Dirichlet's Theorem on Primes in Arithmetic Progression

IMO Training 2006-2007 Phase 3 Level 2

Advanced Number Theory

3, 10 March 2007

Contents

1 Analytic Preliminaries **4**

1.1 The Big Oh Notation 4

1.2 Abel's Identity 5

1.3 Dirichlet Convolution 8

1.4 Generalized Convolution 13

1.5 Another Identity for the Partial Sum of a Dirichlet Convolution 20

2 Algebraic Background **24**

2.1 Groups 24

2.2 Construction of Subgroups 26

2.3 Characters of Finite Abelian Groups 29

2.4 The Character Group 32

2.5 Orthogonality Relations for Characters 33

2.6 Dirichlet Characters 35

2.7 Sums involving Dirichlet Characters 39

2.8 The Nonvanishing of $L(1, \chi)$ for Real Nonprincipal χ 42

3 The Elementary Proof **45**

4 Miscellaneous Problems **55**

Introduction

In 1837, Dirichlet changed the landscape of number theory by introducing methods of analysis. He proved the celebrated theorem on the infinitude of primes in an arithmetic progression. The version of proof given in these notes follow very closely (almost verbatim!) Apostol's exposition (see [1]) who in turn based his presentation on Shapiro (see [3]). Shapiro's proof uses $\sum p^{-1} \log p$ instead of Euler's $\sum p^{-1}$ (so somehow it is more "sensitive" to primes). Nothing is original in these notes! They merely compile the necessary background (so many sections in thses notes seem "ad hoc") and present an elementary proof as found in the above-mentioned literature. This version is considered "elementary" in a sense that it does not use complex analysis (and in particular Dirichlet series). But it does use "Dirichlet characters" (an arithmetical function formulated by Dirichlet to attack the problem) and lots of estimates. So please be patient in going through the first two sections because, in the end, with patience everything will fall into their proper place as in a jigsaw puzzle!

Acknowledgment: Some miscellaneous problems were taken from Dr. Kin Yin Li's unpublished training notes (way back in 2002). The others came from journals with problem solving sections.

Notations

Throughout these notes, we adopt the following notations:

- the letter p will denote a positive prime number
- m, n, h, k will usually be natural numbers (unless otherwise specified);
- x, y will denote real numbers (again unless otherwise specified);
- $[x]$ is the greatest integer not exceeding x;

- $\sum_{n \leq x} f(n) = \sum_{n=1}^{[x]} f(n)$;
- $\sum_{p \leq x} f(p)$ or $\prod_{p \leq x} f(p)$ means p runs over all positive prime numbers less than or equal to x (if a condition is imposed, like $p \equiv 1 (\mathrm{mod}\, 4)$, instead take all primes obeying that condition);
- $\sum_{d|n} f(d)$ means d runs over all positive divisors of n;
- $\log x$ will mean the natural logarithm of x.
- $\varphi(n)$ is Euler's totient function (the number of positive integers not exceeding n which are relatively prime to n).

1 Analytic Preliminaries

1.1 The Big Oh Notation

Definition 1. If $g(x) > 0$ for all $x \geq a$, we write

$$f(x) = O(g(x))$$

(read as $f(x)$ is big oh of $g(x)$) to mean that the quotient $f(x)/g(x)$ is bounded for $x \geq a$; that is, there exists a constant $M > 0$ such that

$$|f(x)| \leq Mg(x) \quad \text{for all } \ x \geq a.$$

An equation of the form

$$f(x) = h(x) + O(g(x))$$

means that $f(x) - h(x) = O(g(x))$.

Below are some properties of the Big Oh notation (mainly taken from Landau [2]):

(a). If $f_1(x) = O(g_1(x))$ and $f_2(x) = O(g_2(x))$, then

$$f_1(x) + f_2(x) = O(g_1(x) + g_2(x)).$$

In particular, if $g_1(x) = g_2(x)$, then

$$f_1(x) + f_2(x) = O(g_1(x)).$$

(b). If $f_1(x) = O(g_1(x))$ and $f_2(x) = O(g_2(x))$, then

$$f_1(x)f_2(x) = O(g_1(x)g_2(x)).$$

(c). If $f(x) \leq g(x)$ for all $x \geq a$, then $O(f(x)) = O(g(x))$.

(d). If $f(t) = O(g(t))$ for all $t \geq a$, then

$$\int_a^x f(t)dt = O\left(\int_a^x g(t)dt\right) \quad \text{for all } \ x \geq a.$$

1.2 Abel's Identity

Theorem 1.1 (**Abel's Identity**). *For any arithmetical function $a(n)$ let $A(x) = \sum_{n \le x} a(n)$, where $A(x) = 0$ if $x < 1$. Assume f has a continuous derivative on the interval $[y, x]$, where $0 < y < x$. Then we have*

$$\sum_{y < n \le x} a(n) f(n) = A(x) f(x) - A(y) f(y) - \int_y^x A(t) f'(t) dt \tag{1}$$

Proof

Let $k = [x]$ and $m = [y]$, so that $A(k) = A(x)$ and $A(m) = A(y)$. Then

$$\begin{aligned}
\sum_{y < n \le x} a(n) f(n) &= \sum_{n=m+1}^{k} \left(A(n) - A(n-1)\right) f(n) \\
&= -\sum_{n=m+1}^{k-1} A(n) \int_n^{n+1} f'(t) dt + A(k) f(k) - A(m) f(m+1) \\
&= -\sum_{n=m+1}^{k-1} \int_n^{n+1} A(t) f'(t) dt + A(k) f(k) - A(m) f(m+1) \\
&= -\int_{m+1}^{k} A(t) f'(t) dt + A(k) f(k) - A(m) f(m+1) \\
&= -\int_{m+1}^{k} A(t) f'(t) dt + A(x) f(x) - A(x) \int_k^x f'(t) dt - A(m) \int_m^{m+1} f'(t) dt + A(m) \int_m^y f'(t) dt - A(y) f(y) \\
&= -\int_{m+1}^{k} A(t) f'(t) dt + A(x) f(x) - \int_k^x A(t) f'(t) dt - \int_y^{m+1} A(t) f'(t) dt - A(y) f(y) \\
&= A(x) f(x) - A(y) f(y) - \int_y^x A(t) f'(t) dt.
\end{aligned}$$

□

Theorem 1.2 (**Euler's Summation Formula**). *If f has a continuous derivative f' on the interval $[y, x]$, where $0 < y < x$, then*

$$\sum_{y<n\leq x} f(n) = \int_y^x f(t)dt + \int_y^x (t - [t])f'(t)dt + f(x)([x] - x) - f(y)([y] - y).$$

Proof

Let $a(n) = 1$ for all n in Abel's Identity. Then $A(x) = [x]$ and we get

$$\sum_{y<n\leq x} f(n) = f(x)[x] - f(y)[y] - \int_y^x [t]f'(t)dt.$$

Combine this with the integration by parts formula

$$\int_y^x tf'(t)dt = xf(x) - yf(y) - \int_y^x f(t)dt$$

and the result follows. □

Definition 2. The ***Riemann zeta function*** $\zeta(s)$ is defined by the equation

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \quad \text{if } s > 1$$

and by the equation

$$\zeta(s) = \lim_{x\to\infty} \left(\sum_{n\leq x} \frac{1}{n^s} - \frac{x^{1-s}}{1-s} \right) \quad \text{if } 0 < s < 1.$$

Theorem 1.3. *We have*

$$\sum_{n\leq x}\frac{1}{n^s}=\frac{x^{1-s}}{1-s}+\zeta(s)+O(x^{-s}).$$

Proof

Take $f(x)=x^{-s}$, where $s>0$, $s\neq 1$, in the Euler's Summation Formula to obtain

$$\begin{aligned}\sum_{n\leq x}\frac{1}{n^s}&=\int_1^x\frac{1}{t^s}dt-s\int_1^x\frac{t-[t]}{t^{s+1}}dt+1-\frac{x-[x]}{x^s}\\&=\frac{x^{1-s}}{1-s}-\frac{1}{1-s}+1-s\int_1^\infty\frac{t-[t]}{t^{s+1}}dt+O(x^{-s}).\end{aligned}$$

Therefore

$$\sum_{n\leq x}\frac{1}{n^s}=\frac{x^{1-s}}{1-s}+C(s)+O(x^{-s}),\tag{2}$$

where

$$C(s)=1-\frac{1}{1-s}-s\int_1^\infty\frac{t-[t]}{t^{s+1}}dt.$$

If $s>1$ the left side of (2) approaches $\zeta(s)$ as $x\to\infty$ and the terms x^{1-s} and x^{-s} both approach 0. Hence $C(s)=\zeta(s)$ if $s>1$.

If $0<s<1$, $x^{-s}\to 0$ and (2) show that

$$\lim_{x\to\infty}\left(\sum_{n\leq x}\frac{1}{n^s}-\frac{x^{1-s}}{1-s}\right)=C(s).$$

Therefore $C(s)$ is also equal to $\zeta(s)$ if $0<s<1$. □

1.3 Dirichlet Convolution

Definition 3. Let f and g be arithmetical functions. Their ***Dirichlet convolution*** (or ***Dirichlet product***) is defined as the arithmetical function h given by

$$h(n) = \sum_{d|n} f(d)g\left(\frac{n}{d}\right).$$

We write $f * g$ for h and $(f * g)(n)$ for $h(n)$.

Definition 4. An arithmetical function f is called ***multiplicative*** if f is not identically zero and if

$$f(mn) = f(m)f(n) \quad \text{whenever} \quad \gcd(m, n) = 1.$$

A multiplicative function f is ***completely multiplicative*** if we also have

$$f(mn) = f(m)f(n) \quad \text{for all} \ \ m, n.$$

Definition 5. The ***Möbius function*** μ is defined as follows:

$$\mu(1) = 1;$$

If $n > 1$, write $n = p_1^{a_1} \cdot \ldots \cdot p_k^{a_k}$. Then

$$\mu(n) = \begin{cases} (-1)^k & \text{if } a_1 = \ldots = a_k, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 6. For every integer $n \geq 1$ we define the ***Mangoldt's function*** $\Lambda(n)$ as follows:

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^m \text{ for some prime } p \text{ and some } m \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 1.4. *If $n \geq 1$ we have*

$$\log n = \sum_{d|n} \Lambda(d). \tag{3}$$

PROOF

The theorem is true for $n = 1$ since both sides are 0. Therefore assume that $n > 1$ and write

$$n = \prod_{k=1}^{r} p_k^{a_k}.$$

Taking logarithms we have

$$\log n = \sum_{k=1}^{r} a_k \log p_k.$$

Now consider the sum on the right of (3). The only nonzero terms in the sum come from those divisors d of the form p_k^m for $m = 1, 2, \ldots, a_k$ and $k = 1, 2, \ldots, r$. Hence

$$\sum_{d|n} \Lambda(d) = \sum_{k=1}^{r} \sum_{m=1}^{a_k} \Lambda(p_k^m) = \sum_{k=1}^{r} a_k \log p_k = \log n.$$

□

Below are some properties of the Dirichlet convolution.

