Conspectus materiae tomi XXXI, fasciculi 3

La revue est consacrée à la Théorie des Nombres
The journal publishes papers on the Theory of Numbers
Die Zeitschrift veröffentlicht Arbeiten aus der Zahlentheorie
Журнал посвящен теории чисел

L'adresse de la Rédaction et de l'échange | Address of the Editorial Board and of the exchange | Die Adresse der Schriftleitung und des Austausches | Адрес редакции и книгообмена

ACTA ARITHMETICA
ul. Śniadeckich 8, 00-950 Warszawa

Les auteurs sont priés d'envoyer leurs manuscrits en deux exemplaires
The authors are requested to submit papers in two copies
Die Autoren sind gebeten um Zusendung von 2 Exemplaren jeder Arbeit
Рукописи статей редакция просит предлагать в двух экземплярах

PRINTED IN POLAND

WROCŁAWSKA DRUKARNIA NAUKOWA

ACTA ARITHMETICA
XXXI (1976)

On the probability that integers chosen according to the binomial distribution are relatively prime

by

J. E. Nymann and W. J. Leahey* (El Paso, Tex.)

Let n be a non-negative integer and denote by N_n the set of integers $0, 1, 2, \ldots, n$. Let P_n be a probability measure on N_n and for k a positive integer denote by P_n^k the k-fold product measure of P_n on N_n^k. Set S_n^k equal to the subset of all elements $(x_1, x_2, \ldots, x_k)$ of N_n^k for which $(x_1, x_2, \ldots, x_k) = 1$. (Here we agree that $(0, 0, \ldots, 0) \neq 1$.) It is well-known that if P_n is the uniform probability measure ($P_n(j) = (n+1)^{-1}$ for all $j \in N_n$), then

$$\lim_{n\to\infty} P_n^2(S_n^2) = 6/\pi^2.$$

It is the object of this paper to show that this also holds in the case where P_n is a binomial distribution, i.e.,

$$P_n(j) = \binom{n}{j} p^j (1-p)^{n-j},$$

where p is some fixed real number, $0 < p < 1$.

In Section 1 we prove some generalities which are of some interest in themselves and which will be useful in Section 2 where we prove our major result.

1. For any positive integer d, let $A_n(d) = \{j \in N_n\colon j \equiv 0 \pmod{d}\}$. We then have the following basic

Lemma 1. *Let P_n be any probability distribution. Then for $n > 1$*

$$P_n^k(S_n^k) = \sum_{d=1}^{n} \mu(d)\left(\{P_n(A_n(d))\}^k - \{P_n(\{0\})\}^k\right).$$

Proof. Let $p_1 < p_2 < \ldots < p_s$ be the primes less than or equal to n. Then, if $\tilde{S}_n^k$ denotes the complement of S_n^k, we have

$$\tilde{S}_n^k = \bigcup_{i=1}^{s} A_n^k(p_i)$$

* The authors wish to thank the referee both for pointing out how to generalize their initial results and at the same time how to shorten the proofs.

where $A_n^k(d)$ denotes the Cartesian product of $A_n(d)$ with itself k times. Therefore

$$P_n^k(S_n^k) = 1 - P_n^k(\tilde{S}_n^k) = 1 - P_n^k\Big(\bigcup_{i=1}^{s} A_n^k(p_i)\Big)$$

$$= 1 - \sum_{r=1}^{s} \sum_{(i_1, i_2, \ldots, i_r)} (-1)^{r-1} P_n^k\big(A_n^k(p_{i_1}) \cap A_n^k(p_{i_2}) \cap \ldots \cap A_n^k(p_{i_r})\big)$$

where the inner sum is taken over all r-tuples $(i_1, i_2, \ldots, i_r)$ such that $1 \leqslant i_1 < i_2 < \ldots < i_r \leqslant s$. Now for $(d_1, d_2) = 1$,

$$A_n^k(d_1) \cap A_n^k(d_2) = A_n^k(d_1 d_2).$$

Hence this last expression can be rewritten as

$$1 + \sum_{r=1}^{s} \sum_{(i_1, i_2, \ldots, i_r)} (-1)^r P_n^k\big(A_n^k(p_{i_1} p_{i_2} \ldots p_{i_r})\big).$$

