Elementary Proof of a Theorem of Bateman

MICHEL BALAZARD AND ABDELHAKIM SMATI

Dedicated to Paul T. Bateman

1. The distribution of values of the Euler totient function $\varphi(n)$ has been investigated from various points of view (*cf.* [7], [4], [2]) . In this paper, we shall study the asymptotic behaviour of the number $N(x)$ of those positive integers n which satisfy $\varphi(n) \leq x$.

The best result up-to-date is the following

Theorem (Bateman 1972, [1]). *For every constant* $c < 1/\sqrt{2}$,

$$N(x) = A\ x + O_c(x\ e^{-c(\log x\ \ \log\log x)^{1/2}}), \tag{1}$$

where $A = \frac{\zeta(2)\zeta(3)}{\zeta(6)} = \prod_p (1 + \frac{1}{p(p-1)})$.

Bateman's proof is analytic: it starts from Perron's integral formula and uses a simple estimate of $|\zeta(s)|$ in the strip $0 < \Re(s) < 1$. As for every result involving only natural numbers and the elementary functions of real analysis, one may ask for a proof using only these elements, and neither complex variables nor Fourier analysis.

The first step in this direction is due to Dressler (*cf.* [3]) , who proved by elementary means in 1970 that $N(x) \sim Ax$. In 1984, Nicolas obtained the elementary error term $O(x/\log x)$ (*cf.* [6]). His starting point is the study of the weighted sum $\sum_{\varphi(n)\leq x} \log \varphi(n)$ and lends itself to generalization. The idea is to estimate the sum $\sum_{\varphi(n)\leq x} \log^k \varphi(n)$ in order to get the error term $O(x(\log x)^{-k})$ in (2).

This generalization has been studied, and completely worked out for $k = 2$, by the second author (*cf.* [8]). Nevertheless, the computations are so intricate that a new approach is needed to go further. During the conference, we presented an elementary proof of a result slightly weaker than (1), namely

$$N(x) = A\ x + O(x\ e^{-c_0(\log x)^{1/2}}), \tag{2}$$

where c_0 is some constant (positive, absolute and computable).

After our talk, G. Tenenbaum convinced us that obtaining Bateman's result by our method would demand only a few more lines of computation. As it seldom occurs that an elementary result reaches the degree of accuracy of the best known analytic one, we will follow Tenenbaum's advice and present a proof of (1) by elementary means.

2. Our starting point is the idea of Dressler. In his proof of $N(x) \sim Ax$, he approximated the Euler function by the truncated function

$$\varphi(n,y) = n \prod_{\substack{p \mid n \\ p \le y}} (1 - \frac{1}{p}), \quad y \ge 2.$$

Denote by $N(x,y)$ the number of positive integers n such that $\varphi(n,y) \le x$. We first give simple inequalities involving $N(x)$ and $N(x,y)$.

Lemma 1. *If x is large enough and $y > 3\log x$, then*

$$N(x,y) \le N(x) \le N(x(1 - \frac{3\log x}{y})^{-1},\ y). \tag{3}$$

Proof: The first inequality follows from $\varphi(n,y) \ge \varphi(n)$. For the second one, observe that the number $\omega(n)$ of prime divisors of n satisfies $\omega(n) \le \log n/\log 2 \le 2\log n$ and that $(1-v)^\alpha \ge 1 - \alpha v$ if $\alpha \ge 1$ and $0 \le v \le 1$. Hence

$$\begin{aligned}\varphi(n) &= \varphi(n,y) \prod_{\substack{p \mid n \\ p > y}} (1 - \frac{1}{p}) \\ &\ge \varphi(n,y)(1 - \frac{1}{y})^{\omega(n)} \\ &\ge \varphi(n,y)(1 - \frac{2\log n}{y}).\end{aligned}$$

If $\varphi(n) \le x$, the classical inequality $\varphi(n) \gg n(\log\log n)^{-1}$ shows that $n \ll x(\log\log x)$ and

$$x \ge \varphi(n) \ge \varphi(n,y)(1 - \frac{3\log x}{y})$$

if x is large enough.

