{cases}$$

and set $X = \sum X_p$, the summation over all primes $p \leq n$, so that $X(x) = v(x)$. Now

$$E[X_p] = \frac{\lfloor n/p \rfloor}{n}$$

As $y-1<\lfloor y\rfloor\le y$
$$E[X_p]=1/p+O(1/n)$$
By linearity of expectation
$$E[X]=\sum_{p\le n}\frac{1}{p}+O(\frac{1}{n})=\ln\ln n+o(1)$$
Now we bound the variance
$$Var[X]\le\ln\ln n+o(1)+\sum_{p\ne q}Cov[X_p,X_q]$$
With p,q distinct primes, $X_pX_q=1$ if and only if $p|x$ and $q|x$ which occurs if and only if $pq|x$. Hence
$$\begin{aligned}&Cov[X_p,X_q]=E[X_p]E[X_q]-E[X_pX_q]\\&=\tfrac{\lfloor n/pq\rfloor}{n}-\tfrac{\lfloor n/p\rfloor}{n}\tfrac{\lfloor n/q\rfloor}{n}\\&\le\tfrac{1}{pq}-(\tfrac{1}{p}-\tfrac{1}{n})(\tfrac{1}{q}-\tfrac{1}{n})\\&\le\tfrac{1}{n}(\tfrac{1}{p}+\tfrac{1}{q})\end{aligned}$$
Thus
$$\sum_{p\ne q}Cov[X_p,X_q]\le\frac{1}{n}\sum_{p\ne q}\frac{1}{p}+\frac{1}{q}=\frac{\pi(n)-1}{n}\sum_p\frac{2}{p}$$
where $\pi(n)\sim\frac{n}{\ln n}$ is the number of primes $p\le n$. So
$$\sum_{p\ne q}Cov[X_p,X_q]<\frac{(n/\ln n)}{n}(\ln\ln n)=o(1)$$
That is, the covariances do not affect the variance, $Var[X]=\ln\ln n+o(1)$ and Chebyschev's Inequality actually gives
$$\Pr[|v(n)-\ln\ln n|>\lambda\sqrt{\ln\ln n}]<\lambda^{-2}+o(1)$$
for any constant λ. $\square$

In a classic paper Paul Erdős and Marc Kac [1940] showed, essentially, that X does behave like a normal distribution with mean and variance $\ln\ln n+o(1)$. Here is their precise result.

The Erdős-Kac Theorem. Let λ be fixed, positive, negative or zero. Then
$$\lim_{n\to\infty}\frac{1}{n}|\{x:1\le x\le n,v(x)\ge\ln\ln n+\lambda\sqrt{\ln\ln n}\}|=\int_\lambda^\infty\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt$$
We do not prove this result here.

3 More Basics

Let X be a nonnegative integral valued random variable and suppose we want to bound $\Pr[X=0]$ given the value $\mu = E[X]$. If $\mu < 1$ we may use the inequality

$$\Pr[X > 0] < E[X]$$

so that if $E[X] \to 0$ then $X = 0$ almost always. (Here we are imagining an infinite sequence of X dependent on some parameter n going to infinity.) But now suppose $E[X] \to \infty$. It does *not* necessarily follow that $X > 0$ almost always. For example, let X be the number of deaths due to nuclear war in the twelve months after reading this paragraph. Calculation of $E[X]$ can make for lively debate but few would deny that it is quite large. Yet we may believe - or hope - that $Pr[X \neq 0]$ is very close to zero. We can sometimes deduce $X > 0$ almost always if we have further information about $Var[X]$.
Theorem 3.1

$$\Pr[X=0] \leq \frac{Var[X]}{E[X]^2}$$

Proof. Set $\lambda = \mu/\sigma$ in Chebyschev's Inequality. Then

$$\Pr[X=0] \leq \Pr[|X-\mu| \leq \lambda\sigma] \leq \frac{1}{\lambda^2} = \frac{\sigma^2}{\mu^2} \ \square$$

