Additive and multiplicative functions with similar global behavior

Jean-Marie De Koninck and Nicolas Doyon

Édition du 16 avril 2015

Abstract

We examine the question of whether, given an additive function f with a limit distribution, one can find a multiplicative function g with the same limit distribution. We show that if an additive function f has a constant asymptotic mean and constant asymptotic variance, one can construct a multiplicative function g with the same properties. It is known that, when $f = \omega$, where $\omega(n)$ stands for the number of distinct prime factors of n, with $\omega(1) = 0$, both the asymptotic mean and variance of $f(n)$ are of the same order, namely $\log\log n$, but we show that no multiplicative function $g(n)$ can have the same mean and variance as $\omega(n)$.

AMS Subject Classification numbers: 11A25, 11N37
Key words: arithmetic function, mean value, variance.

1 Introduction

In an earlier paper, De Koninck, Doyon and Letendre [2] studied the question of how close an additive function can be to a multiplicative function. More precisely, given an additive function f and a multiplicative function g, they investigated the size of $E(f, g; x) := \#\{n \le x : f(n) = g(n)\}$. In the particular case $f = \omega$, where $\omega(n)$ stands for the number of distinct prime factors of n, with $\omega(1) = 0$, they established that given any $\varepsilon > 0$, then $E(\omega, g; x) \gg x/(\log\log x)^{1+\varepsilon}$ for an appropriate choice of g, but that, given any multiplicative function g, then $E(\omega, g; x) = o(x)$ as $x \to \infty$.

In this paper, we examine how close additive functions can be to multiplicative ones, but this time globally. For instance, we ask if it is possible to construct an additive function f and a multiplicative function g that have the same global behavior, namely the same asymptotic mean, variance or limit distribution.

For convenience, let us write $\mathcal{A}$ for the set of all additive functions f such that $f(1) = 0$ and $\mathcal{M}$ for the set of all multiplicative functions g such that $g(1) = 1$. A *strongly additive function* (resp. *strongly multiplicative function*) h is a function in $\mathcal{A}$ (resp. in $\mathcal{M}$) such that $h(p^\alpha) = h(p)$ for all integers $\alpha \ge 1$ and all primes p. We shall write $\mathcal{A}^*$ (resp. $\mathcal{M}^*$) for the set of strongly additive functions (resp. strongly multiplicative functions). We will use $P(n)$ to denote the largest prime factor of the integer $n \ge 2$, setting for convenience $P(1) = 1$. The letters p and q with or without subscript will always denote prime numbers.

The distribution of additive and multiplicative functions has been studied in great depth by several authors, namely Daboussi [1], Galambos [5], Levin and Timofeev [6] and many more. Actually, one can find a large variety of such results in the books of Elliott [4]. In particular, the famous Erdős-Wintner theorem (see for instance Theorem III-4.1 in the book of Tenenbaum [7]) gives necessary and sufficient conditions for a real valued additive function to have a limiting distribution. Despite this, the following natural question does not seem to have been raised before: "Given a real valued additive function f which admits a limiting distribution, can one construct a multiplicative function g with the same limiting distribution?" Given that this question is indeed very difficult to study in its generality, we restrict ourselves to the study of the first two moments of the distribution, namely the mean and the variance.

We say that an arithmetic function h has an *asymptotic mean value* $M(h)$ if the limit

$$M(h) := \lim_{x\to\infty} \frac{1}{x} \sum_{n\le x} h(n)$$

exists.

If an arithmetic function h has an asymptotic mean value $M(h)$, we say that it has an *asymptotic variance* $V(h)$ if the limit

$$V(h) := \lim_{x\to\infty} \frac{1}{x} \sum_{n\le x} (h(n) - M(h))^2$$

exists.

We will prove that one can construct additive functions with arbitrary mean and variance. In other words, given any real numbers a and b, we will show how to construct an additive function with mean value equal to a and variance equal to b. Furthermore, we will show that the same is true for multiplicative functions if $a > 0$.

2 Additive and multiplicative function functions with the same limit distribution

In some very simple instances, one can construct multiplicative and additive functions which have the same limit distribution. First, consider the following example. Let p_0 be a fixed prime number and let f be an additive function defined by $f(p_0^a) = 1$ for each positive integer a and by $f(q^b) = 0$ for each prime $q \neq p_0$ and each positive integer b. Clearly the limit distribution of f is given by

$$\lim_{x\to\infty} \frac{1}{x} \#\{n \le x : f(n) = 1\} = \frac{1}{p_0}$$

and

$$\lim_{x\to\infty} \frac{1}{x} \#\{n \le x : f(n) = 0\} = \frac{p_0 - 1}{p_0}.$$

In order to construct a multiplicative function g with the same limiting distribution, we let T be an infinite set of primes such that

$$\prod_{q\in T}\left(1-\frac{1}{q}\right)=\frac{1}{p_0}.$$

We then define the multiplicative function g on prime powers q^α by

$$g(q^\alpha)=\begin{cases} 0 & \text{if} \quad q\in T,\\ 1 & \text{if} \quad q\notin T,\end{cases}$$

which yields

$$\lim_{x\to\infty}\frac{1}{x}\#\{n\le x: g(n)=0\}=1-\prod_{q\in T}\left(1-\frac{1}{q}\right)=1-\frac{1}{p_0}$$

and

$$\lim_{x\to\infty}\frac{1}{x}\#\{n\le x: g(n)=1\}=1-\lim_{x\to\infty}\frac{1}{x}\#\{n\le x: g(n)=0\}=1-\frac{p_0-1}{p_0}=\frac{1}{p_0},$$

as requested.

We can construct a slightly more complex example. Let p_1 and p_2 be two distinct prime numbers. Let f be an additive function defined by $f(p_1^\alpha)=1$, $f(p_2^\alpha)=1$ and $f(q^\alpha)=0, q\neq p_1,p_2$. The limit distribution of this function f is given by

$$\lim_{x\to\infty}\frac{1}{x}\#\{n\le x: f(n)=0\}=\left(1-\frac{1}{p_1}\right)\left(1-\frac{1}{p_2}\right)=\frac{p_1p_2-p_1-p_2+1}{p_1p_2},$$

$$\lim_{x\to\infty}\frac{1}{x}\#\{n\le x: f(n)=1\}=\left(1-\frac{1}{p_1}\right)\frac{1}{p_2}+\left(1-\frac{1}{p_2}\right)\frac{1}{p_1}=\frac{p_1+p_2-2}{p_1p_2}$$

and

$$\lim_{x\to\infty}\frac{1}{x}\#\{n\le x: f(n)=2\}=\frac{1}{p_1p_2}.$$

In order to construct a multiplicative function g with the same limit distribution we choose a prime p_3 and an infinite set of primes T which does not contain p_3 and such that $\sum_{q\in T}\frac{1}{q}<\infty$. We then define g on prime powers q^α as follows:

$$g(q^\alpha)=\begin{cases} 2 & \text{if} \quad q=p_3,\\ 0 & \text{if} \quad q\in T,\\ 1 & \quad\quad \text{otherwise.}\end{cases}$$

We further set

$$t:=\prod_{q\in T}\left(1-\frac{1}{q}\right).$$

The limit distribution of g is thus given by

$$\lim_{x\to\infty} \frac{1}{x}\#\{n \le x : g(n) = 0\} = 1 - t,$$

$$\lim_{x\to\infty} \frac{1}{x}\#\{n \le x : g(n) = 1\} = t\left(\frac{p_3 - 1}{p_3}\right)$$

and

$$\lim_{x\to\infty} \frac{1}{x}\#\{n \le x : g(n) = 2\} = t\left(\frac{1}{p_3}\right).$$

The distribution of g will be the same as that of f if and only if $p_3 = p_1 + p_2 - 1$ and $t = \dfrac{p_1 + p_2 - 1}{p_1 p_2}$. It is not clear if one can construct other pairs of additive and multiplicative functions with the same limit distribution. In any event, the method used in the above examples cannot be straightforwardly generalized. We believe this challenge in itself could lead to interesting investigations.

3 Additive and multiplicative functions with same mean and variance

Theorem 1. *Given a real number a and a positive real number b and any $\varepsilon > 0$, there exists $f \in \mathcal{A}^*$ such that $M(f) = a$ and $|V(f) - b| < \varepsilon$. Moreover, if a is positive, there exists $g \in \mathcal{M}^*$ such that $|M(g) - a| < \varepsilon$ and $|V(g) - b| < \varepsilon$.*

Proof. Let $\mathcal{S}$ be a finite set of primes whose properties will be revealed later. If f is a strongly additive function such that $f(p) \neq 0$ only if $p \in \mathcal{S}$, then it follows that

$$\begin{aligned} \frac{1}{x}\sum_{n\le x} f(n) &= \frac{1}{x}\sum_{n\le x}\sum_{\substack{p\mid n\\ p\in\mathcal{S}}} f(p) = \frac{1}{x}\sum_{\substack{p\le x\\ p\in\mathcal{S}}} f(p)\left\lfloor\frac{x}{p}\right\rfloor \\ &= \sum_{\substack{p\le x\\ p\in\mathcal{S}}}\frac{f(p)}{p} + \frac{1}{x}\sum_{\substack{p\le x\\ p\in\mathcal{S}}} f(p)\left(\left\lfloor\frac{x}{p}\right\rfloor - \frac{x}{p}\right), \end{aligned}$$

from which we deduce that, if x is sufficiently large,

$$\frac{1}{x}\sum_{n\le x} f(n) = \sum_{p\in\mathcal{S}}\frac{f(p)}{p} + O\left(\frac{1}{x}\right),$$

which implies that the mean value of f exists is given by

$$M(f) = \sum_{p\in\mathcal{S}}\frac{f(p)}{p}. \qquad (3.1)$$

On the other hand, we can show that the variance $V(f)$ of f is given by

$$V(f) = \sum_{p \in \mathcal{S}} \frac{f(p)}{p}\left(1 - \frac{1}{p}\right). \tag{3.2}$$

The proof of (3.2) goes as follows. It is well known that

$$V(f) = \lim_{x \to \infty} \frac{1}{x} \sum_{n \le x} f(n)^2 - M(f)^2 \tag{3.3}$$

Let $P = p_1 p_2 \cdots p_r$. Then we have

$$\begin{aligned} \frac{1}{x}\sum_{n \le x} f(n)^2 &= \frac{1}{x}\sum_{i=1}^{r} \sum_{\substack{p_i m \le x \\ (m, P/p_i)=1}} f(p_i)^2 + \frac{1}{x} \sum_{1 \le i < j \le r} \sum_{\substack{p_i p_j m \le x \\ (m, P/p_i p_j)=1}} (f(p_i) + f(p_j))^2 \\ &\quad + \cdots + \frac{1}{x} \sum_{p_1 p_2 \cdots p_r m \le x} (f(p_1) + \cdots + f(p_r))^2. \end{aligned} \tag{3.4}$$

Expanding the above and letting $x \to \infty$, we find terms of the form $f(p_i)^2$ and terms of the form $f(p_i)f(p_j)$ with $i \ne j$. First consider the terms of the form $f(p_i)^2$. It is easy to see that the coefficient of such a term is

$$\frac{1}{p_1 p_2 \cdots p_r} \prod_{\substack{j=1 \\ j \ne i}}^{r} (1 + (p_j - 1)) = \frac{1}{p_1 p_2 \cdots p_r} \prod_{\substack{j=1 \\ j \ne i}}^{r} p_j = \frac{1}{p_i}. \tag{3.5}$$

On the other hand, in light of (3.1), we have

$$M(f)^2 = \left(\frac{f(p_1)}{p_1} + \cdots + \frac{f(p_r)}{p_r} \right)^2. \tag{3.6}$$

Therefore, using this in (3.3) and in light of expression (3.5), we find that the terms of the form $f(p_i)^2$ along with their coefficients amount in the expansion of $V(f)$ to

$$\sum_{i=1}^{r} \left(\frac{f(p_i)^2}{p_i} - \frac{f(p_i)^2}{p_i^2} \right) = \sum_{i=1}^{r} \frac{f(p_i)^2}{p_i} \left(1 - \frac{1}{p_i}\right).$$

Hence, if we can show that the terms of the form $f(p_i)f(p_j)$ coming from (3.4) are cancelled by those coming from the expansion of (3.6), the proof of (3.2) will be complete. To do so, observe that expanding the right hand side of (3.6), we find that the terms of the form $f(p_i)f(p_j)$ along with their coefficients appear as $2\dfrac{f(p_i)f(p_j)}{p_i p_j}$, while we obtain exactly the same expression by expanding each term in (3.4) and summing up, thus proving our claim and completing the proof of (3.2).

