STATISTICS OF SIEVES AND SQUARE-FREE NUMBERS

GEOFFREY GRIMMETT

ABSTRACT

Let $\mathscr{S} = (s_1, s_2, \ldots)$ be a collection of relatively prime numbers. The asymptotic properties of the process of sieving by $\mathscr{S}$ may be realized in terms of a stationary random process. In the case when $\mathscr{S}$ is the set of squares of the primes, one may make use of this representation to verify a conjecture of R. Hall: in a 'typical' interval of length k, the number S_k of square-free numbers has a probability mass function having order $k^{-\frac{1}{4}}$ in the limit as $k \to \infty$.

1. *Introduction*

The primary objective of this paper is to verify a conjecture of R. Hall concerning the asymptotic distribution of the square-free numbers. The ensuing result, whilst of interest in the analytic theory of numbers, is easiest expressed and proved using the language of probability theory. In expressing this result, we shall make use of a basic fact that, for fixed n, the distribution of square-free numbers in a 'typical interval' $(m+1, m+2, \ldots, m+n)$ of length n may be thought of as the weak limit (in the sense of the weak convergence of probability measures) of a natural collection of measures indexed by the positive integers $\mathbb{N}$. This representation is easy to show, but nevertheless offers as a consequence a considerable simplification over certain more primitive techniques of analysis.

In some ways the square-free numbers are not special, when seen from the point of view of this paper. Similar techniques may in principle be applied to the numbers which escape the action of any 'sieve', so long as the elements of the sieve are relatively prime. Actually the coprimality of the numbers in the sieve is largely irrelevant to certain aspects of the probabilistic analysis which follows, although it is crucial to the number-theoretic interpretation.

Let $\mathscr{S} = (s_1, s_2, \ldots)$ be a sequence of positive integers; we shall assume throughout that $1 < s_1 < s_2 < \ldots$. We shall speak of the 'sieve generated by $\mathscr{S}$' as being the following process. We are provided with a set $\mathscr{G} = (g_1, g_2, \ldots)$ of labels, the label g_i corresponding to the action (to be described next) of the integer s_i. We inspect each integer m ($\geqslant 1$) in turn, and we label m with g_i (for each i) if and only if $s_i | m$. The result is a sequence $G = (G_1, G_2, \ldots)$ of subsets of $\mathscr{G}$, G_m being the set of labels of m. We shall commonly be interested in those integers which escape the action of the sieve: $U(G) = \{m : G_m = \varnothing\}$. If $s_i = p_i^a$, the ath power of the ith prime, then $U(G)$ is the set of 'a-free numbers'.

A central notion of this paper is a variant of the above sieving process. We shall speak of the 'random sieve generated by $\mathscr{S}$' as being the following random process. Let $X_1, X_2, \ldots$ be independent random variables, X_i having probability mass function

$$P(X_i = k) = \begin{cases} 1/s_i & \text{if } 1 \leqslant k \leqslant s_i, \\ 0 & \text{otherwise,} \end{cases}$$

so that X_i is equally likely to take any value in $\{1, 2, \ldots, s_i\}$. (Throughout this

Received 10 May 1989.

1980 *Mathematics Subject Classification* (1985 *Revision*) 11N35.

J. London Math. Soc. (2) 43 (1991) 1–11

paper, P stands for probability and E for expectation.) We now label the positive integers in the following way: for each $i \geqslant 1$, we label with g_i each member of the set $\{X_i + ks_i : 0 \leqslant k < \infty\}$. The outcome of this process is a random vector $\Gamma = (\Gamma_1, \Gamma_2, \ldots)$ of subsets of $\mathcal{G}$, Γ_i being the (random) set of labels of i. Of particular interest is the set of integers which escape this 'sieve with random starting points': $U(\Gamma) = \{m : \Gamma_m = \varnothing\}$. We shall show that the statistics of Γ (and especially $U(\Gamma)$) are useful in understanding asymptotic properties of G (in particular $U(G)$) *whenever the numbers s_i are relatively prime.*

Let $\mathcal{G}^\infty = \mathcal{G}_1 \times \mathcal{G}_2 \times \ldots$, and think of G and Γ as being elements of $\mathcal{G}^\infty$. There are two topologies of interest for $\mathcal{G}^\infty$. The first arises from thinking of $\mathcal{G}^\infty$ as a product space whose components are endowed with the discrete topology, and the second is the topology associated with the metric

$$d(\omega, \omega') = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |1_{\omega_i}(g_j) - 1_{\omega'_i}(g_j)| \, 2^{-i-j} \tag{1.1}$$

for $\omega, \omega' \in \mathcal{G}^\infty$, where 1_A denotes the indicator function of A. It is tempting to work with the second of these, but this route requires more work later, and the main reason for this is that the indicator functions of sets such as $\{\omega \in \mathcal{G}^\infty : \omega_1 = \varnothing\}$ are not continuous with this topology. We shall therefore endow $\mathcal{G}^\infty$ with the former topology, and write $\mathcal{F}$ for the σ-field generated by the open sets.