3. In order to estimate $N(x,y)$, one writes $n = ab$ where, here and throughout this paper, a (resp. b) denotes a generic positive integer whose prime divisors p all satisfy $p \le y$ (resp. $p > y$) . One has $\varphi(n,y) = \varphi(a)b$. Since $b = 1$ or $b > y$, one gets

$$N(x,y) = \sum_{\varphi(a)\le x} 1 + \sum_{\varphi(a)\le x/y} \sum_{1<b\le x/\varphi(a)} 1 \qquad (4)$$
$$= \sum_{\varphi(a)\le x/y} \sum_{1\le b\le x/\varphi(a)} 1 + O(\sum_{\varphi(a)\le x} 1).$$

The estimation of these sums depends on the following three lemmas.

Lemma 2. *Suppose y and k are real numbers so that $y \ge 2$ and $0 \le k \le \frac{1}{3}\log y$. Put $\sigma = 1 - k/\log y$. Then*

$$\sum_a \varphi(a)^{-\sigma} \ll \log y \; e^{O(e^k)}.$$

Proof: The sum $\sum_a \varphi(a)^{-\sigma}$ equals the Euler product

$$\prod_{p\le y}(1 + \frac{1}{(p-1)^\sigma}(1 + \frac{1}{p^\sigma} + \frac{1}{p^{2\sigma}} + \cdots))$$
$$= \prod_{p\le y}(1 + \frac{1}{(p-1)^\sigma} - \frac{1}{p^\sigma}) \prod_{p\le y}(1 + \frac{1}{p^\sigma} + \frac{1}{p^{2\sigma}} + \cdots)$$
$$\ll e^{\sum_{p\le y} p^{-\sigma}} \qquad \text{since } \sigma \ge \frac{2}{3}.$$

Now, since the function $(e^t - 1)/t$ increases with $t > 0$, we have

$$\sum_{p\le y} p^{-\sigma} = \sum_{p\le y} p^{-1} + \sum_{p\le y}(p^{1-\sigma} - 1)p^{-1}$$
$$\le \sum_{p\le y} p^{-1} + (e^k - 1)(\log y)^{-1} \sum_{p\le y} p^{-1}\log p$$
$$\le \log\log y + O(e^k),$$

and the Lemma follows.

Lemma 3. *Suppose $x \ge y \ge 2$ are real numbers and let $u = \log x/\log y$. Then*

$$\sum_{\varphi(a)\le x} 1 \le x \log y \; e^{-u\log u + O(u)},$$

provided that $u \le y^{1/3}$.

Proof: This is a typical application of the by now classical Rankin method. We have, for every positive σ,

$$\sum_{\varphi(a)\le x} 1 \le x^\sigma \sum_a \varphi(a)^{-\sigma}.$$

Choosing $\sigma = 1 - \log u/\log y$ and using Lemma 2 gives the result.

Lemma 4. *With the notations of Lemma 3 one has for every positive* ϵ

$$\sum_{b\le x} 1 = x \prod_{p\le y} (1-\frac{1}{p})(1+O_\epsilon(e^{-(1-\epsilon)u\log u}))$$

provided that $\log y > \sqrt{\log x}$ *(i.e.* $u < \log y$*).*

Proof: Although not stated explicitly by Halberstam and Richert, this is an easy consequence of their proof of the Fundamental Lemma of Brun's Sieve given in [5], pp. 82-83.

4. We now come back to (4). Suppose that x is large enough and that $\log y > \sqrt{\log x}$. Let ϵ be a positive real number. By Lemma 3, the error term in (4) is $O_\epsilon(x \log y \; e^{-(1-\epsilon)u\log u})$.

Moreover, by Lemma 4

$$\sum_{\varphi(a)\le x/y} \; \sum_{1\le b\le x/\varphi(a)} 1 \tag{5}$$
$$= x\prod_{p\le y}(1-\frac{1}{p}) \sum_{\varphi(a)\le x/y} \frac{1}{\varphi(a)}(1+O_\epsilon(e^{-(1-\epsilon)u_a\log u_a})),$$

where $u_a = \log(x/\varphi(a))/\log y$.

