SHARP PROBABILITY ESTIMATES FOR GENERALIZED SMIRNOV STATISTICS

KEVIN FORD

Dedicated to the memory of Walter Philipp

Abstract. We give sharp, uniform estimates for the probability that the empirical distribution function for n uniform-$[0,1]$ random variables stays to one side of a given line.

1. Introduction

Let $U_1, \dots, U_n$ be independent, uniformly distributed random variables in $[0,1]$ and let $u > 0$, $v > 0$. Our goal is to estimate

$$Q_n(u,v) = \mathbf{P}\left(F_n(t) \le \frac{vt+u}{n} \;\; (0 \le t \le 1) \right),$$

where $F_n(t) = \frac{1}{n}\sum_{U_i \le t} 1$ is the associated empirical distribution function. In 1939, N. V. Smirnov introduced the statistic $D_n^+ = \sqrt{n}\sup_{0\le t\le 1}(F_n(t)-t)$ and proved [14] for each fixed $\lambda \ge 0$ the asymptotic formula

$$\mathbf{P}(D_n^+ \le \lambda) = Q_n(\lambda\sqrt{n}, n) \to 1 - e^{-2\lambda^2} \qquad (n \to \infty). \tag{1.1}$$

When $\lambda_0 \le \lambda = O(n^{1/6})$ with *fixed* $\lambda_0 > 0$, sharper forms of (1.1) have been proven by a number of people (e.g. [10]; see also Ch. 9 of [13]), in particular

$$\mathbf{P}(D_n^+ \le \lambda) = 1 - e^{-2\lambda^2}\left(1 - \frac{2\lambda}{3n^{1/2}} + O\left(\frac{\lambda^4+1}{n}\right)\right). \tag{1.2}$$

Here and throughout the Landau $O-$symbol has its usual meaning: $f(\cdot) = O(g(\cdot))$ means $|f| \le cg$ for some constant c, which is independent of the inputs to the function f. Also, $f \ll g$ means $f = O(g)$ and $f \asymp g$ means $f = O(g)$ and $g = O(f)$.

One may ask about the behavior of $Q_n(u,v)$ for a wider range of the variables u, v. The strong Komlós-Major-Tusnády theorem [9] implies

$$|F_n(t) - t - n^{-1/2}B_n(t)| \ll \frac{\log n}{n} \qquad (0 \le t \le 1)$$

with probability $\ge 1 - O(1/n)$, where $B_n(t)$ is a Brownian bridge process. The order $\frac{\log n}{n}$ on the right side is also best possible [9] (see also Ch. 4 of [1]). Since

$$\mathbf{P}\left(\sup_{0\le t\le 1}(B_n(t) - (at+b)) \le 0\right) = 1 - e^{-2b(a+b)},$$

and writing

$$w = u + v - n,$$

Date: November 27, 2007.

2000 Mathematics Subject Classification: Primary 62G30, 60G50.

Research supported by National Science Foundation grants DMS-0301083 and DMS-0555367.

the KMT theorem implies the uniform estimate

$$\begin{aligned}Q_n(u,v) &= O\left(\frac{1}{n}\right) + 1 - e^{-\frac{2(u+O(\log n))(w+O(\log n))}{n}} \\ &= 1 - e^{-2uw/n} + O\left(\frac{(u+w+\log n)\log n}{n}\right).\end{aligned} \tag{1.3}$$

This gives an asymptotic for $Q_n(u,v)$ provided $\frac{u}{\log n} \to \infty$, $\frac{w}{\log n} \to \infty$, and $u+w = o(n/\log n)$ as $n \to \infty$. In the author's recent paper [6] on the distribution of divisors of integers, sharper information was needed for very small u and w. That paper includes a short proof of the crude bound $Q_n(u,v) \ll \frac{(u+1)(w+1)^2}{n}$ uniformly in $n \ge 1$, $u \ge 0$, and $w \ge 0$.

By using different methods, we prove here new uniform estimates, which essentially remove the logarithm terms from the right side of (1.3).

Theorem 1. *Uniformly in $u > 0$, $w > 0$ and $n \ge 1$, we have*

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

In particular, if $u \to \infty$, $w \to \infty$ and $u + w = o(n)$ as $n \to \infty$, then

$$\frac{Q_n(u,v)}{1 - e^{-\frac{2uw}{n}}} \to 1.$$

2. A random walk with a barrier

Exact formulas for $Q_n(u,v)$ are known, which we record below.

Lemma 2.1. *Assume $n \ge 1$ and $v > 0$. Then*

(i) *If $n - v < u \le 1$, then $Q_n(u,v) = \frac{w}{v}(1+u/v)^{n-1}$;*
(ii) *If $n - v < u < n$ and $u \ge 0$, then*

$$Q_n(u,v) = 1 - \frac{w}{v^n} \sum_{u<j\le n} \binom{n}{j} (v+u-j)^{n-j-1}(j-u)^j.$$

Formula (i) is due to H. E. Daniels [2] and (ii) is due to R. Pyke [11]. The case $v = n$ in (ii) was earlier proved by Smirnov [14]. Starting with (ii), one may use a more complicated version of the complex analytic method of Lauwerier [10] to prove Theorem 1. This was carried out in an early version of the author's paper [6], a sketch of which may be found in [3] (the English paper [5] includes a sketch of the argument below). We present below an elementary, probabilistic proof of Theorem 1. Rather than work with (ii), we reinterpret $Q_n(u,v)$ in terms of a random walk.

Lemma 2.2. *Let $X_1, \ldots, X_{n+1}$ be independent random variables, each with density function e^{x-1} if $x \le 1$ and 0 if $x > 1$. Put $S_0 = 0$ and $S_j = X_1 + \cdots + X_j$ for $j \ge 1$. Then*

$$Q_n(u,v) = \mathbf{P}\left[\max_{0\le j\le n} S_j < u \,\middle|\, S_{n+1} = n+1-v\right].$$

Proof. Let $Y_1, \cdots, Y_{n+1}$ be independent random variables with exponential distribution, and let $W_k = Y_1 + \cdots + Y_k$ for $1 \le k \le n+1$. Let $\xi_1, \ldots, \xi_n$ be the order statistics of $U_1, \ldots, U_n$, so that $Q_n(u,v)$ is the probability that $\xi_j \ge \frac{j-u}{v}$ for every j. By a well-known theorem of Rényi

[12], the vectors $(\xi_1, \ldots, \xi_n)$ and $(W_1/W_{n+1}, \ldots, W_n/W_{n+1})$ have identical distributions. Similarly, given that $W_{n+1} = v$, the probability density function of the vector $(W_1/v, \ldots, W_n/v)$ is identically $n!$ on the set $\{(x_1, \ldots, x_n) : 0 \le x_1 \le \cdots \le x_n \le 1\}$. Therefore,

$$Q_n(u, v) = \mathbf{P}\left[\min_{1\le i\le n}(W_i - i) \ge -u \,\middle|\, W_{n+1} = v\right].$$

Putting $X_i = 1 - Y_i$ completes the proof. □

The sequence $0, S_1, S_2, \ldots$ can be thought of as a recurrent random walk on the real line, with $Q_n(u, v)$ being the probability that the walk does not cross a barrier at the point u given that it ends at the point $n + 1 - v$ after $n + 1$ steps. A similar quantity may be defined for a random walk with the X_i having a different distribution. In the paper [4], an analog of Theorem 1 is proven for a general walk whose steps X_i have a continuous or lattice distribution, but valid in a more limited range of the variables. More specifically, under appropriate conditions on X_i, we prove that

$$\mathbf{P}\left[\max_{0\le j\le n-1} S_j < y \,\middle|\, S_n = y - z\right] = 1 - e^{-2yz/n} + O\left(\frac{y + z + 1}{n}\right)$$

uniformly for $0 \le y \le c\sqrt{n}$, $0 \le z \le c\sqrt{n}$ (c being any fixed constant).

Kolmogorov used a relation similar to that in Lemma 2.2 in his seminal 1933 paper [8] on the distribution of the statistic

$$D_n = \sqrt{n} \sup_{0\le t\le 1} |F_n(t) - t|.$$

Specifically, let $\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_n$ be independent random variables with discrete distribution

$$\mathbf{P}[\widetilde{X}_j = r - 1] = \frac{e^{-1}}{r!} \qquad (r = 0, 1, 2, \ldots)$$

and let $\widetilde{S}_j = \widetilde{X}_1 + \cdots + \widetilde{X}_j$ for $j \ge 1$. Like the variables X_i in Lemma 2.2, each $\widetilde{X}_i$ has mean 0 and variance 1. Kolmogorov proved that for integers $u \ge 1$,

$$\begin{aligned}\mathbf{P}(\sup_{0\le t\le 1} |F_n(t) - t| \le u/n) &= \frac{n!e^n}{n^n}\mathbf{P}\left(\max_{0\le j\le n-1} |\widetilde{S}_j| < u, \widetilde{S}_n = 0\right)\\ &= \mathbf{P}\left(\max_{0\le j\le n-1} |\widetilde{S}_j| < u \,\middle|\, \widetilde{S}_n = 0\right).\end{aligned}$$

Small modifications to the proof yield, for *integers* $u \ge 1$ and for $n \ge 2$, that

$$Q_n(u, n) = \mathbf{P}\left(\max_{0\le j\le n-1} \widetilde{S}_j < u \,\middle|\, \widetilde{S}_n = 0\right).$$

When $v \ne n$, however, it does not seem feasible to express $Q_n(u, v)$ in terms of the variables $\widetilde{S}_j$.