We now sample from $\mathcal{G}^\infty$ at random as follows. Let N_n be chosen randomly and uniformly from $\{1, 2, \ldots, n\}$, so that $P(N_n = k) = n^{-1}$ for $1 \leqslant k \leqslant n$, and define $\gamma_n \in \mathcal{G}^\infty$ by $\gamma_n = (G_{N_n}, G_{N_n+1}, \ldots)$; thus γ_n is obtained from G by deleting its first $N_n - 1$ components. Let μ_n be the measure induced on $(\mathcal{G}^\infty, \mathcal{F})$ by the random vector γ_n (that is, $\mu_n(A) = P(\gamma_n \in A)$ for $A \in \mathcal{F}$), and let μ be the measure on $(\mathcal{G}^\infty, \mathcal{F})$ corresponding to Γ. We note that μ is *stationary* in the sense that $\mu(A) = \mu(\tau^{-1}A)$ for any event A, where $\tau(\omega) = (\omega_2, \omega_3, \ldots)$ for any $\omega = (\omega_1, \omega_2, \ldots) \in \mathcal{G}^\infty$.

THEOREM 1. *If the elements of $\mathcal{S}$ are relatively prime and $\sum_i 1/s_i < \infty$, then μ_n converges weakly to μ as $n \to \infty$.*

The assumption that $\sum_i 1/s_i < \infty$ is necessary when working with the discrete topology. To see why, suppose for example that $s_i = p_i$, the ith prime, so that $\sum_i 1/s_i = \infty$. Let $A = \{\omega \in \mathcal{G}^\infty : |\omega_1| = \infty\}$, the set of vectors of $\mathcal{G}^\infty$ whose first component contains infinitely many labels. Then $\mu_n(A) = 0$ for all n, whilst $\mu(A) = 1$ by the Borel–Cantelli lemma (see [3, p. 187]). This assumption may be dispensed with when working with the metric given in (1.1), but the conclusion of the corresponding theorem would then be weaker.

Similar computations to those necessary for the proof of Theorem 1 may be found in [8], but in a rather restricted context. For specific choices of the set $\mathcal{S}$, one may be interested in the *rate* of convergence. We shall see in the proof that, for events C which depend only on the first r components of the realization, it is the case that

$$|\mu_n(C) - \mu(C)| \leqslant r\left(2 + \frac{r}{n}\right) \sum_{j>m} \frac{1}{s_j} + \frac{1}{n} s_1 s_2 \ldots s_m \tag{1.2}$$

for all m ($\geqslant 1$). Much better rates should be derivable in particular cases.

We have as a consequence of Theorem 1 (see [1] for details of weak convergence of probability measures) that

$$\int f d\mu_n \to \int f d\mu \quad \text{as } n \to \infty \tag{1.3}$$

for all bounded continuous functions f on $\mathscr{G}^\infty$, and also that

$$\mu_n(B) \to \mu(B) \quad \text{if } \mu(\partial B) = 0, \tag{1.4}$$

for any $B \in \mathscr{F}$, where ∂B is the topological boundary of B.

We turn now to a concrete application of Theorem 1. Suppose that $s_i = p_i^2$, the square of the ith prime, and let $S_k(m)$ be the number of square-free numbers in the interval $\{m, m+1, \ldots, m+k-1\}$. It follows from a calculation of Mirsky [8] (see also [4]) that

$$p_k(j) = \lim_{n \to \infty} \frac{1}{n} |\{m \in \{1, 2, \ldots, n\} : S_k(m) = j\}| \tag{1.5}$$

exists for all k and j; actually, this is a consequence of the weak convergence theorem above, for the case of the squared primes. The function p_k is a probability mass function on the set $\{0, 1, \ldots, k\}$, and it is known that the mean and variance of the associated distribution satisfy

$$m_k = \sum_j j p_k(j) = \frac{6}{\pi^2} k \tag{1.6}$$

and

$$\sigma_k^2 = \sum_j (j - m_k)^2 p_k(j) = \sigma^2 \sqrt{k} + O(k^{\varepsilon + \frac{1}{5}}) \tag{1.7}$$

as $k \to \infty$, for all $\varepsilon > 0$, where

$$\sigma^2 = \frac{\zeta(\frac{3}{2})}{\pi} \prod_p \left(\frac{p^3 - 3p + 2}{p^3} \right); \tag{1.8}$$

here, ζ is the Riemann zeta function and the product is over all primes p. The remarkable asymptotic relation (1.7) for the variance is due to Hall [4], who has pointed out that the calculation of higher moments seems to be of a different level of difficulty (see also [5]). Hall has made the natural conjecture that

$$M_k = \sup_j p_k(j)$$

has the order $k^{-\frac{1}{4}}$ as $k \to \infty$, and it is our purpose to prove this here.