Now if $p_{i_1} p_{i_2} \ldots p_{i_r} > n$, $A_n^k(p_{i_1} p_{i_2} \ldots p_{i_r}) = \{(0, 0, \ldots, 0)\}$. Hence this last expression is the same as

$$\sum_{d=1}^{n} \mu(d) P_n^k\big(A_n^k(d)\big) + \sum_{p_{i_1} p_{i_2} \ldots p_{i_r} > n} \mu(p_{i_1} p_{i_2} \ldots p_{i_r}) P_n^k\big(\{(0, 0, \ldots, 0)\}\big).$$

Since $\sum_{d | p_1 p_2 \ldots p_s} \mu(d) = 0$,

$$\sum_{p_{i_1} p_{i_2} \ldots p_{i_r} > n} \mu(p_{i_1} p_{i_2} \ldots p_{i_r}) = -\sum_{d=1}^{n} \mu(d).$$

This observation together with $P_n^k\big(A_n^k(d)\big) = \{P_n\big(A_n(d)\big)\}^k$ completes the proof of the lemma.

Corollary 2. *If P_n is the uniform distribution, then*

$$\lim_{n \to \infty} P_n^k(S_n^k) = 1/\zeta(k)$$

for all $k \geqslant 2$. (ζ denotes the Riemann zeta function.)

Proof. Define $\varepsilon_n(d)$ by the equation

$$P_n\big(A_n(d)\big) = d^{-1} + \varepsilon_n(d).$$

It is easy to check that $0 \leqslant \varepsilon_n(d) < n^{-1}$ for all positive integers n and d. By Lemma 1

$$P_n^k(S_n^k) = \sum_{d=1}^{n} \mu(d) \{\big(d^{-1} + \varepsilon_n(d)\big)^k - (n+1)^{-k}\}$$

$$= \sum_{d=1}^{n} d^{-k} \mu(d) + \sum_{j=1}^{k} \binom{k}{j} \sum_{d=1}^{n} d^{j-k} \mu(d) \big(\varepsilon_n(d)\big)^j - (n+1)^{-k} \sum_{d=1}^{n} \mu(d).$$

Now for $k \geqslant 2$, $\sum_{d=1}^{\infty} d^{-k} \mu(d) = 1/\zeta(k)$. On the other hand since $0 \leqslant \varepsilon_n(d) < n^{-1}$, it is not difficult to see that each of the terms

$$\sum_{d=1}^{n} d^{j-k} \mu(d) \big(\varepsilon_n(d)\big)^j, \quad j = 1, 2, \ldots, k,$$

goes to zero as n gets large. This together with $\Big|\sum_{d=1}^{n} \mu(d)\Big| \leqslant n$ establishes the corollary. For another proof of this result see [2].

We now turn our attention to binomial distributions. In particular we wish to show that the conclusion of Corollary 2 holds when P_n is a binomial distribution. For $k \geqslant 3$ this is relatively easy and will be proved shortly. The case $k = 2$, which is the principal result of this paper, is dealt with in Section 2.

From now on P_n will always be understood to be a binomial distribution relative to some fixed parameter p, $0 < p < 1$. We define $\varepsilon_n(d)$ as in the proof of Corollary 2. Thus

$$\varepsilon_n(d) = P_n\big(A_n(d)\big) - d^{-1} = \sum_{k \equiv 0 (d)} \binom{n}{k} p^k (1-p)^{n-k} - d^{-1}$$

for n and d positive integers with $d \leqslant n$.