Set $t := a^2/b$ and $r := \dfrac{a}{\sum_{p\in\mathcal{S}} 1/p}$. Then choose the elements of $\mathcal{S}$ in such a way that $\mathcal{S}$ satisfies the two properties

$$\left| \left(\sum_{p\in\mathcal{S}} \frac{1}{p} \right)^{-1} - \frac{1}{t} \right| < \frac{\varepsilon}{2} \qquad \text{and} \qquad \sum_{p\in\mathcal{S}} \frac{1}{p^2} < \frac{\varepsilon}{2r^2}. \tag{3.7}$$

One way to construct such a set S is as follows. As the smallest element of $\mathcal{S}$, we choose a prime number p_0 which is sufficiently large as to satisfy the two inequalities

$$\begin{aligned} \frac{1}{p_0 - 1} &< \frac{2\varepsilon t^2}{9a^2}, \\ \frac{1}{p_0} &< t. \end{aligned} \tag{3.8}$$

For the next element of $\mathcal{S}$, we choose a prime $p_1 > p_0$ satisfying the condition $\frac{1}{p_0} + \frac{1}{p_1} < t$. Having chosen $p_0, p_1, \ldots, p_{n-1} \in \mathcal{S}$, we choose a prime $p_n > p_{n-1}$ satisfying

$$\frac{1}{p_0} + \frac{1}{p_1} + \cdots + \frac{1}{p_n} < t. \tag{3.9}$$

We stop this iteration as soon as the condition

$$t - \sum_{j=0}^{n} \frac{1}{p_j} < \frac{\varepsilon t^2}{3} \tag{3.10}$$

is satisfied, and then set $\mathcal{S} = \{p_0, p_1, \ldots, p_n\}$. Observe that condition (3.10) implies that

$$\left| \frac{1}{\sum_{p\in\mathcal{S}} \frac{1}{p}} - \frac{1}{t} \right| = \left| \frac{t - \sum_{p\in\mathcal{S}} \frac{1}{p}}{t \sum_{p\in\mathcal{S}} \frac{1}{p}} \right| < \left| \frac{\varepsilon t^2}{3t(t - \frac{\varepsilon t^2}{3})} \right|. \tag{3.11}$$

Now, assuming that $\varepsilon t < 1$ (we can always do so by choosing ε even smaller), we have

$$t - \frac{\varepsilon t^2}{3} > \frac{2t}{3} \qquad \text{and thus} \qquad 3t\left(t - \frac{\varepsilon t^2}{3}\right) > 2t^2,$$

thereby implying that

$$\left| \frac{\varepsilon t^2}{3t(t - \frac{\varepsilon t^2}{3})} \right| < \frac{\varepsilon}{2},$$

which combined with (3.11) yields the first inequality of (3.7).

In order to establish the second inequality of (3.7), we only need to observe that, using (3.8) and (3.9) along with the fact that $2t/3 < \sum_{p\in\mathcal{S}} \frac{1}{p}$, we have

$$\sum_{p\in\mathcal{S}} \frac{1}{p^2} < \sum_{p\geq p_0} \frac{1}{p^2} \leq \frac{1}{p_0 - 1} \leq \frac{2\varepsilon}{9} \frac{t^2}{a^2} = \frac{2\varepsilon}{9} \frac{9}{4} \frac{((2/3)t)^2}{a^2} < \frac{\varepsilon}{2} \frac{\left(\sum_{p\in\mathcal{S}} \frac{1}{p}\right)^2}{a^2} = \frac{\varepsilon}{2r^2}.$$

Now let $f(p) = r$ for each $p \in \mathcal{S}$. It follows immediately from (3.1) that

$$M(f) = r \sum_{p \in \mathcal{S}} \frac{1}{p} = a.$$

On the other hand, from (3.2) we obtain that

$$V(f) = r^2 \sum_{p \in \mathcal{S}} \frac{1}{p} - r^2 \sum_{p \in \mathcal{S}} \frac{1}{p^2} = \frac{a^2}{\sum_{p \in \mathcal{S}} \frac{1}{p}} - r^2 \sum_{p \in \mathcal{S}} \frac{1}{p^2}. \tag{3.12}$$

From the conditions imposed on the set $\mathcal{S}$ in (3.7), we have that

$$r^2 \sum_{p \in \mathcal{S}} \frac{1}{p^2} < \frac{\varepsilon}{2} \tag{3.13}$$

and

$$\left| \frac{a^2}{\sum_{p \in \mathcal{S}} \frac{1}{p}} - b \right| = a^2 \left| \frac{1}{\sum_{p \in \mathcal{S}} \frac{1}{p}} - \frac{1}{t} \right| < \frac{\varepsilon}{2}. \tag{3.14}$$

Gathering (3.12), (3.13) and (3.14) completes the proof of the first part of the theorem.

We now deal with the second part of the theorem. Let $\mathcal{T}$ be a finite set of primes whose elements will be revealed later. Consider a strongly multiplicative function g such that $g(p) \neq 1$ only if p belongs to $\mathcal{T}$. We then have, for any fixed $s > 1$ and letting ζ stand for the Riemann Zeta Function,

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{g(n)}{n^s} &= \prod_p \left(1 + \frac{g(p)}{p^s - 1}\right) = \zeta(s) \prod_p \left(1 - \frac{1}{p^s}\right)\left(1 + \frac{g(p)}{p^s - 1}\right) \\ &= \zeta(s) \prod_p \left(1 + \frac{p^s g(p) - g(p) - p^s + 1}{p^s(p^s - 1)}\right) = \zeta(s) \prod_p \left(1 + \frac{g(p) - 1}{p^s}\right) \\ &= \zeta(s) \sum_{n=1}^{\infty} \frac{h(n)}{n^s}, \end{aligned}$$

say. Observing that

$$\sum_{n=1}^{\infty} \frac{h(n)}{n} = \prod_{p \in \mathcal{T}} \left(1 + \frac{g(p) - 1}{p}\right)$$

converges absolutely (being a finite product), one can use Theorem 6.13 in the book of De Koninck and Luca [3] and conclude that the mean value of g and also of g^2 are given by

$$M(g) = \prod_{p \in \mathcal{T}} \left(1 + \frac{g(p) - 1}{p}\right) \quad \text{and} \quad M(g^2) = \prod_{p \in \mathcal{T}} \left(1 + \frac{g^2(p) - 1}{p}\right). \tag{3.15}$$