THEOREM 2. *There exists an absolute constant γ such that*

$$\left(\frac{1 + o(1)}{\sigma 3 \sqrt{3}} \right) k^{-\frac{1}{4}} \leqslant M_k \leqslant \gamma k^{-\frac{1}{4}} (\log k)^{\frac{1}{2}} \quad \text{as } k \to \infty,$$

where σ is given in (1.8).

It is natural to conjecture that the distribution associated with p_k converges as $k \to \infty$, when suitably normalized, to a non-trivial limit. That is to say, we conjecture that the limit

$$\Psi(x) = \lim_{k \to \infty} \sum_{j \leqslant K_k(x)} p_k(j)$$

exists, where $K_k(x) = m_k + x\sigma_k$, and that Ψ is a probability distribution function. Seen in the light of Theorem 1, this conjecture amounts to the statement that

$(U_n - E(U_n))/(\mathrm{var}(U_n))^{\frac{1}{2}}$ converges in distribution as $n \to \infty$, where U_n is the number of integers in $\{1, 2, \ldots, n\}$ which remain unlabelled in the random sieve generated by $\mathscr{S}$. It is tempting to conjecture further that the process $Y_n(t) = U_{\lfloor nt \rfloor}$ converges weakly as $n \to \infty$, when suitably normalized, the limit being a self-similar process with index $\frac{1}{4}$. It is immediate from the ergodic theorem that

$$\frac{1}{n} U_n \to \frac{6}{\pi^2} \quad \text{as } n \to \infty$$

almost surely, the numerical value $6/\pi^2$ arising as the mean value of U_1:

$$E(U_1) = \prod_p \left(1 - \frac{1}{p^2}\right) = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}.$$

The full power of Theorem 1 is not needed for the proof of Theorem 2; for the latter, we need little more than the results of [**8**].

2. *Proof of Theorem* 1

Suppose that the elements of $\mathscr{S} = (s_1, s_2, \ldots)$ are relatively prime and satisfy

$$\sum_j \frac{1}{s_j} < \infty. \tag{2.1}$$

For any $\omega \in \mathscr{G}^\infty$ we write $\omega^m = (\omega_1^m, \omega_2^m, \ldots)$, where $\omega_i^m = \omega_i \cap \{g_1, g_2, \ldots, g_m\}$, and use a similar notation for any $\psi \in \mathscr{G}_1 \times \mathscr{G}_2 \times \ldots \times \mathscr{G}_k$.

Let $\mathscr{A} = (A_1, A_2, \ldots, A_r) \in \mathscr{G}^r$, and let

$$C(\mathscr{A}) = \{\omega \in \mathscr{G}^\infty : \omega_i = A_i \text{ for } 1 \leqslant i \leqslant r\}. \tag{2.2}$$

The principal calculation lies in showing that

$$\mu_n(C(\mathscr{A})) \to \mu(C(\mathscr{A})) \quad \text{as } n \to \infty \tag{2.3}$$

for all such $\mathscr{A}$. Later, we shall indicate why this suffices.

Let us denote by $C^m(\mathscr{A})$ the event

$$C^m(\mathscr{A}) = \{\omega \in \mathscr{G}^\infty : \omega_i^m = A_i^m \text{ for } 1 \leqslant i \leqslant r\}.$$

Clearly $C(\mathscr{A}) \subseteq C^m(\mathscr{A})$, and so

$$\mu(C(\mathscr{A})) \leqslant \mu(C^m(\mathscr{A})), \qquad \mu_n(C(\mathscr{A})) \leqslant \mu_n(C^m(\mathscr{A})). \tag{2.4}$$

Furthermore, by the stationarity of μ,

$$\begin{aligned} \mu(C^m(\mathscr{A})) - \mu(C(\mathscr{A})) &\leqslant \mu(g_j \in \omega_i \text{ for some } 1 \leqslant i \leqslant r, j > m) \\ &\leqslant r\,\mu(g_j \in \omega_1 \text{ for some } j > m) \\ &\leqslant r \sum_{j>m} \mu(g_j \in \omega_1) \\ &= r \sum_{j>m} \frac{1}{s_j}. \end{aligned} \tag{2.5}$$

Similarly

$$\begin{aligned} \mu_n(C^m(\mathscr{A})) - \mu_n(C(\mathscr{A})) &\leqslant \sum_{i=1}^{r} \sum_{j>m} \mu_n(g_j \in \omega_i) \\ &\leqslant \frac{r}{n} \sum_{j>m} |\{k : s_j \mid k, 1 \leqslant k \leqslant n+r\}| \\ &\leqslant \frac{r(n+r)}{n} \sum_{j>m} \frac{1}{s_j}. \end{aligned} \tag{2.6}$$

Hence

$$|\mu_n(C(\mathscr{A}))-\mu(C(\mathscr{A}))| \leqslant |\mu_n(C^m(\mathscr{A}))-\mu(C^m(\mathscr{A}))|+r\left(2+\frac{r}{n}\right)\sum_{j>m}\frac{1}{s_j}. \tag{2.7}$$

It remains to estimate the difference between $\mu_n(C^m(\mathscr{A}))$ and $\mu(C^m(\mathscr{A}))$.

