

GENERALIZED SMIRNOV STATISTICS AND THE DISTRIBUTION OF PRIME FACTORS

KEVIN FORD

Dedicated to Jean-Marc Deshouillers on the occasion of his 60th birthday

ABSTRACT. We apply recent bounds of the author for generalized Smirnov statistics to the distribution of integers whose prime factors satisfy certain systems of inequalities.

1. INTRODUCTION

For a positive integer n , denote by $p_1 < p_2 < \cdots < p_{\omega(n)}$ the sequence of distinct prime factors of n . In this note, we study integers for which

$$(1.1) \quad \log_2 p_j \geq \alpha j - \beta \quad (1 \leq j \leq \omega(n))$$

or

$$(1.2) \quad \log_2 p_j \leq \alpha j + \beta \quad (1 \leq j \leq \omega(n)),$$

where $\alpha \geq 0$ and $\beta \geq 0$ and $\log_2 y$ denotes $\log \log y$. The distribution of integers satisfying (1.1) is important in the study of the distribution of divisors of integers (see [3]; Ch. 2 of [4]). We present here estimates for

$$N_k(x; \alpha, \beta) = \#\{n \leq x : \omega(n) = k, (1.1)\},$$

$$M_k(x; \alpha, \beta) = \#\{n \leq x : \omega(n) = k, (1.2)\}.$$

It is a relatively simple matter, at least heuristically, to reduce the estimation of $N_k(x; \alpha, \beta)$ and $M_k(x; \alpha, \beta)$ to the estimation of a certain probability connected to Kolmogorov-Smirnov statistics. Let us focus on the upper bound for $N_k(x; \alpha, \beta)$. If we suppose that $p_k \geq x^c$ for some small c , then for each choice of (p_1, \dots, p_{k-1}) , the number of possible p_k is $\ll x/(p_1 \cdots p_{k-1} \log x)$. Since $\sum_{p \leq y} 1/p \approx \log_2 y$, given a well-behaved function f , by partial summation we anticipate that

$$(1.3) \quad \sum_{p_1 < \cdots < p_{k-1} \leq x} \frac{f\left(\frac{\log_2 p_1}{\log_2 x}, \dots, \frac{\log_2 p_{k-1}}{\log_2 x}\right)}{p_1 \cdots p_{k-1}} \approx (\log_2 x)^{k-1} \int_{0 \leq \xi_1 \leq \cdots \leq \xi_{k-1} \leq 1} f(\boldsymbol{\xi}) d\boldsymbol{\xi},$$

where $\boldsymbol{\xi} = (\xi_1, \dots, \xi_{k-1})$.

Let U_1, \dots, U_m be independent, uniformly distributed random variables in $[0, 1]$ and let ξ_1, \dots, ξ_m be their order statistics (ξ_1 is the smallest of the U_i , ξ_2 is the next smallest, etc.). Taking $m = k-1$, the right side of (1.3) is equal to $(\log_2 x)^{k-1}/(k-1)!$ times the expectation of $f(\xi_1, \dots, \xi_{k-1})$. Letting f be 1 if (1.1) holds and 0 otherwise, the expectation of f is the probability that $\xi_j \geq (\alpha j - \beta)/\log_2 x$ for each j .

Date: April 13, 2007.

2000 Mathematics Subject Classification: Primary 11N25; Secondary 62G30.

Research supported by National Science Foundation grant DMS-0555367.

In general, let $Q_m(u, v)$, the probability that $\xi_i \geq \frac{i-u}{v}$ for $1 \leq i \leq m$. Equivalently, if $u \geq 0$ then

$$Q_m(u, v) = \text{Prob} \left(F_n(t) \leq \frac{vt + u}{n} \quad (0 \leq t \leq 1) \right),$$

where $F_m(t) = \frac{1}{m} \sum_{U_i \leq t} 1$ is the associated empirical distribution function. The first estimates for $Q_m(u, v)$ were given in 1939 by N. V. Smirnov [5], who proved for each fixed $\lambda \geq 0$ the asymptotic formula

$$(1.4) \quad Q_m(\lambda\sqrt{m}, m) \rightarrow 1 - e^{-2\lambda^2} \quad (m \rightarrow \infty).$$