Let f_n be the density function for S_n ($n = 1, 2, \ldots$). The Central Limit Theorem for densities (e.g., Theorem 1 in §46 of [7]) implies that for large n and $|x| \ll \sqrt{n}$, $f_n(x) \approx (2\pi n)^{-1/2} e^{-x^2/2n}$. However, there are asymmetries in the distribution for $|x| > \sqrt{n}$. We have

$$f_n(x) = \begin{cases} \frac{(n-x)^{n-1}}{e^{n-x}(n-1)!} & x \le n \\ 0 & x > n \end{cases} \tag{2.1}$$

which is easily proved by induction on n.

Lemma 2.3. *Let $n \ge 2$. Then*

(i) *$f_n(x)$ is unimodular in x, with a maximum value $f_n(1)$, and $f_n(1) \sim \frac{1}{\sqrt{2\pi n}}$;*
(ii) *For $x \ge 0$, $f_n(1+x) \le f_n(1-x)$;*
(iii) *For each real $z \ge 0$, there is a unique number $b = b(n,z)$ satisfying $0 \le b \le z$ and $f_n(1-z) = f_n(1+z-b)$.*

Proof. Item (i) follows from

$$f_n'(x) = \frac{1-x}{n-x} f_n(x) \qquad (x < n) \tag{2.2}$$

and Stirling's formula. For (ii), suppose $0 \le x < n-1$. Then

$$\frac{f_n(1+x)}{f_n(1-x)} = e^{2x}\left(1 - \frac{x}{n-1}\right)^{n-1}\left(1+\frac{x}{n-1}\right)^{-(n-1)} = \exp\left\{-2\sum_{j=1}^{\infty} \frac{x^{2j+1}}{(2j+1)(n-1)^{2j}}\right\} \le 1.$$

Item (iii) follows immediately from (i) and (ii). □

Using properties of $b(n,z)$, we will prove a sharper form of Theorem 1.

Theorem 2. *Suppose $n \ge 1$, $1 \le u \le \frac{n}{10}$, $1 \le w \le \frac{n}{10}$ and let $b = b(n+1, w)$. Then*

$$Q_n(u,v) = 1 - \left(1 - \frac{u(2w-b)}{(n-w+b)(n+w-u)}\right)^n + O\left(\left(\frac{u+w}{n} + \frac{uw^2}{n^2}\right) e^{-\frac{uw}{n+w-u}}\right).$$

3. A recurrence formula

Our principal tool for estimating $Q_n(u,v)$ is a recurrence formula based on the reflection principle for random walks : For $y \ge 0$ and $y \ge x$, a recurrent random walk of n steps that crosses the point y and ends at the point x is about as likely as a random walk which ends at $2y - x$ after n steps. For convenience, define

$$R_n(x,y) = f_n(x)\mathbf{P}\left[\max_{0 \le j \le n-1} S_j < y \middle| S_n = x\right] = \mathbf{D}\left[\max_{0\le j\le n-1} S_j < y, S_n = x\right],$$

where the last expression stands for the density function $\frac{d}{dx}\mathbf{P}[T_{n-1} < y, S_n \le x]$. From the reflection principle we expect that $R_n(x,y) \approx f_n(x) - f_n(2y-x)$. The next lemma gives a precise measure of the accuracy of the reflection principle for our specific random walk.

Lemma 3.1. *For a positive integer $n \ge 2$, real $y > 0$, real x, and real $a \ge 1$,*

$$R_n(x,y) = f_n(x) - f_n(y+a) + \int_0^1 \sum_{k=1}^{n-1} R_k(y+\xi, y)\left(f_{n-k}(a-\xi) - f_{n-k}(x-y-\xi)\right) d\xi. \tag{3.1}$$

Proof. Define $T_j = \max(S_0, \dots, S_j)$. Start with

$$R_n(x,y) = f_n(x) - f_n(y+a) + f_n(y+a) - \mathbf{D}[T_{n-1} \ge y, S_n = x].$$

If $S_n = y + a$, then there is a unique k, $1 \le k \le n-1$, so that $T_{k-1} < y$ and $S_k \ge y$. Thus,

$$\begin{aligned} f_n(y+a) &= \sum_{k=1}^{n-1} \mathbf{D}[T_{k-1} < y, S_k \ge y, S_n = y+a] \\ &= \sum_{k=1}^{n-1} \int_0^1 \mathbf{D}[T_{k-1} < y, S_k = y+\xi, S_n = y+a]\, d\xi \\ &= \sum_{k=1}^{n-1} \int_0^1 R_k(y+\xi, y) f_{n-k}(a-\xi)\, d\xi. \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbf{D}[T_{n-1} \ge y, S_n = x] &= \sum_{k=1}^{n-1} \mathbf{D}[T_{k-1} < y, S_k \ge y, S_n = x] \\ &= \sum_{k=1}^{n-1} \int_0^1 R_k(y+\xi, y) f_{n-k}(x-y-\xi)\, d\xi. \end{aligned}$$

□

In Lemma 3.1, choosing $a \approx y - x - b(n, y-x)$ should make $|f_{n-k}(a-\xi) - f_{n-k}(x-y-\xi)|$ small for small k. Also, we expect $R_k(y+\xi, y)$ to be small, especially for large k, so the integral-sum on the right of (3.1) will be treated as an error term.

The same argument provides an analogous formula when the steps in the random walk have an arbitrary distribution (see [4]).

We next give a crude estimate for $R_n(x, y)$ when $x \ge y$ which will be used on the right side of (3.1).

Lemma 3.2. *If $k \ge 1$, $y \ge 0$, and $0 \le \mu \le 1$, then $R_k(y+\mu, y) \ll \frac{y+1}{k} f_k(y)$.*

Proof. Without loss of generality, suppose $k \ge 10$ and $0 \le y \le \frac{k}{10}$. By Lemma 2.3 (ii), when $1 \le j \le k-1$, $f_j(4-\xi) \le f_j(\mu - \xi)$. By Lemma 3.1 (with $a = 4$ and $x = y + \mu$) and (2.2),

$$R_k(y+\mu, y) \le f_k(y+\mu) - f_k(y+4) = \int_{y+\mu}^{y+4} \frac{t-1}{k-t} f_k(t)\, dt \ll \frac{(y+1) f_k(y)}{k}.$$

□

4. Estimates for $f_n(x)$

Lemma 4.1. *We have*

(i) *If $n \ge 20$ and $0 \le z \le \frac{n}{10}$, then $b(n,z) \le \frac{z}{3}$ and $b(n,z) = \frac{2z^2}{3(n-1)} + O\left(\frac{z^3}{n^2}\right)$;*
(ii) *If $n \ge 1$ and $|x| \le \frac{n}{3}$, then $n^{-1/2} e^{-x^2/n} \ll f_n(x) \ll n^{-1/2} e^{-x^2/3n}$;*
(iii) *If $1 \le h \le H \le 10x^2$, then $f_h(x) h^{-2} \ll f_H(x) H^{-2}$;*
(iv) *If $1 \le k \le n$, then $f_k(x) \ll (n/k)^{1/2} f_n(x)$.*

Proof. First, writing $b = b(n,z)$, we have

$$\left(1 - \frac{2z - b}{n-1+z}\right)^{n-1} = e^{-2z+b}.$$

Under the hypotheses of (i), let $t = \frac{2z-b}{n-1+z}$, so that $0 \le t \le \frac{1}{5}$ by Lemma 2.3 (iii). Then

$$\frac{z}{n-1} = -\frac{\log(1-t)+t}{t} = \frac{t}{2} + \frac{t^2}{3} + \cdots$$

which implies

$$t = 2\left(\frac{z}{n-1}\right) - \frac{8}{3}\left(\frac{z}{n-1}\right)^2 + O\left(\left(\frac{z}{n-1}\right)^3\right).$$

The asymptotic for b follows. Since $\frac{t}{2} + \frac{t^2}{3} + \cdots \le \frac{3}{5}t$, $b \le \frac{z}{3}$ and this proves (i).

Item (ii) is trivial when $n < 100$. When $n \ge 100$, (2.1) and Stirling's formula give

$$f_n(x) \asymp n^{-1/2} e^{x-1} \left(1 - \frac{x-1}{n-1}\right)^{n-1} = n^{-1/2} \exp\left[-\frac{(x-1)^2}{n-1} \sum_{m=2}^{\infty} \frac{1}{m}\left(\frac{x-1}{n-1}\right)^{m-2}\right].$$

Since $|\frac{x-1}{n-1}| \le 0.35$, the sum on m is between $\frac{1}{2}$ and $\frac{2}{3}$, which proves (ii).

Since $f_n(x)n^{-2} \asymp n^{-5/2}$ for $n \ge (\frac{x-1}{10})^2$, it suffices to prove (iii) when $H \le (\frac{x-1}{10})^2$. For $1 \le h \le (\frac{x-1}{10})^2$ and $h > x$,

$$f_h(x)h^{-2} \asymp g(h) := h^{-5/2} e^{x-1} \left(1 - \frac{x-1}{h-1}\right)^{h-1}.$$