With $\mathscr{A}=(A_1, A_2, \dots, A_r)$ as before, and m a positive integer, let

$$U(\mathscr{A}) = \{j: g_j \in A_i^m \text{ for some } i\}$$

and $V(\mathscr{A}) = \{1, 2, \dots, m\} \setminus U(\mathscr{A})$. We write $\delta_j(\mathscr{A}) = \min\{k: g_j \in A_k^m\}$ for $j \in U$. Then $C^m(\mathscr{A})$ may be written as

$$C^m(\mathscr{A}) = \{\omega: X_j(\omega) = \delta_j \text{ for } j \in U, X_j(\omega) > r \text{ for } j \in V\},$$

where

$$X_j(\omega) = \min\{k: g_j \in \omega_k\}.$$

Let $T = s_1 s_2 \dots s_m$. We claim that

$$\mu_T(C^m(\mathscr{A})) = \frac{1}{T}\prod_{j\in V}(s_j - r). \tag{2.8}$$

To see this, let $\eta = (\eta_1, \eta_2, \dots, \eta_m)$ be a sequence of positive integers satisfying $\eta_j \leqslant s_j$ for all j, and let

$$D(\eta) = \{\omega: X_j(\omega) = \eta_j \text{ for } 1 \leqslant j \leqslant m\}.$$

Now $\mu_T(D(\eta))$ is the proportion of integers k such that $1 \leqslant k \leqslant T$ and $s_j \mid (k+\eta_j-1)$ for $1 \leqslant j \leqslant m$. If k_1 and k_2 are two such numbers then $s_j \mid (k_1 - k_2)$ for all j ($\leqslant m$), so that $T \mid (k_1 - k_2)$, and therefore $k_1 = k_2$. Thus there is at most one such value k for any η. On the other hand there are $s_1 s_2 \dots s_m = T$ possible choices for η, and there are exactly T values of k. Therefore, for each η, there exists *exactly* one value of k. Hence $\mu_T(D(\eta)) = 1/T$ for any η. Summing over all η satisfying $\eta_j = \delta_j$ if $j \in U$ and $\eta_j > r$ if $j \in V$, we obtain (2.8). Note also that $\mu_{kT}(C^m(\mathscr{A})) = \mu_T(C^m(\mathscr{A}))$, since every pattern of labels drawn from $(g_1, g_2, \dots, g_m)$ recurs at intervals of length $s_1 s_2 \dots s_m = T$.

There remains a detail. Clearly

$$\begin{aligned}\mu(C^m(\mathscr{A})) &= \prod_{j\in U}\frac{1}{s_j}\prod_{j\in V}\left(1-\frac{r}{s_j}\right)\\ &= \frac{1}{T}\prod_{j\in V}(s_j - r) = \mu_T(C^m(\mathscr{A}))\end{aligned} \tag{2.9}$$

by (2.8). Now $n\mu_n(C^m(\mathscr{A}))$ is monotone in n so that, if $kT \leqslant n < (k+1)T$, then

$$\frac{kT}{n}\mu(C^m(\mathscr{A})) \leqslant \mu_n(C^m(\mathscr{A})) \leqslant \frac{(k+1)T}{n}\mu(C^m(\mathscr{A})).$$

However $n - T < kT \leqslant n$, so that

$$|\mu_n(C^m(\mathscr{A})) - \mu(C^m(\mathscr{A}))| \leqslant \frac{T}{n}\mu(C^m(\mathscr{A})) \leqslant \frac{1}{n}s_1 s_2 \dots s_m. \tag{2.10}$$

We combine (2.7) with (2.10) to obtain

$$|\mu_n(C(\mathscr{A})) - \mu(C(\mathscr{A}))| \leqslant r\left(2+\frac{r}{n}\right)\sum_{j>m}\frac{1}{s_j} + \frac{1}{n}s_1 s_2 \dots s_m. \tag{2.11}$$

Letting $n \to \infty$ and then $m \to \infty$, we conclude that $\mu_n(C(\mathscr{A})) \to \mu(C(\mathscr{A}))$, and (2.3) is proved.

There remain two steps before the proof is complete. First, we have to prove that μ is specified by its values on cylinders $C(\mathscr{A})$, and secondly, we must show that the sequence $(\mu_n : n \geqslant 1)$ is tight. The result will then follow in the usual way (see [**1**, pp. 35 *et seq.*]).

