ON THE THEOREM OF ERDÖS-KAC

by

Tevekkul Mehreliyev

A Thesis Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of
Master of Science

in

Mathematics

Koç University

September 2011

Koç University
Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis
by
Tevekkul Mehreliyev

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Thesis Committee Members:

Assoc. Prof. Emre Alkan (Advisor)..........................
Assoc. Prof. Burak Gürel ....
Asst. Prof. Kazm Büyükboduk ....

Date: 08 September 2011

ABSTRACT

This study presents different proofs and applications of the celebrated Erdös-Kac theorem, named after Paul Erdös and Mark Kac, also known as the fundamental theorem of probabilistic number theory which states that if n is a randomly chosen large integer, then the number of distinct prime factors of n has approximately the normal distribution with mean and variance $\log \log n$.

We first give the original proof from the authors which is so called elementary proof meaning that the complex functin theory is not used. This proof does not give an error term but rather gives an asymptotic result. In the second part we give the proof of A. Renyi and P. Turan which makes use of the standard tools of analytic number theory, Dirichlet series, contour integration. Although the latter method is not elementary, it is much simpler than the original proof and also helps us get an error term besides an asymptotic result.

Finally we use the article " On the Normal Number of Prime Factors of $\phi(n)$ by Paul Erdös and Carl Pomerance where ϕ " is Euler's function. We also give the related result for the divisor function which counts the number of positive divisors for a given integer.

ÖZET

Bu çalışmada, Olasılıksal Sayılar Teorisinin temel teoremi olarak bilinen Erdös-Kac teoreminin farklı ispatlarının yanısıra aynı konuyla ilgili birkaç çalışma verilmiştir. Öncelikle ana teoremi kompleks fonksiyon teorisi kullanmadan elementer metotlarla ispatlayarak hata terimi içermeyen bir asimptotik buluyoruz. İkinci kısımda, olasılıktan gelen bir teoremle, Riemann zeta-fonksiyonun ve genel Dririchlet serilerinin bazı özellikleri kullanılarak bulunan asimptotiğin yanısıra hata terimi de hesaplanmıştır. Son kısımda ise, konuyla ilgili elementer metotlar kullanılarak yapılan iki farklı çalışma incelenmiştir.

ACKNOWLEDGEMENTS

My first gratitude is to Assoc. Prof Emre Alkan, my thesis supervisor, for his support, tolerance and guidance. His efforts and the time he spent were invaluable and the courses we have made were very fruitful.

I would like to thank Assoc. Prof. Burak Gürel and Asst. Prof. Kazım Büyükboduk and for their participation in my thesis committee.

I also would like to express my thankfulness to Prof. Ali Ülger, the director of the graduate studies of Mathematics Department for his support and guidance.

I am thankful to my officemates for the friendly atmosphere they provided.

Finally, my deepest gratitude is to my family, for their endavours, love and support. Their presence made everything easier.

Contents

LIST OF SYMBOLS/ABBREVIATIONS

$\zeta(s)$	The Riemann zeta-function.
$li(x)$	$\int_2^x \frac{du}{\log u}$; the logarithmic integral.
$\Re s$	The real part of the complex number s.
$\Im s$	The imaginary part of the complex number s.
$\Gamma(s)$	The Gamma function.
$\Lambda(n)$	The von-Mangoldt function.
$\phi(n)$	The Euler's totient function.
$\pi(x)$	The number of primes $\leq x$.
$\pi(x; q, a)$	The number of primes $\leq x$ which are $\equiv a \pmod{q}$.
$\psi(x; q, a)$	The sum of $\Lambda(n)$ over $n \leq x$ which are $\equiv a \pmod{q}$.
E	The Euler's constant.
$[x]$	The integer part of x.
x	$x - [x]$; the fractional part of x.
$\mu(n)$	The Möbius mu function.
$f(x) = O(g(x))$	$\lvert f(x)\rvert \leq Cg(x)$ where C is an absolute constant.
$f(x) = o(g(x))$	$\lim f(x)/g(x) = 0$.
$f(x) \ll g(x)$	$f(x) = O(g(x))$.
$f(x) \sim g(x)$	$\lim f(x)/g(x) = 1$.

1 Preliminaries

This chapter includes the basic information needed to understand the text as we frequently will refer in the following chapters. It consists of four main sections and in each of them, we will present the functions and some of their properties that we are going to deal with. We also will introduce some main formulas and tools that are widely used in Analytic Number Theory. All these will be given briefly, without proof, since detailed arguments can be found in [3] or [2].

1.1 Arithmetic Functions.

Definition 1. *A real- or complex-valued function defined on the positive integers is called an arithmetic function.*

We introduce some arithmetic functions which play an important role on distribution of primes.

1. The Möbius function μ is defined as follows:
   $\mu(1) = 1$;
   If $n > 1$, write $n = p_1^{a_1} \cdots p_k^{a_k}$. Then
   $$\mu(n) = \begin{cases} (-1)^k & \text{if } a_1 = a_2 = \cdots a_k = 1, \\ 0 & \text{otherwise.} \end{cases}$$

2. If $n > 1$ the Euler totient $\phi(n)$ is defined to be the number of positive integers not exceeding n which are relatively prime to n; i.e.,
   $$\phi(n) = \sum_{\substack{m=1 \\ (m,n)=1}}^{n} 1.$$

3. The Von Mangoldt function $\Lambda(n)$ is defined as:
   $$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^m \text{ for some prime } p \text{ and some integer } m \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $f(n)$ be an arithmetic function. We usually denote by $F(x)$, the summatory function of $f(n)$

$$F(x) = \sum_{n \leq x} f(n).$$

In analytic number theory, we estimate the averages $\frac{F(x)}{x}$ of arithmetic functions because they are expected to behave more regularly for large x whereas an arithmetic function may behave beyond prediction when n is large. So we are interested in tools for evaluating the averages.

Now let us give the partial summation formula which is one of the most powerful methods for estimating the summatory of arithmetic functions.

Theorem 1.1 (The Partial Summation Formula). *Let x and y be real numbers with $0 < y < x$. Let $f(n)$ be an arithmetic function with summatory function $F(x)$ and $g(t)$ be a function with a continuous derivative on $[y, x]$. Then,*

$$\sum_{y<n\leq x} f(n)g(n) = F(x)g(x) - F(y)g(y) - \int_y^x F(t)g'(t)dt. \tag{1.1}$$

In particular, if $x \geq 2$ and $g(t)$ is continuously differentiable on $[1, x]$, then

$$\sum_{n\leq x} f(n)g(n) = F(x)g(x) - \int_1^x F(t)g'(t)dt. \tag{1.2}$$

This theorem, applied to the functions $f(n) = 1$ and $g(t) = 1/t$ gives

$$\sum_{n\leq x} \frac{1}{n} = \log x + E + r(x) \quad \text{where} \quad |r(x)| < \frac{2}{x}. \tag{1.3}$$

The number E in (1.3) is called the *Euler's constant.*

1.2 Dirichlet Series.

Given an arithmetic function $f(n)$, we define the Dirichlet series associated by f as

$$F(s) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s}.$$

A Dirichlet series can be regarded as a function of the complex variable s, defined in the region in which the series converges. We write the variable s as

$$s = \sigma + it, \quad \text{where} \quad \sigma = \Re s, t = \Im s,$$

and we will use this notation throughout the text.

An important result about Dirichlet series is the Euler product identity when applied to the Dirichlet series.

Theorem 1.2 (Euler Product Identity). *Let f be a multiplicative arithmetic function with Dirichlet series $F(s) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s}$. Assume $F(s)$ converges* *absolutely for $\sigma > \sigma_a$, then we have*

$$F(s) = \prod_p \left(1 + \frac{f(p)}{p^s} + \frac{f(p^2)}{p^{2s}} + \dots\right) \quad \textit{for } \sigma > \sigma_a. \tag{1.4}$$

If f is completely multiplicative, then

$$F(s) = \prod_p \left(1 + \frac{f(p)}{p^s}\right)^{-1} \quad \textit{for } \sigma > \sigma_a. \tag{1.5}$$

The most famous Dirichlet series is the one associated with the function $f(n) = 1$, so-called the Riemann zeta function $\zeta(s)$,

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}. \qquad (\sigma > 1)$$

We initially define $\zeta(s)$ for $\sigma > 1$ but it has an analytic continuation to the half-plane $\sigma > 0$:

$$\zeta(s) = \frac{s}{s-1} - s\int_1^{\infty} \{u\} u^{-s-1} du. \tag{1.6}$$

Moreover, by the Euler product identity (1.15), we have

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1} \quad (\sigma > 1). \tag{1.7}$$

Logarithmic derivative of the identity (1.17) gives the Dirichlet series for $-\frac{\zeta'(s)}{\zeta(s)}$,

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_p \sum_{n=1}^{\infty} \frac{\log p}{p^{ms}} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} = \prod_p \left(1 + \frac{1}{p^s}\right)^{-1} \quad (\sigma > 1). \tag{1.8}$$

Another important property of Dirichlet series is that we can relate them to the summatory functions of arithmetic functions. Finally lets indicate one of the inversion formulas which we will need later.

Theorem 1.3 (Riesz typical means.)**.** *For positive integer m and positive real x put*

$$R_m(x) = \frac{1}{m!} \sum_{k \le x} a_k (\log x/k)^m. \tag{1.9}$$

Then

$$R_m(x) = \frac{1}{2\pi i}\int_{\sigma_0 - i\infty}^{\sigma_0 + i\infty} \alpha(s)\frac{x^s}{s^{k+1}}ds \tag{1.10}$$

where $\alpha(s) = \sum_{n=1}^{\infty} a_n n^{-s}$, $x > 0$ *and* $\sigma_0 > \max(0, \sigma_c)$.

1.3 The Sieve Method.

Let $\Phi(x,z) := \sharp\{n \le x : n \quad \text{is not divisible by any prime} < z\}$ and $P_z := \prod_{p<z} p$. Then

$$\begin{aligned}\Phi(x,z) &= \sum_{n\le x}\sum_{d|(n,P_z)} \mu(d)\\ &= x\sum_{d|P_z}\frac{\mu(d)}{d} + O\left(\sum_{d|P_z}|\mu(d)|\right)\\ &= x\prod_{p<z}(1-\frac{1}{p}) + O(2^z).\end{aligned}$$

Observation: One can notice that

$$\pi(x) \le \Phi(x,z) + \pi(z) \le \Phi(x,z) + z.$$

By means of an elementary inequality $1 - x \le e^{-x}$ for $x > 0$ we obtain

$$\prod_{p<z}(1-\frac{1}{p}) \le \exp(-\sum_{p<z}\frac{1}{p}).$$

Now by choosing $z := c\log x$ for some small constant c and using the fact that $\sum_{p<z}\frac{1}{p} \ge \log\log z + O(1)$ we get

$$\prod_{p<z}(1-\frac{1}{p}) \ll \frac{x}{\log\log x}.$$

And since $2^z < (e^{\log x})^c \ll \frac{x}{\log\log x}$, then we obtain $\pi(x) \ll \frac{x}{\log\log x}$. Although this is a weak result in comparison to the result of Prime Number Theorem, but still better than a trivial bound.

Brun's Pure Sieve: By comparing the coefficients of x^r in the inequality $(1-x)^{-1}(1-x)^w = (1-x)^w$ we can get the equality of the form

$$\sum_{k\le r}(-1)^k\binom{w}{k} = (-1)^r\binom{w-1}{r}.$$

Now let $0 \leq r \leq w(n)-1$. Then

$$\sum_{\substack{d|n \\ w(d)\leq r}} \mu(d) = \sum_{k\leq r} (-1)^k \binom{w(n)}{k} = (-1)^r \binom{w(n)-1}{k}.$$

If we set

$$\Psi_r(n) = \sum_{d|n} \mu_r(d) \quad \text{where} \quad \mu_r(d) = \begin{cases} \mu(d) & \text{if } w(d) \leq r \\ 0 & \text{otherwise.} \end{cases}$$

then we get $\Psi_r(n) = (-1)^r \binom{w(n)-1}{k}$ is ≥ 0 if r is even and ≤ 0 if r is odd. So what we obtain is

$$\Psi_{2r+1}(n) \leq \sum_{d|n} \mu(d) \leq \Psi_{2r}(n)$$

and

$$\begin{aligned} \Psi_{2r+1}(n) &= \sum_{\substack{d|n \\ w(d)\leq 2r}} \mu(d) + \sum_{\substack{d|n \\ w(d)=2r+1}} \mu(d) \\ &= \Psi_{2r}(n) + O\left(\sum_{\substack{d|n \\ w(d)=2r+1}} |\mu(d)| \right) \end{aligned}$$

By the last inequality, by playing with the parity of r, we can easily get that

$$\sum_{d|n} \mu(d) = \Psi_r(n) + O\left(\sum_{\substack{d|n \\ w(d)=r+1}} |\mu(d)| \right)$$

Now, doing all these combinatorical calculations more precisely, one can get for the natural numbers n and r with $r \leq w(n)$, there exists $|\theta| \leq 1$ such that

$$\sum_{d|n} \mu(d) = \sum_{\substack{d|n \\ w(d)\leq r}} \mu(d) + \theta \sum_{\substack{d|n \\ w(d)=r+1}} \mu(d)$$

And putting this into the identity

$$S(A,P,z) = \sum_{a\in A} \sum_{d|(a,P(z))} \mu(d)$$

by choosing $\log z < \frac{c \log z}{\log\log x}$, with c small, we can get

$$S(A, P, z) = XW(z)(1 + o(1))$$

where X denotes the number of elements in the set A,

$$S(A, P, z) := \sharp\{a \in A : \quad \text{a is not divisible by any prime} \quad < z\}$$

and

$$W(z) := \prod_{p|P(z)} (1 - \frac{1}{p}).$$

We should also notice that

$$\prod_{p<z}(1 - \frac{1}{p}) = \frac{e^{-C}}{\log z}\left(1 + O(\frac{1}{\log z})\right)$$

where C is an Euler constant.

1.4 Results from Probability.

An additive function $f(n)$ is called strongly additive if $f(p^a) = f(p)$ for all $a \geq 1$. If $f(n)$ is real-valued and strongly additive, let

$$A(x) = \sum_{p<x} \frac{f(p)}{p}, \quad B(x) = \left(\sum_{p<x} \frac{f(p)^2}{p}\right)^{1/2} \tag{1.11}$$

We have the following theorem:

Theorem 1.4 (Kubilius-Shapiro)**.** *Suppose for each $\epsilon > 0$, we have*

$$\lim_{x\to\infty} \frac{1}{B(x)^2} \sum_{\substack{p \leq x \\ |f(p)| > \epsilon B(x)}} \frac{f(p)^2}{p} = 0 \tag{1.12}$$

then for each real number u,

$$\lim_{x\to\infty} \frac{1}{x} \cdot \sharp\{n \leq x : f(n) - A(x) \leq uB(x)\} = G(u), \tag{1.13}$$

where G(u) is defined as

$$G(u) := (2\pi)^{-1/2} \int_{-\infty}^{u} e^{-t^2/2} dt.$$

What theorem says is that, if (1.12) holds then the normal value for $n \leq x$ of $f(n)$ is $A(x)$ and the standard deviation is $B(x)$.

Continuity theorem for characteristic functions in R_1: For a scalar random variable X the characteristic function is defined as the expected value of e^{itX}, where i is the imaginary unit, and $t \in R$ is the argument of the characteristic function:

$$\varphi_X(t) = E\left[e^{itX}\right] = \int_{-\infty}^{\infty} e^{itX} dF_X(x). \tag{1.14}$$

Here F_X is the cumulative distribution function of X, and the integral is of the RiemannStieltjes kind. In probability theory and statistics, the characteristic function of any random variable completely defines its probability distribution. Thus it provides the basis of an alternative route to analytical results compared with working directly with probability density functions or cumulative distribution functions. In the applications it is sometimes very difficult to investigate directly the convergence of a sequence of distribution functions, while the convergence problem for the corresponding sequence of characteristic functions may be comparatively easy to deal with. Because of the same reason, we will need the following theorem, which is due to Levy:

Theorem 1.5. *We are given a sequence of distributions, with the distribution functions $F_1(x)$, $F_2(x)$, ..., and the characteristic functions $\varphi_1(t)$, $\varphi_2(t)$,... A necessary and sufficient condition for the convergence of the sequence $\{F_n(x)\}$ to a distribution function $F(x)$ is that, for every t, the sequence $\{\varphi_n(t)\}$ converges to a limit $\varphi(t)$, which is continuous for the special value $t = 0$.*

When this condition is satisfied, the limit $\varphi(t)$ is identical with the characteristic function of the limitting distribution function $F(x)$.

Another theorem that comes from probability theory is Turan-Kubilius inequality which is useful for proving results about the normal order of an arithmetic function. Theorem was proved in a special case in 1934 by Paul Turan and generalized in 1956 and 1964 by Jonas Kubilius.

Theorem 1.6 (Turan-Kubilius inequality)**.** *Suppose f is an additive complex-valued arithmetic function, and write p for an arbitrary prime and m for an arbitrary positive integer. Write*

$$A(x) = \sum_{p^m \le x} \frac{f(p^m)}{p^m}(1 - p^{-1})$$

and

$$B(x)^2 = \sum_{p^m \le x} \frac{|f(p^m)|^2}{p^m}$$

Then for $x \geq 2$ we have

$$\frac{1}{x}\sum_{n\leq x}|f(n)-A(x)|^2 \leq 32B(x)^2.$$

In fact in the right part of the inequality above, 32 can be replaced with $2+\epsilon(x)$, where $\epsilon(x)$ is a function that goes to zero when x goes to infinity. But the given version is sufficient for us.

1.5 Primes in arithmetic progression

The following two results about the primes in arithmetic progression, will be beneficial in later chapters.

Theorem 1.7 (The Brun-Titchmarsh inequality)**.** *Let a and k be coprime integers and let x be a positive real number such that $k < x$. Then*

$$\pi(x;k,a) \leq \frac{2x}{\phi(k)\log(x/k)},$$

where ϕ is Euler's totient function.

Theorem 1.8 (The Bombieri-Vinogradov Theorem)**.** *For any $A > 1$ and $Q = x^{1/2}(\log x)^{-B}$, where the constant B only depends on A, one has*

$$\sum_{k\leq Q}\max_{(a,k)=1}\left|\pi(x;k,a)-\frac{\pi(x)}{\phi(k)}\right| \ll x(\log x)^{-A}.$$

2 The Gaussian Law of Errors in the Theory of Additive Number Theoretic Functions

2.1 Introduction.

One of the first results in probabilistic number theory is the theorem of Hardy and Ramanujuan that the normal value of $w(n)$ is $\log\log n$, where $w(n)$ counts the number of distinct prime factors of n. What this statement means is that for each $\epsilon > 0$, the set of n for which

$$|w(n) - \log\log n| < \epsilon \log\log n \tag{2.1}$$

has asymptotic density 1.

A particular simple proof of these result was later given by P. Turan. He showed that

$$\sum_{n\leq x}(w(n) - \log\log x)^2 = x\log\log x + O(x) \tag{2.2}$$

from which (2.1) is an immediate corollary. The method of proof of the asymptotic formula (2.2) was later generalized independently by Turan and Kubilius to give an upper bound for the left hand side where $w(n)$ is replaced by an arbitrary additive function. The significance of the "$\log\log x$" in (2.2) is that it is about

$$\sum_{p\leq x} w(p)p^{-1}$$

where p runs over primes. Similarly the expected value of an arbitrary additive function $g(n)$ should be about

$$\sum_{p\leq x} g(p)p^{-1}.$$

What we will do in this chapter is to formulate the celebrated Erdös-Kac theorem [14] and give the original proof of authors. To begin with let $f(m)$ be an additive number theoretic function, so that $f(mn) = f(m) + f(n)$ if $(m,n) = 1$. Suppose $|f(p)| \leq 1$ and $f(p^\alpha) = f(p)$. Obviously $f(m) = \sum_{p|m} f(p)$. Put

$$A_n = \sum_{p\leq n}\frac{f(p)}{p}, \quad B_n = \left(\sum_{p\leq n}\frac{f(p)^2}{p}\right)^{1/2}.$$

Then we have:

Theorem 2.1 (Erdös-Kac). *If $B_n \to \infty$ as $n \to \infty$ and K_n denotes the* *number of integers m from 1 up to n for which*

$$f(m) < A_n + w\sqrt{2}B_n$$

then

$$\lim_{n\to\infty} \frac{K_n}{n} = \pi^{-1/2} \int_{-\infty}^{w} exp(-u^2)du = D(w).$$

2.2 Preparation for the proof.

To prove the theorem we will need a couple of Lemmas.

Lemma 2.2. *Let*

$$f_l(m) = \sum_{p|m \quad p<l} f(p).$$

Then denoting by σ_l the density of the set of integers m for which $f_l(m) < A_l + w\sqrt{2}B_l$ one has

$$\lim_{l\to\infty} \sigma_l = D(w).$$

The Lemma above is the only "statistical" lemma in the proof. Using this lemma, the main result will be established bu purely number-theoretic methods.The following lemma is just simple application of Brun's pure sieve. So we omit the proof.

Lemma 2.3. *If m_n tends to ∞ (as $n \to \infty$) more rapidly than any fixed power of s_n, then the number of integers from 1 up to m_n which are not divisible by any prime less than s_n is equal to*

$$\frac{m_n e^{-C}}{\log s_n} + o\left(\frac{m_n}{\log s_n}\right),$$

where C denotes Euler's constant.

Now, let $\phi(n)$ represent a function which tends, as $n \to \infty$, to 0 in a such way that $n^{\phi(n)} \to \infty$. The function $n^{\phi(n)}$ will be denoted by α_n and $n^{\sqrt{\phi(n)}}$ by β_n. Let $a_1(n)$, $a_2(n)$,... be the integers whose prime factors are all less than α_n, and let $\psi(m; n)$ be the greatest a_i which divides m. We then have the following:

Lemma 2.4. *The number of integers $m \leq n$ for which $\psi(m; n) = a_i(n)$, where $a_i(n) \leq \beta_n$ is equal to*

$$\frac{e^{-C} n}{a_i(n)\phi(n)\log s_n} + o\left(\frac{n}{a_i(n)\phi(n)\log s_n}\right).$$

Proof. In fact we are looking for the integers $m \leq \frac{n}{a_i(n)}$ which are not divisible by any prime less than $n^{\phi(n)}$. In order to use the Lemma above, with

$m_n = \frac{n}{a_i(n)}$ and $s_n = n^{\phi(n)}$, it's enough to check that

$$\lim_{n\to\infty} \frac{s_n^\alpha}{m_n} = 0$$

for any fixed α. (i.e. $m_n \to \infty$ more rapidly than any fixed power of s_n). But

$$\frac{s_n^\alpha}{m_n} = a_i(n) \cdot n^{\alpha\phi(n)} - 1 \leq n^{\sqrt{\phi(n)}+\alpha\phi(n)-1}$$

If we let $y_n = n^{\sqrt{\phi(n)}+\alpha\phi(n)-1}$, then

$$\log y_n = [\sqrt{\phi(n)} + \alpha\phi(n) - 1] \log n.$$

And since $\phi(n) \to 0$ as $n \to \infty$, we have $\log y_n \to -\infty$, i.e. $y_n \to 0$ as $n \to \infty$. So we are done. □

Lemma 2.5. *The number y of integers $\leq M$ divisible by an for which $a_i(n) > \beta_n$ is less than $bM\sqrt{\phi(n)}$, where b is an absolute constant.*

Proof. We should notice that m is divisible by $a_i(n)$ if and only if $\psi(m;n)$ is divisible by $a_i(n)$, and therefore $\psi(m;n) \geq a_i(n)$. So we have

$$(\beta_n)^y < \prod_{m=1}^{M} \psi(m;n) = \prod_{p<\alpha_n} p^{\sum_{r=1}^\infty [\frac{M}{p^r}]}$$

But

$$\sum_{r=1}^\infty \left[\frac{M}{p^r}\right] \leq M \sum_{r=1}^\infty \left[\frac{1}{p^r}\right] = \frac{M}{p^r} \cdot \frac{1}{1-\frac{1}{p}} \leq \frac{2M}{p}.$$

So we get

$$(\beta_n)^y < \prod_{p<\alpha_n} p^{\frac{2M}{p}}.$$

And since

$$\lg \prod_{p<\alpha_n} p^{\frac{2M}{p}} = 2M \sum_{p<\alpha_n} p^{-1} \log p \sim 2M\phi(n) \log n,$$

we obtain

$$(\beta_n)^y < \prod_{m=1}^{M} \psi(m;n) < \prod_{p<\alpha_n} p^{\frac{2m}{p}} < n^{bM\phi(n)}$$

for some absolute constant b.