We have

$$\frac{d}{dh} \log g(h) = \frac{-5}{2h} + \frac{x-1}{h-x} - \log\left(1 + \frac{x-1}{h-x}\right) > 0,$$

and (iii) follows. If $|x| \le \sqrt{n}$, Lemma 2.3 (i) and part (ii) above imply $f_k(x) \ll k^{-1/2}$ and $f_n(x) \gg n^{-1/2}$. When $|x| > \sqrt{n}$, applying (iii) gives $f_k(x) \ll f_n(x) \ll (n/k)^{1/2} f_n(x)$, proving (iv). ☐

A useful corollary of Lemma 3.2 and Lemma 4.1 (iv) is

$$R_k(u+\xi, u) \ll \frac{n^{1/2} u f_{n+1}(u)}{k^{3/2}} \qquad (1 \le k \le n+1, u \ge 1, 0 \le \xi \le 1). \tag{4.1}$$

Lemma 4.2. *Suppose $n \ge 100$, $1 \le w \le \frac{n}{10}$, $b = b(n+1, w)$ and $0 \le \xi \le 1$.*

(i) *If $w^{3/2} \le h \le n$, then*

$$|f_h(1+w-b-\xi) - f_h(1-w-\xi)| \ll \left(\frac{w}{h} + \frac{w^3}{h^2}\right) f_h(1-w).$$

(ii) *If $2\sqrt{n} \le w \le \frac{n}{10}$ and $1 \le k \le n-3w$, then $f_{n+1-k}(1+w-b-\xi)$ and $f_{n+1-k}(1-w-\xi)$ are each*

$$= f_{n+1}(1-w) \exp\left\{ \sum_{j=n-k}^{n} \left(\frac{1}{2j}\left(1 - \frac{w^2}{j}\right) + O\left(\frac{w^3}{j^3}\right)\right) + O\left(\frac{w}{n}\right)\right\}.$$

Proof. Assume $w^{3/2} \le h \le n$ and write

$$\frac{f_h(1+w-b-\xi)}{f_h(1-w-\xi)} = e^{2w-b}\left(1 - \frac{2w-b}{h-1+w+\xi}\right)^{h-1} = e^E,$$

where, by Lemma 4.1 (i),

$$
\begin{aligned}
E &= (2w-b)\left(1-\frac{h-1}{h-1+w+\xi}-\frac{1}{2}\frac{(h-1)(2w-b)}{(h-1+w+\xi)^2}+O\left(\frac{w^2}{h^2}\right)\right)\\
&= \frac{2w-b}{h-1+w+\xi}\left(w+\xi-\frac{2w-b}{2}\left(1-\frac{w+\xi}{h-1+w+\xi}\right)\right)+O\left(\frac{w^3}{h^2}\right)\\
&\ll \frac{w}{h}+\frac{w^3}{h^2}.
\end{aligned}
$$

By hypothesis, $E \ll 1$ and hence

$$
\begin{aligned}
|f_h(1+w-b-\xi)-f_h(1-w-\xi)| &= f_h(1-w-\xi)|e^E-1| \ll |E| f_h(1-w-\xi)\\
&\le |E| f_h(1-w) \ll \left(\frac{w}{h}+\frac{w^3}{h^2}\right) f_h(1-w).
\end{aligned}
$$

This proves (i).

To prove (ii), we write

$$
\begin{aligned}
\frac{f_{n+1-k}(1+w-b-\xi)}{f_{n+1}(1+w-b)} &= \frac{f_{n+1}(1+w-b-\xi)}{f_{n+1}(1+w-b)}\prod_{j=n-k}^{n}\frac{f_j(1+w-b-\xi)}{f_{j+1}(1+w-b-\xi)}\\
&= e^{A+B_{n-k}+\cdots+B_n},
\end{aligned}
\tag{4.2}
$$

say. By (2.1) and the hypothesis on w, $A \ll \frac{w}{n}$ and

$$
\begin{aligned}
B_j &= 1+(j-1)\log\left(1-\frac{1}{j-w+b+\xi}\right)+\log\left(1+\frac{w-b-\xi}{j-w+b+\xi}\right)\\
&= \frac{1}{j-w+b+\xi}\left(1-\frac{j-1}{2(j-w+b+\xi)}-\frac{(w-b-\xi)^2}{2(j-w+b+\xi)}+O\left(\frac{1}{j}+\frac{w^3}{j^2}\right)\right)\\
&= \frac{1}{j}\left(\frac{1}{2}-\frac{w^2}{2j}\right)+O\left(\frac{w^3}{j^3}\right).
\end{aligned}
$$

Arguing similarly,

$$
\frac{f_{n+1-k}(1-w-\xi)}{f_{n+1}(1-w)} = e^{C+D_{n-k}+\cdots+D_n},
\tag{4.3}
$$

where $C \ll \frac{w}{n}$ and $D_j = \frac{1}{2j}(1-\frac{w^2}{j})+O(w^3/j^3)$. Combining (4.2), (4.3), the above estimates for A, B_j, C and D_j, and the relation $f_{n+1}(1-w)=f_{n+1}(1+w-b)$ concludes the proof of (ii). □