That μ is determined by the finite-dimensional cylinders is easy to see—simple calculations based on the quantities $\mu(C(\mathscr{A}))$ show that, for any measure ν agreeing with μ on the finite-dimensional cylinders, the random variables $X_j(\omega) = \min\{n : g_j \in \omega_n\}$ are independent with mass functions $\nu(X_j = k) = 1/s_j$ for $1 \leqslant k \leqslant s_j$, and therefore $\nu = \mu$.

Finally we show that $(\mu_n : n \geqslant 1)$ is tight. Let $m = (m_1, m_2, \ldots)$ be a strictly increasing sequence of positive integers, and let

$$K(m) = \{\omega \in \mathscr{G}^\infty : \omega_i \subseteq \{g_1, g_2, \ldots, g_{m_i}\} \text{ for all } i\}; \tag{2.12}$$

then $K(m)$ is a compact subset of the topological space $\mathscr{G}^\infty$, endowed with the discrete product topology. Let $\varepsilon > 0$; we shall show that there exists m such that $\mu_n(K(m)) > 1 - \varepsilon$ for all n.

Let $(t_i : i \geqslant 1)$ be a non-decreasing sequence of positive integers such that $t_i \to \infty$ as $i \to \infty$, and

$$\sum_{j=1}^{\infty} \frac{t_j}{s_j} < \infty; \tag{2.13}$$

such a sequence exists since $1/s_j$ is summable. Suppose further that $m_t \to \infty$ sufficiently fast that

$$m_1 \geqslant R, \quad m_{t_j} \geqslant j \quad \text{for all } j, \tag{2.14}$$

where R has been chosen such that

$$\sum_{j \geqslant R} \frac{t_j}{s_j} < \varepsilon. \tag{2.15}$$

Now,

$$\mu_n(K(m)) = \frac{1}{n} \sum_{1 \leqslant i \leqslant n} J_i, \tag{2.16}$$

where J_i is 0 or 1, taking the latter value if and only if $j \leqslant m_{ks_j - i + 1}$ for all $s_j \in \mathscr{S}$ and all $k \geqslant 1$ satisfying $ks_j \geqslant i$. Let

$$I = \bigcup_{\substack{k \geqslant 1 \\ j \geqslant R}} \{ks_j - t_j + 1, ks_j - t_j + 2, \ldots, ks_j\}.$$

We claim that $J_i = 1$ if $i \notin I$. To see this, suppose that $i \notin I$. Let $s_j \in \mathscr{S}$ and let $\kappa = \kappa_j$ be the smallest value of k such that $ks_j \geqslant i$. If $j \geqslant R$ then $\kappa s_j - i \geqslant t_j$, so that $j \leqslant m_{t_j} \leqslant m_{\kappa s_j - i + 1}$ by (2.14); the numbers m_t are non-decreasing, so that a similar inequality is valid for all $k \geqslant \kappa$. On the other hand, if $j < R$ then $j < m_r$ for all $r \geqslant 1$ by (2.14). Thus the claim is valid, and it is then a consequence of (2.16) that

$$\mu_n(K(m)) \geqslant 1 - \frac{1}{n} |I \cap \{1, 2, \ldots, n\}|.$$

However,

$$\frac{1}{n} |I \cap \{1, 2, \ldots, n\}| \leqslant \frac{1}{n} \sum_{j \geqslant R} \frac{nt_j}{s_j} < \varepsilon \qquad \text{by (2.15)}$$

and the proof is finished.

3. *Proof of Theorem* 2

Let $\mathscr{S} = (s_1, s_2, \ldots)$, where $s_i = p_i^2$, the square of the ith prime. For positive integers j and k, let $f_{jk}: \mathscr{G}^\infty \to \{0, 1\}$ be the indicator function of the event that there are exactly j values of $i \in \{1, 2, \ldots, k\}$ with the property that $\omega_i = \varnothing$. Then f_{jk} is a bounded continuous function on the product space $\mathscr{G}^\infty$, so that

$$\int f_{jk}\, d\mu_n \to \int f_{jk}\, d\mu \quad \text{as } n \to \infty,$$

by Theorem 1. Hence

$$\lim_{n\to\infty} \frac{1}{n} |\{m \in \{1, 2, \ldots, n\}: S_k(m) = j\}| = p_k(j)$$

exists, where $S_k(m)$ is the number of square-free numbers in $\{m, m+1, \ldots, m+k-1\}$ and $p_k(j) = \mu(\Sigma_k = j)$, with Σ_k the random variable

$$\Sigma_k(\omega) = |\{i \in \{1, 2, \ldots, k\}: \omega_i = \varnothing\}|.$$

We introduce the indicator variables

$$I_i(\omega) = \begin{cases} 1 & \text{if } \omega_i = \varnothing, \\ 0 & \text{otherwise,} \end{cases}$$