Lemma 3. *$|\varepsilon_n(d)| \ll n^{-1/2}$ uniformly in d.*

Proof. We wish to show that

$$\Big| d \sum_{k \equiv 0 (d)} \binom{n}{k} p^k (1-p)^{n-k} - \sum_{k} \binom{n}{k} p^k (1-p)^{n-k} \Big| \ll d n^{-1/2}.$$

Let us write $\binom{n}{k}'$ for $\binom{n}{k} p^k (1-p)^{n-k}$. The term within the absolute value signs on the left-hand side is the same as

$$\sum_{j=1}^{d-1} \Big\{ \sum_{k \equiv 0 (d)} \binom{n}{k}' - \sum_{k \equiv j (d)} \binom{n}{k}' \Big\}.$$

Set $s = [p(n+1)]$ and let t be the largest integer such that $td \leqslant s$. For given j we look at that portion of the sums within the braces for which $k \leqslant td$. We have

$$\sum_{\substack{k \equiv 0 \\ k \leqslant td}} \binom{n}{k}' - \sum_{\substack{k \equiv j \\ k \leqslant td}} \binom{n}{k}' = \sum_{i=0}^{t} \Big\{ \binom{n}{id}' - \binom{n}{id - d + j}' \Big\}$$

(where to take care of the $i = 0$ term we agree that $\binom{n}{k} = 0$ for $k < 0$).

Since $\binom{n}{k}'$ is non-decreasing for $0 \leqslant k \leqslant s$ this is a sum of the type

$$\sum_{i=0}^{m} (a_i - b_i)$$

with $b_0 \leqslant a_0 \leqslant b_1 \leqslant a_1 \leqslant \ldots \leqslant b_m \leqslant a_m$. The value of such a sum does not exceed its largest term a_m. Hence

$$\sum_{\substack{k=0\\k<td}} \binom{n}{k}' - \sum_{\substack{k\equiv j\\k<td}} \binom{n}{k}' \leqslant \binom{n}{s}'.$$

Similarly the absolute value of the sum of those terms of

$$\sum_{k=0} \binom{n}{k}' - \sum_{k\equiv j} \binom{n}{k}'$$

for which $k \geqslant (t+1)d$ is also bounded by $\binom{n}{s}'$. Finally there is only one term from these sums for which $td < k < (t+1)d$. Hence

$$\Big| \sum_{k=0} \binom{n}{k}' - \sum_{k\equiv j} \binom{n}{k}' \Big| \leqslant 3 \binom{n}{s}'.$$

Therefore

$$\Big| \sum_{j=1}^{d-1} \Big\{ \sum_{k=0} \binom{n}{k}' - \sum_{k\equiv j} \binom{n}{k}' \Big\} \Big| \leqslant 3(d-1) \binom{n}{s}' \leqslant 3d \binom{n}{s}'.$$

The lemma now follows from Stirling's formula.

COROLLARY 4. *If P_n is a binomial distribution, then*

$$\lim_{n\to\infty} P_n^k(S_n^k) = 1/\zeta(k) \quad \textit{for all } k \geqslant 3.$$

Proof. Using the estimate for $\varepsilon_n(d)$ given in Lemma 3, the proof is almost identical to the proof of Corollary 2 and will not be given in detail. We remark that for the case $k = 2$ the estimate $|\varepsilon_n(d)| \ll n^{-1/2}$ is not sufficient to show that the $j = 2$ term, $\sum_{d=1}^{n} \mu(d)(\varepsilon_n(d))^2$, goes to zero as n gets large.

2. In this section we show that $\lim_{n\to\infty} P_n^2(S_n^2) = 6/\pi^2$ $(= 1/\zeta(2))$ when P_n is a binomial distribution. As outlined in Section 1 we must show that the sums

$$\sum_{d=1}^{n} d^{-1}\mu(d)\varepsilon_n(d), \qquad \sum_{d=1}^{n} \mu(d)(\varepsilon_n(d))^2 \quad \text{and} \quad \sum_{d=1}^{n} \mu(d)(1-p)^n$$

go to zero as n gets large. The last sum clearly goes to zero. Hence it is sufficient to show that

$$\sum_{d=1}^{n} d^{-1}|\varepsilon_n(d)| \quad \text{and} \quad \sum_{d=1}^{n} (\varepsilon_n(d))^2$$