The sharpest and most general bounds are due to the author [2]; see also [1]. For convenience, write $w = u + v - n$. Uniformly in $u > 0$, $w > 0$ and $m \geq 1$, we have

$$(1.5) \quad Q_m(u, v) = 1 - e^{-2uw/m} + O\left(\frac{u+w}{m}\right).$$

Moreover,

$$(1.6) \quad Q_m(u, v) \asymp \min\left(1, \frac{uw}{m}\right) \quad (u \geq 1, w \geq 1).$$

See [2] for more information about the history of such bounds and techniques for proving them. A short proof of weaker bounds is given in §11 of [3].

Returning to our heuristic estimation of $N_k(x)$ (and assuming that a similar lower bound holds), we find that

$$N_k(x) \approx \frac{x(\log_2 x)^{k-1}}{(k-1)! \log x} Q_{k-1}\left(\frac{\beta}{\alpha}, \frac{\log_2 x}{\alpha}\right).$$

We have (cf. Theorem 4 in §II.6.1 of [6])

$$(1.7) \quad \pi_k(x) := \#\{n \leq x : \omega(n) = k\} \asymp_A \frac{x(\log_2 x)^{k-1}}{(k-1)! \log x}$$

uniformly for $1 \leq k \leq A \log_2 x$, A being any fixed positive constant. Thus, we anticipate that

$$N_k(x; \alpha, \beta) \asymp Q_{k-1}\left(\frac{\beta}{\alpha}, \frac{\log_2 x}{\alpha}\right) \pi_k(x).$$

Observing that the vectors (ξ_1, \dots, ξ_m) and $(1 - \xi_m, 1 - \xi_{m-1}, \dots, 1 - \xi_1)$ have identical distributions, we have

$$Q_m(u, v) = \text{Prob} \left(\xi_i \leq \frac{u + v - m - 1 + i}{v} \quad (1 \leq j \leq m) \right).$$

Hence, we likewise anticipate that

$$M_k(x; \alpha, \beta) \asymp Q_{k-1} \left(k + \frac{\beta - \log_2 x}{\alpha}, \frac{\log_2 x}{\alpha} \right) \pi_k(x).$$

To make our heuristics rigorous, we must impose some conditions on α and β to ensure among other things that there are integers satisfying (1.1) or (1.2). To that end, we set

$$(1.8) \quad u = \frac{\beta}{\alpha}, \quad v = \frac{\log_2 x}{\alpha}, \quad w = u + v - (k-1) = \frac{\log_2 x + \beta}{\alpha} - k + 1$$

for the estimation of $N_k(x; \alpha, \beta)$ and

$$(1.9) \quad u = k + \frac{\beta - \log_2 x}{\alpha}, \quad v = \frac{\log_2 x}{\alpha}, \quad w = u + v - (k - 1) = \frac{\beta}{\alpha} + 1$$

for the estimation of $M_k(x; \alpha, \beta)$.

Theorem 1. *Suppose $\varepsilon > 0$, $A \geq 1$ and $1 \leq k \leq A \log_2 x$. Assume (1.8), $u \geq 1$, $w \geq 1 + \varepsilon$ and*

$$(1.10) \quad e^{\alpha(w-1)} - e^{\alpha(w-2)} \geq 1 + \varepsilon.$$

Then, for sufficiently large x , depending on ε and A ,

$$N_k(x; \alpha, \beta) \asymp_{\varepsilon, A} \min\left(1, \frac{uw}{k}\right) \pi_k(x),$$

the implied constants depending only on ε and A .

Theorem 2. *Suppose $A \geq 1$ and $1 \leq k \leq A \log_2 x$. Assume (1.9), $u \geq 1$, $w \geq 1$ and that for $1 \leq j \leq k$, there are at least j primes $\leq \exp \exp(\alpha j + \beta)$. Then, for sufficiently large x , depending on A ,*

$$M_k(x; \alpha, \beta) \asymp_A \min\left(1, \frac{uw}{k}\right) \pi_k(x),$$

the implied constants depending only on A .