so that

$$\Sigma_k = \sum_{i=1}^{k} I_i.$$

The measure μ is stationary, so that $m_k = E(\Sigma_k)$ satisfies

$$m_k = kE(I_1) = k \prod_p \left(1 - \frac{1}{p^2}\right) = \frac{k}{\zeta(2)} = \frac{6}{\pi^2} k, \tag{3.1}$$

in agreement with (1.6); here and later, such products and sums are over all primes p. We note that Hall [4] has proved that the variance $\sigma_k^2 = \operatorname{var}(\Sigma_k)$ satisfies

$$\sigma_k^2 = \sigma^2 \sqrt{k} + O(k^{\varepsilon + \frac{1}{3}}) \tag{3.2}$$

as $k \to \infty$, for all $\varepsilon > 0$, where

$$\sigma^2 = \frac{\zeta(\frac{3}{2})}{\pi} \prod_p \left(\frac{p^3 - 3p + 2}{p^3}\right). \tag{3.3}$$

The lower bound in Theorem 2 is an immediate consequence of an application of Chebyshev's inequality (see [3, p. 186]). With $M_k = \sup_j p_k(j)$, we have that

$$\begin{aligned} (1 + 2\alpha\sigma_k) M_k &\geqslant \mu(|\Sigma_k - E\Sigma_k| < \alpha\sigma_k) \\ &\geqslant 1 - \frac{1}{\alpha^2} \quad \text{for } \alpha > 0. \end{aligned}$$

Setting $\alpha = \sqrt{3}$ and using (3.2), we obtain the required lower bound.

The upper bound for M_k is the principal component of Theorem 2, and the idea of its proof is as follows. We divide $\mathscr{S}$ into two parts, $\mathscr{S} = \mathscr{M} \cup \mathscr{N}$, where

$$\mathscr{M} = \{s \in \mathscr{S}: s \leqslant k\}, \qquad \mathscr{N} = \mathscr{S} \setminus \mathscr{M}.$$

For any $\omega \in \mathscr{G}^\infty$, we call an integer $i \in \{1, 2, \ldots, k\}$ *$\mathscr{M}$-free* if $g_j \notin \omega_i$ for all $s_j \in \mathscr{M}$. If k is large then the number of $\mathscr{M}$-free integers does not generally differ greatly from the number Σ_k of integers $i \in \{1, 2, \ldots, k\}$ for which $g_j \notin \omega_i$ for all j. Now, for each $s_j \in \mathscr{N}$, the corresponding label g_j can occur at most once amongst $\omega_1, \omega_2, \ldots, \omega_k$, and g_j occurs exactly once with μ-probability k/s_j. Writing N for the (random) number of occurrences of the g_j (for $s_j \in \mathscr{N}$) amongst the ω_i (for $1 \leqslant i \leqslant k$), we have that

$$N = \sum_{s_j \in \mathscr{N}} J_j,$$

where J_j is the indicator function of the event that g_j occurs somewhere; J_j takes the value 0 or 1, the latter having probability k/s_j. It is not difficult to check that the distribution of N is approximately normal with mean and variance of order $\sqrt{k/\log k}$, and thus the maximum value of its mass function has order $k^{-\frac{1}{4}}(\log k)^{\frac{1}{2}}$. In the words of Richard Hall, the large primes muddy the water to the correct degree. We shall make the above argument rigorous. In doing so we shall make use of the asymptotic relations

$$\sum_{p > x} \frac{1}{p^2} \sim \frac{1}{x \log x} \quad \text{as } x \to \infty, \tag{3.4}$$

$$\sum_{p > x} \frac{1}{p^4} \sim \frac{1}{3x^3 \log x} \quad \text{as } x \to \infty, \tag{3.5}$$

easily derived from [**6**, Theorem 415], together with the estimate $\pi(n) \sim n/\log n$ for the number $\pi(n)$ of primes not exceeding n.

Fix k, and let

$$K_i(\omega) = \begin{cases} 0 & \text{if } g_j \in \omega_i \text{ for some } s_j \in \mathscr{M}, \\ 1 & \text{otherwise,} \end{cases}$$

and write

$$\Psi_k(\omega) = \sum_{i=1}^{k} K_i(\omega),$$

the number of $\mathscr{M}$-free integers in $\{1, 2, \ldots, k\}$. Certainly $\Psi_k \geqslant \Sigma_k$, so that

$$\begin{aligned} \mu(\Psi_k \leqslant \tfrac{1}{2}k) &\leqslant \mu(\Sigma_k \leqslant \tfrac{1}{2}k) \\ &\leqslant \frac{\operatorname{var}(\Sigma_k)}{((6/\pi^2) - \frac{1}{2})^2 k^2} = O(k^{-\frac{3}{2}}) \end{aligned} \tag{3.6}$$

by Chebyshev's inequality and (3.2). Writing T for the set of $\mathscr{M}$-free numbers, we have that

$$\mu(|T| \geqslant \tfrac{1}{2}k) \geqslant 1 - O(k^{-\frac{3}{2}}). \tag{3.7}$$