The contribution of the main terms in (5) amounts to

$$x\prod_{p\le y}(1-\frac{1}{p})\left(\sum_a \frac{1}{\varphi(a)} - \sum_{\varphi(a)>x/y}\frac{1}{\varphi(a)}\right). \tag{6}$$

By Rankin's method,

$$\sum_{\varphi(a)>x/y} \frac{1}{\varphi(a)} \le y^\sigma x^{-\sigma} \sum_a \frac{1}{\varphi(a)^{1-\sigma}} \quad (\sigma > 0).$$

With $\sigma = \log u/\log y$ this is , by Lemma 2,

$$\ll u e^{-u\log u} \log y \; e^{O(u)}$$
$$\ll_\epsilon \log y \; e^{-(1-\epsilon)u\log u}.$$

Thus (6) can be written as

$$x\prod_{p\le y}(1+\frac{1}{p(p-1)}) + O_\epsilon(x e^{-(1-\epsilon)u\log u}).$$

We now turn to the contribution of the error term in (5). Observe that

$$u\log u - u_a\log u_a = \int_{u_a}^{u}(\log v + 1)dv$$
$$\le (\log u + 1)\frac{\log\varphi(a)}{\log y}.$$

Hence

$$\sum_{\varphi(a)\le \frac{x}{y}} \varphi(a)^{-1} e^{-(1-\epsilon)u_a\log u_a}$$
$$\le e^{-(1-\epsilon)u\log u}\sum_a \varphi(a)^{-1+\frac{\log u+1}{\log y}}$$
$$\ll_\epsilon \log y \; e^{-(1-2\epsilon)u\log u} \qquad \text{by Lemma 2.}$$

We summarize these computations in a Lemma.

Lemma 5. *Suppose $x \ge y > e^{\sqrt{\log x}}$ and x large enough. Then*

$$N(x,y) = x\prod_{p\le y}(1+\frac{1}{p(p-1)}) + O_\epsilon(x\log y \; e^{-(1-\epsilon)u\log u})$$

for every positive ϵ, where $u = \log x/\log y$.

5. We are now in a position to use Lemma 1. First, we observe that $\prod_{p\le y}(1+\frac{1}{p(p-1)}) = A + O(\frac{1}{y})$. Then, if x is large enough, $y > 4\log x$ and $\log y > [\log(x(1-\frac{3\log x}{y})^{-1})]^{1/2}$, Lemmas 1 and 5 give

$$N(x) = A\; x + O(x\; y^{-1}\log x) + O_\epsilon(x\; \log y\; e^{-(1-\epsilon)u\log u}).$$

We choose $y = e^{\sqrt{\frac{1}{2}\log x\;\log\log x}}$, so that

$$u\log u = (1+o(1))\sqrt{\frac{1}{2}\log x\;\log\log x},$$

and this completes the proof of (1).

In conclusion, let us point out that the prime number theorem is not required in our proof (neither in Bateman's proof, by the way). This is in contrast to [6] and [8], where the prime number theorem with remainder term is an essential tool.

REFERENCES

[1] P. T. Bateman, The distribution of values of Euler's function, Acta Arith. **21** (1972), 329–345.
[2] H.G. Diamond, The distribution of values of Euler's phi function, Proc. Symp. Pure Maths, AMS **24** (1973), 63–76.
[3] R.E. Dressler, A density which counts multiplicity, Pacific J. Math. **34** (1970), 371–378.
[4] P.Erdös, Some remarks on Euler's φ-function and some related problems, Bull. Amer. Math. Soc. **51** (1945), 540–544.
[5] H. Halberstam, H.-E. Richert, *Sieve Methods*, Academic Press, 1974.
[6] J.-L. Nicolas, Distribution des valeurs de la fonction d' Euler, L' Ens. Math. **30** (1984), 331–338.
[7] I.J.Schoenberg, Über die asymptotische Verteilung reeler Zahlen mod. 1, Math. Z. **28** (1928), 171–199.
[8] A. Smati, Répartition des valeurs de la fonction d' Euler, L' Ens. Math. **35** (1989), 61–76.

Michel Balazard and Abdelhakim Smati
Département de Mathématiques
Faculté des Sciences
123 Av. A. Thomas
87060 Limoges Cedex
France