Remarks. Inequality (1.10) is necessary, since for large k , (1.1) implies

$$\log n \geq \sum_{j=1}^k \log p_j \geq \sum_{j=1}^k e^{\alpha j - \beta} \approx \frac{e^{\alpha k - \beta}}{1 - e^{-\alpha}} = \frac{\log x}{e^{\alpha(w-1)} - e^{\alpha(w-2)}}.$$

The condition $u \geq 1$ in Theorem 1 means that there is no significant restriction on p_1 .

It is a simple matter to apply the estimates for $N_k(x; \alpha, \beta)$ and $M_k(x; \alpha, \beta)$ to problems of the distribution of prime factors of integers where $\omega(n)$ is not fixed. In the following, let $\omega(n, t)$ be the number of distinct prime factors of n which are $\leq t$. It is well-known (cf. Ch. 1 of [4]) that $\omega(n, t)$ has normal order $\log_2 t$. We estimate below the likelihood that $\omega(n, t)$ does not stray too far from $\log_2 t$ in one direction.

Corollary 1. *Uniformly for large x and $1 \leq \beta \leq \sqrt{\log_2 x}$, we have*

$$(1.11) \quad \#\{n \leq x : \forall t, 2 \leq t \leq x, \omega(n, t) \leq \log_2 t + \beta\} \asymp \frac{\beta x}{\sqrt{\log_2 x}}$$

and

$$(1.12) \quad \#\{n \leq x : \forall t, 2 \leq t \leq x, \omega(n, t) \geq \log_2 t - \beta\} \asymp \frac{\beta x}{\sqrt{\log_2 x}}$$

Proof of Corollary 1. The quantity of the left side of (1.11) is $\sum_k N_k(x; 1, \beta)$. Here $u = \beta$, $v = \log_2 x$ and $w = \log_2 x + \beta - k + 1$. By Theorem 1 and (1.7),

$$\sum_{\log_2 x - 2\sqrt{\log_2 x} \leq k \leq \log_2 x - \sqrt{\log_2 x}} N_k(x; 1, \beta) \gg \frac{\beta x}{\sqrt{\log_2 x}},$$

since $\pi_k(x) \asymp x/\sqrt{\log_2 x}$ for $|k - \log_2 x| \leq 2\sqrt{\log_2 x}$. This proves the lower bound in (1.11). For the upper bound, we note that if $k > \log_2 x + \beta$, then $N_k(x; 1, \beta) = 0$. Hence, by Theorem 1 and (1.7),

$$\begin{aligned} \sum_k N_k(x; 1, \beta) &\ll \sum_{k \leq \log_2 x + \beta - 2} \frac{\beta(\log_2 x + \beta - k + 1)}{k} \pi_k(x) + \sum_{\log_2 x + \beta - 2 < k \leq \log_2 x + \beta} \pi_k(x) \\ &\ll \frac{\beta x}{\sqrt{\log_2 x}}. \end{aligned}$$

This proves the upper bound in (1.11).

The quantity on the left side of (1.12) is $\sum_k M_k(x; 1, \beta - 1)$. Here $v = \log_2 x$, $u = \beta + k - \log_2 x$ and $w = \beta$. By Theorem 2,

$$\sum_{\log_2 x + \sqrt{\log_2 x} \leq k \leq \log_2 x + 2\sqrt{\log_2 x}} M_k(x; 1, \beta - 1) \gg \frac{\beta x}{\sqrt{\log_2 x}},$$

proving the lower bound in (1.12). Also by Theorem 2,

$$\sum_{\log_2 x - \beta + 1 < k \leq 10 \log_2 x} M_k(x; 1, \beta - 1) \ll \frac{\beta x}{\sqrt{\log_2 x}}.$$