5. proof Theorem 2

Without loss of generality, suppose $n \ge n_0$, where n_0 is a large absolute constant. We apply Lemma 3.1 with $a = 1+w-b$, where $b = b(n+1,w)$, obtaining

$$
R_{n+1}(n+1-v,u) = f_{n+1}(u+1-w) - f_{n+1}(u+1+w-b) + \sum_{k=1}^{n}\Delta_k,
\tag{5.1}
$$

where

$$
|\Delta_k| \le \max_{0\le\xi\le1} R_k(u+\xi,u)\big|f_{n+1-k}(1+w-b-\xi)-f_{n+1-k}(1-w-\xi)\big|.
$$

If $n \le u^2$, then $f_k(u)/k \ll f_{n+1}(u)/n$ for $1 \le k \le n$ by Lemma 4.1 (iii). If $n > u^2$, then we have

$$\frac{f_k(u)}{k} \ll \begin{cases} u^{-1} f_{\lfloor u^2 \rfloor}(u) \asymp u^{-3} & \text{if } k \le u^2 \\ k^{-3/2} & \text{if } k > u^2 \end{cases}$$

by Lemma 4.1 (ii), (iii). In both cases,

$$\sum_{k=1}^{n} \frac{f_k(u)}{k} \ll \left(1 + \frac{n^{1/2}}{u}\right) f_{n+1}(u). \tag{5.2}$$

Suppose that $1 \le w \le 2\sqrt{n}$, so that $b = O(1)$. We will prove that

$$\sum_{k=1}^{n} |\Delta_k| \ll \frac{u+w}{n} f_{n+1}(u) \ll \frac{u+w}{n^{1/2}} f_{n+1}(u) f_{n+1}(1-w) \quad (1 \le w \le 2\sqrt{n}). \tag{5.3}$$

The second inequality follows from the first and Lemma 4.1 (ii). Let $h = n+1-k$, $h_0 = \lfloor w^{3/2} \rfloor$ and $h_1 = \lfloor w^2/10 \rfloor$. Choose $n_0 \ge 2^{10}$ so that $h_0 \le n/2$. For $1 \le h \le h_0$, (4.1) and Lemma 4.1 (ii) give

$$\Delta_{n+1-h} \ll \frac{u f_{n+1}(u)}{n} \max_{0 \le \xi \le 1} \left(f_h(1+w-b-\xi) + f_h(1-w-\xi)\right) \ll \frac{u f_{n+1}(u)}{n w^3}.$$

If $h_0 < h \le h_1$, then (4.1), Lemma 4.2 (i) and Lemma 4.1 (ii),(iv) imply

$$\Delta_{n+1-h} \ll \frac{u f_{n+1}(u)}{n} \left(\frac{w}{h} + \frac{w^3}{h^2}\right) f_h(1-w) \ll \frac{u f_n(u)}{n} \left(\frac{w}{h_1} + \frac{w^3}{h_1^2}\right) f_{h_1}(1-w) \ll \frac{u f_{n+1}(u)}{n w^2}.$$

When $h_1 < h \le \frac{n}{2}$, (4.1) and Lemma 4.2 (i) imply

$$\Delta_{n+1-h} \ll \frac{u f_{n+1}(u)}{n} \frac{w}{h} f_h(1-w) \ll \frac{u w f_{n+1}(u)}{n w^{3/2}}.$$

Summing on $h \le \frac{n}{2}$ we obtain

$$\sum_{1 \le h \le \frac{n}{2}} |\Delta_{n+1-h}| \ll \frac{u f_{n+1}(u)}{n}. \tag{5.4}$$

Lemma 3.2, Lemma 4.2 (i) and (5.2) imply

$$\sum_{n/2 < h \le n} |\Delta_{n+1-h}| \ll \frac{uw}{n^{3/2}} \sum_{1 \le k < n/2+1} \frac{f_k(u)}{k} \ll \frac{u+w}{n} f_{n+1}(u).$$

Combined with (5.4), this proves (5.3).

Next, suppose $2\sqrt{n} < w \le \frac{n}{10}$ and set

$$K = \left\lfloor \min\left(n - C_0 w, \frac{n^3}{w^3}\right) \right\rfloor,$$

where C_0 is a large absolute constant. When $1 \le k \le K$, apply Lemma 3.2 and Lemma 4.2 (ii), observing that for each $j \le n$, $\frac{1}{2j}(1 - w^2/j) \le -\frac{w^2}{3j^2} \le -\frac{w^2}{3n^2}$. If $k \le n/2$, then

$$\Delta_k \ll u \frac{f_k(u)}{k} \left(\frac{w}{n} + \frac{k w^3}{n^3}\right) e^{-kw^2/(10n^2)} f_{n+1}(1-w). \tag{5.5}$$