We generally apply this result in asymptotic terms.
Corollary 3.2

$$\text{If } Var[X] = o(E[X]^2) \text{ then } X > 0 \text{ a.a.}$$

The proof of Theorem 3.1 actually gives that for any $\epsilon > 0$

$$\Pr[|X - E[X]| \geq \epsilon E[X]] \leq \frac{Var[X]}{\epsilon^2 E[X]^2}$$

and thus in asymptotic terms we actually have the stronger
Corollary 3.3

$$\text{If } Var[X] = o(E[X]^2) \text{ then } X \sim E[X] \text{ a.a.}$$

Suppose again $X = X_1 + \ldots + X_m$ where X_i is the indicator random variable for event A_i. For indices i, j write $i \sim j$ if $i \neq j$ and the events A_i, A_j are not independent. We set (the sum over ordered pairs)

$$\Delta = \sum_{i \sim j} \Pr[A_i \wedge A_j]$$

Note that when $i \sim j$

$$Cov[X_i, X_j] = E[X_i X_j] - E[X_i]E[X_j] \leq E[X_i X_j] = \Pr[A_i \wedge A_j]$$

and that when $i \neq j$ and not $i \sim j$ then $Cov[X_i, X_j] = 0$. Thus

$$Var[X] \leq E[X] + \Delta$$

Corollary 3.4 If $E[X] \to \infty$ and $\Delta = o(E[X]^2)$ then $X > 0$ almost always. Furthermore $X \sim E[X]$ almost always.

Let us say $X_1, \ldots, X_m$ are *symmetric* if for every $i \neq j$ there is an automorphism of the underlying probability space that sends event A_i to event A_j. Examples will appear in the next section. In this instance we write

$$\Delta = \sum_{i \sim j} \Pr[A_i \wedge A_j] = \sum_i \Pr[A_i] \sum_{j \sim i} \Pr[A_j | A_i]$$

and note that the inner summation is independent of i. We set

$$\Delta^* = \sum_{j \sim i} \Pr[A_j | A_i]$$

where i is any fixed index. Then

$$\Delta = \sum_i \Pr[A_i] \Delta^* = \Delta^* \sum_i \Pr[A_i] = \Delta^* E[X]$$

Corollary 3.5 If $E[X] \to \infty$ and $\Delta^* = o(E[X])$ then $X > 0$ almost always. Furthermore $X \sim E[X]$ almost always.

The condition of Corollary 4 has the intuitive sense that conditioning on any specific A_i holding does not substantially increase the expected number $E[X]$ of events holding.

4 Random Graphs

The definition of the random graph $G(n,p)$ and of "threshold function" are given in Chapter 10, Section 1. The results of this section are generally surpassed by those of Chapter 10 but they were historically the first results and provide a good illustration of the second moment. We begin with a particular example.
Theorem 4.1 The property $\omega(G) \geq 4$ has threshold function $n^{-2/3}$.
Proof. For every 4-set S of vertices in $G(n,p)$ let A_S be the event "S is a clique" and X_S its indicator random variable. Then

$$E[X_S] = \Pr[A_S] = p^6$$

as six different edges must all lie in $G(n,p)$. Set

$$X = \sum_{|S|=4} X_S$$

so that X is the number of 4-cliques in G and $\omega(G) \geq 4$ if and only if $X > 0$. Linearity of Expectation gives

$$E[X] = \sum_{|S|=4} E[X_S] = \binom{n}{4} p^6 \sim \frac{n^4 p^6}{24}$$

When $p(n) << n^{-2/3}$, $E[X] = o(1)$ and so $X = 0$ almost surely.

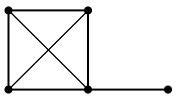
Now suppose $p(n) >> n^{-2/3}$ so that $E[X] \to \infty$ and consider the Δ^* of Corollary 3.5. (All 4-sets "look the same" so that the X_S are symmetric.) Here $S \sim T$ if and only if $S \neq T$ and S, T have common edges - i.e., if and only if $|S \cap T| = 2$ or 3. Fix S. There are $O(n^2)$ sets T with $|S \cap T| = 2$ and for each of these $\Pr[A_T | A_S] = p^5$. There are $O(n)$ sets T with $|S \cap T| = 3$ and for each of these $\Pr[A_T | A_S] = p^3$. Thus

$$\Delta^* = O(n^2 p^5) + O(np^3) = o(n^4 p^6) = o(E[X])$$

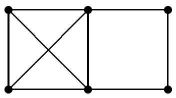
since $p >> n^{-2/3}$. Corollary 3.5 therefore applies and $X > 0$, i.e., there *does* exist a clique of size 4, almost always. $\square$

The proof of Theorem 4.1 appears to require a fortuitous calculation of Δ^*. The following definitions will allow for a description of when these calculations work out.
Definitions. Let H be a graph with v vertices and e edges. We call $\rho(H) = e/v$ the *density* of H. We call H *balanced* if every subgraph H' has $\rho(H') \leq$