If $\omega(n) = k > 10 \log_2 x$, then the number, $\tau(n)$, of divisors of n satisfies $\tau(n) \geq 2^{\omega(n)} \geq (\log x)^6$. Since $\sum_{n \leq x} \tau(n) \sim x \log x$, the number of $n \leq x$ with $\omega(n) > 10 \log_2 x$ is $O(x/\log^5 x)$. By (1.7), the number of $n \leq x$ with $\log_2 x - \beta - 4 < k \leq \log_2 x - \beta + 1$ is $O(x/\sqrt{\log_2 x})$. Finally, suppose $k \leq \log_2 x - \beta - 4$. The number of $n \leq x$ for which $d^2 | n$ for some $d > \log x$ is $O(x \sum_{d > \log x} 1/d^2) = O(x/\log x)$. If there is no such d , then by (1.2),

$$\log n \leq 2 \log_2 x + \sum_{j=1}^k \log p_j \leq 2 \log_2 x + \sum_{j=1}^k e^{j+\beta-1} \leq 2 \log_2 x + 2e^{k+\beta-1} \leq \frac{1}{2} \log x,$$

thus $n \leq \sqrt{x}$. This completes the proof of the upper bound in (1.12).

Our methods for proving Theorems 1 and 2 are borrowed from [3], especially sections 8, 10 and 12 therein. The tools there are adequate for making precise the heuristic argument outlined above when the function f is monotonic in each variable, even if f is discontinuous. We provide details only for Theorem 1. In lower bound for $M_k(x; \alpha, \beta)$, we may need to fix several of the smallest prime factors of n , but otherwise the details of the proof of Theorem 2 are very similar.

2. CERTAIN PARTITIONS OF THE PRIMES

We describe in this section certain partitions of the primes which will be needed in the proof of Theorems 1 and 2. The constructions are similar to those given in §4 and §8 of [3].

Let $\lambda_0 = 1.9$ and inductively define λ_j to be the largest prime such that

$$\sum_{\lambda_{j-1} < p \leq \lambda_j} \frac{1}{p} \leq 1.$$

In particular, $\lambda_1 = 3$ and $\lambda_2 = 109$. By Mertens' estimate, $\log_2 \lambda_j = j + O(1)$. Let G_j be the set of primes in $(\lambda_{j-1}, \lambda_j]$ for $j \geq 1$. Then there is an absolute constant K so that if $p \in G_j$ then $|\log_2 p - j| \leq K$.

Next, let $Q \geq e^{10}$ and $\gamma = 1/\log Q$. If $p \leq Q$, then $p^\gamma \leq e$, hence $p^\gamma \leq 1 + (e - 1)\gamma \log p$. By Merten's estimates,

$$\sum_{\substack{p \leq Q \\ f \geq 1}} \frac{1}{p^{f(1-\gamma)}} = O(1) + \sum_{p \leq Q} \left(\frac{1}{p} + (e - 1)\gamma \frac{\log p}{p} \right) = \log_2 Q + O(1).$$

It follows for an absolute constant K' , independent of Q , that the set of primes $p \leq Q$ may be partitioned into at most $\frac{1}{2} \log_2 Q + K'$ sets E_j so that (i) for each j ,

$$\sum_{\substack{p \in E_j \\ f \geq 1}} \frac{1}{p^{f(1-\gamma)}} \leq 2$$

and (ii) for $p \in E_j$, $|\log_2 p - 2j| \leq K'$. We stipulate that the above sum is ≤ 2 rather than ≤ 1 in order to accommodate the prime 2.

3. PROOF OF THEOREM 1 UPPER BOUND

Without loss of generality, suppose that k is large, $uw \leq k/10$, and $n \geq x/\log x$. We have $v \leq 1.1k$ and consequently $\alpha \geq 1/(1.1A)$. Also, by (1.1),

$$\log_2 p_k \geq \alpha k - \beta = \frac{k - u}{v} \log_2 x \geq 0.8 \log_2 x.$$

We may suppose $p_k^2 \nmid n$, as the number of $n \leq x$ with $p_k^2 | n$ is $O(x \exp(-(\log x)^{0.8})) = O(x/\log x)$. For brevity, write $x_\ell = x^{1/e^\ell}$. For some integer ℓ satisfying $\ell \geq 0$ and $\exp(\alpha k - \beta) \leq x_\ell$, we have $x_{\ell+1} < p_k \leq x_\ell$. With ℓ fixed, given p_1, \dots, p_{k-1} with exponents f_1, \dots, f_{k-1} , the number of possibilities for p_k is

$$\ll \frac{x}{p_1^{f_1} \cdots p_{k-1}^{f_{k-1}} \log x_\ell} \ll \frac{x^{1-\gamma/2} e^\ell}{(p_1^{f_1} \cdots p_{k-1}^{f_{k-1}})^{1-\gamma} \log x},$$

where $\gamma = 1/\log x_\ell$. This follows for $\ell \geq 1$ since $p_1^{f_1} \cdots p_{k-1}^{f_{k-1}} \geq x/(p_k \log x) > x^{1/2}$. We conclude that

$$(3.13) \quad N_k(x; \alpha, \beta) \ll \frac{x}{\log x} \sum_{\ell} e^{\ell - \frac{1}{2}e^\ell} \sum_{\substack{p_1 < \cdots < p_{k-1} \leq x_\ell \\ f_1, \dots, f_{k-1} \geq 1 \\ (1.1)}} (p_1^{f_1} \cdots p_{k-1}^{f_{k-1}})^{1-\gamma}.$$

Consider the intervals E_j defined in the previous section corresponding to $Q = x_\ell$. Put $J = \lfloor \frac{1}{2} \log_2 x_\ell + K' \rfloor$ and define j_1, \dots, j_{k-1} by $p_i \in E_{j_i}$. Let \mathcal{J} denote the set of tuples (j_1, \dots, j_{k-1}) so that $1 \leq j_1 \leq \cdots \leq j_{k-1} \leq J$ and such that $j_i \geq 2(\alpha i - \beta - K')$ for every i . Given p_1, \dots, p_{k-1} , let b_j be the number of p_i in E_j , for $1 \leq j \leq J$. The contribution to the inner sum of (3.13) from those tuple of primes with a fixed (j_1, \dots, j_{k-1}) is

$$\begin{aligned} &\leq \prod_{j=1}^J \frac{1}{b_j!} \left(\sum_{p \in E_j, f \geq 1} \frac{1}{p^{f(1-\gamma)}} \right)^{b_j} \\ &\leq \frac{2^{k-1}}{b_1! \cdots b_J!}. \end{aligned}$$

We observe that $1/(b_1! \cdots b_J!)$ is the volume of the region $(y_1, \dots, y_{k-1}) \in \mathbb{R}^{k-1}$ satisfying $0 \leq y_1 \cdots \leq y_{k-1} \leq J$ and $j_i - 1 < y_i \leq j_i$ for each i (there are b_j numbers y_i in each interval

$(j-1, j]$). Making the change of variables $\xi_i = y_i/J$ and summing over all possible vectors $(j_1, \dots, j_{k-1}) \in \mathcal{J}$, we find that the inner sum in (3.13) is

$$\begin{aligned} &\leq (2J)^{k-1} \text{Vol} \left\{ 0 \leq \xi_1 \leq \dots \leq \xi_{k-1} \leq 1 : \xi_i \geq \frac{2(\alpha i - \beta - K' - 1)}{J} \ (1 \leq i \leq k-1) \right\} \\ &\leq \frac{(\log_2 x + 2K')^{k-1}}{(k-1)!} Q_{k-1} \left(\frac{\beta + K' + 1}{\alpha}, \frac{J}{2\alpha} \right) \\ &\ll_A \frac{(\log_2 x)^{k-1}}{(k-1)!} \frac{(u + K'/\alpha)(w + K'/\alpha)}{k} \\ &\ll_A \frac{uw}{k} \frac{(\log_2 x)^{k-1}}{(k-1)!}, \end{aligned}$$

where we have used (1.6). By (3.13), summing on ℓ and using (1.7) completes the proof.

4. PROOF OF THEOREM 1 LOWER BOUND

First, we assume $k \geq 2$, since if $k = 1$ then $N_1(x; \alpha, \beta) = \pi_1(x) + O(\log x)$ trivially as $\beta \geq \alpha$ (we assume $u \geq 1$; powers of 2 may not be counted in $N_1(x; \alpha, \beta)$). Also, we may assume that $\alpha \geq 1/2A$. If $\alpha < 1/2A$, then $N_k(x; \alpha, \beta) \geq N_k(x; 1/2A, 1/2A)$ and we prove below that $N_k(x; 1/2A, 1/2A) \gg \pi_k(x)$ (here $u = 1$, $v \geq 2k$ and $w \geq k$).