In order to prove that

$$|M(g) - a| < \varepsilon, \tag{3.16}$$

we only need to prove that

$$\left| \sum_{p \in \mathcal{T}} \log \left(1 + \frac{g(p) - 1}{p} \right) - \log a \right| < \frac{\varepsilon}{2a} \tag{3.17}$$

for all positive $\varepsilon < a$. To show this, we will prove that if x and y are positive real numbers, then the fact that $|\log x - \log y| < \frac{\varepsilon}{2y}$ with $\varepsilon < y$ does indeed imply that $|x - y| < \varepsilon$. But this follows immediately from the following chain of deductions

$$\begin{aligned}
-\frac{\varepsilon}{2y} &< \log \left(\frac{x}{y} \right) < \frac{\varepsilon}{2y}, \\
\exp \left(-\frac{\varepsilon}{2y} \right) &< \exp \left(\log \left(\frac{x}{y} \right) \right) < \exp \left(\frac{\varepsilon}{2y} \right), \\
\exp \left(-\frac{\varepsilon}{2y} \right) - 1 &< \frac{x}{y} - 1 < \exp \left(\frac{\varepsilon}{2y} \right) - 1, \\
-\frac{\varepsilon}{2y} &< \frac{x}{y} - 1 < \frac{\varepsilon}{y}, \\
-\frac{\varepsilon}{2} &< x - y < \varepsilon,
\end{aligned}$$

where we used the inequality $e^t - 1 < 2t$ which holds for all $t \in (0, 1/2)$ and in particular for $t = \varepsilon/y$. Now observe that, by definition,

$$V(g) = M((g - M(g))^2) = M(g^2) - (M(g))^2 = M(g^2) - a^2. \tag{3.18}$$

It follows from this that the condition $|V(g) - b| < \varepsilon$ is equivalent to

$$|M(g^2) - (a^2 + b)| < \varepsilon. \tag{3.19}$$

Arguing as we did above, in order to prove (3.19), we only need to prove, in light of (3.15), that

$$\left| \sum_{p \in \mathcal{T}} \log \left(1 + \frac{g^2(p) - 1}{p} \right) - \log(a^2 + b) \right| < \frac{\varepsilon}{2(a^2 + b)}. \tag{3.20}$$

Set $u := \dfrac{(\log a)^2}{\log(a^2 + b) - 2 \log a}$ and $v := \dfrac{\log(a^2 + b)}{\log a} - 1$. Observe that the denominator of u is nonzero (in fact, positive) and that $v > 1$, all because $b > 0$. We then choose the elements of $\mathcal{T}$ in such a way that $\mathcal{T}$ satisfies the following three conditions:

(i) $\displaystyle \left|\sum_{p\in\mathcal{T}} \frac{1}{p} - u\right| < \min\left(\frac{\varepsilon}{4(v-1)a}, \frac{\varepsilon}{4(v^2-1)(a^2+b)}\right)$;

(ii) $\displaystyle \left|\sum_{p\in\mathcal{T}} \left(\frac{v-1}{p} - \log\left(1+\frac{v-1}{p}\right)\right)\right| < \frac{\varepsilon}{4a}$;

(iii) $\displaystyle \left|\sum_{p\in\mathcal{T}} \left(\frac{v^2-1}{p} - \log\left(1+\frac{v^2-1}{p}\right)\right)\right| < \frac{\varepsilon}{4(a^2+b)}$.

Then, making use of (ii) and then (i), we obtain that

$$\begin{aligned}
\left|\sum_{p\in\mathcal{T}} \log\left(1+\frac{v-1}{p}\right) - \log a\right| &< \left|\sum_{p\in\mathcal{T}} \frac{v-1}{p} - \log a\right| + \frac{\varepsilon}{4a} \\
&< |u(v-1) - \log a| + \frac{\varepsilon}{4a} + \frac{\varepsilon}{4a} \\
&= \frac{\varepsilon}{2a},
\end{aligned}$$

where we used the fact that $u(v-1) = \log a$. Hence, setting $g(p) = 1$ if $p \in \mathcal{T}$ and $g(p) = v$ if $p \notin \mathcal{T}$, inequality (3.17) follows, which implies (3.16), thus establishing the first conclusion of the second part of the theorem.

We now turn to the estimation of $V(g)$. Making use of (iii) and then (i), we easily obtain that

$$\left|\sum_{p\in\mathcal{T}} \log\left(1+\frac{v^2-1}{p}\right) - u(v^2-1)\right| < \frac{\varepsilon}{2(a^2+b)}.$$

Given that $u(v^2-1) = \log(a^2+b)$, we have thus proved (3.20) and therefore, recalling (3.18), that

$$|M(g^2) - (a^2+b)| < \varepsilon,$$

which, because of (3.18), establishes that

$$|V(g) - b| < \varepsilon,$$

as requested. □

Observe that Theorem 1 implies in particular that if an additive function has finite mean and finite variance, there exists a strongly multiplicative function with the same mean and variance.

In what follows, we consider what happens when the mean $M_f(x) := \dfrac{1}{x}\displaystyle\sum_{n\le x} f(n)$ and variance $V_f(x) := \dfrac{1}{x}\displaystyle\sum_{n\le x} (f(n) - M_f(x))^2$ of an additive function f are both slowly increasing functions of x.

4 Additive and multiplicative functions whose mean value behaves like $\log\log n$

In the previous section, we showed that if an additive function f has finite mean and variance, one can construct a multiplicative function with the same mean and variance. We now investigate whether this can also be done if the mean value of f is a function tending to infinity. For example, consider the classical additive function $f = \omega$. It is well known that $f(n)$ has both an asymptotic mean and variance of $\log\log n$. In this section, we construct a multiplicative function g with same asymptotic mean value while in the next section we show that if the mean value of a multiplicative $g(n)$ is of order $\log\log n$, then its variance is necessarily much larger. But first we construct a multiplicative function g such that

$$\sum_{n\le x} g(n) = x\log\log x + O(x\log\log\log x). \tag{4.1}$$

Observe that a more precise estimate is known for $\sum_{n\le x}\omega(n)$, namely

$$\sum_{n\le x}\omega(n) = x\log\log x + \beta x + O\left(\frac{x}{\log x}\right), \tag{4.2}$$

where $\beta \approx 0.2644$ (see for instance Section 6.7 in the book of De Koninck and Luca [3]).