Let N be the number of times that some g_j (with $s_j \in \mathscr{N}$) belongs to some ω_i (with $i \in T$). Conditional on T, N can be expressed as

$$N = \sum_{s_j \in \mathscr{N}} L_j, \tag{3.8}$$

where

$$L_j(\omega) = \begin{cases} 1 & \text{if } g_j \in \omega_i \text{ for some } i \in T, \\ 0 & \text{otherwise;} \end{cases}$$

it is easy to see that

$$\mu(L_j = 1 \mid T) = |T|/s_j. \tag{3.9}$$

Now $\Sigma_k = |T| - R$, where

$$R = |\{i \in T : g_j \in \omega_i \text{ for some } s_j \in \mathscr{N}\}|.$$

Hence

$$\begin{aligned}\mu(\Sigma_k = j) &= \sum_t \mu(\Sigma_k = j \,|\, |T| = t)\,\mu(|T| = t)\\ &\leqslant \mu(|T| < \tfrac{1}{2}k) + \sum_{t \geqslant \frac{1}{2}k} \mu(R = t-j \,|\, |T| = t)\,\mu(|T| = t)\\ &\leqslant O(k^{-\frac{3}{2}}) + \eta_k \qquad (3.10)\end{aligned}$$

by (3.7), where

$$\eta_k = \sup\{\mu(R = j \,|\, |T| = t) : 0 \leqslant j \leqslant t, \tfrac{1}{2}k \leqslant t \leqslant k\}.$$

The theorem will therefore be proved once we have shown that

$$\eta_k \leqslant \gamma k^{-\frac{1}{4}} (\log k)^{\frac{1}{2}} \qquad (3.11)$$

for some constant γ.

Fix t such that $\frac{1}{2}k \leqslant t \leqslant k$. We shall show (3.11) in two steps. First we shall explore the asymptotic distribution of N for large k, and then we shall show that the distributions of N and R are sufficiently close to one another. For ease of notation, we shall write μ^t, E^t, var^t for the corresponding mappings conditional on $|T| = t$. We have from (3.8) and (3.9) that

$$E^t(N) = t \sum_{s_j \in \mathscr{N}} \frac{1}{s_j} = t \sum_{p > \sqrt{k}} \frac{1}{p^2} \sim 2\beta \frac{\sqrt{k}}{\log k} \quad \text{as } k, t \to \infty \qquad (3.12)$$

by (3.4), where $\beta = \beta(t,k) = t/k$. Similarly,

$$\begin{aligned}\mathrm{var}^t(N) &= \sum_{s_j \in \mathscr{N}} \frac{t}{s_j}\left(1 - \frac{t}{s_j}\right) = t \sum_{p > \sqrt{k}} \frac{1}{p^2} - t^2 \sum_{p > \sqrt{k}} \frac{1}{p^4}\\ &\sim \tfrac{2}{3}(3\beta - \beta^2) \frac{\sqrt{k}}{\log k} \quad \text{as } k, t \to \infty. \qquad (3.13)\end{aligned}$$

Let Φ_{tk} be the normal distribution function with mean $E^t(N)$ and variance $\mathrm{var}^t(N)$. It is a consequence of the Berry–Esséen bounds (see [2, p. 544]) that

$$\begin{aligned}\sup_x |\mu^t(N \leqslant x) - \Phi_{tk}(x)| &\leqslant 12 \frac{E^t(N)}{\{\mathrm{var}^t(N)\}^{\frac{3}{2}}}\\ &\sim \gamma_1 \frac{\beta}{3\beta - \beta^2} \frac{(\log k)^{\frac{1}{2}}}{k^{\frac{1}{4}}}\end{aligned}$$

for some constant γ_1. Now $\frac{1}{2} \leqslant \beta \leqslant 1$, so that we may choose γ_1 such that

$$\sup_x |\mu^t(N \leqslant x) - \Phi_{tk}(x)| \leqslant \gamma_1 \frac{(\log k)^{\frac{1}{2}}}{k^{\frac{1}{4}}}, \qquad (3.14)$$

uniformly in appropriate values of t. We have from (3.14) that

$$\begin{aligned}\mu^t(N = i) &= \mu^t(N \leqslant i) - \mu^t(N \leqslant i-1)\\ &\leqslant |\mu^t(N \leqslant i) - \Phi_{tk}(i)| + |\mu^t(N \leqslant i-1) - \Phi_{tk}(i-1)|\\ &\quad + |\Phi_{tk}(i) - \Phi_{tk}(i-1)|\\ &\leqslant 2\gamma_1 \frac{(\log k)^{\frac{1}{2}}}{k^{\frac{1}{4}}} + \frac{\gamma_2}{\sqrt{\mathrm{var}^t(N)}}\\ &\leqslant \gamma_3 \frac{(\log k)^{\frac{1}{2}}}{k^{\frac{1}{4}}} \qquad (3.15)\end{aligned}$$

for some absolute constants γ_2 and γ_3.