(a). Dirichlet convolution is commutative and associative.

(b). If f is an arithmetical function with $f(1) \neq 0$ there is a unique arithmetical function f^{-1}, called the ***Dirichlet inverse*** of f, such that

$$f * f^{-1} = f^{-1} * f = I$$

where $I(n) = \left[\frac{1}{n}\right]$. Moreover, f^{-1} is given by the recursion formulas

$$f^{-1}(n) = \frac{1}{f(1)}, \quad f^{-1}(n) = \frac{-1}{f(n)} \sum_{\substack{d|n \\ d<n}} f\left(\frac{n}{d}\right) f^{-1}(d) \quad \text{for } n > 1.$$

PROOF

Given f, we shall show that the equation $(f * f^{-1})(n) = I(n)$ has a unique solution for the function values $f^{-1}(n)$. For $n = 1$ we have to solve the equation $(f * f^{-1})(1) = I(1)$ which reduces to $f(1)f^{-1}(1) = 1$. Since $f(1) \neq 0$ there is one and only one solution, namely $f^{-1}(1) = 1/f(1)$. Assume now that the function values $f^{-1}(k)$ have been uniquely determined for all $k < n$. Then we have to solve the equation $(f * f^{-1})(n) = I(n)$, or

$$\sum_{d|n} f\left(\frac{n}{d}\right) f^{-1}(d) = 0.$$

This can be written as

$$f(1)f^{-1}(n) + \sum_{\substack{d|n \\ d<n}} f\left(\frac{n}{d}\right) f^{-1}(d).$$

If the values $f^{-1}(d)$ are known for all divisors $d < n$, there is a uniquely determined value for $f^{-1}(n)$, namely,

$$f^{-1}(n) = \frac{1}{f(1)}, \quad f^{-1}(n) = \frac{-1}{f(n)} \sum_{\substack{d|n \\ d<n}} f\left(\frac{n}{d}\right) f^{-1}(d),$$

since $f(1) \neq 0$. This establishes the existence and uniqueness of f^{-1} by induction. □

(c). The Dirichlet convolution of two multiplicative functions is multiplicative.

(d). The Dirichlet inverse of a multiplicative function is multiplicative.

Theorem 1.5 (**Möbius Inversion Formula**). *We have*

$$f(n) = \sum_{d|n} g(d) \quad \textit{if and only if} \quad g(n) = \sum_{d|n} f(d)\mu\left(\frac{n}{d}\right).$$

Now we use the Möbius Inversion Formula to express $\Lambda(n)$ in terms of the logarithm.

Theorem 1.6. *If $n \geq 1$ we have*

$$\Lambda(n) = \sum_{d|n} \mu(d) \log \frac{n}{d} = -\sum_{d|n} \mu(d) \log d.$$

Proof

Inverting (3) by the Möbius Inversion Formula we obtain

$$\begin{aligned}
\Lambda(n) &= \sum_{d|n} \mu(d) \log \frac{n}{d} \\
&= \log n \sum_{d|n} \mu(d) - \sum_{d|n} \mu(d) \log d \\
&= I(n) \log n - \sum_{d|n} \mu(d) \log d.
\end{aligned}$$

Since $I(n) \log n = 0$ for all n the proof is complete. □

Theorem 1.7. *Let f be multiplicative. Then f is completely multiplicative if and only if*

$$f^{-1}(n) = \mu(n)f(n) \quad \textit{for all } \ n \geq 1.$$

Proof

Let $g(n) = \mu(n)f(n)$. If f is completely multiplicative we have

$$(g * f)(n) = \sum_{d|n} \mu(d)f(d)f\left(\frac{n}{d}\right) = f(n)\sum_{d|n} \mu(d) = f(n)I(n) = I(n)$$

since $f(1) = 1$ and $I(n) = 0$ for $n > 1$. Hence $g = f^{-1}$.

Conversely, assume $f^{-1}(n) = \mu(n)f(n)$. To show that f is completely multiplicative it is enough to prove that $f(p^a) = (f(p))^a$ for prime powers. The equation $f^{-1}(n) = \mu(n)f(n)$ implies that

$$\sum_{d|n} \mu(d)f(d)f\left(\frac{n}{d}\right) = 0 \quad \text{for all } \ n > 1.$$

Hence, taking $n = p^a$ we have

$$\mu(1)f(1)f(p^a) + \mu(p)f(p)f(p^{a-1}) = 0,$$

from which we find $f(p^a) = f(p)f(p^{a-1})$. This implies $f(p^a) = (f(p))^a$, so f is completely multiplicative. □

1.4 Generalized Convolution

Definition 7. Let F be a real or complex-valued function defined on the positive real axis $(0, +\infty)$ such that $F(x) = 0$ for $0 < x < 1$ and let α be any arithmetical function. Then the sum G given by

$$G(x) = \sum_{n \leq x} \alpha(n) F\left(\frac{x}{n}\right)$$

is called the ***generalized convolution*** of α and F and is denoted by $\alpha \circ F$. If $F(x) = 0$ for all nonintegral x, the restriction of F to the integers is an arithmetical function and we find that $(\alpha \circ F)(m) = (\alpha * F)(m)$ for all integers $m \geq 1$, so the operation $\circ$ can be regarded as a generalization of the Dirichlet convolution $*$. The operation $\circ$ is, in general, neither commutative nor associative. However, the following theorem serves as a useful substitute for the associative law.

Theorem 1.8. *For any arithmetical functions α and β we $\alpha \circ (\beta \circ F) = (\alpha * \beta) \circ F$.*

Proof
For $x > 0$ we have

$$\begin{aligned}
(\alpha \circ (\beta \circ F))(x) &= \sum_{n \leq x} \alpha(n) \sum_{m \leq x/n} \beta(m) F\left(\frac{x}{mn}\right) \\
&= \sum_{mn \leq x} \alpha(n)\beta(m) F\left(\frac{x}{mn}\right) \\
&= \sum_{k \leq x} \left(\sum_{n|k} \alpha(n) \beta\left(\frac{k}{n}\right) \right) F\left(\frac{x}{k}\right) \\
&= ((\alpha * \beta) \circ F)(x).
\end{aligned}$$

□

Next we note that the identity function

$$I(n) = \left[\frac{1}{n}\right]$$

for the Dirichlet convolution is also a left identity for the operation $\circ$. That is, we have

$$(I \circ F)(x) = \sum_{n \le x} \left[\frac{1}{n}\right] F\left(\frac{x}{n}\right) = F(x).$$

Now we use this fact along with the associative property to prove the following inversion formula.

Theorem 1.9 (**Generalized Inversion Formula**). *If α has a Dirichlet inverse α^{-1} then we have*

$$G(x) = \sum_{n \le x} \alpha(n) F\left(\frac{x}{n}\right) \quad \textit{if and only if} \quad F(x) = \sum_{n \le x} \alpha^{-1}(n) G\left(\frac{x}{n}\right).$$

Proof
If $G = \alpha \circ F$ then

$$\alpha^{-1} \circ G = \alpha^{-1} \circ (\alpha \circ F) = (\alpha^{-1} * \alpha) \circ F = I \circ F = F.$$

The converse is similarly proved. □

Theorem 1.10 (**Generalized Möbius Inversion Formula**). *If α is completely multiplicative then we have*

$$G(x) = \sum_{n \le x} \alpha(n) F\left(\frac{x}{n}\right) \quad \textit{if and only if} \quad F(x) = \sum_{n \le x} \mu(n)\alpha(n) G\left(\frac{x}{n}\right).$$

Proof
In this case $\alpha^{-1}(n) = \mu(n)\alpha(n)$. □

Theorem 1.11. *If $h = f * g$, let*

$$H(x) = \sum_{n \le x} h(n), \quad F(x) = \sum_{n \le x} f(n), \quad G(x) = \sum_{n \le x} g(n).$$

Then we have

$$H(x) = \sum_{n \le x} f(n) G\left(\frac{x}{n}\right) = \sum_{n \le x} g(n) F\left(\frac{x}{n}\right). \tag{4}$$

Proof

Let

$$U(x) = \begin{cases} 0 & \text{if } 0 < x < 1, \\ 1 & \text{if } x \ge 1. \end{cases}$$

Then $F = f \circ U$, $G = g \circ U$. So, by Theorem 1.8

$$f \circ G = f \circ (g \circ U) = (f * g) \circ U = H$$

and

$$g \circ F = g \circ (f \circ U) = (g * f) \circ U = H.$$

□

If $g(n) = 1$ for all n then $G(x) = [x]$, and (4) gives us the following theorem.

Theorem 1.12. *If $F(x) = \sum_{n \le x} f(n)$ we have*

$$\sum_{n \le x} \sum_{d \mid n} f(d) = \sum_{n \le x} f(n) \left[\frac{x}{n}\right] = \sum_{n \le x} F\left(\frac{x}{n}\right). \tag{5}$$

Theorem 1.13. *For $x \geq 1$ we have*

$$\sum_{n \leq x} \Lambda(n) \left[\frac{x}{n}\right] = \log[x]! \tag{6}$$

Proof

From (5) we have

$$\sum_{n \leq x} \Lambda(n) \left[\frac{x}{n}\right] = \sum_{n \leq x} \sum_{d|n} \Lambda(d) = \sum_{n \leq x} \log n = \log[x]!.$$

□

Next we use the Euler's Summation Formula to determine an asymptotic formula for $\log[x]!$.