$\rho(H)$. We call H *strictly balanced* if every proper subgraph H' has $\rho(H') < \rho(H)$.
Examples. K_4 and, in general, K_k are strictly balanced. The graph



is not balanced as it has density $7/5$ while the subgraph K_4 has density $3/2$. The graph



is balanced but not strictly balanced as it and its subgraph K_4 have density $3/2$.
Theorem 4.2 Let H be a balanced graph with v vertices and e edges. Let $A(G)$ be the event that H is a subgraph (not necessarily induced) of G. Then $p = n^{-v/e}$ is the threshold function for A.
Proof. We follow the argument of Theorem 4.1 For each v-set S let A_S be the event that $G|_S$ contains H as a subgraph. Then

$$p^e \leq \Pr[A_S] \leq v!p^e$$

(Any particular placement of H has probability p^e of occuring and there are at most $v!$ possible placements. The precise calculation of $\Pr[A_S]$ is, in general, complicated by overlapping of potential copies of H.) Let X_S be the indicator random variable for A_S and

$$X = \sum_{|S|=v} X_S$$

so that A holds if and only if $X > 0$. Linearity of Expectation gives

$$E[X] = \sum_{|S|=v} E[X_S] = \binom{n}{v} \Pr[A_S] = \Theta(n^v p^e)$$

If $p << n^{-v/e}$ then $E[X] = o(1)$ so $X = 0$ almost always.

Now assume $p >> n^{-v/e}$ so that $E[X] \to \infty$ and consider the Δ^* of Corollary 3.5 (All v-sets look the same so the X_S are symmetric.) Here $S \sim T$ if and only if $S \neq T$ and S, T have common edges - i.e., if and only if $S \cap T = i$ with $2 \leq i \leq v - 1$. Let S be fixed. We split

$$\Delta^* = \sum_{T \sim S} \Pr[A_T|A_S] = \sum_{i=2}^{v-1} \sum_{|T \cap S|=i} \Pr[A_T|A_S]$$

For each i there are $O(n^{v-i})$ choices of T. Fix S, T and consider $\Pr[A_T|A_S]$. There are $O(1)$ possible copies of H on T. Each has - since, critically, H is balanced - at most $\frac{ie}{v}$ edges with both vertices in S and thus at least $e - \frac{ie}{v}$ other edges. Hence

$$\Pr[A_T|A_S] = O(p^{e-\frac{ie}{v}})$$

and

$$\begin{aligned} \Delta^* &= \textstyle\sum_{i=2}^{v-1} O(n^{v-i} p^{e-\frac{ie}{v}}) \\ &= \textstyle\sum_{i=2}^{v-1} O((n^v p^e)^{1-\frac{i}{v}}) \\ &= \textstyle\sum_{i=2}^{v-1} o(n^v p^e) \\ &= o(E[X]) \end{aligned}$$

since $n^v p^e \to \infty$. Hence Corollary 3.5 applies. □

Theorem 4.3 In the notation of Theorem 4.2 if H is *not* balanced then $p = n^{-v/e}$ is *not* the threshold function for A.

Proof. Let H_1 be a subgraph of H with v_1 vertices, e_1 edges and $e_1/v_1 > e/v$. Let α satisfy $v/e < \alpha < v_1/e_1$ and set $p = n^{-\alpha}$. The expected number of copies of H_1 is then $o(1)$ so almost always $G(n,p)$ contains no copy of H_1. But if it contains no copy of H_1 then it surely can contain no copy of H. □

The threshold function for the property of containing a copy of H, for general H, was examined in the original papers of Erdős and Rényi. (Erdős,Rényi [1960] still provides an excellent introduction to the theory of Random Graphs.) Let H_1 be that subgraph with maximal density $\rho(H_1) = e_1/v_1$. (When H is balanced we may take $H_1 = H$.) They showed that $p = n^{-v_1/e_1}$ is the threshold function. We do not show this here though it follows fairly straightforwardly from these methods.

We finish this section with two strengthenings of Theorem 4.2.