tend to zero for large n. The estimate of Lemma 3 shows that $\sum_{d=1}^{n} d^{-1}|\varepsilon_n(d)|$ is bounded by a constant times $n^{-1/2}\log n$ and hence goes to zero. If we had an estimate of the type $|\varepsilon_n(d)| \ll 1/d$, then, using it together with $|\varepsilon_n(d)| \ll n^{-1/2}$, we would have that the second sum is also of the order of $n^{-1/2}\log n$. This estimate is not correct however. For example, if p is rational, and $n = p^{-1}k$, then $\varepsilon_n(k)$ is of the order of $1/\sqrt{k}$ rather than $1/k$. As it turns out, however, this sort of situation does not occur too often. The plan of our proof is essentially as follows. We will show that an estimate of the type $|\varepsilon_n(d)| \ll 1/d$ holds for a rather large set of d's; in fact, for a set of d's which is roughly of the order $n - n^{3/4}$. For the remaining d's (roughly $n^{3/4}$ in number) we use the estimate $|\varepsilon_n(d)| \ll n^{-1/2}$. This will be enough to show that the term $\sum_{d=1}^{n} (\varepsilon_n(d))^2$ goes to zero as n gets large. We need the following lemmas.

LEMMA 5.

$$\sum_{|k-pn|>pn^{3/4}} \binom{n}{k}' \ll n^{-1}.$$

Proof. We refer to Theorem A(i) of Section 18.1 of [1]. Using the notation of that theorem let X_k be the random variable which takes on value $1-p$ with probability p and value $-p$ with probability $1-p$. We have then $s = \sqrt{np(1-p)}$ and $c = a/s$ where $a = \max\{p, 1-p\}$. If we take $\varepsilon = \{p/(1-p)\}^{1/2} n^{1/4}$, then $\varepsilon c \ll n^{-1/4}$, and hence $\varepsilon c \leqslant 1$ for n sufficiently large. According to the theorem then

$$\mathrm{P}\{S/s > \varepsilon\} < \exp\Big\{-\frac{\varepsilon^2}{2}\Big(1 - \frac{\varepsilon c}{2}\Big)\Big\}$$

or

$$\mathrm{P}\{S > pn^{3/4}\} < \exp\Big\{\frac{-pn^{1/2}}{2(1-p)}\Big(1 - \frac{an^{-1/4}}{1-p}\Big)\Big\}.$$

But $\mathrm{P}\{S > pn^{3/4}\}$ is exactly $\sum_{k-pn>pn^{3/4}} \binom{n}{k}'$ and hence $\sum_{k-pn>pn^{3/4}} \binom{n}{k}' \ll n^{-1}$. Replacing X_k by $-X_k$ and applying the theorem again gives $\sum_{k-pn>-pn^{3/4}} \binom{n}{k}' \ll n^{-1}$ from which the lemma follows.

COROLLARY 6. *If $d > p(n+n^{3/4})$, then $|\varepsilon_n(d)| \ll d^{-1}$ uniformly in d.*

LEMMA 7. *Let K_n be the number of integers d which satisfy $pn^{3/4} \leqslant d \leqslant p(n-n^{3/4})$ and which have the property that some multiple of them lies in the interval $\big(p(n-n^{3/4}), p(n+n^{3/4})\big)$. Then*

$$K_n \ll n^{3/4}\log n.$$

Proof. Let $u = pn$, $v = pn^{3/4}$, and let $s = [(u+v)/v]$. Suppose that $kd \in (u-v, u+v)$. Then we must have $2 \leqslant k \leqslant s$. For each such k we ask how many possible d's are there such that $kd \in (u-v, u+v)$. Such d's must necessarily lie in the interval $\big((u-v)/k, (u+v)/k\big)$ and hence there are no more than $(2v/k)+1$ of them. We have then that K_n is bounded by

$$\sum_{k=2}^{s}\big((2v/k)+1\big) \leqslant 2v\log s + (s-1) \leqslant 2pn^{3/4}\log(n^{1/4}+1) + n^{1/4} \ll n^{3/4}\log n.$$