Theorem 1.14. *If $x \geq 2$ we have*

$$\log[x]! = x \log x - x + O(\log x), \tag{7}$$

and hence

$$\sum_{n \leq x} \Lambda(n) \left[\frac{x}{n}\right] = x \log x - x + O(\log x). \tag{8}$$

Proof

Taking $f(t) = \log t$ in Euler's Summation Formula we obtain

$$\begin{aligned} \sum_{n \leq x} \log n &= \int_1^x \log t dt + \int_1^x \frac{t - [t]}{t} dt - (x - [x]) \log x \\ &= x \log x - x + 1 + \int_1^x \frac{t - [t]}{t} dt + O(\log x). \end{aligned}$$

This proves (7) since

$$\int_1^x \frac{t-[t]}{t}dt = O\left(\int_1^x \frac{1}{t}dt\right) = O(\log x),$$

and (8) follows from (6). □

The next theorem is a consequence of (8).

Theorem 1.15. *For $x \geq 2$ we have*

$$\sum_{p\leq x} \left[\frac{x}{p}\right] \log p = x \log x + O(x). \tag{9}$$

Proof

Since $\Lambda(n) = 0$ unless n is a prime power we have

$$\sum_{n\leq x} \left[\frac{x}{n}\right] \Lambda(n) = \sum_{p} \sum_{\substack{m=1 \\ p^m \leq x}}^{\infty} \left[\frac{x}{p^m}\right] \Lambda(p^m).$$

Now $p^m \leq x$ implies $p \leq x$. Also $[x/p^m] = 0$ if $p > x$ so we can write the last sum as

$$\sum_{p\leq x} \sum_{m=1}^{\infty} \left[\frac{x}{p^m}\right] \log p = \sum_{p\leq x} \left[\frac{x}{p}\right] \log p + \sum_{p\leq x} \sum_{m=2}^{\infty} \left[\frac{x}{p^m}\right] \log p.$$

Next we prove that the last sum is $O(x)$. We have

$$\begin{aligned}
\sum_{p\le x}\log p\sum_{m=2}^{\infty}\left[\frac{x}{p^m}\right] &\le \sum_{p\le x}\log p\sum_{m=2}^{\infty}\frac{x}{p^m}\\
&= x\sum_{p\le x}\log p\sum_{m=2}^{\infty}\left(\frac{1}{p}\right)^m\\
&= x\sum_{p\le x}\log p\cdot\frac{1}{p^2}\cdot\frac{1}{1-\frac{1}{p}}\\
&= x\sum_{p\le x}\frac{\log p}{p(p-1)}\\
&\le x\sum_{n=2}^{\infty}\frac{\log n}{n(n-1)}\\
&= O(x)
\end{aligned}$$

Hence we have shown that

$$\sum_{n\le x}\left[\frac{x}{n}\right]\Lambda(n)=\sum_{p\le x}\left[\frac{x}{p}\right]\log p+O(x),$$

which, when used with (8) proves (9). □

To see why

$$\sum_{n=2}^{\infty} \frac{\log n}{n(n-1)} = O(1)$$

observe that

$$\begin{aligned}
\sum_{n=2}^{N} \frac{\log n}{n(n-1)} &= \sum_{n=2}^{N} \left(\frac{1}{n-1} - \frac{1}{n} \right) \log n \\
&= \sum_{n=2}^{N} \frac{\log n}{n-1} - \sum_{n=2}^{N} \frac{\log n}{n} \\
&= \log 2 + \sum_{n=2}^{N-1} \frac{1}{n} \left(\log(n+1) - \log n \right) - \frac{\log N}{N} \\
&< \log 2 + \sum_{n=2}^{N-1} \frac{1}{n} \cdot \frac{1}{n} \\
&< \log 2 + \sum_{n=2}^{N-1} \frac{1}{(n-1)n} \\
&= \log 2 + 1 - \frac{1}{N-1} \\
&< \log 2 + 1.
\end{aligned}$$

1.5 Another Identity for the Partial Sum of a Dirichlet Convolution

As in Theorem 1.11 we write

$$H(x) = \sum_{n \le x} h(n), \quad F(x) = \sum_{n \le x} f(n), \quad G(x) = \sum_{n \le x} g(n)$$

so

$$H(x) = \sum_{n \le x} \sum_{d|n} f(d) g\left(\frac{n}{d}\right) = \sum_{\substack{q,d \\ qd \le x}} f(d)g(q).$$

Theorem 1.16. *If a and b are positive real numbers such that $ab = x$, then*

$$\sum_{\substack{q,d \\ qd \le x}} f(d)g(q) = \sum_{n \le a} f(n) G\left(\frac{x}{n}\right) + \sum_{n \le b} g(n) F\left(\frac{x}{n}\right) - F(a)G(b). \tag{10}$$

Proof

The sum $H(x)$ on the left of (10) is extended over the lattice points of the hyperbolic region R bounded by $qd = x$, $q, d \ge 1$ Note that the point (a, b) lies on the hyperbola $qd = x$. Now consider the regions bounded by $d \le a$ (call it R_1) and $q \le b$ in R (call it R_2). Add up the contributions from these two regions (namely $R_1 \cup R_2$) and subtract contributions from the overlapping regions (namely $R_1 \cap R_2$). Then

$$H(x) = \sum_{d \le a} \sum_{q \le x/d} f(d)g(q) + \sum_{q \le b} \sum_{d \le x/q} f(d)g(q) - \sum_{d \le a} \sum_{q \le b} f(d)g(q).$$

□

Theorem 1.17 (**Shapiro's Tauberian Theorem**). *Let* $\{a(n)\}$ *be a nonnegative sequence such that*

$$\sum_{n\le x} a(n)\left[\frac{x}{n}\right] = x\log x + O(x) \quad \textit{for all } x \ge 1. \tag{11}$$

Then the following statements hold.

(a). *There is a constant* $A > 0$ *such that*

$$\sum_{n\le x} a(n) \le Ax \quad \textit{for all } x \ge 1.$$

(b). *For* $x \ge 1$, *we have*

$$\sum_{n\le x} \frac{a(n)}{n} = \log x + O(1).$$

Proof

Let

$$S(x) = \sum_{n\le x} a(n), \quad T(x) = \sum_{n\le x} a(n)\left[\frac{x}{n}\right].$$

To prove (a), we need first to show that

$$S(x) - S\left(\frac{x}{2}\right) \le T(x) - 2T\left(\frac{x}{2}\right). \tag{12}$$

Now observe that

$$\begin{aligned} T(x) - 2T\left(\frac{x}{2}\right) &= \sum_{n\le x}\left[\frac{x}{n}\right]a(n) - 2\sum_{n\le \frac{x}{2}}\left[\frac{x}{2n}\right]a(n) \\ &= \sum_{n\le \frac{x}{2}}\left(\left[\frac{x}{n}\right] - 2\left[\frac{x}{2n}\right]\right)a(n) + \sum_{\frac{x}{2}<n\le x}\left[\frac{x}{n}\right]a(n). \end{aligned}$$

Since $[2y] - 2[y]$ is either 0 or 1, the first sum is nonnegative, so

$$T(x) - 2T\left(\frac{x}{2}\right) \geq \sum_{\frac{x}{2}<n\leq x} \left[\frac{x}{n}\right] a(n) = \sum_{\frac{x}{2}<n\leq x} a(n) = S(x) - S\left(\frac{x}{2}\right).$$

This proves (12). But (11) implies

$$T(x) - 2T\left(\frac{x}{2}\right) = x\log x + O(x) - 2\left(\frac{x}{2}\log\frac{x}{2} + O(x)\right) = O(x).$$

Hence (12) implies

$$S(x) - S\left(\frac{x}{2}\right) = O(x).$$

This means that there is some constant $K > 0$ such that

$$S(x) - S\left(\frac{x}{2}\right) \leq Kx \quad \text{for all } \ x \geq 1.$$

Replace x by $x/2, x/4, \ldots$ to get

$$S\left(\frac{x}{2}\right) - S\left(\frac{x}{4}\right) \leq K\frac{x}{2},$$
$$S\left(\frac{x}{4}\right) - S\left(\frac{x}{8}\right) \leq K\frac{x}{4},$$

and so on. Note that $S(x/2^n) = 0$ when $2^n > x$. Adding these inequalities we get

$$S(x) \leq Kx\left(1 + \frac{1}{2} + \frac{1}{4} + \cdots\right) = 2Kx.$$

This proves (a) with $A = 2K$.

Next we prove (b). Observe that

$$\left[\frac{x}{n}\right] = \frac{x}{n} + O(1).$$

Then

$$\begin{aligned} T(x) &= \sum_{n\le x} \left[\frac{x}{n}\right] a(n) \\ &= \sum_{n\le x} \left(\frac{x}{n} + O(1)\right) a(n) \\ &= x \sum_{n\le x} \frac{a(n)}{n} + O\left(\sum_{n\le x} a(n)\right) \\ &= x \sum_{n\le x} \frac{a(n)}{n} + O(x), \end{aligned}$$

by part (a). Hence

$$\sum_{n\le x} \frac{a(n)}{n} = \frac{1}{x} T(x) + O(1) = \log x + O(1).$$

This proves (b). □

Theorem 1.18. *For all $x \ge 1$ we have*

$$\sum_{p\le x} \frac{\log p}{p} = \log x + O(1).$$

Proof

Let

$$a(n) = \begin{cases} \log p & \text{if } n \text{ is a prime } p, \\ 0 & \text{otherwise.} \end{cases}$$

Since $a(n) \ge 0$, (9) shows that the hypothesis of Shapiro's Theorem is satisfied. □

2 Algebraic Background

The proof of Dirichlet's Theorem will require a knowledge of certain arithmetical functions called Dirichlet characters. Although the study of Dirichlet characters can be done without any knowledge of groups, the introduction of a minimal amount of group theory places the theory of Dirichlet characters in a more natural setting and simplifies some of the discussion.

2.1 Groups

Definition 8. A ***group*** G is a nonempty set of elements together with a binary operation, which we denote by $\cdot$, such that the following postulates are satisfied:

(a). *Closure.* For every $a, b \in G$, $a \cdot b$ is also in G.

(b). *Associativity.* For every $a, b, c \in G$, we have $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

(c). *Existence of Identity.* There is a unique element $e \in G$, called the identity, such that $a \cdot e = e \cdot a = a$.

(d). *Existence of Inverses.* For every $a \in G$, there is a unique $b \in G$, such that $a \cdot b = b \cdot a = e$. This b is usually denoted by a^{-1}.

Note that we usually omit the dot and write ab for $a \cdot b$.

Definition 9. A group G is called ***abelian*** if every pair of elements commute; that is, if $ab = ba$ for all $a, b \in G$.