Let T be a sufficiently large constant, depending on ε and A , and put

$$C = e^{3T+2K+10}.$$

We first prove the theorem in the case that

$$(4.14) \quad e^{\alpha(w-1)} - e^{\alpha(w-2)} \geq C.$$

Notice that

$$(4.15) \quad \alpha j - \beta = \log_2 x - \alpha(w + k - 1 - j).$$

In particular,

$$\alpha k - \beta = \log_2 x - \alpha(w - 1) \leq \log_2 x - \log C.$$

Let $J = \lfloor \log_2 x - K - \log T - 2 \rfloor$. Recall the definition of the numbers λ_j and sets G_j from section 2. Consider squarefree n satisfying (1.1), with $p_{k-1} \leq \lambda_J$ and for which

$$p_1 \cdots p_{k-1} \leq x^{1/2}.$$

Also take p_k so that $x/2 < n \leq x$. Given p_1, \dots, p_{k-1} , the number of possible p_k is $\gg x/(p_1 \cdots p_{k-1} \log x)$. Put $b_1 = \dots = b_{T-1} = 0$ and for $T \leq j \leq J$, suppose $b_j \leq \min(T(j - T - 1), T(J - j + 1))$ for each j . Suppose there are exactly b_j primes p_i in the set G_j for $1 \leq j \leq J$. By the definition of J ,

$$\sum_{i=1}^{k-1} \log p_i \leq T e^{J+K} \sum_{r=1}^{k-1} r e^{1-r} < 3T e^{J+K} \leq \frac{1}{2} \log x,$$

as required. Define the numbers j_i by $p_i \in G_{j_i}$. The inequalities (1.1) will be satisfied if

$$(4.16) \quad j_i \geq \alpha i - \beta + K \quad (1 \leq i \leq k-1).$$

This is possible since

$$\alpha(k-1) - \beta = \log_2 x - \alpha w \leq \log_2 x - 2K - 3T - 10 < J - T - 1.$$

With (j_1, \dots, j_{k-1}) fixed (so that b_1, \dots, b_J are fixed), the sum of $1/p_1 \cdots p_{k-1}$ is

$$\begin{aligned}
&= \prod_{j=T}^J \frac{1}{b_j!} \left(\sum_{p_1 \in G_j} \frac{1}{p_1} \sum_{\substack{p_2 \in G_j \\ p_2 \neq p_1}} \frac{1}{p_2} \cdots \sum_{\substack{p_{b_j} \in G_j \\ p_{b_j} \notin \{p_1, \dots, p_{b_j-1}\}}} \frac{1}{p_{b_j}} \right) \\
&\geq \prod_{j=T}^J \frac{1}{b_j!} \left(\sum_{p \in G_j} \frac{1}{p} - \frac{b_j - 1}{\lambda_{j-1}} \right)^{b_j} \\
&\geq \prod_{j=T}^J \frac{1}{b_j!} \left(1 - \frac{b_j}{\lambda_{j-1}} \right)^{b_j} \\
&\geq \prod_{j=T}^J \frac{1}{b_j!} \left(1 - \frac{T(j-T+1)}{\exp \exp(j-1-K)} \right)^{T(j-T+1)} \\
&\geq \frac{1/2}{b_T! \cdots b_J!}
\end{aligned}$$

if T is large enough. The right side is $1/2$ of the volume of the region of $(y_1, \dots, y_{k-1}) \in \mathbb{R}^{k-1}$ satisfying $0 \leq y_1 \leq \cdots \leq y_{k-1} \leq J - T + 1$ and $j_i - T \leq y_i \leq j_i + 1 - T$ for each i . Set $H = J - T + 1$. Assume that

$$(4.17) \quad j_{mT+1} \geq T + m, \quad j_{k-1-mT} \leq J - m \quad (\text{integers } m \geq 1),$$

so that $b_j \leq \min(T(j-T+1), T(J-j+1))$ for each j . Making the substitution $\xi_i = y_i/H$ and summing over all tuples (j_1, \dots, j_{k-1}) yields

$$(4.18) \quad N_k(x; \alpha, \beta) \gg \frac{xH^k}{\log x} \text{Vol}(R) \gg_A \frac{x(\log_2 x)^k}{\log x} \text{Vol}(R),$$

where, by (4.16) and (4.17), R is the set of $\boldsymbol{\xi}$ satisfying (i) $0 \leq \xi_1 \leq \cdots \leq \xi_{k-1} \leq 1$, $\xi_i \geq (\alpha i - \beta + K - T)/H$ for each i , (ii) $\xi_{mT+1} \geq m/H$ and $\xi_{k-1-mT} \leq 1 - m/H$ for each positive integer m .