Hence we get

$$(\beta_n)^y = n^{\sqrt{\phi(n)}y} < n^{bM\phi(n)}. \quad \text{i.e.} \quad y < bM\sqrt{\phi(n)}.$$

We are done. ∎

Corollary 2.6. *The density of the integers which are divisible by an $a_i(n) > \beta_n$ is less than $b\sqrt{\phi(n)}$.*

Lemma 2.7. *Denote by l_n the number of integers from 1 up to n for which*

$$f_{\alpha_n}(m) < A_{\alpha_n} + w\sqrt{2}B_{\alpha_n} \quad (i)$$

Then

$$\lim_{n\to\infty} \frac{l_n}{n} = D(w).$$

Proof. Divide the integers from 1 up to n which satisfy (i) into classes E_1, E_2,... so that m belongs to E_i if and only if $\psi(m; n) = a_i(n)$. And denote by $|E_i|$ the number of integers in E_i. Then one obviously has

$$l_n = \sum_i |E_i| = \sum_{a_i(n)\le\beta_n} |E_i| + \sum_{a_i(n)>\beta_n} |E_i|.$$

By Lemma 2.5 we have

$$\sum_{a_i(n)>\beta_n} |E_i| < bn\sqrt{\phi(n)}.$$

Therefore it's safficient to prove that

$$\frac{1}{n} \sum_{a_i(n)\le\beta_n} |E_i| \to D(w) \quad \text{as} \quad n \to \infty$$

On the other hand by Lemma 2.4 we have

$$\sum_{a_i(n)\le\beta_n} |E_i| = \left(\frac{e^{-C}n}{\phi(n)\log n} + o\left(\frac{n}{\phi(n)\log n}\right)\right) \sum_{a_i\le\beta_n}{}^{'} \frac{1}{a_i(n)} \quad \text{(ii)}$$

where the dash in the summation indicates that it's extended over the a_i's satisfying $f_{\alpha_n}(a_i) < A_{\alpha_n} + w\sqrt{2}B_{\alpha_n}$. (We should notice that $f_{\alpha_n}(a_i(n)) = f_{\alpha_n}(m)$ for all $m \in E_i$). Now, in order to evaluate $\sum'$, divide all the integers into classes F_1, F_2,... having the property that $m \in F_i$ if and only if $\psi(m; n) = a_i(n)$ and let $\{F_i\}$ denote the density of the set F_i. Consider now the set $\cup' F_i$ where the dash in union has the same meaning as above. By putting $l = \alpha_n$ and using Lemma 2.1 we have $\{\cup' F_i\} \to D(w)$ as $n \to \infty$ or $\{\cup' F_i\} = D(w) + o(1)$. Now

$$\cup' F_i = \left(\cup_{a_i\le\beta_n}{}' F_i\right) \cup \left(\cup_{a_i>\beta_n}{}' F_i\right) \quad \text{(iii)}$$

and by Lemma 2.5

$$\left\{ \cup_{a_i > \beta_n}{}^{\text{`}} F_i \right\} < b\sqrt{\phi(n)} \quad \text{(iv)}$$

Furthermore there is only a finite number of a_i's which are less than β_n and therefore $\left\{ \cup_{a_i \leq \beta_n}{}^{\text{`}} F_i \right\} = \cup_{a_i \leq \beta_n}{}^{\text{`}} \{F_i\}$. But

$$\{F_i\} = \frac{1}{a_i(n)} \left(\frac{e^{-C} n}{\phi(n) \log n} + o\left(\frac{n}{\phi(n) \log n} \right) \right) \quad \text{(v)}$$

so we have

$$\left\{ \cup_{a_i \leq \beta_n}{}^{\text{`}} F_i \right\} = \left( \frac{e^{-C}}{\phi(n) \log n} + o\left( \frac{1}{\phi(n) \log n} \right) \right) \sum_{a_i \leq \beta_n}{}^{\text{`}} \frac{1}{a_i(n)}$$

Finally by (iii), iv and v we get;

$$D(w) - b\sqrt{\phi(n)} < \left(\frac{e^{-C}}{\phi(n) \log n} + o\left(\frac{1}{\phi(n) \log n} \right) \right) \sum_{a_i \leq \beta_n}{}^{\text{`}} \frac{1}{a_i(n)} < D(w) + o(1).$$

Combining this with (ii) we get the desired result. □

2.3 Proof of the Main Theorem.

Now we are ready to prove the theorem. Since $f(m) = \sum_{p|m} f(p)$ and

$$f_{\alpha_n}(m) = \sum_{\substack{p|m \\ p < \alpha_n}} f(p),$$

for $m \leq n$, we have

$$|f(m) - f_{\alpha_n}(m)| = \left| \sum_{\substack{p|m \\ \alpha_n \leq p \leq n}} f(p) \right|.$$

But as $|f(p)| \leq 1$ we have that $|f(m) - f_{\alpha_n}(m)|$ is less than the number of those prime divisors of m which are $\geq \alpha_n = n^{\phi(n)}$. And since $(\alpha_n)^{\frac{1}{\phi(n)}} = n$ we get $|f(m) - f_{\alpha_n}(m)| < \frac{1}{\phi(n)}$. Also

$$|A_n - A_{\alpha_n}| = |\sum_{\alpha_n \leq p \leq n} p^{-1} f(p)| \leq \sum_{p<n} \frac{1}{p} - \sum_{p<\alpha_n} \frac{1}{p} \sim \left[\log\log n - \log\log n^{\phi(n)} \right]$$

therefore

$$|A_n - A_{\alpha_n}| < -C_1 \log \phi(n) = C_1 \log \frac{1}{\phi(n)} \quad \text{for some} \quad C_1 > 0.$$

Similarly

$$|B_n - B_{\alpha_n}| < -C_2 \log \phi(n) = C_2 \log \frac{1}{\phi(n)} \quad \text{for some} \quad C_2 > 0.$$

Now choose $\phi(n)$ so that $\frac{1}{\phi(n)} = o(B_n)$. Since $|B_n - B_{\alpha_n}| < C_2 \log \frac{1}{\phi(n)}$ and $\frac{1}{\phi(n)} = o(B_n)$ we have $B_n \sim B_{\alpha_n}$. Letting $m \leq n$ satisfy $f(m) < A_n + w\sqrt{2}B_n$ we have

$$\begin{aligned} f_{\alpha_n}(m) - \frac{1}{\phi(n)} &< f(m) < A_n + w\sqrt{2}B_n \\ &< A_{\alpha_n} + w\sqrt{2}B_{\alpha_n} + C_1 \log \frac{1}{\phi(n)} + C_2 \log \frac{1}{\phi(n)} \end{aligned}$$

and since $\frac{1}{\phi(n)} = o(B_n)$, for sufficintly large n

$$f_{\alpha_n}(m) < A_{\alpha_n} + w\sqrt{2}B_{\alpha_n} + \epsilon\sqrt{2}B_{\alpha_n}$$

Similarly for sufficiently large n and for $m \leq n$ satisfying $f_{\alpha_n}(m) < A_{\alpha_n} + (w-\epsilon)\sqrt{2}B_{\alpha_n}$, we have

$$\begin{aligned} f(m) &< A_n + w\sqrt{2}B_n + \frac{1}{\phi(n)} + C_1 \log \frac{1}{\phi(n)} + C_2 \log \frac{1}{\phi(n)} - \epsilon\sqrt{2}B_{\alpha_n} \\ &< A_n + w\sqrt{2}B_n \quad \text{since} \quad B_n \sim B_{\alpha_n}. \end{aligned}$$

What we get at last is

$$D(w-\epsilon) \leq \liminf \frac{K_n}{n} \leq \limsup \frac{K_n}{n} \leq D(w+\epsilon)$$

And since $\epsilon > 0$ is arbitrary, then the proof is completed.

The immediate corollary is in the case of $\omega(n)$, where ω counts distinct prime factors of n.

Corollary 2.8. *If $\omega(n)$ denotes the number of prime divisors of m, and K_n the number of those integers from 1 up to n for which $\omega(n) < \log\log n + w\sqrt{2\log\log n}$, then*

$$\lim_{n\to\infty} \frac{K_n}{n} = \pi^{-1/2} \int_{-\infty}^{w} \exp(-u^2) du.$$

3 Analytic Proof of Erdös-Kac Theorem.

3.1 Introduction.

The aim of this chapter is to give a new proof of the theorem of P. Erdös and M. Kac concerning the function $\Omega(n)$. To remember what the theorem states; if $N_n(\Omega, x)$ denotes the number of those natural numbers $k \leq n$ for which

$$\frac{\Omega(k) - \log\log n}{\sqrt{\log\log n}} < x$$

then we have

$$\lim_{n\to\infty} \frac{N_n(\Omega, x)}{n} = \phi(x),$$

where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du.$$

In other words, the random variable ξ_n, which assumes the value $\Omega(1)$, $\Omega(2)$,..., $\Omega(n)$, each with the same probability $1/n$, is, for $n \to \infty$, asymptotically normally distributed with mean value $\log\log n$ and standard deviation $\sqrt{\log\log n}$.

W. J. Leveque introduced certain modifications of the proof of Erdös and Kac and obtained the following improvement of their result:

$$\frac{N_n(\Omega, x)}{n} = \phi(x) + O\left(\frac{\log\log\log n}{\sqrt[4]{\log\log n}}\right).$$

Le Veque conjectured that the error term is actually of order $1/\sqrt{\log\log n}$.

3.2 Analytic proof of the theorem.

Put by definition $\Omega(1) = 0$. We prove the following

Theorem 3.1 (Erdös-Kac)**.** *Let us denote by $N_n(\Omega, x)$ the number of those positive integers $k \leq n$ for which*

$$\frac{\Omega(k) - \log\log n}{\sqrt{\log\log n}} < x$$

Then putting

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du,$$

we have

$$\lim_{n\to\infty} \frac{N_n(\Omega, x)}{n} = \phi(x) \quad (-\infty < x < \infty).$$

The proof is due to A. Renyi and P. Turan [13]. The strategy is to pass from the distribution function

$$F_n(x) = P\left(\frac{\Omega(k) - \log\log n}{\sqrt{\log\log n}} < x\right) = \frac{N_n(\Omega, x)}{n}.$$

to the characteristic function

$$\varphi_n(u) = \frac{1}{n}\sum_{k\le n} e^{iu\left(\frac{\Omega(k)-\log\log n}{\sqrt{\log\log n}}\right)}$$

and then show that $\varphi_n(u) \to e^{-u^2/2}$, as $n \to \infty$.

Proof. Consider the Dirichlet series

$$\lambda(s, u) = \sum_{n=1}^{\infty} \frac{e^{iu\Omega(n)}}{n^s} \tag{3.1}$$

where u is real and $s = \sigma + it$ a complex variable. The series on the right of (3.1) is convergent for $\sigma > 1$. As $e^{iu\Omega(n)}$ is (completely) multiplicative, i.e.,

$$e^{iu\Omega(nm)} = e^{iu\Omega(n)}e^{iu\Omega(m)} \tag{3.2}$$

for any pair n, m of natural numbers, it follows that

$$\lambda(s, u) = \prod \frac{1}{(1 - e^{iu}/p^s)}, \tag{3.3}$$

where p runs over all primes. Now let us put

$$\mu(s, u) = \frac{\lambda(s, u)}{(\zeta(s))^{e^{iu}})} \tag{3.4}$$

where

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod \frac{1}{(1 - 1/p^s)} \tag{3.5}$$

is the zeta-function of Riemann, and the product on the right of (3.5) is extended over all primes p and $\log\zeta(s)$ is real for $\sigma > 1$.

