Cramer (1) discusses a simple stochastic model for the distribution of primes. Let there be a sequence of independent trials of an event $S_n$ with

$$Pr(S_n) = \frac{1}{\log n}$$

The numbers $P_1, P_2, \ldots P_m, \ldots$ of which $P_m$ is the $m$th value of $n$ for which $S_n$ occurs, will then with probability one have a limiting density $1/\log n$, like the primes. A number of other conclusions follow from the strong law of large numbers, for example that with probability one

$$\limsup_{n \to \infty} \frac{P_{m+1} - P_m}{\log^2 P_m} = 1$$

The model is however quite artificial in that it has the prime number theorem built into it ad hoc and in its assumption of independence. Primes are not independent in a statistical sense. The occurrence of an unusually long run of composite numbers leads one to expect a compensating increase in the number of primes later on.

A more natural stochastic model is that of the random sieve. In the sieve of Eratosthenes we sieve out the multiples of every number which is not a multiple of some earlier sieving number. We define the random sieve as follows: Check the number 2, and then with probability $\frac{1}{2}$ strike out each subsequent number. If $P_2$ is the first number not stricken out, check it and strike out each number thereafter with probability $1/P_2$. $P_3$ is the next number not stricken out, and we use $1/P_3$ as the probability with which to strike out each subsequent number, etc. The set of all possible sequences of numbers checked (sequences of random sieving numbers) contains the sequence of primes as one of its most probable members. Another typical member is the sequence of "lucky numbers" (2). $S_n$ now stands for the proposition that $n$ is a sieving number, and we have the following recurrence relation:

$$Pr(S_{n+1}) = Pr(S_n) - \frac{Pr(S_n)^2}{n}$$

Proof: Let $T_n$ be the contradictory of $S$. Let $S_n^*$ be the proposition that $n$ is not sieved out by any sieving number less than $n - 1$, and let $T_n^*$ be its contradictory. Then
\[ T_{n+1} = T_n T_{n+1} \vee S_n T^*_n + 1 \vee S_n T_{n+1} S^*_n + 1 \]

But
\[ \Pr(T_n T_{n+1}) = \Pr(T_n)^2, \quad \Pr(S_n T^*_n + 1) = \Pr(S_n) \Pr(T_n), \]
and
\[ \Pr(S_n T_{n+1} S^*_n + 1) = \Pr(S_n)^2/n. \]

Since the alternatives are mutually exclusive, the recurrence relation follows.

We solve (3) as follows:

Substitute \( g_n = 1/\Pr(S_n) \), obtaining
\[ g_{n+1} - g_n = \frac{1}{n - 1/g_n} \]
which may be solved recursively, using \( g_n = \sum 1/n = L(n) \) as the first trial solution. This gives \( g_n = L(n) + o(1) \) and thus
\[ \Pr(S_n) \sim 1/\log n \]

Thus the random sieve gives asymptotically the result assumed ad hoc by Cramer. The events \( S \) are not, moreover, independent. In fact, it is obvious that
\[ \Pr(T_n T_{n+1} \ldots T_n r-1) = \Pr(T_n)^r \]
which shows that the interdependence is negative, in conformity with our comment about the primes. The result does not affect the validity of (2), however, unless it strengthens it. For if in (6) we put \( r = c(\log^2 n) \), it is easy to see that the occurrence of a run of that length without sieving number is asymptotically of probability \( 1/n^c \). From the convergence of \( \sum 1/n^c \) for \( c > 1 \) it follows that with probability one the number of runs of such length is finite. Hence \( \log^2 P_n \) is almost certainly an upper bound, from some \( P_n \) on, to the interval \( P_{n+1} - P_n \). Because of interdependence it is more difficult to prove that this is false for any \( c > 1 \).

It is not difficult to define and solve various random-sieve problems analogous to those of multiplicative number theory, for example the relative frequency of pairs of sieving numbers separated by a given interval, the expected number of "divisors" of any number, etc.

In the case of the lucky numbers 1, 3, 7, 8, 13, 15, \ldots, we start with odd numbers and sieve out first every third odd number, leaving 7 as the next lucky. Then we sieve out every seventh of the remaining numbers, etc. In this case nothing is known about the asymptotic density, so we are in the position, say, of Gauss viz a viz the prime distribution. A randomization of this process is,
however, just the random sieve again, and from this fact we immediately conjecture that the asymptotic density of luckies is $1/\log n$ rather than, say, $B/\log n$ where $B \neq 1$. *

There is one direction in which the random sieve may facilitate something more than conjectures. The distribution of numbers prime to the first $m$ primes, or that of pairs of such numbers separated by a constant interval (e.g. 2) is in all likelihood more regular in a certain sense than the corresponding random sieve distribution, and if this is true it implies a number of results somewhat stronger than those that have been obtained, e.g. the infinity of twin primes.

Suppose that the first $m$ random sieving numbers are the first $m$ primes, $P_r = p_r \quad r = 1, 2, \ldots m$. Then the distribution of the number of numbers not sieved out from a sequence of $N$ consecutive numbers $> p_m$ is given by the binomial distribution, with probability $Q_m = (1 - 1/2)(1 - 1/3) \ldots (1 - 1/p_m)$. This same distribution may be expressed in a different way as the sum of $2^m$ random variables (non-independent). Let $N(i)$ be the number of numbers sieved out by the sieving number $p_i$, $N(i,j)$ those sieved out both by $p_i$ and $p_j$, etc. Those not sieved out will be given by the well-known combinatoric formula

$$N = \sum_i N(i) + \sum_{i,j} N(i,j) \ldots + (-1)^m N(1,2,3,\ldots,m).$$

For the sieve of Eratosthenes, on the other hand, we have a precisely similar expression. In this case the distribution of numbers prime to $2 \cdot 3 \cdot 5 \cdot \ldots p_m$ is periodic modulo $K_m$, and we pick our sequence of $N$ numbers, $N << K_m$, at random from a period of length $K_m$. We can calculate moments for the distribution of the number given by (7), both for the random and the Eratosthenes case. If as seems likely we can prove that moments of even order for the Eratosthenes distribution are smaller than the corresponding moments of the random sieve distribution, then it will follow that the longest interval between numbers prime to $K_m$ is of the order of $p_m \log p_m$, a stronger result than has been obtained by other methods. A similar argument applied to twins prime to $K_m$ would, if valid, establish the infinity of twin primes. It is not hard to prove the inequality for second moments, but the problem of proof for moments of order $2k$ remains.

* * * * *

* Subsequently verified by W.E. Briggs and the author, and independently by P. Erdos. See "The Lucky Number Theorem" to be published in this magazine.

References: