Functiones et Approximatio
XXXVII.1 (2007), 119–129

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.
Keywords: Smirnov statistics, prime factors.

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

$$\log_2 p_j \geqslant \alpha j - \beta \quad (1 \leqslant j \leqslant \omega(n)) \tag{1.1}$$

or

$$\log_2 p_j \leqslant \alpha j + \beta \quad (1 \leqslant j \leqslant \omega(n)), \tag{1.2}$$

where $\alpha \geqslant 0$ 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

$$N_k(x;\alpha,\beta) = \#\{n \leqslant x : \omega(n) = k,\ (1.1)\},$$
$$M_k(x;\alpha,\beta) = \#\{n \leqslant x : \omega(n) = k,\ (1.2)\}.$$

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 $p_k \geqslant x^c$ for some small c, then for each choice of $(p_1,\dots,p_{k-1})$,

---

2000 Mathematics Subject Classification: primary 11N25; secondary 62G30.
Research supported by National Science Foundation grant DMS-0555367.

the number of possible p_k is $\ll x/(p_1 \cdots p_{k-1} \log x)$. Since $\sum_{p \leqslant y} 1/p \approx \log_2 y$, given a well-behaved function f, by partial summation we anticipate that

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

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

Let $U_1, \ldots, U_m$ be independent, uniformly distributed random variables in $[0,1]$ and let $\xi_1, \ldots, \xi_m$ 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 $f(\xi_1, \ldots, \xi_{k-1})$. Letting f be 1 if (1.1) holds and 0 otherwise, the expectation of f is the probability that $\xi_j \geqslant (\alpha j - \beta)/\log_2 x$ for each j.

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

$$Q_m(u,v) = \text{Prob}\left(F_m(t) \leqslant \frac{vt+u}{m} \quad (0 \leqslant t \leqslant 1)\right),$$

where $F_m(t) = \frac{1}{m}\sum_{U_i \leqslant t} 1$ 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 $\lambda \geqslant 0$ the asymptotic formula

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

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

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

Moreover,

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

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])

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

uniformly for $1 \leqslant k \leqslant A \log_2 x$, 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 $(\xi_1, \dots, \xi_m)$ and $(1-\xi_m, 1-\xi_{m-1}, \dots, 1-\xi_1)$ have identical distributions, we have

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

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

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

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

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

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

Theorem 1. *Suppose* $\varepsilon > 0$, $A \geqslant 1$ *and* $1 \leqslant k \leqslant A \log_2 x$. *Assume* (1.8), $\beta \geqslant 0$, $\alpha - \beta \leqslant A$, $w \geqslant 1 + \varepsilon$ *and*

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

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

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

the implied constants depending only on ε *and* A.

Theorem 2. *Suppose* $A \geqslant 1$ *and* $1 \leqslant k \leqslant A \log_2 x$. *Assume* (1.9), $u \geqslant 1$, $w \geqslant 0$ *and that for* $1 \leqslant j \leqslant k$, *there are at least* j *primes* $\leqslant \exp\exp(\alpha j + \beta)$. *Then, for sufficiently large* x, *depending on* A,

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

the implied constants depending only on A.

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

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

The condition $\alpha-\beta \leqslant A$ 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 $\leqslant t$. 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 $0 \leqslant \beta \leqslant \sqrt{\log_2 x}$, we have*

$$\#\{n \leqslant x : \forall t, 2 \leqslant t \leqslant x, \omega(n,t) \leqslant \max(0, \log_2 t + \beta)\} \asymp \frac{(\beta+1)x}{\sqrt{\log_2 x}} \tag{1.11}$$

and

$$\#\{n \leqslant x : \forall t, 2 \leqslant t \leqslant x, \omega(n,t) \geqslant \log_2 t - \beta\} \asymp \frac{(\beta+1)x}{\sqrt{\log_2 x}} \tag{1.12}$$

Proof. 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),

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

since $\pi_k(x) \asymp x/\sqrt{\log_2 x}$ for $|k-\log_2 x| \leqslant 2\sqrt{\log_2 x}$. 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),

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

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,

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

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

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

If $\omega(n)=k>10\log_2 x$, then the number, $\tau(n)$, of divisors of n satisfies $\tau(n)\geqslant 2^{\omega(n)}\geqslant(\log x)^6$. Since $\sum_{n\leqslant x}\tau(n)\sim x\log x$, the number of $n\leqslant x$ with $\omega(n)>10\log_2 x$ is $O(x/\log^5 x)$. By (1.7), the number of $n\leqslant x$ with $\log_2 x-\beta-4<k\leqslant\log_2 x-\beta+1$ is $O(x/\sqrt{\log_2 x})$. Finally, suppose $k\leqslant\log_2 x-\beta-4$. The number of $n\leqslant x$ 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),

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

thus $n\leqslant\sqrt{x}$. 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

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

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 $j\geqslant 1$. Then there is an absolute constant K so that if $p\in G_j$ then $|\log_2 p-j|\leqslant K$.