A consequence of (4.1) and (4.2) is that $g(n)$ and $\omega(n)$ have the same mean value.

4.1 A glimpse at a truncated function

Before we exhibit a multiplicative function g satisfying (4.1), we introduce a truncated strongly multiplicative function of special interest.

Let $2 = p_1 < p_2 < \cdots$ stand for the sequence of all primes. Let g stand for the strongly multiplicative function defined on the primes p by $g(p) = 0$ if $p = 2$ and on all other primes p_r, for $r \ge 2$, by

$$g(p_r) = 1 + \frac{1}{\sum_{1\le k<r} 1/p_k}.$$

In other words, g can be defined as

$$g(n) := \begin{cases} 0 & \text{if} \quad n \text{ is even,} \\ \prod_{p\mid n}\left(1 + \frac{1}{\sum_{q<p} 1/q}\right) & \text{if} \quad n \text{ is odd.} \end{cases} \tag{4.3}$$

Then, for each positive integer j, we introduce the truncated strongly multiplicative function

$$g_j(n) := \prod_{\substack{p\mid n\\ p\le p_j}} g(p). \tag{4.4}$$

Observe that for any positive real number x, if $j \geq \pi(x)$, we have

$$\sum_{n \leq x} g(n) = \sum_{n \leq x} g_j(n).$$

The following result provides a surprisingly accurate and simple expression for the sum $g_j(1) + g_j(2) + \cdots + g_j(N)$ when N is a multiple of $p_1 p_2 \cdots p_j$.

Theorem 2. *If $N \in \mathbb{N}$ is such that $p_1 p_2 p_3 \cdots p_j \mid N$, then*

$$\sum_{n \leq N} g_j(n) = N\left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{p_j}\right). \tag{4.5}$$

Proof. Let j be an arbitrary positive integer. Setting $P := p_2 p_3 \cdots p_j$, we then obtain, letting ϕ stand for the Euler totient function,

$$\begin{aligned} \sum_{n \leq N} g_j(n) &= \sum_{\substack{n \leq N \\ n \text{ odd}}} \prod_{\substack{p \mid n \\ p \leq p_j}} g(p) = \sum_{d \mid P} g(d) \sum_{\substack{m \leq N/d \\ (m, 2P/d) = 1}} 1 \\ &= \sum_{d \mid P} g(d) \frac{N}{d} \frac{\phi(2P/d)}{2P/d} = N \frac{\phi(P)}{2P} \sum_{d \mid P} \frac{g(d)}{\phi(d)}. \end{aligned} \tag{4.6}$$

On the other hand, we can prove by induction on j that, for $P = p_2 p_3 \cdots p_j$, we have

$$\frac{\phi(P)}{2P} \sum_{d \mid P} \frac{g(d)}{\phi(d)} = \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{p_j}. \tag{4.7}$$

Indeed, we first observe that in the case $j = 2$, (4.7) boils down to

$$\frac{3-1}{6}\left(1 + \frac{1 + 1/2}{2}\right) = \frac{1}{2} + \frac{1}{3},$$

which is clearly true. So, assuming that (4.7) holds for $P = p_2 p_3 \cdots p_{j-1}$, let us prove that it must then hold for $P = p_2 p_3 \cdots p_j$. Setting $A = \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{p_{j-1}}$, we have, using our induction hypothesis,

$$\begin{aligned} \frac{\phi(P)}{2P} \sum_{d \mid P} \frac{g(d)}{\phi(d)} &= \frac{\phi(p_2 \cdots p_{j-1})}{2 p_2 \cdots p_{j-1}} \cdot \frac{\phi(p_j)}{p_j} \cdot \sum_{d \mid p_2 \cdots p_{j-1}} \frac{g(d)}{\phi(d)} \left(1 + \frac{g(p_j)}{\phi(p_j)}\right) \\ &= A \cdot \frac{\phi(p_j)}{p_j} \left(1 + \frac{1 + 1/A}{p_j - 1}\right) \\ &= \frac{A(p_j - 1)}{p_j} + \frac{A(1 + 1/A)(p_j - 1)}{(p_j - 1) p_j} \\ &= \frac{A p_j + 1}{p_j} = A + \frac{1}{p_j} = \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{p_j}, \end{aligned}$$

which clearly establishes (4.7). Now, combining (4.6) and (4.7), estimate (4.5) follows immediately. □

4.2 The average value of $g(n)$

We now prove estimate (4.1) which we state as a theorem.

Theorem 3. *Let g stand for the strongly multiplicative function defined in (4.3). Then,*

$$\sum_{n\le x} g(n) = x\log\log x + O(x\log\log\log x).$$

Proof. We first introduce the multiplicative function h such that $g(n)=\sum_{d|n}\mu^2(d)h(d)$. Hence, we can define the multiplicative function h on prime powers p^α by

$$h(p^\alpha)=\begin{cases} 0 & \text{if} \quad \alpha\ge 2,\\ -1 & \text{if} \quad p^\alpha = 2,\\ g(p)-1 & \text{if} \quad p^\alpha \text{ is odd,}\end{cases}$$

so that, since $h(n)=O(1)$,

$$\begin{aligned}
\sum_{n\le x} g(n) &= \sum_{n\le x}\sum_{d|n}\mu^2(d)h(d) = \sum_{d\le x}\mu^2(d)h(d)\left[\frac{x}{d}\right]\\
&= x\sum_{d\le x}\frac{\mu^2(d)h(d)}{d}+O(x)\\
&= x\left(\sum_{\substack{d\le x\\ d \text{ even}}}\frac{\mu^2(d)h(d)}{d}+\sum_{\substack{d\le x\\ d \text{ odd}}}\frac{\mu^2(d)h(d)}{d}\right)+O(x)\\
&= x\sum_{\substack{2d_1\le x\\ d_1 \text{ odd}}}\frac{\mu^2(d_1)h(2d_1)}{2d_1}+x\sum_{\substack{d\le x\\ d \text{ odd}}}\frac{\mu^2(d)h(d)}{d}+O(x)\\
&= -\frac{x}{2}\sum_{\substack{d_1\le x/2\\ d_1 \text{ odd}}}\frac{\mu^2(d_1)h(d_1)}{d_1}+x\sum_{\substack{d\le x\\ d \text{ odd}}}\frac{\mu^2(d)h(d)}{d}+O(x)\\
&= \frac{x}{2}\sum_{\substack{d\le x\\ d \text{ odd}}}\frac{\mu^2(d)h(d)}{d}+O(x), \qquad (4.8)
\end{aligned}$$

where we used the fact that

$$\left|\sum_{x/2<d_1\le x}\frac{\mu^2(d_1)h(d_1)}{d_1}\right| \ll \sum_{x/2<d_1\le x}\frac{1}{d_1}=O(1).$$