Definition 10. A group G is called ***finite*** if G is a finite set. In this case the number of elements in G is called the ***order*** of G and is denoted by $|G|$.

Definition 11. A nonempty subset G' of a group G which is itself a group, under the same operation, is called a ***subgroup*** of G.

Examples of groups and subgroups are given below:

(a). *Trivial subgroups.* Every group G has at least two subgroups, G itself and the set $\{e\}$ consisting of the identity alone.

(b). *Integers under addition.* The set of all integers is an abelian group with $+$ as the operation and 0 as the identity. The inverse of n is $-n$.

(c). *Complex numbers under multiplication.* The set of all non-zero complex numbers is an abelian group with ordinary multiplication of complex numbers as the operation and 1 as the identity. The inverse of z is the reciprocal $1/z$. The set of all complex numbers of absolute value 1 is a subgroup.

(d). *The nth roots of unity.* The set $\{1, \varepsilon, \varepsilon^2, \ldots, \varepsilon^{n-1}\}$, where $\varepsilon = e^{\frac{2\pi i}{n}}$, together with the ordinary multiplication of complex numbers, is a finite group of order n.

The following elementary theorems concern an arbitrary group G. Unless otherwise stated, G is not required to be abelian nor finite.

Theorem 2.1. *If $a, b, c \in G$ satisfy $ac = bc$ or $ca = cb$, then $a = b$.*

Proof
In the first case multiply each side on the right by c^{-1} and use associativity. In the second case multiply on the left by c^{-1} and use associativity. □

Theorem 2.2. *In any group G, we have*

(a). $e^{-1} = e$.
(b). *For every $a \in G$,* $(a^{-1})^{-1} = a$.
(c). *For all $a, b \in G$,* $(ab)^{-1} = b^{-1}a^{-1}$.
(d). *For all $a, b \in G$, the equation $ax = b$ has the unique solution $x = a^{-1}b$; the equation $ya = b$ has the unique solution $y = ba^{-1}$.*

Definition 12. If $a \in G$ we define a^n for any integer n by the following relations:

$$a^0 = e, \quad a^n = aa^{n-1}, \quad a^{-n} = (a^{-1})^n \quad \text{for } n > 0.$$

Theorem 2.3. *If $a \in G$, any two powers of a commute, and for all integers m and n we have*

$$a^m a^n = a^{m+n} = a^n a^m \quad \textit{and} \quad (a^m)^n = a^{mn} = (a^n)^m.$$

Moreover, if a and b commute we have $a^n b^n = (ab)^n$.

Theorem 2.4. *If G' is a nonempty subset of a group G, then G is a subgroup if and only if G' satisfies the following:*

(a). *If $a, b \in G'$, then $ab \in G'$.*

(b). *If $a \in G'$, then $a^{-1} \in G'$.*

2.2 Construction of Subgroups

A subgroup of a given group G can always be constructed by choosing any element a in G and forming the set of all its powers a^n, $n = 0, \pm 1, \pm 2, \ldots$. This set clearly satisfies the postulates of a group. It is called the ***cyclic subgroup generated by*** a and is denoted by $\langle a \rangle$.

Note that $\langle a \rangle$ is abelian, even if G is not. If $a^n = e$ for some positive integer n there will be a smallest $n > 0$ with this property and the subgroup $\langle a \rangle$ will be a finite group of order n,

$$\langle a \rangle = \left\{ a, a^2, \ldots, a^{n-1}, a^n = e \right\}.$$

The integer n is also called the ***order of the element*** a. The next theorem shows us that every element of a finite group has a finite order.

Theorem 2.5. *If G is finite and $a \in G$, then there is a positive integer $n \leq |G|$ such that $a^n = e$.*

Proof

Let $g = |G|$. Then at least two of the following $g + 1$ elements of G must be equal:

$$e, a, a^2, \ldots, a^g.$$

Suppose that $a^r = a^s$, where $0 \leq s < r \leq g$. Then we have

$$e = a^r \left(a^s\right)^{-1} = a^{r-s}.$$

This proves the theorem with $n = r - s$. □

If G' is a subgroup of a finite group G, then for any element a in G there is an integer n such that $a^n \in G'$. If a is already in G' we simply take $n = 1$. If $a \notin G'$ we can take n to be the order of a, since $a^n = e \in G'$. However, there may be a smaller positive power of a which lies in G'. By the Well-ordering Principle, there is a *smallest* positive integer n such that $a^n \in G'$. We call this integer the ***indicator*** of a.

Theorem 2.6. *Let G' be a subgroup of a finite abelian group G, where $G' \neq G$. Choose an element a in G, $a \notin G'$, and let h be the indicator of a in G'. Then the set of products*

$$G'' = \{\ xa^k\ :\ x \in G' \text{ and } k = 0, 1, \ldots, h-1\ \}$$

is a subgroup of G which contains G'. Moreover, the order of G'' is h times that of G', that is,

$$|G''| = h\,|G'|\,.$$

Proof

Choose two elements in G'', say xa^k and ya^j, where $x, y \in G'$ and $0 \leq k, j < h$. Since G is abelian the product of the elements is

$$(xy)a^{k+j} \tag{13}$$

Now $k + j = qh + r$ where $0 \leq r < h$. Hence

$$a^{k+j} = a^{qh+r} = a^{qh}a^r = za^r$$

where $z = a^{qh} = (a^h)^q \in G'$ since $a^h \in G'$. Therefore the element in (13) is $(xyz)a^r = wa^r$, where $w \in G'$ and $0 \leq r < h$. This proves that G is closed.

Next we prove that the inverse of each element in G'' is also in G''. Choose an arbitrary element in G'', say xa^k. If $k = 0$ then the inverse is x^{-1} which is in G''. If $0 < k < h$ the inverse is the element

$$(x^{-1}(a^h)^{-1})a^{h-k}$$

which is again in G''. This proves that G'' is indeed a group. Obviously G'' contains G'.

Finally, we determine the order of G''. Let $m = |G'|$. As x runs through the m elements of G' and k runs through the h integers $0, 1, \ldots, h-1$ we obtain mh products xa^k. If we show that all these are *distinct*, then G'' has order mh. Now consider two of these products, say xa^k and ya^j and assume that

$$xa^k = ya^j \quad \text{with } 0 \leq j \leq k < h.$$

Then $a^{k-j} = x^{-1}y$ and $0 \leq k - j < h$. Since $x^{-1}y \in G'$ we must have $a^{k-j} \in G'$ so $k = j$ and hence $x = y$. □

2.3 Characters of Finite Abelian Groups

Definition 13. Let G be an arbitrary group. A complex-valued function f defined on G is called a ***character*** of G if f has the multiplicative property

$$f(ab) = f(a)f(b)$$

for all $a, b \in G$, and if $f(c) \neq 0$ for some $c \in G$.

Theorem 2.7. *If f is a character of a finite group G with identity element e, then $f(e) = 1$ and each function value $f(a)$ is a root of unity. In fact, if $a^n = e$ then $(f(a))^n = 1$.*

Proof

Choose $c \in G$ such that $f(c) \neq 0$. Since $ce = c$ we have

$$f(c)f(e) = f(c)$$

so $f(e) = 1$. If $a^n = e$ then $(f(a))^n = f(a^n) = f(e) = 1$. □

Every group G has at least one character, namely the function which is identically 1 on G. This is called the *principal* character. The next theorem tells us that there are further characters if G is abelian and has finite order greater than 1.

Theorem 2.8. *A finite abelian group G of order n has exactly n distinct characters.*

Proof

In Theorem 2.6 we learned how to construct, from a given subgroup $G' \neq G$, a new subgroup G'' containing G' and at least one more element a not in G'. We use the symbol $\langle\, G'; a \,\rangle$ to denote the subgroup G'' constructed in Theorem 2.6. Thus

$$\langle\, G'; a \,\rangle = \{\, xa^k \;:\; x \in G' \text{ and } 0 \leq k < h \,\}$$

where h is the indicator of a in G'.

Now we apply the construction repeatedly, starting with the subgroup $\{e\}$ which we denote by G_1. If $G_1 \neq G$, we let a_1 be an element of G other than e and define $G_2 = \langle\, G_1; a_1 \,\rangle$. If $G_2 \neq G$, we let a_2 be an element of G which is not in G_2 and define $G_3 = \langle\, G_2; a_2 \,\rangle$. Continue the process to obtain a finite set of elements $a_1, a_2, \ldots, a_t$ and a corresponding set of subgroups $G_1, G_2, \ldots, G_{t+1}$ such that

$$G_{r+1} = \langle\, G_r; a_r \,\rangle \quad \text{with} \;\; G_1 \subset G_2 \subset \cdots \subset G_{t+1} = G.$$

The process must terminate in a finite number of steps since the given group is finite and each G_{r+1} contains more elements than its predecessor G_r. We consider such a chain of subgroups and prove the theorem by induction, showing that if it is true for G_r it must also be true for G_{r+1}.

It is clear that there is only one character for G_1, namely the function which is identically 1. Assume, therefore, that G_r has order m and that there are exactly m distinct characters for G_r. Consider $G_{r+1} = \langle\, G_r; a_r \,\rangle$ and let h be the indicator of a_r in G_r, that is, the smallest positive integer such that $a_r^h \in G_r$. We shall show that there are exactly h different ways to extend each character of G_r to obtain a character of G_{r+1}, and that each character of G_{r+1} is the extension of some character of G_r. This will prove that G_{r+1} has exactly mh characters, and since mh is also the order of G_{r+1} this will prove the theorem by induction on r.

A typical element in G_{r+1} has the form

$$x a_r^k \quad \text{where} \;\; x \in G_r \;\; \text{and} \;\; 0 \leq k < h.$$

Suppose for the moment that it is possible to extend a character f of G_r to G_{r+1}. Call this extension $\tilde{f}$ and let us see what can be said about $\tilde{f}(x a_r^k)$. The multiplicative property requires

$$\tilde{f}(x a_r^k) = \tilde{f}(x)\tilde{f}(a_r)^k.$$

But $x \in G_r$ so $\tilde{f}(x) = f(x)$ and the foregoing equation implies

$$\tilde{f}(x a_r^k) = f(x)\tilde{f}(a_r)^k.$$

This tells us that $\tilde{f}(xa_r^k)$ is determined as soon as $\tilde{f}(a_r)$ is known.