It remains to estimate from below the volume of R . Let S be the set of $\boldsymbol{\xi}$ satisfying (i), so that

$$\text{Vol}(S) = \frac{Q_{k-1}(\mu, \nu)}{(k-1)!}, \quad \mu = \frac{\beta + T - K}{\alpha}, \quad \nu = \frac{H}{\alpha}.$$

If $T \geq K$, then $\mu \asymp_A u$. By the definition of C and J ,

$$\mu + \nu - (k-1) = \frac{J - K + 1 + \beta}{\alpha} - (k-1) = w - \frac{\log T + O(1)}{\alpha} \geq \frac{w}{2} \geq 1.$$

Hence, by (1.6),

$$(4.19) \quad \text{Vol}(S) \gg \frac{f}{(k-1)!}, \quad f = \min(1, uw/k).$$

The implied constant in (4.19) does not depend on T , but the inequality does require that T be sufficiently large.

For a positive integer m , let

$$\begin{aligned}
V_1(m) &= \text{Vol}\{\boldsymbol{\xi} \in S : \xi_{mT+1} < m/H\}, \\
V_2(m) &= \text{Vol}\{\boldsymbol{\xi} \in S : \xi_{k-1-mT} > 1 - m/H\}.
\end{aligned}$$

We have by (1.6),

$$\begin{aligned}
V_1(m) &\leq \frac{(m/H)^{mT+1}}{(mT+1)!} \text{Vol}\{0 \leq \xi_{mT+2} \leq \cdots \leq \xi_{k-1} \leq 1 : \xi_i \geq \frac{i-\mu}{\nu} \ (mT+2 \leq i \leq k-1)\} \\
&= \frac{(m/H)^{mT+1}}{(mT+1)!} \frac{Q_{k-2-mT}(\mu - (mT+1), \nu)}{(k-2-mT)!} \\
&\ll \frac{(m/H)^{mT+1}}{(mT+1)!} \frac{\mu(\mu + \nu - (k-1))}{(k-mT)(k-2-mT)!} \\
&\ll \frac{fk(m/H)^{mT+1}}{(k-mT)(mT+1)!(k-2-mT)!} \\
&\leq \frac{f}{(k-1)!} \frac{(km/H)^{mT+1}}{(mT+1)!} \frac{k}{k-mT}.
\end{aligned}$$

Since $k/H \ll_A 1$ and $r! \geq (r/e)^r$, it follows from (4.19) that for large enough T ,

$$\sum_m V_1(m) \leq \frac{1}{4} \text{Vol}(S).$$

Similarly,

$$V_2(m) \leq \frac{Q_{k-2-mT}(\mu, \nu)}{(k-2-mT)!} \frac{(m/H)^{mT+1}}{(mT+1)!}.$$

By (1.6),

$$Q_{k-2-mT}(\mu, \nu) \ll \min\left(1, \frac{\mu(\mu + \nu - (k-1) + mT + 1)}{k-mT}\right) \ll \frac{mTkf}{k-mT}.$$

Hence, if T is large enough then

$$\sum_m V_2(m) \leq \frac{1}{4} \text{Vol}(S).$$

We therefore have, for T large enough,

$$\text{Vol}(R) \geq \text{Vol}(S) - \sum_{m \geq 1} (V_1(m) + V_2(m)) \gg_A \frac{f}{(k-1)!}.$$

Together with (4.18) and (1.7), this completes the proof under the assumption (4.14).