When $n/2 < k \le K$,

$$\Delta_k \ll \frac{uf_k(u)f_{n+1}(1-w)}{k} e^{-\frac{w^2}{6(n-k)}} \left(\exp\left\{ C_1 \left(\frac{w^3}{(n-k)^2} + \frac{w}{n} \right) \right\} - 1 \right)$$

for an absolute constant C_1. If in addition $n - k \ge w^{3/2}$, then

$$e^{-\frac{w^2}{6(n-k)}} \left(\exp\left\{ C_1 \left(\frac{w^3}{(n-k)^2} + \frac{w}{n} \right) \right\} - 1 \right) \ll \left(\frac{w^3}{(n-k)^2} + \frac{w}{n} \right) e^{-\frac{w^2}{6(n-k)}},$$

which implies (5.5). If $C_0 w \le n - k < w^{3/2}$ and we take $C_0 = 20C_1$, then

$$e^{-\frac{w^2}{6(n-k)}} \left(\exp\left\{ C_1 \left(\frac{w^3}{(n-k)^2} + \frac{w}{n} \right) \right\} - 1 \right) \ll e^{-\frac{w^2}{12(n-k)}} \ll n^{-3} e^{-\frac{kw^2}{10n^2}},$$

and (5.5) follows in this case as well.

By Lemma 4.1 (iv), (5.2) and (5.5),

$$\begin{aligned} \sum_{k \le K} |\Delta_k| &\ll u f_{n+1}(1-w) \left[\frac{w}{n} \sum_{k \le K} \frac{f_k(u)}{k} + \frac{w^3 f_{n+1}(u)}{n^{5/2}} \sum_{k=1}^{\infty} k^{-1/2} e^{-kw^2/(10n^2)} \right] \\ &\ll f_{n+1}(1-w) f_{n+1}(u) \left(\frac{uw^2}{n^{3/2}} + \frac{w}{n^{1/2}} \right). \end{aligned} \tag{5.6}$$

When $k > K$, we combine Lemma 4.1 (i), (iii) and Lemma 4.2 (ii) to obtain

$$\begin{aligned} f_{n+1-k}(1+w-b-\xi) + f_{n+1-k}(1-w-\xi) &\ll f_{n+1-K}(1+w-b-\xi) + f_{n+1-K}(1-w-\xi) \\ &\ll e^{-Kw^2/(10n^2)} f_{n+1}(1-w). \end{aligned}$$

Together with (4.1), this gives

$$\sum_{K<k\le n} |\Delta_k| \ll u n^{1/2} f_{n+1}(u) f_{n+1}(1-w) e^{-Kw^2/(10n^2)} \sum_{K<k\le n} \frac{1}{k^{3/2}}.$$

If $2\sqrt{n} < w \le n^{2/3}$, then $K = \lfloor n - 3w \rfloor$ and

$$e^{-Kw^2/(10n^2)} \sum_{K<k\le n} \frac{1}{k^{3/2}} \ll \frac{w}{n^{3/2}} e^{-w^2/(20n)} \ll \frac{w^2}{n^2}.$$

If $n^{2/3} < w \le \frac{n}{10}$, then $K \ge n^3/2w^3$ and

$$e^{-Kw^2/(10n^2)} \sum_{K<k\le n} \frac{1}{k^{3/2}} \ll (n^3/w^3)^{-1/2} e^{-n/20w} \ll \frac{w^2}{n^2}.$$

Therefore,

$$\sum_{K<k\le n} |\Delta_k| \ll f_{n+1}(1-w) f_{n+1}(u) \frac{uw^2}{n^{3/2}}.$$

Combined with (5.6), we have

$$\sum_{k=1}^{n} |\Delta_k| \ll f_{n+1}(1-w) f_{n+1}(u) \left(\frac{uw^2}{n^{3/2}} + \frac{w}{n^{1/2}} \right) \qquad (2\sqrt{n} < w \le n/10). \tag{5.7}$$

Combining (5.1), (5.3) and (5.7) with Lemma 2.2, in all cases we have

$$Q_n(u,v) = 1 - \frac{f_{n+1}(u+1+w-b)}{f_{n+1}(u+1-w)} + O\left(\frac{n^{1/2} f_{n+1}(1-w) f_{n+1}(u)}{f_{n+1}(u+1-w)} \left[\frac{u+w}{n} + \frac{uw^2}{n^2} \right] \right).$$

By the definition of b,

$$\begin{aligned} \frac{f_{n+1}(u+1+w-b)}{f_{n+1}(u+1-w)} &= \frac{f_{n+1}(u+1+w-b) f_{n+1}(1-w)}{f_{n+1}(1+w-b) f_{n+1}(u+1-w)} \\ &= \left(1 - \frac{u(2w-b)}{(n-w+b)(n+w-u)} \right)^n. \end{aligned}$$