It remains to evaluate $\sum_{\substack{d\le x\\ d \text{ odd}}}\frac{\mu^2(d)h(d)}{d}$. On the one hand, since it is clear that

$$\sum_{\substack{d\le x\\ d \text{ odd}}}\frac{\mu^2(d)h(d)}{d}\le\prod_{3\le p\le x}\left(1+\frac{h(p)}{p}\right),$$

it follows from (4.8) that

$$\sum_{n\le x} g(n) \le \frac{x}{2}\prod_{3\le p\le x}\left(1+\frac{h(p)}{p}\right)+O(x). \tag{4.9}$$

Now, for each integer $j \ge 1$, let g_j be the strongly multiplicative function defined in (4.4). Observe that, for any $j \ge 1$, if n is even, $g_j(n) = g(n) = 0$, while for p odd, $g(p) > 1$, in which case $h(p) > 0$. Hence, arguing as in (4.8), it follows that, for any $j \ge 1$,

$$\begin{aligned}\sum_{n\le x} g(n) &\ge \sum_{n\le x} g_j(n) = \sum_{n\le x}\sum_{\substack{d\le x\\ P(d)\le p_j}} \mu^2(d)h(d) = \sum_{\substack{d\le x\\ P(d)\le p_j}} \mu^2(d)h(d)\left\lfloor\frac{x}{d}\right\rfloor\\ &= x\sum_{\substack{d\le x\\ P(d)\le p_j}}\frac{\mu^2(d)h(d)}{d}+O(x) = \frac{x}{2}\sum_{\substack{d\le x\\ d \text{ odd}\\ P(d)\le p_j}}\frac{\mu^2(d)h(d)}{d}+O(x). \end{aligned}\tag{4.10}$$

It follows from (4.10) that, for any real number y,

$$\sum_{n\le x} g(n) \ge \frac{x}{2}\sum_{\substack{d\le x\\ d \text{ odd}\\ P(d)\le y}}\frac{\mu^2(d)h(d)}{d}+O(x). \tag{4.11}$$

Now it is clear that

$$\sum_{\substack{d\le x\\ d \text{ odd}\\ P(d)\le y}}\frac{\mu^2(d)h(d)}{d} = \prod_{3\le p\le y}\left(1+\frac{h(p)}{p}\right)-\sum_{\substack{d> x\\ d \text{ odd}\\ P(d)\le y}}\frac{\mu^2(d)h(d)}{d} = \prod_{3\le p\le y}\left(1+\frac{h(p)}{p}\right)-Q_y(x), \tag{4.12}$$

say. We now choose $y = \exp\{\log x/(2\log\log x)\}$, thus allowing us to write, since $h(n) = O(1)$,

$$Q_y(x) \le \sum_{\substack{d>x\\ P(d)\le y}}\frac{1}{d}. \tag{4.13}$$

Using partial summation and the known estimate

$$\Psi(x,y) := \#\{n\le x : P(n)\le y\} \ll x\exp\left\{-\frac{1}{2}\frac{\log x}{\log y}\right\}$$

(see for instance Theorem 9.5 in the book of De Koninck and Luca [3]), we obtain

$$\sum_{\substack{d>x\\ P(d)\le y}}\frac{1}{d} = \left.\frac{\Psi(t,y)}{t}\right|_x^\infty + \int_x^\infty \frac{\Psi(t,y)}{t^2}\,dt$$

$$
\begin{aligned}
&\ll \exp\left\{-\frac{1}{2}\frac{\log x}{\log y}\right\} + \int_x^\infty \frac{\Psi(t,y)}{t^2}\,dt \\
&\ll \exp\left\{-\frac{1}{2}\frac{\log x}{\log y}\right\} + \int_x^\infty \exp\left\{-\frac{1}{2}\frac{\log t}{\log y}\right\}\frac{dt}{t} \\
&= \exp\left\{-\frac{1}{2}\frac{\log x}{\log y}\right\} + \int_{\log x}^\infty \exp\left\{-\frac{1}{2}\frac{u}{\log y}\right\} du \\
&= \exp\left\{-\frac{1}{2}\frac{\log x}{\log y}\right\} + \exp\left\{-\frac{1}{2}\frac{\log x}{\log y}\right\}\cdot 2\log y \\
&\ll \exp\left\{-\frac{1}{2}\frac{\log x}{\log\log x}\right\}\cdot\frac{\log x}{\log\log x} \\
&\ll \frac{1}{\log x}\cdot\frac{\log x}{\log\log x} = \frac{1}{\log\log x}. \qquad (4.14)
\end{aligned}
$$

Gathering estimates (4.12), (4.13) and (4.14), we obtain

$$
\frac{x}{2}\sum_{\substack{d\le x\\ d\ \text{odd}\\ P(d)\le y}} \frac{\mu^2(d)h(d)}{d} = \frac{x}{2}\prod_{3\le p\le y}\left(1+\frac{h(p)}{p}\right) + O\left(\frac{x}{\log\log x}\right).
$$

which in light of (4.11) implies that

$$
\sum_{n\le x} g(n) \ge \frac{x}{2}\prod_{3\le p\le y}\left(1+\frac{h(p)}{p}\right) + O(x). \qquad (4.15)
$$

Observe that, using Mertens' estimate, we have

$$
|h(p)| \le g(p) - 1 = \left(\sum_{q<p}\frac{1}{q}\right)^{-1} \ll \frac{1}{\log\log p}. \qquad (4.16)
$$

On the other hand, using Problem 4.2 in the book of De Koninck and Luca [3]), we have that for some absolute constant c,

$$
\sum_{3\le p\le x}\frac{1}{p\log\log p} = \log\log\log x + c + o(1) \qquad (x\to\infty). \qquad (4.17)
$$