What are the possible values of $\tilde{f}(a_r)$? Let $c = a_r^h$. Since $c \in G_r$ we have $\tilde{f}(c) = f(c)$, and since $\tilde{f}$ is multiplicative we also have $\tilde{f}(c) = \tilde{f}(a_r)^h$. Hence

$$\tilde{f}(a_r)^h = f(c),$$

so $\tilde{f}(a_r)$ is one of the hth roots of $f(c)$. There are at most h choices for $\tilde{f}(a_r)$.

These observations tell us how to define $\tilde{f}$. If f is a given character of G_r we choose one of the hth roots of $f(c)$, where $c = a_r^h$, and define $\tilde{f}(a_r)$ to be this root. Then we define $\tilde{f}$ on the rest of G_{r+1} by the equation

$$\tilde{f}(xa_r^k) = f(x)\tilde{f}(a_r)^k. \tag{14}$$

The h choices for $\tilde{f}(a_r)$ are all different so this gives us h different ways to define $\tilde{f}(xa_r^k)$. Now we verify that the function $\tilde{f}$ so defined has the required multiplicative property. From (14) we find

$$\begin{aligned}\tilde{f}(xa_r^k \cdot ya_r^j) &= \tilde{f}(xy \cdot a_r^{k+j}) \\ &= f(xy)\tilde{f}(a_r)^{k+j} \\ &= f(x)f(y)\tilde{f}(a_r)^k\tilde{f}(a_r)^j \\ &= \tilde{f}(xa_r^k)\tilde{f}(ya_r^j)\end{aligned}$$

so $\tilde{f}$ is a character of G_{r+1}. No two of the extensions $\tilde{f}$ and $\tilde{g}$ can be identical on G_{r+1} because the functions f and g which they extend would then be identical on G_r. Therefore each of the m characters of G_r can be extended in h different ways to produce a character of G_{r+1}. Moreover, if φ is any character of G_{r+1} then its restriction to G_r is also a character of G_r, so the extension process produces all the characters of G_{r+1}. □

2.4 The Character Group

In what follows G will be assumed to be a finite abelian group of order n. The principal character of G will be denoted by f_1. The others, denoted by $f_2, f_3, \ldots, f_n$, are called non-principal characters. They have the property that $f(a) \neq 1$ for some $a \in G$.

Theorem 2.9. *If multiplication of characters is defined by the relation*

$$(f_i f_j)(a) = f_i(a) f_j(a)$$

for each $a \in G$, then the set of characters of G forms an abelian group of order n. We denote this group by $\hat{G}$. The identity element of $\hat{G}$ is the principal character f_1. The inverse of f_i is the reciprocal $1/f_i$.

Proof
Straightforward. □

For each character f we have $|f(a)| = 1$. Hence the reciprocal $1/f(a)$ is equal to the complex conjugate $\overline{f(a)}$. Thus, the function $\overline{f}$ defined by $\overline{f}(a) = \overline{f(a)}$ is also a character of G. Moreover, we have

$$\overline{f}(a) = \frac{1}{f(a)} = f(a^{-1})$$

for every $a \in G$.

2.5 Orthogonality Relations for Characters

Let $a_1, a_2, \ldots, a_n \in G$ and denote by $A = A(G)$ the $n \times n$ matrix (a_{ij}) whose element a_{ij} in the ith row and jth column is $a_{ij} = f_i(a_j)$.

We will prove that the matrix A has an inverse and then use this fact to deduce the so-called orthogonality relations for characters. First we determine the sum of the entries in each row of A.

Theorem 2.10. *The sum of the entries in the ith row of A is given by*

$$\sum_{r=1}^{n} f_i(a_r) = \begin{cases} n & \text{if } f_i \text{ is the principal character,} \\ 0 & \text{otherwise.} \end{cases}$$

Proof

Let S denote the sum in question. If $f_i = f_1$, then each term of the sum is 1 and $S = n$. If $f_i \neq f_1$, there is an element b in G for which $f_i(b) \neq 1$. As a_r runs through the elements of G so does the product ba_r. Hence

$$S = \sum_{r=1}^{n} f_i(ba_r) = f_i(b) \sum_{r=1}^{n} f_i(a_r) = f_i(b)S.$$

Therefore $S(1 - f_i(b)) = 0$. Since $f_i(b) \neq 1$ it follows that $S = 0$. □

Now we use this theorem to show that A has an inverse.

Theorem 2.11. *Let A^* denote the conjugate transpose of the matrix A. Then we have $AA^* = nI$, where I is the $n \times n$ identity matrix. Hence $n^{-1}A^*$ is the inverse of A.*

Proof

Let $B = AA^*$. The entry b_{ij} in the ith row and the jth column of B is given by

$$b_{ij} = \sum_{r=1}^{n} f_i(a_r)\overline{f}_j(a_r) = \sum_{r=1}^{n} (f_i\overline{f}_j)(a_r) = \sum_{r=1}^{n} f_k(a_r),$$

where $f_k = f_i\overline{f}_j = f_i/f_j$. Now $f_i/f_j = f_1$ if and only if $i = j$. Hence by Theorem 2.10 we have

$$b_{ij} = \begin{cases} n & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

In other words, $B = nI$. □

Next we use the fact that a matrix commutes with its inverse to deduce the orthogonality relations for characters.

Theorem 2.12. *We have*

$$\sum_{r=1}^{n} \overline{f}_r(a_i) f_r(a_j) = \begin{cases} n & \text{if } a_i = a_j, \\ 0 & \text{if } a_i \neq a_j. \end{cases} \tag{15}$$

Proof

The relation $AA^* = nI$ implies $A^*A = nI$. But the element in the ith row and the jth column of A^*A is the sum on the left of (15). □

Since $\overline{f}_r(a_i) = f_r(a_i)^{-1} = f_r(a_i^{-1})$, the general term of the sum in (15) is equal to

$$f_r(a_i^{-1}) f_r(a_j) = f_r(a_i^{-1} a_j).$$

Therefore the orthogonality relations can be expressed as follows:

$$\sum_{r=1}^{n} f_r(a_i^{-1} a_j) = \begin{cases} n & \text{if } a_i = a_j, \\ 0 & \text{if } a_i \neq a_j. \end{cases}$$

When a_i is the identity element e we obtain the following result.

Theorem 2.13. *The sum of the entries in the jth column of A is given by*

$$\sum_{r=1}^{n} f_r(a_j) = \begin{cases} n & \text{if } a_j = e, \\ 0 & \text{otherwise.} \end{cases}$$

2.6 Dirichlet Characters

The foregoing discussion dealt with characters of an arbitrary finite abelian group G. Now we specialize G to be the group of reduced residue classes modulo a fixed positive integer k. First we prove that these residue classes do, indeed, form a group if multiplication is suitably defined.

We recall that that reduced residue system modulo k is a set of $\varphi(k)$ integers

$$\left\{a_1, a_2, \ldots, a_{\varphi(k)}\right\}$$

incongruent modulo k, each of which is relatively prime to k. For each integer a the corresponding residue class $\hat{a}$ is the set of all integers congruent to a modulo k, that is,

$$\hat{a} = \{x : x \equiv a \pmod{k}\}.$$

Multiplication of residue classes is defined by the relation

$$\hat{a} \cdot \hat{b} = \widehat{ab}. \tag{16}$$

That is, the product of two residue classes $\hat{a}$ and $\hat{b}$ is the residue class of the product ab.

Theorem 2.14. *With multiplication defined by* (16)*, the set of reduced residue classes modulo k is a finite abelian group of order $\varphi(k)$. The identity is the residue class $\hat{1}$. The inverse of $\hat{a}$ is the residue class $\hat{b}$ where $ab \equiv 1 \pmod{k}$.*

Proof

The closure property is automatically satisfied because of the way multiplication of residue classes was defined. The class $\hat{1}$ is clearly the identity element. If $\gcd(a, k) = 1$ there is a unique b such that $ab \equiv 1 \pmod{k}$. Hence the inverse of $\hat{a}$ is $\hat{b}$. Finally, it is clear that the group is abelian and that its order is $\varphi(k)$. □

Definition 14. Let G be a group of reduced residue classes modulo k. Corresponding to each character f of G we define an arithmetical function $\chi = \chi_f$ as follows:

$$\chi(n) = \begin{cases} f(\hat{n}) & \text{if } \gcd(n,k) = 1, \\ 0 & \text{if } \gcd(n,k) > 1. \end{cases}$$

The function χ is called a ***Dirichlet character modulo*** k. The principal character χ_1 is that which has the properties

$$\chi_1(n) = \begin{cases} 1 & \text{if } \gcd(n,k) = 1, \\ 0 & \text{if } \gcd(n,k) > 1. \end{cases}$$

Theorem 2.15. *There are $\varphi(k)$ distinct Dirichlet characters modulo k, each of which is completely multiplicative and periodic wih period k. That is, we have*

$$\chi(mn) = \chi(m)\chi(n) \quad \textit{for all } m, n \tag{17}$$

and

$$\chi(n+k) = \chi(n) \quad \textit{for all } n.$$

Conversely, if χ is completely multiplicative and periodic with period k, and if $\chi(n) = 0$ whenever $\gcd(n,k) > 1$, then χ is one of the Dirichlet characters modulo k.

Proof

There are $\varphi(k)$ characters f for the group G of reduced residue classes modulo k, hence $\varphi(k)$ characters χ_f modulo k. The multiplicative property (17) of χ_f follows from that of f when both m and n are relatively prime to k. If one of m or n is not relatively prime to k then neither is mn, hence both sides of (17) are zero. The periodicity property follows from the fact that $\chi_f(n) = f(\hat{n})$ and that $a \equiv b \pmod{k}$ implies $\gcd(a,k) = \gcd(b,k)$.

To prove the converse we note that the function f defined on the group G by the equation

$$f(\hat{n}) = \chi(n) \quad \text{if} \quad \gcd(n,k) = 1$$

is a character of G, so χ is a Dirichlet character modulo k. □

Example 1. When $k = 1$ or $k = 2$ then $\varphi(k) = 1$ and the only Dirichlet character is the principal character χ_1. For $k \geq 3$, there are at least two Dirichlet characters since $\varphi(k) \geq 2$. The following tables display all the Dirichlet characters for $k = 3, 4$ and 5, respectively.

n	1	2	3
$\chi_1(n)$	1	1	0
$\chi_2(n)$	1	−1	0

n	1	2	3	4
$\chi_1(n)$	1	0	1	0
$\chi_2(n)$	1	0	−1	0

n	1	2	3	4	5
$\chi_1(n)$	1	1	1	1	0
$\chi_2(n)$	1	−1	−1	1	0
$\chi_3(n)$	1	i	$-i$	−1	0
$\chi_4(n)$	1	$-i$	i	−1	0

To fill these tables we use the fact that $\chi(n)^{\varphi(k)} = 1$ whenever $\gcd(n,k) = 1$, so $\chi(n)$ is a $\varphi(k)$th root of unity. We also note that if χ is a character modulo k so is the complex conjugate $\overline{\chi}$. This information is enough to complete the tables for $k = 3$ and $k = 4$.