We now state and prove our main result as

THEOREM 8. *Let P_n be a binomial distribution. Then*

$$\lim_{n\to\infty} P_n^2(S_n^2) = 6/\pi^2.$$

Proof. According to the ideas outlined at the beginning of this section we must show that

$$(1) \qquad \lim_{n\to\infty}\sum_{d=1}^{n}\big(\varepsilon_n(d)\big)^2 = 0.$$

Let $n_1 = [pn^{3/4}]$, $n_2 = [p(n-n^{3/4})]$, and $n_3 = [p(n+n^{3/4})]$. The sum in (1) can then be written as

$$(2) \qquad \sum_{d=1}^{n} = \sum_{d=1}^{n_1} + \sum_{d=n_1+1}^{n_2} + \sum_{d=n_2+1}^{n_3} + \sum_{d=n_3+1}^{n}.$$

(We assume, of course, that n is large enough so that $n_1 \leqslant n_2$ and $n_3 \leqslant n$.) We will examine each of these sums separately.

By Lemma 3

$$\sum_{d=1}^{n_1}\big(\varepsilon_n(d)\big)^2 \ll n_1 n^{-1} \leqslant n^{3/4}n^{-1} = n^{-1/4}$$

and hence the first term on the right-hand side of (2) goes to zero as n gets large. A similar argument works for the third sum on the right-hand side of (2).

By Corollary 6 $|\varepsilon_n(d)| \ll d^{-1}$ for $d > p(n+n^{3/4})$. Hence for the fourth sum

$$\sum_{d=n_3+1}^{n}\big(\varepsilon_n(d)\big)^2 \ll n^{-1/2}\sum_{d=n_3+1}^{n} d^{-1} \leqslant n^{-1/2}\log n.$$

Therefore this sum goes to zero as n gets large.

The second sum on the right-hand side of (2) is somewhat more difficult to deal with. We break it into two parts:

$$(3) \qquad \sum_{d=n_1+1}^{n_2} = \sum_{d=n_1+1}^{n_2}{}' + \sum_{d=n_1+1}^{n_2}{}''$$

where the summation with the prime on it is taken over those d's which have the property that some multiple of them lies in the interval $\big(p(n-n^{3/4}), p(n+n^{3/4})\big)$ and the double primed summation is taken over the remaining d's. By Lemma 7 we have

$$\sum_{d=n_1+1}^{n_2}{}'\big(\varepsilon_n(d)\big)^2 \ll (n^{3/4}\log n)n^{-1} = n^{-1/4}\log n.$$

Hence the single primed sum goes to zero with large n. We now examine the double primed sum. Recall that

$$\varepsilon_n(d) = \sum_{k\equiv 0(d)}\binom{n}{k}' - d^{-1}.$$

For the d's in question we have by Lemma 5

$$\sum_{k\equiv 0(d)}\binom{n}{k}' = \sum_{\substack{k\equiv 0(d)\\ |k-pn|>n^{3/4}}}\binom{n}{k}' \leqslant \sum_{|k-pn|>n^{3/4}}\binom{n}{k}' \ll n^{-1}.$$

Hence for these d's $|\varepsilon_n(d)| \ll d^{-1}$. Thus for the double primed sum

$$\sum_{d=n_1+1}^{n_2}{}''\big(\varepsilon_n(d)\big)^2 \ll n^{-1/2}\sum_{d=n_1+1}^{n_2} d^{-1} \leqslant n^{-1/2}\log n.$$

This completes the proof of Theorem 8.

References

[1] M. Loeve, *Probability Theory*, Princeton, New Jersey, 1963.

[2] J. E. Nymann, *On the probability that k positive integers are relatively prime*, J. Number Theory 4 (1972), pp. 469–473.

THE UNIVERSITY OF TEXAS AT EL PASO
El Paso, Texas

Received on 18. 10. 1974
and in revised form on 30. 6. 1975 (627)