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

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

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

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

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

3. Proof of Theorem 1 upper bound

Without loss of generality, suppose that k is large, $(u+1)w \leqslant k/10$, and $n \geqslant x/\log x$. We have $v \leqslant 1.1k$ and consequently $\alpha \geqslant 1/(1.1A)$. Also, by (1.1),

$$\log_2 p_k \geqslant \alpha k - \beta = \frac{k-u}{v}\log_2 x \geqslant \frac{9}{11}\log_2 x.$$

We may suppose $p_k^2 \nmid n$, as the number of $n \leqslant x$ with $p_k^2 | n$ is $O(x\exp(-(\log x)^{\frac{9}{11}}))$ $= O(\pi_k(x)/k)$. For brevity, write $x_\ell = x^{1/e^\ell}$. For some integer ℓ satisfying $\ell \geqslant 0$ and $\exp\exp(\alpha k - \beta) \leqslant x_\ell$, we have $x_{\ell+1} < p_k \leqslant x_\ell$. With ℓ fixed, given $p_1, \ldots, p_{k-1}$ with exponents $f_1, \ldots, f_{k-1}$, 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 $\ell \geqslant 1$ from $p_1^{f_1}\cdots p_{k-1}^{f_{k-1}} \geqslant x/(p_k \log x) > x^{1/2}$. We conclude that

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

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

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

We observe that $1/(b_1!\cdots b_J!)$ is the volume of the region $(y_1,\cdots,y_{k-1})\in\mathbb{R}^{k-1}$ satisfying $0\leqslant y_1\leqslant\cdots\leqslant y_{k-1}\leqslant J$ and $j_i-1<y_i\leqslant j_i$ 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 $(j_1,\ldots,j_{k-1})\in\mathcal{J}$, we find that the inner sum in (3.13) is

$$\leqslant(2J)^{k-1}\mathrm{Vol}\left\{0\leqslant\xi_1\leqslant\cdots\leqslant\xi_{k-1}\leqslant1:\ \xi_i\geqslant\frac{(\alpha i-\beta-K'-A-2)}{2J}\ (1\leqslant i\leqslant k-1)\right\}$$
$$\leqslant\frac{(\log_2 x+2K')^{k-1}}{(k-1)!}Q_{k-1}\left(\frac{\beta+K'+A+2}{\alpha},\frac{2J}{\alpha}\right)$$
$$\ll_A\frac{(\log_2 x)^{k-1}}{(k-1)!}\ \frac{(u+1)w}{k},$$

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 $k\geqslant2$, since if $k=1$ then $N_1(x;\alpha,\beta)=\pi_1(x)+O(\log x)$ trivially as $A+\beta\geqslant\alpha$ (powers of primes $\leqslant e^{\alpha-\beta}$ are not counted in $N_1(x;\alpha,\beta)$). Also, we may assume that $\alpha\geqslant1/2A$. If $\alpha<1/2A$, then $N_k(x;\alpha,\beta)\geqslant N_k(x;1/2A,0)$ and we prove below that $N_k(x;1/2A,0)\gg\pi_k(x)$ (here $u=0$, $v\geqslant2k$ and $w\geqslant k$).

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

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

Notice that

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

In particular,

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

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 $p_{k-1}\leqslant\lambda_J$ and for which

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

Also take p_k so that $x/2<n\leqslant x$. Given $p_1,\ldots,p_{k-1}$, the number of possible p_k is $\gg x/(p_1\cdots p_{k-1}\log x)$. Put $b_1=\cdots=b_{T-1}=0$ and for $T\leqslant j\leqslant J$, suppose

$b_j \leqslant \min(T(j-T-1), T(J-j+1))$. Suppose there are exactly b_j primes p_i in the set G_j for $1 \leqslant j \leqslant J$. By the definition of J,

$$\sum_{i=1}^{k-1} \log p_i \leqslant Te^{J+K} \sum_{r=1}^{k-1} re^{1-r} < 3Te^{J+K} \leqslant \frac{1}{2}\log x,$$

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

$$j_i \geqslant \alpha i - \beta + K \qquad (1 \leqslant i \leqslant k-1). \tag{4.16}$$

This is possible since by (4.14)

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

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

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

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

$$j_{mT+1} \geqslant T + m, \quad j_{k-1-mT} \leqslant J - m \qquad (\text{integers } m \geqslant 1), \tag{4.17}$$

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

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

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

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

If $T \geqslant K + A$, then $\mu \asymp_A (u+1)$. By the definition of C and J, if T is large enough then

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

Hence, by (1.6),

$$\mathrm{Vol}(S) \gg \frac{f}{(k-1)!}, \qquad f = \min(1, (u+1)w/k). \tag{4.19}$$

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

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

We have by (1.6),

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

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

$$\sum_m V_1(m) \leqslant \frac{1}{4}\mathrm{Vol}(S).$$

Similarly,

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

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

$$\sum_m V_2(m) \leqslant \frac{1}{4}\mathrm{Vol}(S).$$

We therefore have, for T large enough,

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

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

It remains to consider the case

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

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

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

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

$$\alpha j - \beta \leqslant \log_2 p_j \leqslant \alpha j - \beta + \log(1+\varepsilon/2) \qquad (k-B \leqslant j \leqslant k-1) \tag{4.20}$$

Then, by (4.15),

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

Assume also that

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

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

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

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),

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

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 empirigue, *Rec. Math. N.S.*, 6(48):3–36, 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.

Address: Department of Mathematics, 1409 West Green Street, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
E-mail: `ford@math.uiuc.edu`
Received: 28 November 2006; **revised:** 13 April 2007