When $k = 5$ we have $\varphi(5) = 4$ so the possible values of $\chi(n)$ are ± 1 and $\pm i$ when $\gcd(n, 5) = 1$. Also $\chi(2)\chi(3) = \chi(6) = \chi(1) = 1$ so $\chi(2)$ and $\chi(3)$ are reciprocals. Since $\chi(4) = \chi(2)^2$ this information suffices to fill the table for $k = 5$. The following tables display all the Dirichlet characters modulo 6 and 7 (where $\omega = e^{\frac{i\pi}{3}}$)

n	1	2	3	4	5	6
$\chi_1(n)$	1	0	0	0	1	0
$\chi_2(n)$	1	0	0	0	-1	0

n	1	2	3	4	5	6	7
$\chi_1(n)$	1	1	1	1	1	1	0
$\chi_2(n)$	1	1	-1	1	-1	-1	0
$\chi_3(n)$	1	ω^2	ω	$-\omega$	$-\omega^2$	-1	0
$\chi_4(n)$	1	ω^2	$-\omega$	$-\omega$	ω^2	1	0
$\chi_5(n)$	1	$-\omega$	ω^2	ω^2	$-\omega$	1	0
$\chi_6(n)$	1	$-\omega$	$-\omega^2$	ω^2	ω	-1	0

In our discussion of Dirichlet's theorem on primes in an arithmetic progression we shall make use of the following orthogonality relation for characters modulo k.

Theorem 2.16. *Let $\chi_1, \ldots, \chi_{\varphi(k)}$ denote the $\varphi(k)$ Dirichlet characters modulo k. Let m and n be two integers with $\gcd(n, k) = 1$. Then we have*

$$\sum_{r=1}^{\varphi(k)} \chi_r(m)\overline{\chi}_r(n) = \begin{cases} \varphi(k) & \text{if } m \equiv n \pmod{k}, \\ 0 & \text{if } m \not\equiv n \pmod{k}. \end{cases}$$

Proof

If $\gcd(m, k) = 1$, take $a_i = \hat{n}$ and $a_j = \hat{m}$ in the orthogonality relation of Theorem 2.12 and note that $\hat{m} = \hat{n}$ if and only if $m \equiv n \pmod{k}$. If $\gcd(m, k) > 1$ each term in the sum vanishes and $m \not\equiv n \pmod{k}$. □

2.7 Sums involving Dirichlet Characters

In what follows, we will discuss certain sums which occur in the proof of Dirichlet's Theorem on primes in arithmetical progression.

The first theorem refers to a nonprincipal character χ modulo k, but the proof is also valid if χ is any arithmetical function with bounded partial sums.

Theorem 2.17. *Let χ be any nonprincipal character modulo k, and let f be a nonnegative function which has a continuous negative derivative $f'(x)$ for all $x \geq x_0$. Then if $y \geq x \geq x_0$, we have*

$$\sum_{x<n\leq y} \chi(n)f(n) = O(f(x)). \tag{18}$$

If, in addition, $f(x) \to 0$ as $x \to \infty$, then the infinite series

$$\sum_{n=1}^{\infty} \chi(n)f(n)$$

converges and we have, for $x \geq x_0$

$$\sum_{n\leq x} \chi(n)f(n) = \sum_{n=1}^{\infty} \chi(n)f(n) + O(f(x)). \tag{19}$$

Proof

Let $A(x) = \sum_{n\leq x} \chi(n)$. Since χ is nonprincipal we have

$$A(k) = \sum_{n=1}^{k} \chi(n) = 0.$$

By periodicity it follows that $A(nk) = 0$ for $n = 2, 3, \ldots$, hence $|A(x)| < \varphi(k)$ for all x. In other words, $A(x) = O(1)$.

Now we use Abel's Identity to express (18) as an integral. This gives us

$$\begin{aligned}\sum_{x<n\le y}\chi(n)f(n) &= f(y)A(y)-f(x)A(x)-\int_x^y A(t)f'(t)dt\\ &= O(f(y))+O(f(x))+O\left(\int_x^y(-f'(t))dt\right)\\ &= O(f(x)).\end{aligned}$$

This proves (18). If $f(x)\to 0$ as $x\to\infty$ then (18) shows that the series

$$\sum_{n=1}^{\infty}\chi(n)f(n)$$

converges because of the Cauchy convergence criterion. To prove (19) we simply note that

$$\sum_{n=1}^{\infty}\chi(n)f(n)=\sum_{n\le x}\chi(n)f(n)+\lim_{y\to\infty}\sum_{x<n\le y}\chi(n)f(n).$$

Because of (18) the limit on the right is $O(f(x))$. □

Now we apply Theorem 2.17 successively with

$$f(x) = \frac{1}{x}, \quad f(x) = \frac{\log x}{x}, \quad f(x) = \frac{1}{\sqrt{x}} \quad \text{for } x \geq 1$$

to obtain the following result.

Theorem 2.18. *If χ is any nonprincipal character modulo k and if $x \geq 1$ we have*

$$\sum_{n \leq x} \frac{\chi(n)}{n} = \sum_{n=1}^{\infty} \frac{\chi(n)}{n} + O\left(\frac{1}{x}\right), \tag{20}$$

$$\sum_{n \leq x} \frac{\chi(n) \log n}{n} = \sum_{n=1}^{\infty} \frac{\chi(n) \log n}{n} + O\left(\frac{\log x}{x}\right), \tag{21}$$

$$\sum_{n \leq x} \frac{\chi(n)}{\sqrt{n}} = \sum_{n=1}^{\infty} \frac{\chi(n)}{\sqrt{n}} + O\left(\frac{1}{\sqrt{x}}\right). \tag{22}$$

□

2.8 The Nonvanishing of $L(1,\chi)$ for Real Nonprincipal χ

We denote by $L(1,\chi)$ the sum of the series in (20). Thus

$$L(1,\chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n}.$$

In the proof of Dirichlet's Theorem we need to know that $L(1,\chi) \neq 0$ when χ is a nonprincipal character. We prove this here for real nonprincipal characters. First we consider the divisor sum of $\chi(n)$.

Theorem 2.19. *Let χ be any real-valued character modulo k and let*

$$A(n) = \sum_{d|n} \chi(d).$$

Then $A(n) \geq 0$ for all n, and $A(n) \geq 1$ if n is a square.

Proof

For prime powers we have

$$A(p^a) = \sum_{t=0}^{a} \chi(p^t) = 1 + \sum_{t=1}^{a} \chi(p)^t.$$

Since χ is real-valued the only possible values for $\chi(p)$ are $0, 1, -1$.

If $\chi(p) = 0$ then $A(p^a) = 1$; if $\chi(p) = 1$ then $A(p^a) = a+1$; $\chi(p) = -1$ then

$$A(p^a) = \begin{cases} 0 & \text{if } a \text{ is odd,} \\ 1 & \text{if } a \text{ is even.} \end{cases}$$

In any case $A(p^a) \geq 1$ if a is even.

Now if $n = p_1^{a_1} \cdot \ldots \cdot p_r^{a_r}$ then $A(n) = A(p_1^{a_1}) \cdot \ldots \cdot A(p_r^{a_r})$ since A is multiplicative. Each factor $A(p_i^{a_i}) \geq 0$ hence $A(n) \geq 0$. Also, if n is a square then each exponent a_i is even, so each factor $A(p_i^{a_i}) \geq 1$ hence $A(n) \geq 1$. □

Theorem 2.20. *For any real-valued nonprincipal character χ modulo k, let*

$$A(n) = \sum_{d|n} \chi(d) \quad \textit{and} \quad B(n) = \sum_{n \le x} \frac{A(n)}{\sqrt{n}}.$$

Then we have

(a). $B(x) \to \infty$ *as* $x \to \infty$.
(b). $B(x) = 2\sqrt{x}L(1,\chi) + O(1)$ *for all* $x \ge 1$.

Therefore $L(1,\chi) \neq 0$.

Proof
To prove part (a) we use Theorem 2.19 to write

$$B(x) \ge \sum_{\substack{n \le x \\ n=m^2}} \frac{1}{\sqrt{n}} = \sum_{m \le \sqrt{x}} \frac{1}{m}.$$

The last sum tends to ∞ as $x \to \infty$ since the harmonic series $\sum \frac{1}{m}$ diverges.
To prove part (b) we write

$$B(x) = \sum_{n \le x} \frac{1}{\sqrt{n}} \sum_{d|n} \chi(d) = \sum_{\substack{q,d \\ qd \le x}} \frac{\chi(d)}{\sqrt{qd}}.$$

Now we invoke Theorem 1.16 which states that

$$\sum_{\substack{q,d \\ qd \le x}} f(d)g(q) = \sum_{n \le a} f(n)G\left(\frac{x}{n}\right) + \sum_{n \le b} g(n)F\left(\frac{x}{n}\right) - F(a)G(b)$$

where $ab = x$, $F(x) = \sum_{n \le x} f(n)$ and $G(x) = \sum_{n \le x} g(n)$. We take $a = b = \sqrt{x}$ and let

$$f(n) = \frac{\chi(n)}{\sqrt{n}} \quad \text{and} \quad g(n) = \frac{1}{\sqrt{n}}$$

to obtain

$$B(x) = \sum_{\substack{q,d \\ qd \le x}} \frac{\chi(d)}{\sqrt{qd}} = \sum_{n \le \sqrt{x}} \frac{\chi(n)}{\sqrt{n}} G\left(\frac{x}{n}\right) + \sum_{n \le \sqrt{x}} \frac{1}{\sqrt{n}} F\left(\frac{x}{n}\right) - F(\sqrt{x})G(\sqrt{x}). \tag{23}$$