We shall need one more estimate for N, for use later. Another application of Chebyshev's inequality yields

$$\mu^t(|N-E^t(N)| > \tfrac{1}{2}E^t(N)) \leqslant 4\frac{\mathrm{var}^t(N)}{E^t(N)^2} \sim \frac{2(3\beta-\beta^2)}{3\beta^2}\frac{\log k}{\sqrt{k}},$$

so that

$$\mu^t\left(N \geqslant \frac{4\sqrt{k}}{\log k}\right) \leqslant \gamma_4 \frac{\log k}{\sqrt{k}} \tag{3.16}$$

for some absolute constant γ_4 and all large k.

We turn our attention to the random variable R, which we think of as the number of distinct targets hit when N bullets are fired (independently of each other) at targets chosen at random from a collection of t. Clearly

$$\begin{aligned}\mu^t(R=j) &= \sum_i \mu^t(R=j\,|\,N=i)\,\mu^t(N=i) \\ &\leqslant \gamma_4\frac{\log k}{\sqrt{k}} + \sum_{j\leqslant i\leqslant (4\sqrt{k})/\log k} \mu^t(R=j\,|\,N=i)\,\mu^t(N=i)\end{aligned} \tag{3.17}$$

for all large k, by (3.16). Elementary calculations (see for example [**6**, p. 5]) show that

$$E^t(R\,|\,N=i) = i - \frac{i^2}{2t} + o(i^2/t) \tag{3.18}$$

and

$$\mathrm{var}^t(R\,|\,N=i) = \frac{i^2}{2t} + o(i^2/t). \tag{3.19}$$

Suppose henceforth that $i \leqslant 4\sqrt{k}/\log k$, and note from (3.18) and (3.19) that $H = i-R$ has conditional mean $E^t(H\,|\,N=i) = i^2/(2t)+o(i^2/t)$ and conditional variance $\mathrm{var}^t(H\,|\,N=i) = i^2/(2t)+o(i^2/t)$. Therefore

$$\begin{aligned}\mu^t(R=j\,|\,N=i) &\leqslant \mu^t(H \geqslant i-j\,|\,N=i) \\ &\leqslant \frac{i^2/t}{(i-j-1)^2} \qquad \text{if } i-j \geqslant 2 \\ &\leqslant \frac{32}{(\log k)^2}\frac{1}{(i-j-1)^2} \qquad \text{since } t \geqslant \tfrac{1}{2}k\end{aligned}$$

by Chebyshev's inequality, for all large k. Hence the summation in (3.17) is no larger than

$$\mu^t(j \leqslant N \leqslant j+1) + \frac{32}{(\log k)^2}\sum_{h=1}^{\infty}\frac{1}{h^2}\mu^t(N=j+h+1) \leqslant \gamma_5\frac{(\log k)^{\frac{1}{2}}}{k^{\frac{1}{4}}} \quad \text{for } 0 \leqslant j \leqslant t$$

by (3.15), for some constant γ_5. We obtain via (3.17) a bound of the form of (3.11) as required.

Acknowledgement. The author heard from Richard Hall of the question dealt with by Theorem 2, and acknowledges with pleasure advice on related number-theoretic matters contained in the ensuing correspondence.

References

1. P. BILLINGSLEY, *Convergence of probability measures* (Wiley, New York, 1968).
2. W. FELLER, *An introduction to probability theory and its applications*, Vol. 2, 2nd edition (Wiley, New York, 1971).
3. G. R. GRIMMETT and D. R. STIRZAKER, *Probability and random processes* (Clarendon Press, Oxford, 1982).
4. R. R. HALL, 'Squarefree numbers on short intervals', *Mathematika* 29 (1982) 7–17.
5. R. R. HALL, 'The distribution of squarefree numbers', *J. Reine Angew. Math.* 394 (1989) 107–117.
6. G. H. HARDY and E. M. WRIGHT, *An introduction to the theory of numbers* (Clarendon Press, Oxford, 1938).
7. V. F. KOLCHIN, B. A. SEVASTYANOV and V. P. CHISTYAKOV, *Random allocations* (Wiley, New York, 1978).
8. L. MIRSKY, 'Arithmetical pattern problems connected with r-free integers', *Proc. London Math. Soc.* 50 (1948) 497–508.

School of Mathematics
University Walk
Bristol BS8 1TW