Theorem 4.3 Let H be strictly balanced with v vertices, e edges and a automorphisms. Let X be the number of copies of H in $G(n,p)$. Assume $p >> n^{-v/e}$. Then almost always

$$X \sim \frac{n^v p^e}{a}$$

Proof. Label the vertices of H by $1, \ldots, v$. For each ordered $x_1, \ldots, x_v$ let $A_{x_1,\ldots,x_v}$ be the event that $x_1, \ldots, x_v$ provides a copy of H in that order. Specifically we define

$$A_{x_1,\ldots,x_v} : \{i,j\} \in E(H) \Rightarrow \{x_i, x_j\} \in E(G)$$

We let $I_{x_1,\ldots,x_v}$ be the corresponding indicator random variable. We define an equivalence class on v-tuples by setting $(x_1, \ldots, x_v) \equiv (y_1, \ldots, y_v)$ if there is an automorphism σ of $V(H)$ so that $y_{\sigma(i)} = x_i$ for $1 \le i \le v$.Then

$$X = \sum I_{x_1,\ldots,x_v}$$

gives the number of copies of H in G where the sum is taken over one entry from each equivalence class. As there are $(n)_v/a$ terms

$$E[X] = \frac{(n)_v}{a} E[I_{x_1,\ldots,x_v}] = \frac{(n)_v p^e}{a} \sim \frac{n^v p^e}{a}$$

Our assumption $p >> n^{-v/e}$ implies $E[X] \to \infty$. It suffices therefore to show $\Delta^* = o(E[X])$. Fixing $x_1, \ldots, x_v$,

$$\Delta^* = \sum_{(y_1,\ldots,y_v) \sim (x_1,\ldots,x_v)} \Pr[A_{(y_1,\ldots,y_v)} | A_{(x_1,\ldots,x_v)}]$$

There are $v!/a = O(1)$ terms with $\{y_1, \ldots, y_v\} = \{x_1, \ldots, x_v\}$ and for each the conditional probability is at most one (actually, at most p), thus contributing $O(1) = o(E[X])$ to Δ^*. When $\{y_1, \ldots, y_v\} \cap \{x_1, \ldots, x_v\}$ has i

elements, $2 \leq i \leq v-1$ the argument of Theorem 4.2 gives that the contribution to Δ^* is $o(E[X])$. Altogether $\Delta^* = o(E[X])$ and we apply Corollary 3.5 $\square$

Theorem 4.4 Let H be *any* fixed graph. For every subgraph H' of H (including H itself) let $X_{H'}$ denote the number of copies of H' in $G(n,p)$. Assume p is such that $E[X_{H'}] \to \infty$ for every H'. Then

$$X_H \sim E[X_H]$$

almost always.

Proof. Let H have v vertices and e edges. As in Theorem 4.3 it suffices to show $\Delta^* = o(E[X])$. We split Δ^* into a finite number of terms. For each H' with w vertice and f edges we have those $(y_1, \ldots, y_v)$ that overlap with the fixed $(x_1, \ldots, x_v)$ in a copy of H'. These terms contribute, up to constants,

$$n^{v-w}p^{e-f} = \Theta\left(\frac{E[X_H]}{E[X_{H'}]}\right) = o(E[X_H])$$

to Δ^*. Hence Corollary 3.5 does apply. $\square$

5 Clique Number

Now we fix edge probability $p = \frac{1}{2}$ and consider the clique number $\omega(G)$. We set

$$f(k) = \binom{n}{k} 2^{-\binom{k}{2}},$$

the expected number of k-cliques. The function $f(k)$ drops under one at $k \sim 2\log_2 n$. (Very roughly, $f(k)$ is like $n^k 2^{-k^2/2}$.)

Theorem 5.1 Let $k = k(n)$ satisfy $k \sim 2\log_2 n$ and $f(k) \to \infty$. Then almost always $\omega(G) \geq k$.

Proof. For each k-set S let A_S be the event "S is a clique" and X_S the corresponding indicator random variable. We set

$$X = \sum_{|S|=k} X_S$$

so that $\omega(G) \geq k$ if and only if $X > 0$. Then $E[X] = f(k) \to \infty$ and we examine the Δ^* of Corollary 3.5. Fix S and note that $T \sim S$ if and only if $|T \cap S| = i$ where $2 \leq i \leq k-1$. Hence

$$\Delta^* = \sum_{i=2}^{k-1} \binom{k}{i} \binom{n-k}{k-i} 2^{\binom{i}{2} - \binom{k}{2}}$$

and so

$$\frac{\Delta^*}{E[X]} = \sum_{i=2}^{k-1} g(i)$$

where we set

$$g(i) = \frac{\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}} 2^{\binom{i}{2}}$$

Observe that $g(i)$ may be thought of as the probability that a randomly chosen T will intersect a fixed S in i points times the factor increase in $\Pr[A_T]$ when it does. Setting $i = 2$,

$$g(2) = 2\frac{\binom{k}{2}\binom{n-k}{k-2}}{\binom{n}{k}} \sim \frac{k^4}{n^2} = o(1)$$

At the other extreme $i = k - 1$

$$g(k-1) = \frac{k(n-k)2^{-(k-1)}}{\binom{n}{k}2^{-\binom{n}{k}}} \sim \frac{2kn2^{-k}}{E[X]}$$

As $k \sim 2\log_2 n$ the numerator is $n^{-1+o(1)}$. The denominator approaches infinity and so $g(k-1) = o(1)$. Some detailed calculation (which we omit) gives that the remaining $g(i)$ are also negligible so that Corollary 3.5 applies. $\square$

Theorem 5.1 leads to a strong concentration result for $\omega(G)$. For $k \sim 2\log_2 n$

$$\frac{f(k+1)}{f(k)} = \frac{n-k+1}{k+1}2^{-k} = n^{-1+o(1)} = o(1)$$

Let $k_0 = k_0(n)$ be that value with $f(k_0) \geq 1 > f(k_0+1)$. For "most" n the function $f(k)$ will jump from a large $f(k_0)$ to a small $f(k_0+1)$. The probability that G contains a clique of size k_0+1 is at most $f(k_0+1)$ which will be very small. When $f(k_0)$ is large Theorem 5.1 implies that G contains a clique of size k_0 with probability nearly one. Together, with very high probability $\omega(G) = k_0$. For some n one of the values $f(k_0), f(k_0+1)$ may be of moderate size so this argument does not apply. Still one may show a strong concentration result found independently by Bollobás, Erdős [1976] and Matula [1976].

Corollary 5.2 There exists $k = k(n)$ so that

$$\Pr[\omega(G) = k \text{ or } k+1] \to 1$$

6 Distinct Sums

A set $x_1, \ldots, x_k$ of positive integers is said to have distinct sums if all sums

$$\sum_{i \in S} x_i, \qquad S \subseteq \{1, \ldots, k\}$$

are distinct. Let $f(n)$ denote the maximal k for which there exists a set

$$\{x_1, \ldots, x_k\} \subset \{1, \ldots, n\}$$

with distinct sums. The simplest example of a set with distinct sums is $\{2^i : i \leq \log_2 n\}$. This example shows

$$f(n) \geq 1 + \lfloor \log_2 n \rfloor$$

Erdős has offered \$300. for a proof or disproof that

$$f(n) \leq \log_2 n + C$$

for some constant C. From above, as all $2^{f(n)}$ sums are distinct and less than nk

$$2^{f(n)} < nk$$

and so

$$f(n) < \log_2 n + \log_2 \log_2 n + O(1)$$

Examination of the second moment gives a modest improvement. Fix $\{x_1, \ldots, x_k\} \subset \{1, \ldots, n\}$ with distinct sums. Let $\epsilon_1, \ldots, \epsilon_k$ be independent with

$$\Pr[\epsilon_i = 1] = \Pr[\epsilon_i = 0] = \frac{1}{2}$$

and set

$$X = \epsilon_1 x_1 + \ldots + \epsilon_k x_k$$

(We may think of X as a random sum.) Set

$$\mu = E[X] = \frac{x_1 + \ldots + x_k}{2}$$

and $\sigma^2 = Var[X]$. We bound

$$\sigma^2 = \frac{x_1^2 + \ldots + x_k^2}{4} \leq \frac{n^2 k}{4}$$

so that $\sigma \leq n\sqrt{k}/2$. By Chebyschev's Inequality for any $\lambda > 1$

$$\Pr[|X - \mu| \geq \lambda n\sqrt{k}/2] \leq \lambda^{-2}$$

Reversing,

$$1 - \frac{1}{\lambda^2} \leq \Pr[|X - \mu| < \lambda n\sqrt{k}/2]$$

But X has any particular value with probability either zero or 2^{-k} since, critically, a sum can be achieved in at most one way. Thus

$$\Pr[|X - \mu| < \lambda n\sqrt{k}/2] \leq 2^{-k}\lambda n\sqrt{k}$$

and

$$n \geq \frac{2^k}{\sqrt{k}}\frac{1 - \lambda^{-2}}{\lambda}$$

While $\lambda = \sqrt{3}$ gives optimal results any choice of $\lambda > 1$ gives Theorem 6.1

$$f(k) \leq \log_2 n + \frac{1}{2}\log_2\log_2 n + O(1)$$