ON RANDOM SEQUENCES OF INTEGERS[1]

BY C. J. EVERETT AND P. R. STEIN

Communicated by Gian-Carlo Rota, August 4, 1969

A "single transit" method is given for the computer-production of random sequences of integers with prescribed expected density, employing a "Poisson series of trials", which is faster and more flexible than "sieve devices". The mathematical probability theory underlying the method is developed, and some special cases are studied with regard to gap distribution and the "Goldbach property". We state here, without proof, the principal results; the details will be found in a forthcoming Los Alamos report (LA-4268).

1. The Poisson series of trials. Let $\{P_n;\ n=1,\ 2,\cdots\}$ be a sequence of numbers $0<P_n\leqq1$, such that $E(N)\equiv\sum_1^N P_n\to\infty$. The corresponding Poisson series of trials consists in setting $\nu_n=1$ or 0 with probability P_n or $Q_n\equiv1-P_n$, respectively, at the same time accepting or rejecting n as a member of the random sequence B. $B(N)=\nu_1+\cdots+\nu_N$, $N\geqq1$, thus counts the number of $n\leqq N$ in B, and the law of large numbers insures that the probability

$$(1)\qquad P\{\,|\,B(N)-E(N)\,|\,<\rho E(N)\}\geqq1-1/\rho^2E(N)\to1$$

where ρ is any preassigned relative error.

If $F(N)$ is any function asymptotic to $E(N)$, we have equally well

$$(2)\quad P\{\,|\,B(N)-F(N)\,|\,<3\rho F(N)\}\geqq1-1/\rho^2(1-\rho)F(N)\to1.$$

2. Two practical devices. In practice, we are *given* a distribution function $F(N)$ and require a sequence $P(n)$ producing sequences satisfying (2). This is easily done if $F(N)$ is the value at $x=N\geqq i$ of a positive function $F(x)$ such that

(A) $F(x)\to\infty$; $F'(x)$ exists, is continuous, nonincreasing, with $0<F'(x)\leqq1$ for $x\geqq i$. For then $P(n)\equiv F'(n+i-1)$, $n\geqq1$, gives the desired result.

As an example, we note that $P(n)=2/\log\,(2n+7)$ generates random sequences π of odd integers $(2n+1\in\pi$ iff $n\in B)$ in the prime-like (expected) distribution $F(2N+1)=\int_2^{2N+1}d\xi/\log\,\xi$. In similar fashion, $P(n)=3/\log\,(3n+20)$ produces sequences π' of odd integers

---

$\not\equiv 0$ mod 3, in prime-like distribution, if for odd $n \in B$ we admit $3n - 2$ to π', while for even $n \in B$, we assign $3n - 1$ to π'.

In applying the preceding device we may encounter needlessly complicated derivatives $F'(x)$ (e.g., consider $F(x) = x/\log x$). In such cases the following remark is sometimes helpful. Suppose the required $F(N)$ is the value at $x = N \geq i$ of a function $F(x) = x/f(x)$ such that:

(B$_1$) $f(x) \geq 1$ is nondecreasing, and $f'(x)$ exists and is continuous for $x \geq i$.

(B$_2$) For $F_1(x) \equiv \int_i^x d\xi/f(\xi)$ and $F_2(x) \equiv \int_i^x \xi f'(\xi) d\xi/f^2(\xi)$, one has $F_1 \to \infty$ and $F_2/F_1 \to 0$. (Note that $F_2 \to \infty$ and $xf'/f \to 0$ suffice for the latter.)

Then F_1 satisfies (A), and $P(n) = 1/f(n+i-1)$ generates sequences B for which (2) holds, relative to the *given* $F(x)$. As an example, one may consider $F(N) = N/\log N \log \log N$.

3. **Gap expectations.** Just as $y_0 = \nu_1 + \cdots + \nu_N$ counts the number of 1's in a random vector $(\nu_1, \cdots, \nu_N)$, so do functions $y_2 = \nu_1\nu_2 + \nu_2\nu_3 + \cdots + \nu_{N-1}\nu_N$, $y_4 = \nu_1(1 - \nu_2)\nu_3 + \cdots + \nu_{N-2}(1 - \nu_{N-1})\nu_N$, $\cdots$ count the number of appearances of (1, 1), (101), $\cdots$. These functions have the expectations $\bar{y}_0 = P_1 + \cdots + P_N$, $\bar{y}_2 = P_1P_2 + \cdots + P_{N-1}P_N$, $\bar{y}_4 = P_1Q_2P_3 + \cdots + P_{N-2}Q_{N-1}P_N$, $\cdots$, and are monotone decreasing if P_n is nonincreasing. For the case $P_n = 1/2$ (coin-tossing) it is obvious that $\bar{y}_{2n} = (N - n)/2^{n+1}$.

For the prime-like sequences π of §2, $\bar{y}_{2n}$ is the expectation of a "gap" of length $2n$ (e.g., $\bar{y}_2$ is the expected number of "twins") and this motivates our notation. For such sequences we have of course the monotonicity $\bar{y}_0 > \bar{y}_2 > \cdots$. The preference of actual primes for gap 6 is well known, and it is interesting to note that the sequences π' of §2, avoiding odd multiples of 3, also show such a preference. The gap expectations for the latter are found to be $\bar{y}_2 = P_2P_3 + \cdots + P_{2M-2}P_{2M-1}$, $\bar{y}_4 = P_1P_2 + \cdots + P_{2M-1}P_{2M}$, $\bar{y}_6 = P_1Q_2P_3 + \cdots + P_{2M-2}Q_{2M-1}P_{2M}$, $\cdots$.

4. **The Goldbach property.** Let S be the set of all random sequences $(\nu_1, \nu_2, \cdots)$ of 0's and 1's generated by P_n as in §1, and denote by F_N the subset such that for *every* (i, j) with $i < j$ and $i + j = N + 1$, the pair (ν_i, ν_j) is *not* (1, 1). Clearly F_N represents a sum of 3^H or $2 \cdot 3^H$ diadic intervals, where $H = N/2$ for N even, or $H = (N-1)/2$ for N odd, with the total probability

$$p_N = P(F_N) = \prod_1^H (1 - P_iP_{N+1-i}).$$

THEOREM. *If B is a random sequence of integers such that P_n is the chance that $n \in B$, and $\sum p_N$ converges, then almost every sequence has some initial index I such that every $N+1 > I$ is the sum of two distinct integers of B.*

For the prime-like sequences π or π' of §2, we have the interesting

COROLLARY. *For almost every sequence of "random-primes" every sufficiently large even integer is a sum of two distinct integers of the sequence.*

LOS ALAMOS SCIENTIFIC LABORATORY, UNIVERSITY OF CALIFORNIA, LOS ALAMOS, NEW MEXICO 87544