By Theorem 1.16,

$$G(x) = \sum_{n \le x} \frac{1}{\sqrt{n}} = 2\sqrt{x} + A + O\left(\frac{1}{\sqrt{x}}\right)$$

where A is a constant, and by Theorem 2.18, Equation (22), we have

$$F(x) = \sum_{n \le x} \frac{\chi(n)}{\sqrt{n}} = B + O\left(\frac{1}{\sqrt{x}}\right)$$

where $B = \sum_{n=1}^{\infty} \frac{\chi(n)}{\sqrt{n}}$. Since

$$F(\sqrt{x})G(\sqrt{x}) = 2Bx^{\frac{1}{4}} + O(1),$$

Equation (23) gives us

$$\begin{aligned} B(x) &= \sum_{n \le \sqrt{x}} \frac{\chi(n)}{\sqrt{n}} \left(2\sqrt{\frac{x}{n}} + A + O\left(\sqrt{\frac{n}{x}}\right)\right) + \sum_{n \le \sqrt{x}} \frac{1}{\sqrt{n}} \left(B + O\left(\sqrt{\frac{n}{x}}\right)\right) - 2Bx^{\frac{1}{4}} + O(1) \\ &= 2\sqrt{x} \sum_{n \le \sqrt{x}} \frac{\chi(n)}{n} + A \sum_{n \le \sqrt{x}} \frac{\chi(n)}{\sqrt{n}} + O\left(\frac{1}{\sqrt{x}} \sum_{n \le \sqrt{x}} |\chi(n)|\right) + B \sum_{n \le \sqrt{x}} \frac{1}{\sqrt{n}} + O\left(\frac{1}{\sqrt{x}} \sum_{n \le \sqrt{x}} 1\right) - 2Bx^{\frac{1}{4}} + O(1) \\ &= 2\sqrt{x} L(1, \chi) + O(1). \end{aligned}$$

This proves part (b). Now it is clear that parts (a) and (b) together imply $L(1, \chi) \neq 0$. □

3 The Elementary Proof

In Theorem 1.18, we derived the asymptotic formula

$$\sum_{p\le x}\frac{\log p}{p}=\log x+O(1), \tag{24}$$

Theorem 3.1. *If $k>0$ and $\gcd(h,k)=1$, then, for all $x>1$,*

$$\sum_{\substack{p\le x\\ p\equiv h\,(\mathrm{mod}\,k)}}\frac{\log p}{p}=\frac{1}{\varphi(k)}\log x+O(1), \tag{25}$$

where the sum is extended over those primes p less than or equal to x which are congruent to h mod k.

Since $\log x\to\infty$ as $x\to\infty$, Theorem 3.1 implies that there are infinitely many primes $p\equiv h\ (\mathrm{mod}\ k)$, hence infinitely many in the progression $nk+h$, where $n=0,1,\ldots$.

Note that the principal term on the right of (25) is independent of h. Therefore (25) not only implies Dirichlet's Theorem but it also shows that the primes in each of the $\varphi(k)$ reduced residue classes mod k make the same contribution to the principal term in (24).

The proof of Theorem 3.1 will be presented through a sequence of lemmas which we have collected together in the next subsection to reveal the plan of the proof. Throughout the section we adopt the following notation.

The positive integer k represents a fixed modulus, and h is a fixed integer relatively prime to k. The $\varphi(k)$ Dirichlet characters mod k are denoted by $\chi_1,\chi_2,\ldots,\chi_{\varphi(k)}$ with χ_1 denoting the principal character. For $\chi\neq\chi_1$ we write $L(1,\chi)$ and $L'(1,\chi)$ for the sums of the following series:

$$L(1,\chi)=\sum_{n=1}^{\infty}\frac{\chi(n)}{n},\quad L'(1,\chi)=-\sum_{n=1}^{\infty}\frac{\chi(n)\log n}{n}.$$

The convergence of each of these series was shown in Theorem 2.18. Moreover, in Theorem 2.20 we proved that $L(1,\chi)\neq 0$ if χ is real-valued.

The Plan of the Proof

Lemma 3.1. *For $x > 1$, we have*

$$\sum_{\substack{p \le x \\ p \equiv h \,(\mathrm{mod}\, k)}} \frac{\log p}{p} = \frac{1}{\varphi(k)} \log x + \frac{1}{\varphi(k)} \sum_{r=2}^{\varphi(k)} \overline{\chi}_r(h) \sum_{p \le x} \frac{\chi_r(p) \log p}{p} + O(1).$$

It is clear that Lemma 3.1 will imply Theorem 3.1 if we show that

$$\sum_{p \le x} \frac{\chi(p) \log p}{p} = O(1) \tag{26}$$

for each $\chi \neq \chi_1$. The next lemma expresses this sum in a form which is not extended over primes.

Lemma 3.2. *For $x > 1$ and $\chi \neq \chi_1$ we have*

$$\sum_{p \le x} \frac{\chi(p) \log p}{p} = -L'(1, \chi) \sum_{n \le x} \frac{\mu(n)\chi(n)}{n} + O(1).$$

Therefore Lemma 3.2 will imply (26) if we show that

$$\sum_{n \le x} \frac{\mu(n)\chi(n)}{n} = O(1). \tag{27}$$

This, in turn, will be deduced from the following lemma.

Lemma 3.3. *For $x > 1$ and $\chi \neq \chi_1$ we have*

$$L(1, \chi) \sum_{n \le x} \frac{\mu(n)\chi(n)}{n} = O(1). \tag{28}$$

If $L(1,\chi) \neq 0$ we can cancel $L(1,\chi)$ in (28) to obtain (27). Therefore, the proof of Dirichlet's Theorem depends ultimately on the non-vanishing of $L(1,\chi)$ for all $\chi \neq \chi_1$. As already remarked, this was proved for real $\chi \neq \chi_1$ in Theorem 2.20 so it remains to prove that $L(1,\chi) \neq 0$ for all $\chi \neq \chi_1$ which take complex as well as real values.

For this purpose we let $N(k)$ denote the number of nonprincipal characters $\chi \bmod k$ such that $L(1,\chi) = 0$. If $L(1,\chi) = 0$, then $L(1,\overline{\chi}) = 0$ and $\chi \neq \overline{\chi}$ since χ is not real. Therefore the characters χ for which $L(1,\chi) = 0$ occur in conjugate pairs, so $N(k)$ is *even*. Our goal is to prove that $N(k) = 0$, and this will be deduced from the following asymptotic formula.

Lemma 3.4. *For $x > 1$ we have*

$$\sum_{\substack{p \leq x \\ p \equiv 1 \,(\mathrm{mod}\, k)}} \frac{\log p}{p} = \frac{1 - N(k)}{\varphi(k)} \log x + O(1). \tag{29}$$

If $N(k) \neq 0$ then $N(k) \geq 2$ since $N(k)$ is even, hence the coefficient of $\log x$ in (29) is negative and the right-hand side approaches $-\infty$ as $x \to \infty$. This is a contradiction since all the terms on the left are positive. Therefore Lemma 3.4 implies that $N(k) = 0$. The proof of Lemma 3.4, in turn, will be based on the following asymptotic formula.

Lemma 3.5. *For $\chi \neq \chi_1$ and $L(1,\chi) = 0$ we have*

$$L'(1,\chi) \sum_{n \leq x} \frac{\mu(n)\chi(n)}{n} = \log x + O(1)$$

Proof of Lemma 3.1

To prove Lemma 3.1 we begin with the asymptotic formula

$$\sum_{p\le x}\frac{\log p}{p}=\log x+O(1)$$

and extract those terms in the sum arising from primes $p\equiv h \pmod k$. The extraction is done with the aid of the orthogonality relation for Dirichlet characters, namely

$$\sum_{r=1}^{\varphi(k)}\chi_r(m)\overline{\chi}_r(n)=\begin{cases}\varphi(k) & \text{if } m\equiv n \pmod k,\\ 0 & \text{if } m\not\equiv n \pmod k.\end{cases}$$

This is valid for $\gcd(n,k)=1$. We take $m=p$ and $n=h$, where $\gcd(h,k)=1$, then multiply both members by $p^{-1}\log p$ and sum over all $p\le x$ to obtain

$$\sum_{p\le x}\sum_{r=1}^{\varphi(k)}\chi_r(p)\overline{\chi}_r(h)\frac{\log p}{p}=\varphi(k)\sum_{\substack{p\le x\\ p\equiv h\,(\mathrm{mod}\,k)}}\frac{\log p}{p}. \tag{30}$$

In the sum on the left we isolate those terms involving only the principal character χ_1 and rewrite (30) in the form

$$\varphi(k)\sum_{\substack{p\le x\\ p\equiv h\,(\mathrm{mod}\,k)}}\frac{\log p}{p}=\overline{\chi}_1(h)\sum_{p\le x}\frac{\chi_1(p)\log p}{p}+\sum_{r=2}^{\varphi(k)}\overline{\chi}_r(h)\sum_{p\le x}\frac{\chi_r(p)\log p}{p} \tag{31}$$

Now $\overline{\chi}_1(h)=1$ and $\chi_1(p)=0$ unless $\gcd(p,k)=1$, in which case $\chi_1(p)=1$. Hence the first term on the right of (31) is given by

$$\sum_{\substack{p\le x\\ \gcd(p,k)=1}}\frac{\log p}{p}=\sum_{p\le x}\frac{\log p}{p}-\sum_{\substack{p\le x\\ p|k}}\frac{\log p}{p}=\sum_{p\le x}\frac{\log p}{p}+O(1), \tag{32}$$

since there are only a finite number of primes which divide k. Combining (32) with (31) we obtain

$$\varphi(k) \sum_{\substack{p \le x \\ p \equiv h \,(\mathrm{mod}\, k)}} \frac{\log p}{p} = \sum_{p \le x} \frac{\log p}{p} + \sum_{r=2}^{\varphi(k)} \overline{\chi}_r(h) \sum_{p \le x} \frac{\chi_r(p) \log p}{p} + O(1). \tag{33}$$

Using (24) and dividing by $\varphi(k)$ we obtain Lemma 3.1. □

Proof of Lemma 3.2

We begin with the sum

$$\sum_{n\le x}\frac{\chi(n)\Lambda(n)}{n},$$

where $\Lambda(n)$ is Mangoldt's function, and express this sum in two ways. First we note that the definition of $\Lambda(n)$ gives us

$$\sum_{n\le x}\frac{\chi(n)\Lambda(n)}{n}=\sum_{p\le x}\sum_{\substack{a=1\\ p^a\le x}}^{\infty}\frac{\chi(p^a)\log p}{p^a}.$$