Evidently for $\sigma > 1$

$$\begin{aligned}\log \mu(s,u) &= \log \lambda(s,u) - e^{iu} \log \zeta(s) \\ &= -\sum_p \log(1 - \frac{e^{iu}}{p^s}) - e^{iu} \sum_p \log(1 - \frac{1}{p^s}) \\ &= \sum_{p=2}^{\infty} \sum_{k=1}^{\infty} \frac{e^{iuk}}{kp^{ks}} - e^{iu} \sum_{p=2}^{\infty} \sum_{k=1}^{\infty} \frac{1}{kp^{ks}} \\ &= \sum_{p=2}^{\infty} \sum_{k=2}^{\infty} \frac{e^{iu}(e^{iu(k-1)-1})}{kp^{ks}}.\end{aligned}$$

As the last series above converges uniformly for $\sigma \geq \frac{1}{2} + \epsilon$ where $\epsilon > 0$ is arbitrary, it follows that, for any fixed real value of u, $\mu(s,u)$ is a regular function of s in the open half-plane $\sigma > \frac{1}{2}$. Later on we shall need the following estimation

$$|\log \mu(s,u)| \leq |u| \tag{3.6}$$

for $s = \sigma + it$, $\sigma \geq 1$, which follows from the equality for $\log \mu(s,u)$ above and the estimation

$$\begin{aligned}\frac{|e^{iu(k-1)} - 1|}{k} &= \frac{|e^{\frac{iu(k-1)}{2}} - e^{-\frac{iu(k-1)}{2}}|}{k} \\ &= \frac{|2i \sin \frac{u(k-1)}{2}|}{k} \\ &\leq \frac{|u|(k-1)}{k} \leq |u|\end{aligned}$$

where we used the fact that $|\sin x| \leq |x|$ for all x real.

Now putting

$$S(n,u) = \sum_{k=1}^{n} e^{iu\Omega(k)} \log \frac{n}{k} \tag{3.7}$$

by (1.10), taking $m = 1$, $x = n$ and $a_k = {}^{iu\Omega(k)}$, we have

$$S(n,u) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{n^s \lambda(s,u) ds}{s^2} \tag{3.8}$$

where $c > 1$.

In what follows we will always suppose $|u| \leq \pi/6$, which implies $\cos u \geq \frac{1}{2}$. Let us effect the decomposition

$$\lambda(s,u) = \frac{\mu(s,u)}{(s-1)^{e^{iu}}} + \mu(s,u) \left((\zeta(s))^{e^{iu}} - \frac{1}{(s-1)^{e^{iu}}} \right) \tag{3.9}$$

with $\log(s-1)$ real for $s > 1$ and put

$$I_1 = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{n^s \mu(s,u) ds}{s^2 (s-1)^{e^{iu}}} \tag{3.10}$$

and

$$I_2 = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{n^s \mu(s,u)}{s^2} \left((\zeta(s))^{e^{iu}} - \frac{1}{(s-1)^{e^{iu}}} \right) ds. \tag{3.11}$$

Then we have

$$S(n,u) = I_1 + I_2. \tag{3.12}$$

Let us consider first I_2. The integrand is regular for $s = \sigma + it$, $\sigma \geq 1$, except for $s = 1$, but it is continuous at this point also, because $(s-1)\zeta(s)$ is regular and equal to 1 at $s = 1$. Thus we have

$$\lim_{s \to 1} \left| \frac{[\zeta(s)(s-1)]^{e^{iu}} - 1}{(s-1)^{e^{iu}}} \right| = \lim_{s \to 1} \left| \frac{[\zeta(s)(s-1)]^{\cos u} - 1}{(s-1)^{\cos u}} \right|$$

and since for $0 \leq \alpha \leq 1$

$$\lim_{t \to 0} \frac{(1+t)^\alpha - 1}{t^\alpha}$$

exists, then

$$\frac{[\zeta(s)(s-1)]^{e^{iu}} - 1}{(s-1)^{e^{iu}}} \tag{3.13}$$

is also continuous at $s = 1$, though of course it has a branching point there. Now we aim to push the path of integration to the line $s = 1 + it$ $(-\infty < t < \infty)$. But

$$\left| \int_{c+iT}^{c+i\infty} \frac{n^s \mu(s,u)}{s^2} \left((\zeta(s))^{e^{iu}} - \frac{1}{(s-1)^{e^{iu}}} \right) ds \right|$$
$$\leq \int_T^\infty \frac{n^c |\mu(c+it,u)|}{t^2} \left((|\zeta(c+it))|^{\cos u} + \frac{1}{T^{\cos u}} \right) dt$$

and the last expression goes to zero as $T \to \infty$. Similarly

$$\left| \int_{c-i\infty}^{c-iT} \frac{n^s \mu(s,u)}{s^2} \left((\zeta(s))^{e^{iu}} - \frac{1}{(s-1)^{e^{iu}}} \right) ds \right| \to 0$$

as $T \to \infty$. Also we should notice that on the line segment $s = \sigma + iT$, where $1 \le \sigma \le c$, the given integral is

$$\left| \int_{c+iT}^{1+iT} \frac{n^s \mu(s,u)}{s^2} \left((\zeta(s))^{e^{iu}} - \frac{1}{(s-1)^{e^{iu}}} \right) ds \right| \le \int_c^1 \frac{n^\sigma |\mu(\sigma+iT,u)|}{T^2} \left((|\zeta(\sigma+iT))|^{\cos u} + \frac{1}{T^{\cos u}} \right) d\sigma$$

and the last expression goes to zero as $T \to \infty$. Similar for on the line segment $s = \sigma - iT$ where again $1 \le \sigma \le c$.

So we are allowed to push the path of integration to the line $s = 1 + it$. Now we will apply partial integration in such a manner that n^s is chosen as the factor to be integrated. We will need the following well-known estimates for Riemann-zeta function on the present line.

$$|\zeta(1+it)| = O(\log t), \tag{3.14}$$

$$\left| \frac{\zeta'(1+it}{\zeta(1+it)} \right| = O(\log t). \tag{3.15}$$

After the integration by parts, we will first need to estimate

$$\frac{1}{2\pi i} \frac{n^s}{\log n} \frac{\mu(s,u)}{s^2} \left((\zeta(s))^{e^{iu}} - \frac{1}{(s-1)^{e^{iu}}} \right)$$

at the end points $1+iT$, $1-iT$. But

$$\left| \frac{1}{2\pi i} \frac{n^s}{\log n} \frac{\mu(s,u)}{s^2} \left((\zeta(s))^{e^{iu}} - \frac{1}{(s-1)^{e^{iu}}} \right) \right|_{s=1+iT} \ll \frac{n}{\log n} \frac{1}{T^2} \left(|\zeta(1+iT)|^{\cos u} + \frac{1}{T^{\cos u}} \right)$$

where, by (3.14), the last expression goes to 0 as $T \to \infty$. Similarly for the point $s = 1 - iT$.

Therefore, in order to complete the estimation of I_2 we need to handle

$$\frac{1}{2\pi i} \int_{1-iT}^{1+iT} \frac{n^s}{\log n} \left(\frac{\mu(s,u)}{s^2} \right)' \left((\zeta(s))^{e^{iu}} - \frac{1}{(s-1)^{e^{iu}}} \right) ds \tag{3.16}$$

$$+ \frac{1}{2\pi i} \int_{1-iT}^{1+iT} \frac{n^s}{\log n} \frac{\mu(s,u)}{s^2} \left((\zeta(s))^{e^{iu}} - \frac{1}{(s-1)^{e^{iu}}} \right)' ds. \tag{3.17}$$

The first part above is

$$\left|\frac{1}{2\pi i}\int_{1-iT}^{1+iT}\frac{n^s}{\log n}\left(\frac{\mu'(s,u)}{s^2}-2\frac{\mu(s,u)}{s^3}\right)\left((\zeta(s))^{e^{iu}}-\frac{1}{(s-1)^{e^{iu}}}\right)ds\right|$$
$$\ll\int_{-T}^{T}\frac{n}{\log n}\frac{1}{1+t^2}\left|(\zeta(s))^{e^{iu}}-\frac{1}{(s-1)^{e^{iu}}}\right|dt.$$

Now, by dividing the interval $|t|\leq T$ into three pieces

$$\int_{|t|\leq T}=\int_{|t|\leq 1}+\int_{1<|t|\leq T}$$

and using the continuity of (3.13) around $t=0$, we get

$$\frac{1}{2\pi i}\int_{1-iT}^{1+iT}\frac{n^s}{\log n}\left(\frac{\mu(s,u)}{s^2}\right)'\left((\zeta(s))^{e^{iu}}-\frac{1}{(s-1)^{e^{iu}}}\right)ds\ll\frac{n}{\log n}$$

Now, to estimate (3.17), which is in fact

$$\frac{1}{2\pi i}\int_{1-iT}^{1+iT}\frac{n^s}{\log n}\frac{\mu(s,u)}{s^2}\left(e^{iu}(\zeta'(s))^{e^{iu}-1}+e^{iu}\frac{1}{(s-1)^{e^{iu}+1}}\right)ds$$

we use the same arguments above to get

$$I_2=O(\frac{n}{\log n}),$$

where the O-sign uniformly in $-\frac{1}{6}\pi\leq u\leq\frac{1}{6}$.

Let us now turn to the investigation of I_1. Clearly we have

$$I_1=I_{11}+I_{12}\tag{3.18}$$

where

$$I_{11}=\frac{\mu(1,u)}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{n^s ds}{s^2(s-1)^{e^{iu}}}\tag{3.19}$$

and

$$I_{12}=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{n^s(s-1)^{1-e^{iu}}}{s^2}\left(\frac{\mu(s,u)-\mu(1,u)}{s-1}\right)ds.\tag{3.20}$$

First we should notice that

$$\frac{\mu(s,u)-\mu(1,u)}{s-1}$$

is regular, and bounded for the half-plane $\Re s\geq 1$. Further $\Re(1-e^{iu})\geq 0$, so it's reasonable to transform the path of integration of I_{12} to the line $\Re s=1$.

We will just follow the same arguments we did before:

$$\left| \int_{c+iT}^{c+i\infty} \frac{n^s (s-1)^{1-e^{iu}}}{s^2} \left(\frac{\mu(s,u) - \mu(1,u)}{s-1} \right) \right|$$
$$\ll n^c \int_T^\infty \frac{t^{1-\cos u}}{t^2} \frac{1}{t} dt,$$

and the last expression goes to 0 as $T \to \infty$. We can get the similar result on the line segment $s = c - it$ where $t \geq T$.

On the other hand, on the line segment from $c + iT$ to $1 + iT$, we have

$$\left| \int_{c+iT}^{1+iT} \frac{n^s (s-1)^{1-e^{iu}}}{s^2} \left(\frac{\mu(s,u) - \mu(1,u)}{s-1} \right) \right|$$
$$\ll \int_c^1 \frac{n^\sigma T^{1-\cos u}}{T^2} \frac{1}{T} d\sigma,$$

which again goes to 0 as $T \to \infty$. Since we can do the similar staff on the line segment from $1 - iT$ to $c - iT$, the transformation of the path of integration to the line $\Re s = 1$ is justified.

Now, applying again partial integration in such a manner that n^s is chosen as the factor to be integrated, we aim to obtain again uniformly in u

$$I_{12} = O(\frac{n}{\log n}). \tag{3.21}$$

To do this, firstly we need to estimate

$$\frac{1}{2\pi i} \frac{n^s (s-1)^{1-e^{iu}}}{s^2} \left(\frac{\mu(s,u) - \mu(1,u)}{s-1} \right)$$

at the points $1 + iT$ and $1 - iT$. But

$$\left| \frac{1}{2\pi i} \frac{n^s (s-1)^{1-e^{iu}}}{s^2} \left(\frac{\mu(s,u) - \mu(1,u)}{s-1} \right) \right|_{s=1+iT}$$
$$\ll \frac{n}{\log n} \frac{T^{1-\cos u}}{T^2} \frac{1}{T}$$

which goes to 0 as $T \to \infty$. Similar result can be obtained easily at the point $s = 1 - iT$. So it remains only to estimate

$$\int_{1-iT}^{1+iT} \frac{n^s}{\log n} \left(\frac{(s-1)^{1-e^{iu}}}{s^2} \right)' \left(\frac{\mu(s,u) - \mu(1,u)}{s-1} \right) ds$$
$$+ \int_{1-iT}^{1+iT} \frac{n^s}{\log n} \frac{(s-1)^{1-e^{iu}}}{s^2} \left(\frac{\mu(s,u) - \mu(1,u)}{s-1} \right)' ds$$

In the first part above, we first divide the range of integration as following

$$\int_{1-iT}^{1+iT} = \int_{1-i}^{1+i} + \int_{1+i}^{1+iT} + \int_{1-iT}^{1-i}. \tag{3.22}$$

and then calculate separately. In the first range

$$\int_{1-i}^{1+i} \frac{n^s}{\log n} \left(\frac{(s-1)^{1-e^{iu}}}{s^2} \right)' \left(\frac{\mu(s,u) - \mu(1,u)}{s-1} \right) ds$$
$$\ll \int_{-1}^{1} \frac{n}{\log n} \frac{1}{t^{\cos u}} dt \ll \frac{n}{\log n}.$$

The estimations for other two ranges in (3.22) can be done similarly. Finally we handle up

$$\int_{1-iT}^{1+iT} \frac{n^s}{\log n} \frac{(s-1)^{1-e^{iu}}}{s^2} \left(\frac{\mu(s,u) - \mu(1,u)}{s-1} \right)' ds \tag{3.23}$$

by again dividing the range of integration into the same pieces as in (3.22) and use similar arguments to get (3.21).

Now, as regards I_{11}, we have

$$I_{11} = \mu(1,u)(I_{111} - I_{112}) \tag{3.24}$$

where

$$I_{111} = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{n^s ds}{(s-1)^{e^{iu}}} \tag{3.25}$$

and

$$I_{112} = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{n^s (s-1)^{1-e^{iu}(s+1)}}{s^2} ds. \tag{3.26}$$

The integral I_{112} can be transformed again to the line $\Re s = 1$ and by integrating partially we obtain as before uniformly in u

$$I_{112} = O\left(\frac{n}{\log n} \right). \tag{3.27}$$

On the other hand, by transforming the integral I_{111} and using the well known integral representation of the Γ-function,

$$\Gamma(z) = \int_0^\infty e^{-u} u^{z-1} du \quad (\Re z > 1),$$

further the functional equation

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin \pi z},$$

we obtain

$$I_{111} = \frac{n(\log n)^{e^{iu}-1}}{\Gamma(e^{iu})}. \tag{3.28}$$

Collecting our results we obtain by virtue of (3.7), (3.12), (3.18), (3.19), (3.21), (3.24), (3.27), and (3.28) uniformly for $-\frac{1}{6}\pi \le u \le \frac{1}{6}\pi$

$$S(n,u) = n\frac{\mu(1,u)}{\Gamma(e^{iu})}(\log n)^{e^{iu}-1} + O\left(\frac{n}{\log n}\right). \tag{3.29}$$

Let us now put

$$s(n,u) = \sum_{k \le n} e^{iuV(k)}; \tag{3.30}$$

then trivially

$$|s(y,u) - s(x,u)| \le |y - x|. \tag{3.31}$$

Since

$$S(n,u) = \sum_{k \le n} e^{iuV(k)}(\log n - \log k) = \int_1^n \frac{s(x,u)}{x} dx, \tag{3.32}$$

we have for any $\lambda > 0$

$$\begin{aligned} s(n,u) &= \frac{\int_n^{n+n\lambda} \frac{s(n,u)}{x} dx}{\log(1+\lambda)} \\ &= \frac{\int_n^{n+n\lambda} \frac{s(x,u)}{x} dx + \int_n^{n+n\lambda} \frac{s(n,u)-s(x,u)}{x} dx}{\log(1+\lambda)} \\ &= \frac{S(n+\lambda n, u) - S(n,u) + \int_n^{n+n\lambda} \frac{s(n,u)-s(x,u)}{x} dx}{\log(1+\lambda)}. \end{aligned}$$

Thus from (3.31) uniformly

$$s(n,u) = \frac{S(n(1+\lambda),u) - S(n,u)}{\log(\lambda+1)} + O\left(\frac{\int_n^{n+n\lambda} \frac{x-n}{x} dx}{\log(\lambda+1)}\right)$$

and since

$$\frac{1}{\log(\lambda+1)} = \frac{1}{\lambda} + O(\lambda) \tag{3.33}$$

we have

$$s(n,u) = \frac{S(n(1+\lambda),u) - S(n,u)}{\log(\lambda+1)} + O(\lambda n). \tag{3.34}$$

But if $0 < \lambda \leq \frac{1}{2}$

$$\begin{aligned} &\frac{(\lambda+1)(\log n(\lambda+1))^{e^{iu-1}} - (\log n)^{e^{iu-1}}}{\log(\lambda+1)} \\ &= (\log n)^{e^{iu-1}} \left(1 + O(\lambda) + O(\frac{|u|}{\log n})\right), \end{aligned} \tag{3.35}$$

the O-estimates being uniform in n, u and λ. To see (3.35), we notice that, the left part of the equality is in fact

$$(\log n)^{e^{iu}-1} \left[\frac{(\lambda+1)}{\log(1+\lambda)} \left(1 - \frac{\log(1+\lambda)}{\log n + \log(1+\lambda)}\right)^{1-e^{iu}} - \frac{1}{\log(1+\lambda)} \right].$$

But since

$$\left(1 - \frac{\log(1+\lambda)}{\log n + \log(1+\lambda)}\right)^{1-e^{iu}} \ll 1 - \frac{\log(1+\lambda)}{\log n + \log(1+\lambda)}(1-\cos u),$$

and

$$1 - \cos u \ll |u|$$

we verify (3.35). Now,

$$\begin{aligned} s(n,u) = n\frac{\mu(1,u)}{\Gamma(e^{iu})} \cdot \frac{(\lambda+1)(\log n(\lambda+1))^{e^{iu-1}} - (\log n)^{e^{iu-1}}}{\log(\lambda+1)} \\ + \frac{1}{\log(1+\lambda)} O(\frac{n}{\log n}) + O(n\lambda). \end{aligned}$$

But since

$$\frac{1}{\log(1+\lambda)} O(\frac{n}{\log n}) \ll (\frac{1}{\lambda} + \lambda)\frac{n}{\log n} \ll \frac{n}{\lambda \log n},$$

and choosing

$$\lambda = \left(\frac{|u|}{\log n}\right)^{1/2},$$

we obtain uniformly in u

$$\frac{s(n,u)}{n} = \frac{\mu(1,u)}{\Gamma(e^{iu})} \cdot (\log n)^{e^{iu}-1} \left(1 + O\left(\left(\frac{|u|}{\log n}\right)^{1/2}\right)\right) + O\left(\frac{1}{\sqrt{|u| \log n}}\right). \tag{3.36}$$

Now, we should notice that, $s(n,u)/n$ is the characteristic function of the probability distribution of a random variable ξ_n, which takes on the values $\Omega(1)$, $\Omega(2)$, ..., $\Omega(n)$ with probability $1/n$. Therefore, in order to complete the proof, by Theorem 1.5, it suffices to show that putting

$$\varphi_n(u) = \frac{s(n, u/\sqrt{\log\log n})e^{-iu\sqrt{\log\log n}}}{n} \tag{3.37}$$

we have

$$\lim_{n\to\infty} \varphi_n(u) = e^{-u^2/2}. \tag{3.38}$$

But

$$\mu(1, \frac{u}{\sqrt{\log\log n}}) \quad \text{and} \quad \Gamma(e^{i\frac{u}{\sqrt{\log\log n}}})$$

go to 1 as $n \to \infty$. Also

$$e^{\log\log n \cdot e^{i\frac{u}{\sqrt{\log\log n}}}} \cdot e^{-iu\sqrt{\log\log n}} \to e^{-u^2/2}.$$

Thus the theorem is proved. □

3.3 Proof of the conjecture of LeVeque

In this section we shall prove

Theorem 3.2 (Conjecture of LeVeque)**.** . *Let $N_n(V,x)$ denote the number of those natural numbers $k \le n$ for which*

$$\frac{\Omega(k) - \log\log n}{\sqrt{\log\log n}} < x.$$

Then we have uniformly in x

$$\frac{N_n(\Omega, x)}{n} = \phi(x) + O\left(\frac{1}{\sqrt{\log\log n}}\right).$$

Proof. In order to prove the theorem, we follow the same method as that used in the proof of the previous theorem. The only difference consists in the fact that now, as we want to estimate the rate of convergence of $(1/n)N_n(V,x)$ to $\phi(x)$, we have to consider the rate of convergence of $\varphi_n(u)$ defined by (3.37), to $e^{-u^2/2}$ and apply the following theorem of C. G. Esseen [11]:

If $F(x)$ and $G(x)$ are two distribution functions, $G(x)$ exists for all x and $|G(x)| \leq A$,

$$f(u) = \int_{-\infty}^{+\infty} e^{iux} dF(x) \quad \text{and} \quad g(u) = \int_{-\infty}^{+\infty} e^{iux} dG(x)$$

denote the characteristic functions of the two distribution functions respectively, and the following condition is satisfied:

$$\int_{-T}^{+T} \left| \frac{f(u) - g(u)}{u} \right| du < \epsilon, \tag{3.39}$$

then for $-\infty < x < +\infty$

$$|F(x) - G(x)| < K \left(\epsilon + \frac{A}{T} \right),$$

where K is an absolute constant. Let us verify the fulfilment of the condition (3.38) with $G(x) = \phi(x)$ (which implies $A = 1/\sqrt{2\pi}$),

$$F(x) = \frac{N_n(\Omega, x)}{n}, \quad T = \frac{\pi}{6}\sqrt{\log\log n}, \quad \epsilon = \frac{c}{\sqrt{\log\log n}}$$

where $c > 0$ is a constant. Now, since $f(u) = \varphi_n(u)$ and $g(u) = e^{-u^2/2}$, we have only to prove that

$$\int_{-\pi\sqrt{\log\log n}/6}^{+\pi\sqrt{\log\log n}/6} \left| \frac{\varphi_n(u) - e^{-u^2/2}}{u} \right| du = O\left(\frac{1}{\sqrt{\log\log n}} \right).$$

We put

$$\int_{-\pi\sqrt{\log\log n}/6}^{+\pi\sqrt{\log\log n}/6} \left| \frac{\varphi_n(u) - e^{-u^2/2}}{u} \right| du = \Delta_1 + \Delta_2 \tag{3.40}$$

where

$$\Delta_1 = \int_{|u| \leq 1/\sqrt{\log\log n}} \left| \frac{\varphi_n(u) - e^{-u^2/2}}{u} \right| du \tag{3.41}$$

and

$$\Delta_2 = \int_{1/\sqrt{\log\log n} \leq |u| \leq \pi\sqrt{\log\log n}/6} \left| \frac{\varphi_n(u) - e^{-u^2/2}}{u} \right| du. \tag{3.42}$$

Let us consider first Δ_1. Evidently, putting $a = 1/\sqrt{\log\log n}$, we have

$$\int_{-a}^{+a}\left|\frac{\varphi_n(u)-e^{-u^2/2}}{u}\right|du \le \int_{-a}^{+a}\left|\frac{1-\varphi_n(u)}{u}\right|du + \int_{-a}^{+a}\left|\frac{1-e^{-u^2/2}}{u}\right|du. \tag{3.43}$$