Combining (4.16) and (4.17), we may conclude that

$$
\sum_{3\le p\le x}\frac{h(p)}{p} \ll \sum_{3\le p\le x}\frac{1}{p\log\log p} = \log\log\log x + c + o(1) \qquad (x\to\infty).
$$

It follows from this estimate that

$$
\log\left(\prod_{y\le p\le x}\left(1+\frac{h(p)}{p}\right)\right) = \sum_{y\le p\le x}\log\left(1+\frac{h(p)}{p}\right)
$$

$$\begin{aligned}
&= \sum_{y \le p \le x} \frac{h(p)}{p} + O\left(\sum_{y \le p \le x} \frac{1}{p^2}\right) \\
&= \sum_{y \le p \le x} \frac{h(p)}{p} + O\left(\frac{1}{y}\right) \\
&\ll \log\log\log x - \log(\log\log x - \log(2\log\log x)) + O\left(\frac{1}{y}\right) \\
&= O\left(\frac{\log\log\log x}{\log\log x}\right). \qquad (4.18)
\end{aligned}$$

Using (4.18) in (4.15), we obtain

$$\sum_{n \le x} g(n) \ge \frac{x}{2} \prod_{3 \le p \le x} \left(1 + \frac{h(p)}{p}\right)\left(1 + O\left(\frac{\log\log\log x}{\log\log x}\right)\right). \tag{4.19}$$

Combining (4.9) and (4.19), it follows that

$$\begin{aligned}
\sum_{n \le x} g(n) &= \left(1 + O\left(\frac{\log\log\log x}{\log\log x}\right)\right)\frac{x}{2}\prod_{3 \le p \le x}\left(1 + \frac{h(p)}{p}\right) \\
&= \frac{x}{2}\prod_{3 \le p \le x}\left(1 + \frac{h(p)}{p}\right) + O(x\log\log\log x). \qquad (4.20)
\end{aligned}$$

Finally, we can prove by induction on $\pi(x)$ that

$$\prod_{3 \le p \le x}\left(1 + \frac{h(p)}{p}\right) = 2\sum_{p \le x}\frac{1}{p}. \tag{4.21}$$

Indeed, we first observe that (4.21) holds in the case $\pi(x) = 2$, that is, that

$$1 + \frac{g(3) - 1}{3} = 2\left(\frac{1}{2} + \frac{1}{3}\right),$$

which is clearly true since $g(3) = 1 + \frac{1}{1/2} = 3$. So, assuming that (4.21) holds for $\pi(x) = r - 1$, let us prove that it must then hold for $\pi(x) = r$. Setting $A = \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{p_{r-1}}$ and using the induction hypothesis, we have

$$\begin{aligned}
\prod_{3 \le p \le x}\left(1 + \frac{h(p)}{p}\right) &= \left(2\sum_{p \le p_{r-1}}\frac{1}{p}\right)\left(1 + \frac{h(p_r)}{p_r}\right) \\
&= 2A\left(1 + \frac{1/A}{p_r}\right) \\
&= 2\left(A + \frac{1}{p_r}\right)
\end{aligned}$$

$$= 2\left(\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{p_r}\right),$$

which clearly establishes (4.21).

Thus, using (4.21) in (4.20) and the well known estimate $\sum_{p\le x} 1/p = \log\log x + O(1)$, we obtain

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

thus completing the proof of Theorem 3. □

5 Multiplicative functions with a large variance

In this section we show that if $g(n)$ is a real multiplicative function with asymptotic mean of the order of $\log\log n$ then its variance is necessarily much larger. This will imply that it is impossible to construct a multiplicative function g with the same asymptotic mean and variance as ω. In fact, we prove the following more general result.

Theorem 4. *Let $g(n)$ be a real valued multiplicative function and set*

$$M_g(x) := \frac{1}{x}\sum_{n\le x} g(n) \quad \text{and} \quad V_g(x) := \frac{1}{x}\sum_{n\le x}\left(g(n) - M_g(x)\right)^2.$$

If $M_g(x)\to\infty$ as $x\to\infty$, then $V_g(x) \gg_g (M_g(x))^2$.

Proof. Since $M_g(x)$ tends to infinity with x, there exists at least one prime p_0 such that $g(p_0) > 1$. For a large real number x, define I as the open interval

$$I := I(x) = \left(\frac{M_g(x)}{\sqrt{g(p_0)}}, \sqrt{g(p_0)}M_g(x)\right).$$

Let n be an integer smaller than x/p_0 and coprime to p_0. Then either $g(n)\notin I$ or $g(n)\in I$. In this second case, that is if $g(n)\in I$, we have that

$$g(np_0) = g(n)g(p_0) > \frac{M_g(x)}{\sqrt{g(p_0)}}\cdot g(p_0) = M_g(x)\cdot\sqrt{g(p_0)} \notin I.$$

This means that

$$g(n)\notin I \qquad \text{or} \qquad g(np_0)\notin I. \tag{5.1}$$

Now observe that

$$\#\{n\le x/p_0 : (n,p_0)=1\} = x\frac{p_0-1}{p_0^2} + O(1),$$

which implies that, using (5.1),

$$\begin{aligned}
\#\{m \le x : g(m) \notin I\} &\ge \#\{n \le x/p_0 : (n, p_0) = 1, g(n) \notin I\} \\
&\qquad + \quad \#\{n \le x : p_0 \mid n, g(n) \notin I\} \\
&\ge x\frac{p_0 - 1}{p_0^2} + O(1). \qquad (5.2)
\end{aligned}$$

On the other hand,

$$V_g(x) = \frac{1}{x}\sum_{m\le x} (g(m) - M_g(x))^2 \ge \frac{1}{x} \sum_{\substack{m\le x \\ g(m)\notin I}} (g(m) - M_g(x))^2 . \qquad (5.3)$$

From the definition of I, we have that $g(m) \notin I$ implies that

$$|g(m) - M_g(x)| > \left(1 - \frac{1}{\sqrt{g(p_0)}}\right) M_g(x),$$

which substituted in (5.3) and using (5.2) yields

$$\begin{aligned}
V_g(x) &\ge \frac{1}{x} \sum_{\substack{m\le x \\ g(m)\notin I}} \left(1 - \frac{1}{\sqrt{g(p_0)}}\right)^2 (M_g(x))^2 \\
&= \left(1 - \frac{1}{\sqrt{g(p_0)}}\right)^2 (M_g(x))^2 \frac{1}{x} \#\{m \le x : g(m) \notin I\} \\
&\ge \left(\frac{p_0 - 1}{p_0^2} + O\left(\frac{1}{x}\right)\right)\left(1 - \frac{1}{\sqrt{g(p_0)}}\right)^2 (M_g(x))^2 \\
&\ge c(M_g(x))^2
\end{aligned}$$

for some positive constant c which depends on g, thus completing the proof of Theorem 4. □

6 Numerical computations

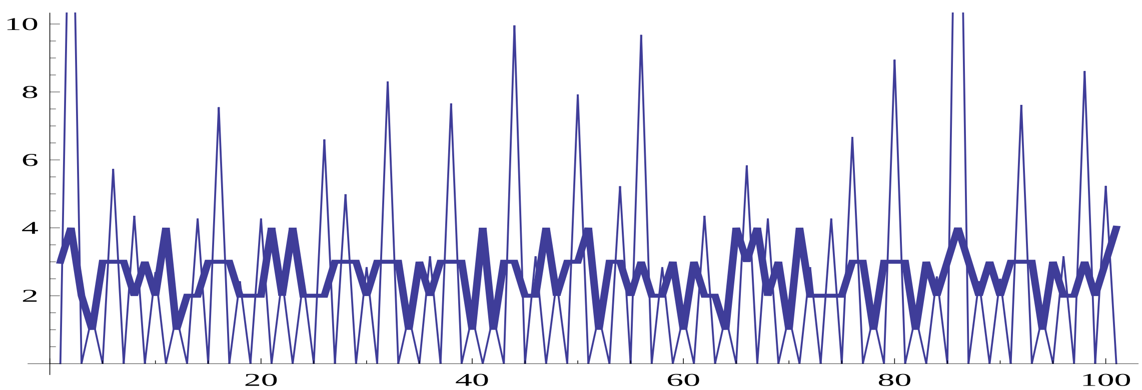
Let $g \in \mathcal{M}^*$ be the function defined in (4.3). Then, set

$$S_g(x) := \sum_{n\le x} g(n) \qquad \text{and} \qquad S_\omega(x) := \sum_{n\le x} \omega(n).$$

The whole interest in the definition of g is that $S_g(x)$ is close to $S_\omega(x)$ as was shown in Section 4, namely by comparing estimates (4.1) and (4.2), and as the following table seems to indicate.

x	$S_\omega(x)$	$\lfloor S_g(x) \rfloor$	$S_g(x)/S_\omega(x)$
10	11	11	0.9850
10^2	171	156	1.0962
10^3	2126	1895	1.1219
10^4	24300	21615	1.1242
10^5	266400	237775	1.1204
10^6	2853708	2560376	1.1146
10^7	30130317	27177837	1.1086

Nevertheless, the spread of the function g is much larger than the spread of the function ω. For instance, when $10^4 \le n \le 10^4 + 100$, the function $\omega(n)$ takes only the values 1 through 4 while $g(n)$ ranges from 0 to 20. This is partially shown in the following two curves representing the values of $\omega(n)$ (thick curve) and of $g(n)$ (ordinary curve), respectively, for $10^4 \le n \le 10^4 + 100$.



On the other hand, as was pointed out in Theorem 4 and as the following table seems to indicate, the variance $V_g(x)$ is larger than $(M_g(x))^2$.

x	$A = V_g(x)$	$B = (M_g(x))^2$	A/B
10	1.52402	1.24718	1.22197
10^2	3.91748	2.44219	1.60408
10^3	7.21603	3.59243	2.00868
10^4	10.5352	4.67219	2.25488
10^5	13.8471	5.65369	2.44921
10^6	17.1836	6.55553	2.62124
10^7	20.4718	7.38635	2.77158

7 Final remarks

Our results suggest that for arithmetic functions whose mean value behaves as $\lambda(x)$, where $\lambda(x)$ is a slowly increasing function tending to $+\infty$ with x, the distribution of the values of multiplicative functions cannot be as "narrow" as that of some additive functions. Perhaps this could even serve as a characterization of additive functions.

In fact, we conjecture that for multiplicative functions, the ratio (standard deviation)/(mean value) cannot tend to zero as it is the case for additive functions f for which $f(p)$ increases slowly enough, for instance like the ω function.

Finally, the problem of deciding whether given an additive function which has a limiting distribution, one can or cannot construct a multiplicative function with the same limiting distribution remains an open question. Here, we showed that it is not always the case for functions with finite support (thus with discontinuous distributions). The general issue of characterizing the set of distributions which are limiting distributions of additive or multiplicative functions as well as the intersection between these two sets, seems to be a very deep problem requiring new ideas.

References

[1] H. Daboussi, *Sur les fonctions multiplicatives ayant une valeur moyenne non nulle*, Bu11. Soc. Math. France **109** (1981), 183–205.

[2] J.M. De Koninck, N. Doyon and P. Letendre, *On the proximity of additive and multiplicative functions*, Functiones at Approximatio (to appear).

[3] J.M. De Koninck and F. Luca, *Analytic Number Theory: Exploring the Anatomy of Integers*, Graduate Studies in Mathematics, Vol. 134, American Mathematical Society, Providence, Rhode Island, 2012.

[4] P.D.T.A. Elliott, *Probabilistic Number Theory I, Mean Value Theorems*, Springer-Verlag, Berlin, 1979; *Probabilistic Number Theory II: Central Limit Theorems*, Springer-Verlag, 1980.

[5] J. Galambos, *Distribution of additive and multiplicative functions*, in The Theory of Arithmetic Functions, Proceedings of Conf. at Western Michigan Univ., 1971, Volume 251, pp. 127–139, Springer Lecture Notes-Mathematics, Berlin, 1972.

[6] B.V. Levin and N.M. Timofeev, *On the distribution of values of additive functions*, Acta Arithmetica **26**, no. 4, (1974/1975), 333–364.

[7] G. Tenenbaum, *Introduction à la théorie analytique des nombres*, Collection Échelles, Belin, 2008.

Jean-Marie De Koninck
Dép. de mathématiques et de statistique
Université Laval
Québec
Québec G1V 0A6
Canada
jmdk@mat.ulaval.ca

Nicolas Doyon
Dép. de mathématiques et de statistique
Université Laval
Québec
Québec G1V 0A6
Canada
nicolas.doyon@mat.ulaval.ca

JMDK, le 16 avril 2015; fichier: Add-mult-globally-16-avril-2015.tex