We separate the terms with $a=1$ and write

$$\sum_{n\le x}\frac{\chi(n)\Lambda(n)}{n}=\sum_{p\le x}\frac{\chi(p)\log p}{p}+\sum_{p\le x}\sum_{\substack{a=2\\ p^a\le x}}^{\infty}\frac{\chi(p^a)\log p}{p^a} \tag{34}$$

Since $|\chi(p^a)|\le 1$, it follows that the second sum on the right is majorized by

$$\sum_{p}\log p\sum_{a=2}^{\infty}\frac{1}{p^a}=\sum_{p}\frac{\log p}{p(p-1)}<\sum_{n=2}^{\infty}\frac{\log n}{n(n-1)}=O(1),$$

so (34) gives us

$$\sum_{p\le x}\frac{\chi(p)\log p}{p}=\sum_{n\le x}\frac{\chi(n)\Lambda(n)}{n}+O(1). \tag{35}$$

Now we recall that

$$\Lambda(n)=\sum_{d|n}\mu(d)\log\frac{n}{d},$$

hence

$$\sum_{n\le x}\frac{\chi(n)\Lambda(n)}{n}=\sum_{n\le x}\frac{\chi(n)}{n}\sum_{d|n}\mu(d)\log\frac{n}{d}.$$

In the last sum we write $n=cd$ and use the multiplicative property of χ to obtain

$$\sum_{n\le x}\frac{\chi(n)\Lambda(n)}{n}=\sum_{d\le x}\frac{\mu(d)\chi(d)}{d}\sum_{c\le x/d}\frac{\chi(c)\log c}{c}. \tag{36}$$

Since $x/d\ge 1$, in the sum over c we may use Equation (21) of Theorem 2.18 to obtain

$$\sum_{c\le x/d}\frac{\chi(c)\log c}{c}=-L'(1,\chi)+O\left(\frac{\log x/d}{x/d}\right).$$

Equation (36) now becomes

$$\sum_{n\le x}\frac{\chi(n)\Lambda(n)}{n}=-L'(1,\chi)\sum_{d\le x}\frac{\mu(d)\chi(d)}{d}+O\left(\sum_{d\le x}\frac{1}{d}\frac{\log x/d}{x/d}\right). \tag{37}$$

The sum in the O-term is

$$\frac{1}{x}\sum_{d\le x}(\log x-\log d)=\frac{1}{x}\left([x]\log x-\sum_{d\le x}\log d\right)=O(1)$$

since

$$\sum_{d\le x}\log d=\log[x]!=x\log x+O(x).$$

Therefore (37) becomes

$$\sum_{n\le x}\frac{\chi(n)\Lambda(n)}{n}=-L'(1,\chi)\sum_{d\le x}\frac{\mu(d)\chi(d)}{d}+O(1)$$

which, with (35), proves Lemma 3.2. □

Proof of Lemma 3.3

We use the Generalized Möbius Inversion Formula proved in Theorem 1.10 which states that if α is completely multiplicative we have

$$G(x) = \sum_{n \le x} \alpha(n) F\left(\frac{x}{n}\right) \quad \text{if and only if} \quad F(x) = \sum_{n \le x} \mu(n)\alpha(n) G\left(\frac{x}{n}\right). \tag{38}$$

We take $\alpha(n) = \chi(n)$ and $F(x) = x$ to obtain

$$x = \sum_{n \le x} \mu(n)\chi(n) G\left(\frac{x}{n}\right) \tag{39}$$

where

$$G(x) = \sum_{n \le x} \chi(n)\frac{x}{n} = x \sum_{n \le x} \frac{\chi(n)}{n}.$$

By Equation (20) of Theorem 2.18 we can write

$$G(x) = xL(1, \chi) + O(1).$$

Using this in (39) we find

$$x = \sum_{n \le x} \mu(n)\chi(n) \left(\frac{x}{n} L(1, \chi) + O(1)\right) = xL(1, \chi) \sum_{n \le x} \frac{\mu(n)\chi(n)}{n} + O(x).$$

Now we divide by x to obtain Lemma 3.3. □

Proof of Lemma 3.5

We prove Lemma 3.5 and then use it to prove Lemma 3.4. Once again we make use of the Generalized Möbius Inversion Formula (38). This time we take $F(x) = x \log x$ to obtain

$$x \log x = \sum_{n \leq x} \mu(n)\chi(n) G\left(\frac{x}{n}\right) \tag{40}$$

where

$$G(x) = \sum_{n \leq x} \chi(n) \frac{x}{n} \log \frac{x}{n} = x \log x \sum_{n \leq x} \frac{\chi(n)}{n} - x \sum_{n \leq x} \frac{\chi(n) \log n}{n}.$$

Now use Equations (20) and (21) of Theorem 2.18 to get

$$\begin{aligned} G(x) &= x \log x \left(L(1, \chi) + O\left(\frac{1}{x}\right)\right) - x \left(L'(1, \chi) + O\left(\frac{\log x}{x}\right)\right) \\ &= xL'(1, \chi) + O(\log x) \end{aligned}$$

since we are assuming $L(1, \chi) = 0$. Hence (40) gives us

$$\begin{aligned} x \log x &= \sum_{n \leq x} \mu(n)\chi(n) \left(\frac{x}{n} L'(1, \chi) + O\left(\log \frac{x}{n}\right)\right) \\ &= xL'(1, \chi) \sum_{n \leq x} \frac{\mu(n)\chi(n)}{n} + O\left(\sum_{n \leq x} (\log x - \log n)\right) \\ &= xL'(1, \chi) \sum_{n \leq x} \frac{\mu(n)\chi(n)}{n} + O(x), \end{aligned}$$

which when we divide by x we obtain Lemma 3.5. □

Proof of Lemma 3.4

We use Lemma 3.1 with $h = 1$ to get

$$\sum_{\substack{p \le x \\ p \equiv 1 \,(\mathrm{mod}\, k)}} \frac{\log p}{p} = \frac{1}{\varphi(k)} \log x + \frac{1}{\varphi(k)} \sum_{r=2}^{\varphi(k)} \sum_{p \le x} \frac{\chi_r(p) \log p}{p} + O(1). \tag{41}$$

In the sum over p on the right we use Lemma 3.2 which states that

$$\sum_{p \le x} \frac{\chi_r(p) \log p}{p} = -L'(1, \chi_r) \sum_{n \le x} \frac{\mu(n)\chi_r(n)}{n} + O(1).$$

If $L'(1, \chi_r) \neq 0$, Lemma 3.3 shows that the right member of the foregoing equation is $O(1)$. But if $L(1, \chi_r) = 0$ then Lemma 3.5 implies

$$-L'(1, \chi_r) \sum_{n \le x} \frac{\mu(n)\chi_r(n)}{n} = -\log x + O(1).$$

Therefore the sum on the right of (41) is

$$\frac{1}{\varphi(k)} \left(-N(k) \log x + O(1) \right),$$

so (41) becomes

$$\sum_{\substack{p \le x \\ p \equiv 1 \,(\mathrm{mod}\, k)}} \frac{\log p}{p} = \frac{1 - N(k)}{\varphi(k)} \log x + O(1).$$

This proves Lemma 3.4 and therefore also Theorem 3.1. □

As remarked earlier, Theorem 3.1 implies Dirichlet's Theorem.

Theorem 3.2 (**Dirichlet**). *If $k > 0$ and $\gcd(h, k) = 1$ there are infinitely many primes in the arithmetic progression $nk + h$, $n = 0, 1, 2, \ldots$.*

4 Miscellaneous Problems

Problem 1

Dirichlet's Theorem implies the following statement: If h and $k > 0$ are any two integers with $\gcd(h, k) = 1$, there exists at least one prime of the form $kn + h$. Prove that this statement also implies Dirichlet's Theorem.

Problem 2

Prove that for any positive integers n, N there are blocks of consecutive integers of length greater than N, with the property that each of their totients is divisible by n.

Problem 3

Prove that for any positive integer n an arithmetic progression exists in which the first two terms are primes, the first n terms are pairwise relatively prime and there are infinitely many primes in the progression.

Problem 4

Prove that if $f(x)$ is a polynomial with rational coefficients such that $f(p)$ is prime for every prime p, then either $f(x) = x$ for all x or $f(x)$ is the same prime constant for all x.

Problem 5

Let m and n be fixed integers greater than 1, n odd. Suppose n is a quadratic residue modulo p for all sufficiently large prime numbers $p \equiv -1 \pmod{2^m}$. Show that n is a square.

Problem 6

Find all positive integers N that are quadratic residues modulo all primes greater than N.

Problem 7

Let p_n be the nth prime number. For every N, prove that there exists a positive integer k such that both p_{k-1} and p_{k+1} are outside the interval $[\, p_k - N, p_k + N \,]$.

Problem 8

Construct an infinite set S of primes with the following property: If $p, q \in S$, then $\gcd(\frac{1}{2}(p-1), \frac{1}{2}(q-1)) = \gcd(p, q-1) = \gcd(p-1, q) = 1$.

Problem 9

Let $s(n)$ denote the smallest r such that -1 is a sum of r squares $\bmod n$. Show that $s(n)$ can be computed as follows:

$$s(n) = \begin{cases} 1 & \text{if } 4 \nmid n \text{ and } p \nmid n \text{ for all primes } p \equiv 3 \pmod 4, \\ 2 & \text{if } 4 \nmid n \text{ and } p \mid n \text{ for some prime } p \equiv 3 \pmod 4, \\ 3 & \text{if } 4 \mid n \text{ and } 8 \nmid n, \\ 4 & \text{if } 8 \mid n. \end{cases}$$

Problem 10

Let $b_1 < b_2 < b_3 < \cdots$ be distinct positive integers expressible as sums of two squares of integers. Prove that for any given positive integer d the equality $b_{n+1} - b_n = d$ holds for infinitely many n.

References

[1] T. Apostol, *Introduction to Analytic Number Theory*, Undergraduate Texts in Mathematics. Springer-Verlag, New York-Heidelberg-Berlin, 1976.

[2] E. Landau, *Handbuch der Lehre von der Verteilung der Primzahlen*, Leipzig: Teubner. Reprinted by Chelsea, 1953.

[3] H. N. Shapiro, On Primes in Arithmetic Progression II, *Ann. of Math.*, 52(1952) 231-243.