Also, by Stirling's formula,

$$\begin{aligned} \frac{n^{1/2} f_{n+1}(1-w) f_{n+1}(u)}{f_{n+1}(u+1-w)} &= \frac{n^{1/2}(n+1)^n}{e^{n+1} n!} \left(\frac{(n+1-u)(n+w)}{(n+1)(n+w-u)} \right)^n \\ &\ll \left(1 - \frac{u(w-1)}{(n+1)(n+w-u)} \right)^n \ll e^{-\frac{uw}{n+w-u}}, \end{aligned}$$

which concludes the proof of Theorem 2. □

6. PROOF THEOREM 1

We may assume $0 \le u \le \delta n$ and $0 \le w \le \delta n$ for a small, fixed, positive δ. If $0 \le u \le 1$ and $0 \le w \le \delta n$, Lemma 2.1 (i) implies $Q_n(u,v) \ll w/n$. When $0 \le w \le 1$ and $1 \le u \le \delta n$, Lemma 3.2 implies $Q_n(u,v) \ll \frac{u}{n}$. When $1 \le u \le \delta n$ and $1 \le w \le \delta n$, we may assume that n is large. The error term in Theorem 2 is

$$\ll \frac{u+w}{n} + \frac{w}{n} \cdot \frac{uw}{n} e^{-\frac{uw}{2n}} \ll \frac{u+w}{n}.$$

When $uw > n^{4/3}$, the main terms are

$$1 - O(e^{-\frac{1}{2}n^{1/3}}) = 1 - e^{-\frac{2uw}{n}} + O\left(\frac{1}{n}\right).$$

When $uw \le n^{4/3}$, the main terms are, by Lemma 4.1 (i),

$$\begin{aligned} &= 1 - \exp\left[-\frac{u(2w-b)n}{(n-w+b)(n+w-u)} + O\left(\frac{(uw)^2}{n^3} \right) \right] \\ &= 1 - \exp\left[-\frac{2uw}{n} \left(1 + O\left(\frac{u+w}{n} \right) \right) \right] + O\left(\frac{(uw)^2}{n^3} \right) \\ &= 1 - e^{-\frac{2uw}{n}} + O\left(\frac{u+w+(uw)^{1/2}}{n} \right) \\ &= 1 - e^{-\frac{2uw}{n}} + O\left(\frac{u+w}{n} \right). \end{aligned}$$

□

Acknowledgments. The author expresses thanks to Valery Nevzorov, Walter Philipp, Steven Portnoy, and Jon Wellner for helpful conversations. The author also thanks the referee for suggestions on improving the exposition.

References

[1] M. Csörgő and P. Révész. *Strong Approximations in probability and statistics.* Academic Press, 1981.

[2] H. E. Daniels. The statistical theory of the strength of bundles of threads. I. *Proc. Roy. Soc. London. Ser. A.*, 183:405–435, 1945.

[3] K. Ford. Du théorème de Kolmogorov sur les distributions empiriques à la théorie des nombres. In *L'héritage de Kolmogorov en mathématiques*, pages 111–120. Editions Belin, Paris, 2004. (French).

[4] K. Ford. Sharp probability estimates for random walks with barriers. 2006. Preprint available on the ArXiv at `http://front.math.ucdavis.edu/math.PR/0610450`.

[5] K. Ford. From Kolmogorov's theorem on empirical distribution to number theory. In *Kolmogorov's heritage in mathematics.* Editions Belin / Springer-Verlag, Paris, 2007. (English).

[6] K. Ford. The distribution of integers with a divisor in a given interval. *Annals of Math.*, 2008. to appear. Preprint available on the ArXiv at `http://front.math.ucdavis.edu/math.NT/0401223`.

[7] B. V. Gnedenko and A. N. Kolmogorov. *Limit distributions for sums of independent random variables.* Translated from the Russian, annotated, and revised by K. L. Chung. With appendices by J. L. Doob and P. L. Hsu. Revised edition. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills., Ont., 1968.

[8] A. N. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione (on the empirical determination of a distribution law). *Giorn. Ist. Ital. Attuar.*, 4:83–91, 1933.

[9] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's and the sample DF. i. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32:111–131, 1975.

[10] H. A. Lauwerier. The asymptotic expansion of the statistical distribution of N. V. Smirnov. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 2:61–68, 1963. German.

[11] R. Pyke. The supremum and infimum of the Poisson process. *Ann. Math. Statist.*, 30:568–576, 1959.

[12] A. Rényi. On the theory of order statistics. *Acta Math. Acad. Sci. Hung.*, 4:191–232, 1953.

[13] G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[14] 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.

Department of Mathematics, 1409 West Green Street, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

E-mail address: `ford@math.uiuc.edu`