Generally if $f(u) = \int_{-\infty}^{+\infty} e^{iux}dF(x)$, then

$$\int_{-a}^{+a}\left|\frac{1-f(u)}{u}\right|du \le 2a\sqrt{\int_{-\infty}^{+\infty} x^2 dF(x)}. \tag{3.44}$$

Thus

$$\int_{-a}^{+a}\left|\frac{1-\varphi_n(u)}{u}\right|du = \left(\frac{1}{\sqrt{\log\log n}}\right). \tag{3.45}$$

To see the last equality, we should notice that, if X is a random variable which takes the value

$$\frac{\Omega(k)-\log\log n}{\sqrt{\log\log n}}$$

for each $1 \le k \le n$ with probability $\frac{1}{n}$, then

$$\int_{-\infty}^{+\infty} x^2 dF(x) = E(X^2) = \frac{1}{n}\sum_{k\le n}\frac{(\Omega(k)-\log\log n)^2}{\log\log n}$$

which is bounded by Turan. As

$$e^{-u^2/2} = \sum_{m=0}^{\infty}\frac{(-1)^m u^{2m}}{2^m m!},$$

we have

$$\left|\frac{1-e^{-u^2/2}}{u}\right| = o\left(\frac{1}{\sqrt{\log\log n}}\right).$$

Therefore

$$\int_{-a}^{+a}\left|\frac{1-e^{-u^2/2}}{u}\right|du = o\left(\frac{1}{\sqrt{\log\log n}}\right),$$

which follows that

$$\Delta_1 = O\left(\frac{1}{\sqrt{\log\log n}}\right).$$

Let us now turn to the estimation of Δ_2. Owing to the inequality

$$|e^{iz}-1-iz+z^2/2| \le |z|^3/6,$$

valid for real z, we have from (3.36)

$$\left|\frac{\varphi_n(u) - e^{-u^2/2}}{u}\right| \leq \frac{1}{|u|^{3/2}} O\left(\frac{1}{\sqrt{\log n}}\right) + A(u)$$

where

$$A(u) = \frac{e^{-u^2/2}}{|u|}\left[\left(1 + \left(\frac{|u|}{\sqrt{\log\log n}}\right)\right) e^{\vartheta|u|^3/6\sqrt{\log\log n}} - 1\right] \quad \text{and} \quad |\vartheta| \leq 1.$$

Thus

$$\Delta_2 \leq O\left(\frac{1}{\log^{1/3} n}\right) + \int_{1/\sqrt{\log\log n}}^{\pi\sqrt{\log\log n}/6} A(u)du. \tag{3.46}$$

In order to estimate the integral on the right of (3.46) we remark that for $|u| \leq \sqrt[6]{\log\log n}$ we have

$$e^{\vartheta|u|^3/\sqrt{\log\log n}} = 1 + O\left(\frac{|u|^3}{\sqrt{\log\log n}}\right),$$

which implies

$$\int_{1/\sqrt{\log\log n}}^{\sqrt[6]{\log\log n}} A(u)du = O\left(\frac{1}{\sqrt{\log\log n}}\right). \tag{3.47}$$

On the other hand for $\sqrt[6]{\log\log n} < |u| \leq \pi\sqrt{\log\log n}$ we have

$$-\frac{u^2}{2} + \frac{\vartheta|u|^3}{6\sqrt{\log\log n}} \leq -\frac{u^2}{4} \quad \text{for} |\vartheta| \leq 1,$$

and thus we obtain

$$\int_{1/\sqrt{\log\log n}}^{\pi\sqrt{\log\log n}/6} A(u)du.$$

Thus we have completed the proof of Theorem 3.2.

□

4 On the Normal Number of Prime Factors of $\phi(n)$

4.1 Introduction.

Denote by $\Omega(n)$ the total number of prime factors of n, counting multiplicity. For each $x \geq 3$, u, let

$$G(x,u) = \frac{1}{x} \cdot \sharp \left\{ n \leq x : \Omega(n) \leq \log\log x + u(\log\log x)^{1/2} \right\}.$$

Then as a corollary of Erdös-Kac theorem, we have

$$\lim_{x\to\infty} G(x,u) = G(u) := (2\pi)^{-1/2} \int_{-\infty}^{u} e^{-t^2/2} dt,$$

the Gaussian normal distribution.

The problem that we consider here is the corresponding problem for the additive function $\Omega(\phi(n))$. What we prove is that

$$\lim_{x\to\infty} \frac{1}{x} \cdot \sharp \left\{ n \leq x : \Omega(\phi(n)) \leq \frac{1}{2}(\log\log x)^2 + \frac{u}{\sqrt{3}}(\log\log x)^{3/2} \right\} = G(u)$$

Thus the normal number of prime factors of $\phi(n)$ is $\frac{1}{2}(\log\log n)^2$ and the "standard deviation" is $3^{-1/2}(\log\log x)^{3/2}$.

To do this, we will need the estimation of the sums

$$\sum_{p\leq x} \Omega(p-1) \quad \text{and} \quad \sum_{p\leq x} \Omega(p-1)^2.$$

Estimations will be simple applications of the Bombieri-Vinogradov and Brun-Titchmarsh theorems. The proof is due to P. Erdös and C. Pomerance [12].

4.2 The number of prime factors of a shifted prime.

For any y we define the completely additive function $\Omega_y(n)$, the total number of prime factors $p \leq y$ of n, counting multiplicity. The letters p, q, r always denote primes. Let $P(n)$ denote the largest prime factor of n.

Lemma 4.1. *If* $3 \leq y \leq x$, *then*

$$\sum_{p\leq x} \Omega_y(p-1) = \frac{x\log\log y}{\log x} + O(\frac{x}{\log x}),$$

where the implied constant is uniform.

Proof. We have $(\pi(x,k,l) = \sum_{\substack{p\le x\\ p\equiv l(k)}} 1)$

$$\begin{aligned}\sum_{p\le x}\Omega_y(p-1) &= \sum_{p\le x}\sum_{\substack{q^a|(p-1)\\ q\le y}} 1 = \sum_{\substack{q^a\\ q\le y}}\pi(x,q^a,1))\\ &= \sum_{q\le y}\pi(x,q,1) + \sum_{\substack{q^a,a\ge 2\\ q\le y}}\\ &= S_1 + S_2, \quad \text{say.}\end{aligned}$$

In order to evaluate S_1 we consider two ranges for the prime q: $q \le \min\{y, x^{1/3}\}$ and $\min\{y, x^{1/3}\} < q < y$. Of course, depending on the size of y, the latter range may be vacuous. Considering the fact

$$\sum_{q\le y}\frac{1}{\phi(q)} = \log\log y + O(1),$$

we estimate the sum in the first range by the Bombierri-Vinogradov theorem by

$$\begin{aligned}\sum_{q\le\min\{y,x^{1/3}\}}\pi(x,q,1) &= \sum_{q\le\min\{y,x^{1/3}\}}\frac{\mathrm{li}(x)}{\phi(q)} + O(\frac{x}{\log^2 x})\\ &= \frac{x\log\log y}{\log x} + O(\frac{x}{\log x}).\end{aligned}$$

Therefore it's enough to show that the second range for q in S_1 and all of S_2 contribute only $O(\frac{x}{\log x})$ to the sum.

For the second range in S_1

$$\begin{aligned}\sum_{\min\{y,x^{1/3}\}<q\le y}\pi(x,q,1) &\le \sum_{q>x^{1/3}}\pi(x,q,1)\\ &= \sum_{p\le x}\sum_{\substack{q|(p-1)\\ q>x1/3}} 1\\ &\le 2\pi(x) = O(\frac{x}{\log x}).\end{aligned}$$

We also break S_2 into two ranges: $q^a \le x^{1/3}$ and $x^{1/3} < q^a \le x$. By the Brun-Titchmarsh theorem, the first range is

$$\sum_{\substack{q^a\le x^{1/3},a\ge 2\\ q\le y}}\pi(x,q^a,1) \ll \frac{x}{\log x}\sum_{q^a,a\ge 2}\frac{1}{\phi(a)} \ll \frac{x}{\log x},$$

where in the last step we used

$$\frac{1}{\phi(a)} = \frac{1}{q^a(1-\frac{1}{q})} \leq \frac{2}{q^a}$$

and the fact that the sum

$$\sum_{\substack{q^a \\ a\geq 2}} \frac{1}{q^a}$$

is convergent.

For the second part of S_2 by using the trivial bound $\pi(x, q^a, 1) \leq \frac{x}{q^a}$ and an easy fact that $\sum_{n>x^{1/3}} \frac{1}{n} \ll x^{-1/6}$ we get

$$\sum_{\substack{q^a>x^{1/3}, a\geq 2 \\ q\leq y}} \pi(x, q^a, 1) \leq \sum \frac{x}{q^a} \ll x^{5/6}.$$

We thus have proved the lemma. □

Lemma 4.2. *If $3 \leq y \leq x$, then*

$$\sum_{p\leq x} \Omega_y(p-1)^2 = \frac{x(\log\log y)^2}{\log x} + O(\frac{x\log\log y}{\log x}),$$

where the implied constant is uniform.

Proof. Let u range over the integers with $w(u) = 2$ and $P(u) \leq y$. Then

$$\begin{aligned}\sum_{p\leq x} \Omega_y(p-1)^2 &= \sum_{p\leq x} \sum_{\substack{q^a||(p-1) \\ q\leq y}} a^2 + 2\sum_{p\leq x} \sum_{u|(p-1)} 1 \\ &= S_3 + S_4, \quad \text{say.}\end{aligned}$$

We have

$$\begin{aligned}S_3 &= \sum_{p\leq x} \Omega_y(p-1) + \sum_{p\leq x} \sum_{\substack{q^a||(p-1) \\ q\leq y, a\geq 2}} (a^2-a) \\ &\leq \sum_{p\leq x} \Omega_y(p-1) + \sum_{\substack{q^a\leq x^{1/3} \\ q\leq y, a\geq 2}} (a^2-a)\pi(x, q^a, 1) \\ &+ \sum_{\substack{q^a>x^{1/3} \\ q\leq y, a\geq 2}} (a^2-a)\pi(x, q^a, 1).\end{aligned}$$

By Lemma (4.1), the first sum is $\ll \frac{x\log\log y}{\log x}$. For the middle sum we use

Brun-Titchmarsh theorem and get

$$\sum_{\substack{q^a \le x^{1/3} \\ q \le y, a \ge 2}} (a^2 - a)\pi(x, q^a, 1) \ll \frac{x}{\log x} \sum_{a \ge 2} a^2 \sum_{q \le y} \frac{1}{q^a} \ll \frac{x \log\log y}{\log x}.$$

And we use the trivial estimate for the last sum:

$$\sum_{\substack{q^a > x^{1/3} \\ q \le y, a \ge 2}} (a^2 - a)\pi(x, q^a, 1) \ll \sum_{\substack{q^a > x^{1/3} \\ q \le y, a \ge 2}} a^2 \frac{x}{q^a} \ll x^{2/3} \sum_{a \le \log x} a^2$$
$$\ll x^{2/3}(\log x)^2 \ll \frac{x \log\log y}{\log x}.$$

For S_4, we write

$$S_4 = S_{4,1} + S_{4,2}$$

where in $S_{4,1}$ neither prime power in u exceeds $x^{1/6}$ and in $S_{4,2}$ at least one power in u exceeds $x^{1/6}$. We have

$$S_{4,1} = 2 \sum_{\substack{1 < q^a, r^b < x^{1/6} \\ q < r \le y}} \pi(x, q^a r^b, 1).$$

By using Bombieri-Vinogradov theorem, we get

$$S_{4,1} = 2\mathrm{li}(x) \sum_{\substack{1 < q^a, r^b < x^{1/6} \\ q < r \le y}} \frac{1}{\phi(q^a r^b)} + O(\frac{x}{\log^2 x})$$
$$= \sum_{\substack{q^a < x^{1/6} \\ q \le y}} \frac{1}{\phi(q^a)} \sum_{\substack{r^b < x^{1/6} \\ r \le y}} \frac{1}{\phi(r^b)} + O(\frac{x}{\log^2 x})$$

Now since

$$\sum_{\substack{q^a < x^{1/6} \\ q \le y}} \frac{1}{\phi(q^a)} = \log\log y + O(1)$$

then,

$$S_{4,1} = \frac{x(\log\log y)^2}{\log x} + O(\frac{x}{\log^2 x}).$$

Finally, for $S_{4,2}$ we have

$$\begin{aligned} S_{4,2} &= 2\sum_{\substack{q^a \\ q\le y}} \sum_{r^b > x^{1/6}} \pi(x, q^a r^b, 1) \\ &= 2\sum_{\substack{q^a \\ q\le y}} \sum_{r > x^{1/6}} \pi(x, q^a r^b, 1) + 2\sum_{\substack{q^a \\ q\le y}} \sum_{\substack{r^b > x^{1/6} \\ b\ge 2}} \pi(x, q^a r^b, 1) \end{aligned}$$

In the first expression above there are at most six r's in the inner sum, so the expression is $\ll \sum_{\substack{q^a \\ q\le y}} \pi(x, q^a, 1)$ which is in fact equal to $\sum_{p\le y} \Omega_y(p-1)$. For the second part, we use the trivial bound and get

$$\ll x \sum_{\substack{q^a \\ q\le y}} \frac{1}{q^a} \sum_{\substack{r^b > x^{1/6} \\ b\ge 2}} \frac{1}{r^b} \ll x^{11/12} \log\log y \ll \frac{x \log\log y}{\log x}.$$

And this completes proof. □

Lemma 4.3. *If $3 \le y \le x$, then*

$$\sum_{p\le x} \frac{\Omega_y(p-1)}{p} = \log\log x \log\log y - \frac{1}{2}(\log\log y)^2 + O(\log\log x),$$

where the implied constant is uniform.