It remains to consider the case

$$1 + \varepsilon \leq e^{\alpha(w-1)} - e^{\alpha(w-2)} \leq C.$$

Since $w \geq 1 + \varepsilon$ and $\alpha \geq 1/2A$, we find that $\alpha \ll_{\varepsilon, A} 1$ and $w \ll_{\varepsilon, A} 1$. Hence, is x is large enough,

$$k = u + v - w + 1 \geq v - w \geq \frac{\log_2 x}{4A}.$$

Let B be a large integer depending on ε . Suppose that

$$(4.20) \quad \alpha j - \beta \leq \log_2 p_j \leq \alpha j - \beta + \log(1 + \varepsilon/2) \quad (k - B \leq j \leq k - 1)$$

Then, by (4.15),

$$\begin{aligned} \sum_{j=k-B}^{k-1} \log p_j &\leq (1 + \varepsilon/2) (e^{-\alpha w} + e^{-\alpha(w+1)} + \dots + e^{-\alpha(w+B-1)}) \log x \\ &< (1 + \varepsilon/2) \left(\frac{1}{e^{\alpha(w-1)} - e^{\alpha(w-2)}} - e^{-\alpha(w-1)} \right) \log x. \end{aligned}$$

Assume also that

$$(4.21) \quad \sum_{j=1}^{k-B-1} \log p_j \leq \frac{\varepsilon/2}{e^{\alpha(w-1)} - e^{\alpha(w-2)}} \log x.$$

If in addition $\alpha k - \beta \leq \log_2 p_k \leq \alpha k - \beta + \log(1 + \varepsilon/2)$, then by (1.10),

$$\log n = \sum_{j=1}^k \log p_j \leq \frac{\varepsilon/2 + 1 + \varepsilon/2}{e^{\alpha(w-1)} - e^{\alpha(w-2)}} \log x \leq \log x,$$

as required. Thus, given p_1, \dots, p_{k-1} satisfying (4.20) and (4.21), the number of p_k is $\gg x/(p_1 \cdots p_{k-1} \log x)$. If B is large enough, there is great flexibility in choosing p_1, \dots, p_{k-B-1} , since by (4.15),

$$\sum_{j=1}^{k-B-1} e^{\alpha j - \beta} \leq \frac{e^{-\alpha(B+1)}}{e^{\alpha(w-1)} - e^{\alpha(w-2)}} \log x,$$

which is small compared with the right side of (4.21). By the same argument used to give a lower bound for the sum of $1/(p_1 \cdots p_{k-1})$ under the assumption (4.14), we obtain

$$\sum_{p_1, \dots, p_{k-B-1}} \frac{1}{p_1 \cdots p_{k-B-1}} \gg_{A, \varepsilon} \frac{f(\log_2 x)^{k-B-1}}{(k-B-1)!}.$$

Also, since $k \gg_A \log_2 x$, we have

$$\sum_{p_{k-B}, \dots, p_{k-1}} \frac{1}{p_{k-B} \cdots p_{k-1}} \gg_{\varepsilon, B} 1 \gg_{\varepsilon, A} (\log_2 x)^B \frac{(k-B-1)!}{(k-1)!}.$$

The proof is again completed by applying (1.7).

REFERENCES

- [1] K. Ford. From Kolmogorov's theorem on empirical distribution to number theory. In *Kolmogorov's legacy in mathematics*. Editions Belin / Springer-Verlag, Paris, 2007.
- [2] K. Ford. Sharp probability estimates for generalized Smirnov statistics. *Monatshefte Math. (to appear)*, 2007. ArXiv math.PR/0609224.
- [3] K. Ford. The distribution of integers with a divisor in a given interval. *Annals of Math. (to appear)*, 2008. ArXiv math.NT/0401223.
- [4] R. R. Hall and G. Tenenbaum. *Divisors*, volume 90 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1988.
- [5] N. V. Smirnov (Smirnoff). Sur les écarts de la courbe de distribution empirique. *Rec. Math. N.S.*, 6(48):3-26, 1939. Russian. French summary.

- [6] G. Tenenbaum. *Introduction to analytic and probabilistic number theory*, volume 46 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995. Translated from the second French edition (1995) by C. B. Thomas.

DEPARTMENT OF MATHEMATICS, 1409 WEST GREEN STREET, UNIVERSITY OF ILLINOIS AT URBANA-
CHAMPAIGN, URBANA, IL 61801, USA
E-mail address: `ford@math.uiuc.edu`