Proof. By partial summation we have

$$\sum_{p\le x} \frac{\Omega_y(p-1)}{p} = \frac{1}{x}\sum_{p\le x} \Omega_y(p-1) + \int_2^x \frac{1}{t^2} \sum_{p\le t} \Omega_y(p-1) dt.$$

Now by Lemma (4.1), the last expression becomes

$$O\left(\frac{\log\log y}{\log x}\right) + \int_2^y \frac{\log\log t}{t \log t} dt + \int_y^x \frac{\log\log y}{t\log t} dt + O\left(\int_2^x \frac{dt}{t\log t}\right)$$

and by simple calculations we get

$$\log\log x \log\log y - \frac{1}{2}(\log\log y)^2 + O(\log\log x).$$

We are done. □

Lemma 4.4. *If $3 \le y \le x$, then*

$$\sum_{p\le x} \frac{\Omega_y(p-1)^2}{p} = \log\log x (\log\log y)^2 - \frac{2}{3}(\log\log y)^3 + O(\log\log x \log\log y),$$

where the implied constant is uniform.

Proof. Again by partial summation we have

$$\sum_{p\le x}\frac{\Omega_y(p-1)^2}{p}=\frac{1}{x}\sum_{p\le x}\Omega_y(p-1)+\int_2^x\frac{1}{t^2}\sum_{p\le t}\Omega_y(p-1)^2dt.$$

Now by Lemma (4.2), the last expression is

$$O\left(\frac{(\log\log y)^2}{\log x}\right)+\int_2^y\frac{(\log\log t)^2}{t\log t}dt+\int_y^x\frac{(\log\log y)^2}{t\log t}dt+O\left(\int_y^x\frac{\log\log y}{t\log t}dt\right)$$

and by simple calculations we get

$$\sum_{p\le x}\frac{\Omega_y(p-1)^2}{p}=\log\log x(\log\log y)^2-\frac{2}{3}(\log\log y)^3+O(\log\log x\log\log y).$$

So the proof is completed. □

Lemma 4.5. *If $2\le k\le x$, then*

$$\sum_{\substack{p\le x\\ p\equiv 1(k)}}\frac{1}{p}=\frac{\log\log x}{\phi(k)}+O\left(\frac{\log k}{\phi(k)}\right),$$

where the implied constant is uniform.

4.3 The normal number of prime factors of $\phi(n)$.

Now, in order to use the theorem of Kubilius-shapiro we need **strongly** additive function, but $\Omega(\phi(n))$ is not strongly additive. Instead we define

$$f(n)=\sum_{p|n}\Omega(p-1). \tag{4.1}$$

Then $f(n)$ is strongly additive and one can easily check that

$$\Omega(\phi(n))=f(n)+\Omega(n)-w(n).$$

Theorem 4.6. *For every real number u we have*

$$\lim_{x\to\infty}\frac{1}{x}\cdot\sharp\left\{n\le x:\Omega(\phi(n))-\frac{1}{2}(\log\log x)^2\le\frac{u}{\sqrt{3}}(\log\log x)^{3/2}\right\}=G(u), \tag{4.2}$$

where $G(u)=(2\pi)^{-1/2}\int_{-\infty}^u e^{-t^2/2}dt$.

Proof. We apply the Kubilius-Shapiro theorem to the strongly additive function $f(n)$ defined in (4.1). We have

$$A(x) = \sum_{p \le x} \frac{\Omega(p-1)}{p} = \frac{1}{2}(\log\log x)^2 + O(\log\log x)$$

by Lemma 4.3 (with $y = x$). Also

$$B(x)^2 = \sum_{p \le x} \frac{\Omega(p-1)^2}{p} = \frac{1}{3}(\log\log x)^3 + O((\log\log x)^2)$$

by Lemma 4.4 (with $y = x$). Thus to apply the Kubilius-Shapiro theorem to $f(n)$ it remains to verify (1.12). Let $\epsilon > 0$ be fixed and define a totally multiplicative function

$$\alpha(p) = \begin{cases} 1 & \text{if } \Omega(\phi(p)) > \epsilon B(x) \\ 0 & \text{otherwise.} \end{cases}$$

We will show that (since $f(p) = \Omega(p-1)$)

$$\frac{1}{(\log\log x)^3} \left| \sum_{p \le x} \alpha(p) \frac{\Omega^2(p-1)}{p} \right| \to 0 \quad \text{as} \quad x \to \infty.$$

But by Cauchy-schwarz inequality

$$\begin{aligned} \frac{1}{(\log\log x)^3} \left| \sum_{p \le x} \alpha(p) \frac{\Omega^2(p-1)}{p} \right| &\le \frac{1}{(\log\log x)^3} \left(\sum_{p \le x} \frac{\alpha(p)}{p} \right)^{1/2} \left(\sum_{p \le x} \frac{\Omega^4(p-1)}{p} \right)^{1/2} \\ &\le \frac{1}{(\log\log x)^3} \left(\sum_{p \le x} \frac{\Omega^4(p-1)}{p} \right)^{1/2}. \quad (\star) \end{aligned}$$

To justify the last step above, we need to show that

$$\sum_{p \le x} \frac{\alpha(p)}{p} \ll 1.$$

But

$$\begin{aligned} \sum_{p \le x} (\Omega(p-1) - \log\log x)^2 &= \sum_{p \le x} \Omega^2(p-1) - 2\log\log x \sum_{p \le x} \Omega(p-1) \\ &\quad + (\log\log x)^2 \pi(x) \ll \frac{x \log\log x}{\log x}. \end{aligned}$$

Therefore

$$\sum_{p\le x}\alpha(p) \ll \frac{x}{\log x(\log\log x)^2}.$$

And by partial summation,

$$\sum_{p\le x}\frac{\alpha(p)}{p} \ll \int_2^x \frac{dt}{t\log t(\log\log t)^2} \ll 1.$$

Finally by using the estimation

$$\sum_{p\le x}\frac{\Omega^4(p-1)}{p} \ll (\log\log x)^5$$

in ($\star$), we get the desired result. Thus (1.12) is verified and, by the Kubilius-Shapiro theorem, we have (4.2) with $f(n)$ in place of $\Omega(\phi(n))$. But $\Omega(\phi(n)) = f(n) + \Omega(n) - w(n)$ and $\Omega(n) - w(n)$ is normally $o(\log\log n)$ by the Hardy-Ramanujan theorem. We therefore may replace $f(n)$ with $\Omega(\phi(n))$. □

The situation with the function $w(\phi(n))$ is the same, but the treatment is less routine, notably because $w(\phi(n))$ is not additive. As one might expect, though, the difference $\Omega(\phi(n))-w(\phi(n))$ is usually not large (compared with $\Omega(\phi(n))$), so we can obtain the same result for $w(\phi(n))$.

Theorem 4.7. *For every real number u we have*

$$\lim_{x\to\infty}\frac{1}{x}\cdot\sharp\left\{n\le x : w(\phi(n)) - \frac{1}{2}(\log\log x)^2 \le \frac{u}{\sqrt{3}}(\log\log x)^{3/2}\right\} = G(u).$$

Proof. If we can show that, but for $o(x)$ choice of $n\le x$,

$$\Omega(\phi(n)) - w(\phi(n)) = o((\log\log x)^{3/2}).$$

In fact, we shall show the stronger result, that but for $o(x)$ choices of $n\le x$

$$\Omega(\phi(n)) - w(\phi(n)) = O(\log\log x\log\log\log\log x). \qquad (4.3)$$

Let $w_y(n)$ denote the number of distinct prime factors of n which do not exceed y. In order to restrict ourselves to bounding $\Omega_y(\phi(n)) - w_y(\phi(n))$, we should first show that but for $o(x)$ choices of $n\le x$

$$\Omega(\phi(n)) - \Omega_y(\phi(n)) = w(\phi(n)) - w_y(\phi(n)). \qquad (4.4)$$

We apply the Turan-Kubilius inequality to the additive function $\Omega_y(\phi(n))$.

We have

$$E_y(x) := \sum_{p^k \le x} \frac{\Omega_y(\phi(p^k))}{p^k}(1 - \frac{1}{p})$$

$$= \sum_{p \le x} \frac{\Omega_y(p-1)}{p} + O\left(\sum_{\substack{p^k \le x \\ k>1}} \frac{\Omega(p-1)p^{k-1}}{p^k}\right)$$

$$= \log\log x \log\log y - \frac{1}{2}(\log\log y)^2 + O(\log\log x),$$

by Lemma 4.3 and

$$D_y(x)^2 := \sum_{p^k \le x} \frac{\Omega_y(\phi(p^k))^2}{p^k} = \sum_{p \le x} \frac{\Omega_y(p-1)^2}{p} + O\left(\sum_{\substack{p^k \le x \\ k>1}} \frac{\Omega_y(\phi(p^k))^2}{p^k}\right)$$

$$= \log\log x(\log\log y)^2 - \frac{2}{3}(\log\log y)^3 + O(\log\log x \log\log y)$$

by Lemma (4.4). Therefore, by the Turan-Kubilius inequality,

$$\sum_{n \le x}(\Omega_y(\phi(n)) - E_y(x))^2 \le 32xD_y(x)^2. \tag{4.5}$$

Now by taking $y = (\log\log x)^2$, we have

$$E_y(x) = \log\log x \log\log\log\log x + O(\log\log x),$$

$$D_y(x)^2 = \log\log x(\log\log\log\log x)^2 + O(\log\log x \log\log\log\log x).$$

Therefore, by (4.5), the number of $n \le x$ with $\Omega_y(\phi(n)) > 2\log\log x \log\log\log\log x$ is $O(x/\log\log x) = o(x)$. We thus have but for $o(x)$ choices of $n \le x$

$$0 \le \Omega_y(\phi(n)) - w_y(\phi(n)) \le 2\log\log x \log\log\log\log x. \tag{4.6}$$

We now show that but for $o(x)$ choices of $n \le x$ we have (4.4). Suppose $p^2|\phi(n)$ where $p > y$ and $n \le x$. There are four possibilities:

(*i*) $p^3|n$,

(*ii*) there is some $q|n$ with $q \equiv 1 \bmod p^2$,

(*iii*) there are distinct q_1, q_2 with $q_1q_2|n$ and $q_1 \equiv q_2 \equiv 1 \bmod p$,

(*iv*) $n|p$ and there is some $q|n$ with $q \equiv 1 \bmod p$

The number of $n \le x$ in the first case is at most

$$\sum_{p>y} x/p^3 = o(x/y^2) = o(x).$$

The number of $n \le x$ in the second case is, by Lemma (4.5), at most

$$\sum_{p>y} \sum_{\substack{q \equiv 1 (p^2) \\ q \le x}} \frac{x}{q} = \sum_{p>y} \frac{x \log\log x}{\phi(p^2)} + O(\sum_{p>y} \frac{x \log p}{p^2})$$

$$= O(\frac{x \log\log x}{y \log y}) + O(\frac{x}{y}) = o(x)$$

where we used

$$\sum_{p>y} \frac{1}{p(p-1)} \ll \frac{1}{y \log y} \quad \text{and} \quad \sum_{p>y} \frac{\log p}{p^2} \ll \frac{1}{y}$$

which follow easily from partial summation.

The number of $n \le x$ in the third case is, by Lemma (4.5), at most

$$\sum_{p>y} \sum_{\substack{q_1 \equiv q_2 \equiv 1 (p) \\ q_1 < q_2 \le x}} \frac{x}{q_1 q_2} \le \frac{1}{2} x \sum_{p>y} \left(\sum_{\substack{q \equiv 1 (p) \\ q \le x}} \frac{1}{q} \right)^2$$

$$= \frac{1}{2} x \sum_{p>y} \left(\frac{\log\log x}{\phi(p)} + O(\frac{\log p}{p}) \right)^2$$

$$= O\left(\frac{x (\log\log x)^2}{y \log y} \right) + O\left(\frac{x \log\log x}{y} \right)$$

$$+ O\left(\frac{x \log\log y}{y} \right) = o(x)$$

where we used the fact

$$\sum_{p>y} \frac{\log^2 p}{p^2} \ll \frac{\log\log y}{y}$$

which again follows by partial summation.

Finally, in the fourth case, the number of $n \le x$ is, by Lemma (4.5) and partial summation, at most

$$\sum_{p>y} \sum_{\substack{q \equiv 1 (p) \\ q \leq x}} \frac{x}{pq} = \sum_{p>y} \frac{1}{p} \left(\frac{\log \log x}{\phi(p)} + O(\frac{\log p}{\phi(p)}) \right)$$

$$= O\left(\frac{x \log \log x}{y \log y} \right) + O(\frac{x}{y}) = o(x).$$

This estimate completes the proof that (4.4) holds for all but $o(x)$ choices of $n \leq x$. Combined with (4.6), we have (4.3) and thus the theorem. □

5 Distribution of values of the arithmetic function $d(n)$

The following result is due to M. Kac [9].

Theorem 5.1. *Denote by $r_n(w)$ the number of integers $m \leq n$ such that*

$$d(m) \leq 2^{\log\log n + w(2\log\log n)^{1/2}}$$

then we have

$$\lim_{n\to\infty} \frac{r_n(w)}{n} = \pi^{-1/2} \int_{-\infty}^{w} e^{-u^2} du = D(w),$$

where $d(n)$ denotes the number of divisors of n.

The proof is based on the Erdös-Kac theorem and two lemmas below:

Lemma 5.2. *If $f(m) \geq 0$ is such that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} f(m) \tag{5.1}$$

is finite, if $\lim_{n\to\infty} g(n) = \infty$ and if $p(n)$ denotes the number of positive integers $m \leq n$ for which $f(m) \leq g(m)$ then

$$\lim_{n\to\infty} \frac{p(n)}{n} = 1 \quad as \quad n \to \infty.$$

Proof. Let $P_n = \{m : m \leq n \quad \text{and} \quad f(m) \leq g(n)\}$. Therefore by definition we have $|P_n| = p(n)$. Then

$$\frac{1}{n} \sum_{m=1}^{n} f(m) = \frac{1}{n} \left\{ \sum_{m\in P_n} f(m) + \sum_{m\notin P_n} f(m) \right\} > \frac{1}{n} \sum_{m\notin P_n} = \frac{n-p(n)}{n} g(n).$$

Since

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} f(m)$$

is finite and $\lim_{n\to\infty} g(n) = \infty$, we get

$$\limsup \frac{n-p(n)}{n} = \liminf \frac{n-p(n)}{n} = 0.$$

□

Lemma 5.3. *The mean value*

$$M\left\{\frac{d(m)}{2^{w(m)}}\right\} = \lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n}\frac{d(m)}{2^{w(m)}}$$

exists and is finite.

Proof. If

$$\frac{d(m)}{2^{w(m)}} = f\star U \quad \text{then} \quad f = \frac{d(\cdot)}{2^{w(\cdot)}}\star\mu.$$

Since f is multiplicative then it's enough to consider

$$f(p^j) = \sum_{m|p^j}\mu(m)\frac{d(m|p^j)}{2^{w(m|p^j)}} = \begin{cases} 0 & \text{if } j=1 \\ 1/2 & \text{if } j\le 2. \end{cases}$$

So we get

$$f(n) = \begin{cases} 0 & \text{if there exists } p \text{ such that } p|n \text{ but } p^2 \dagger n \\ \frac{1}{2^{w(n)}} & \text{if } p|n \Rightarrow p^2|n. \end{cases}$$

And also

$$\begin{aligned}\frac{1}{n}\sum_{m=1}^{n}\frac{d(m)}{2^{w(m)}} &= \frac{1}{n}\sum_{m=1}^{n}\sum_{d|n} f(d) \\ &= \frac{1}{n}\sum_{d=1}^{n} f(d)\left[\frac{n}{d}\right] \\ &= \sum_{d=1}^{n}\frac{f(d)}{d} + O\left(\frac{1}{n}\sum_{d=1}^{n}|f(d)|\right)\end{aligned}$$

Now since the error term is $\ll 1$, we have only to show that

$$\sum_{d=1}^{n}\frac{f(d)}{d} \tag{5.2}$$

converges. In fact if we able to show (5.2), then by Kroneker's theorem $\lim_{n\to\infty}\frac{1}{n}\sum_{d\le n} f(d) = 0$. (i.e. the error term will go to 0). If we set $S = \left\{n : p|n \Rightarrow p^2|n\right\}$ then since $\frac{f(d)}{d} \le \frac{1}{d}$, it's enough to show that $\sum_{n\in S}\frac{1}{n}$ converges. But since the series

$$\sum_{p}\frac{1}{p(p-1)}$$

is convergent, then

$$\prod_p \left(1 + \frac{1}{p^2} + \frac{1}{p^3} + ...\right) = \prod_p \left(1 + \frac{1}{p(p-1)}\right)$$

is also a convergent product. And this completes the proof. $\square$

Proof of the Theorem: Let w be an arbitrary real number and $\epsilon > 0$. Put

$$f(m) = \frac{d(m)}{2^{w(m)}} \quad \text{and} \quad g(n) = 2^{\epsilon(2 \log\log n)^{1/2}}.$$

Let F_n be the set of positive integers $m \le n$ for which

$$w(m) \le \log\log n + (w - \epsilon)(2 \log\log n)^{1/2},$$

G_n the set of positive integers $m \le n$ for which $f(m) \le g(n)$ and H_n the set of positive integers $m \le n$ for which

$$d(m) \le 2^{\log\log n + w(2\log\log n)^{1/2}}.$$

We should observe that if $m \in F_n \cap G_n$ then $m \in H_n$. Hence

$$F_n \cap G_n \subset H_n.$$

The number of elements in F_n is $k_n(w - \epsilon)$; in G_n, $p(n)$; and in H_n, $r_n(w)$. Thus the number of elements in $F_n \cap G_n$ is $\ge k_n(w - \epsilon) - (n - p(n))$. This is because of the fact $A \backslash B^c \subseteq A \cap B$, for any two sets A and B. Finally

$$k_n(w - \epsilon) - (n - p(n)) \le r_n(w).$$

On the other hand for every m, $2^{w(m)} \le d(m)$ (the equality occurs only if m is square-free) and therefore $H_n \subset F_n$ or $r_n(w) \le k_n(w)$. The inequalities combined give

$$k_n(w - \epsilon) - (n - p(n)) \le r_n(w) \le k_n(w).$$

But as $n \to \infty$, by Erdös-Kac theorem, we have

$$k_n(w - \epsilon) \to D(w - \epsilon)$$
$$k_n(w) \to D(w).$$

Also, by Lemma 5.2 and 5.3

$$\frac{n - p(n)}{n} \to 0$$

Hence

$$D(w - \epsilon) \leq \liminf_{n\to\infty} \frac{r_n(w)}{n} \leq \limsup_{n\to\infty} \frac{r_n(w)}{n} \leq D(w).$$

Since ϵ is arbitrary and $D(w)$ is a continuous function of w,

$$\lim_{n\to\infty} \frac{r_n(w)}{n} = D(w)$$

And this ends the proof of the theorem.

6 Concluding Remarks

1. In chapter 3, it can be shown by using a theorem of P. Erdös [7] and L.G Sathe [8] that the result obtained in Theorem 3.2 is the best possible in the sense that $O(1/\sqrt{\log\log n})$ cannot be replaced by $o(1/\sqrt{\log\log n})$ uniformly in x. Nevertheles the remainder term can be improved in the sense that its dependence on x can be investigated.

2. Let $\lambda(n)$ denote the Charmichael function. Then, one can show that the random variable ξ_n, which assumes the values $\Omega(\lambda(1))$, $\Omega(\lambda(2))$,..., $\Omega(\lambda(n))$, each with the same probability $1/n$, for $n \to \infty$, asymptotically normally distributed with mean value $\log\log n$ and standard deviation $\sqrt{\log\log n}$.

3. The result in chapter 5 has been improved by LeVeque [10] to

$$\frac{r_n(w)}{n} = D(w) + O\left(\frac{\log\log\log n}{\sqrt[4]{\log\log n}}\right),$$

and by Kubilius to

$$\frac{r_n(w)}{n} = D(w) + O\left(\frac{\log\log\log n}{\sqrt{\log\log n}}\right).$$

References

[1] Davenport, H., Multiplicative number theory, 3rd edition, Springer, 2000.

[2] Montgomery, H. L. and Vaughan R., Multiplicative number theory, I. Classical theory. Cambridge, 2007.

[3] Apostol, T. M., Introduction to Analytic Number Theory, Springer, 1976.

[4] Nathanson, M. B., Elementary Methods in Number Theory, Springer, 2000.

[5] Tenenbaum, G., Introduction to Analytic and Probabilistic Number Theory, Cambridge, 1995.

[6] Goldfeld, D.,The Elementary proof of the Prime Number Theorem, An Historical Perspective, Pages 179–192 in Number Theory, New York Seminar 2003, eds. D. and G. Chudnovsky, M. Nathanson, Springer-Verlag, New York, 2004.

[7] Erdös, P., On the integers having exactly k prime factors, Ann. of Math. 49 (1948), p.$53 - 56$.

[8] Sathe, L. G., On a problem of Hardy on the distribution of integers having a given number of prime factors, $I - IV$, J. Indian Math. Soc. 17 (1953), p. $63 - 82$, $83 - 141$, 18 (1954), p.$27 - 42$, $43 - 81$.

[9] Kac, M., Note on the distribution of values of the arithmetic function $d(n)$, Bull. Amer. Math. Soc. 47 (1947), p. $815 - 817$.

[10] LeVeque, W. J., On the size of certain number-theoretic functions, Trans. Amer. Math. Soc. 66 (1949), p. $157 - 162$.

[11] Esseen, C. G., Fourier analysis of distribution functions, A mathematical study of the Laplace-Gaussian law, Acta Math. 77 (1945), p. $1 - 125$.

[12] Erdös, P. and Pomerance C., The normal number of prime factors of $\phi(n)$, Rocky Mtn. J. Math. 15 (1985), $729 - 731$.

[13] Renyi, A. and Turan, P., On a theorem of Erdös-Kac, Acta Arith. 4 (1958), pp $71 - 84$.

[14] Erdös, P. and Kac, M. On the Gaussian law of errors in the theory of additive number theoretic function, Amer. J. Math. 62, no. 1 (1940), pp $738 - 742$.