The Annals of Mathematical Statistics
1972, Vol. 43, No. 3, 865-870

THE EFFICIENT CONSTRUCTION OF AN UNBIASED RANDOM SEQUENCE[1]

BY PETER ELIAS

Massachusetts Institute of Technology

We consider procedures for converting input sequences of symbols generated by a stationary random process into sequences of independent, equiprobable output symbols, measuring the efficiency of such a procedure when the input sequence is finite by the expected value of the ratio of output symbols to input symbols. For a large class of processes and a large class of procedures we give an obvious information-theoretic upper bound to efficiency. We also construct procedures which attain this bound in the limit of long input sequences without making use of the process parameters, for two classes of processes. In the independent case we generalize a 1951 result of von Neumann and 1970 results of Hoeffding and Simons for independent but biased binary input, gaining a factor of 3 or 4 in efficiency. In the finite-state case we generalize a 1968 result of Samuelson for two-state binary Markov input, gaining a larger factor in efficiency.

1. Introduction. In 1951, von Neumann [5] described a procedure for generating an output sequence $z_1 z_2 \cdots z_m \cdots$ of statistically independent and equiprobable binary digits from an input sequence $x_1 x_2 \cdots x_n \cdots$ generated by a process $X_v(p)$ which chooses x_n from $\{0, 1\}$ with independence and with uniform bias: for all n, $x_n = 1$ with probability p, $x_n = 0$ with probability $q = 1 - p$, p unknown but fixed, $0 < p < 1$. von Neumann used on each of the pairs $x_1 x_2$, $x_3 x_4$, $\cdots$ the mapping

$$(1) \qquad 00 \to \Lambda\,, \qquad 01 \to 0\,, \qquad 10 \to 1\,, \qquad 11 \to \Lambda$$

where Λ represents no output digit. He defined the efficiency of this procedure as the expected number of output digits per input digit. For each input pair the probability of generating a non-null output digit z is $2pq$, so the efficiency is just $2pq/2 = pq$, which is $\frac{1}{4}$ at $p = q = \frac{1}{2}$ and less elsewhere. The map (1) is independent of the value of p, the output 0's and 1's are statistically independent and equiprobable for any p in (0, 1), but the efficiency depends on p.

Hoeffding and Simons [2] for the same process $X_v(p)$ investigate the mean delay or waiting time—i.e., the mean number or x's needed to generate the first non-null z. The von Neumann algorithm has expected delay 4: they find 3 as a lower bound for the class of strategies they consider and find a strategy of mean delay 3.10 at $p = q = \frac{1}{2}$. Their strategies also produce independent and equiprobable z_m for all p in (0,1).

Samuelson [3] starts with $x_1 x_2 \cdots x_n \cdots$ which are generated by a stationary two-state Markov process $X_s(p_0, p_1)$, in which $\Pr\{x_n = 1 \mid x_{n-1} = 0\} = p_0$ and

Received June 18, 1970; revised February 19, 1971.

[1] Research sponsored by the Joint Services Electronics Program (Contract DA28-043-AMC-02536(E)).

$\Pr\{x_n = 1 \mid x_{n-1} = 1\} = p_1$ are distinct but independent of n. He gives several mappings, of which the most efficient, credited to John W. Pratt, is

$$00 \to \Lambda\,, \qquad 01 \to \Lambda\,, \qquad 10 \to 0\,, \qquad 11 \to 1\,. \tag{2}$$

This mapping produces output 0's and 1's which are all generated in the 1-state of the process, and therefore all have the fixed probability p_1 that the output digit is a 1. The output digits of (2) can be used as inputs for the von Neumann mapping (1) to produce independent and equiprobable output digits. The efficiency of (2) is $\leqq \frac{1}{2}$, but $\geqq \frac{1}{4}$ if the more probable conditioning state is selected; applying first (2) and then (1) gives an efficiency $\leqq \frac{1}{8}$.

We concentrate on maximum efficiency rather than on minimum delay. For any stationary discrete-valued input random process X which meets a finiteness condition, and for any nonrandom mapping procedure which maps the output of X into independent equiprobable 0's and 1's we give the obvious informational upper bound to efficiency. For the independent but biased binary input process $X_v(p)$ of von Neumann and the binary Markov input process $X_s(p_0, p_1)$ of Samuelson, we construct procedures which approach the upper bound with increasing length of input sequence, and do so for any value of the process parameters (p in the independent case, p_0 and p_1 in the Markov case). In both cases, as p, p_0 and p_1 approach $\frac{1}{2}$, the efficiency approaches 1. Extension to nonbinary independent or Markov processes and to a more general class of finite-state input processes and to equiprobable and independent but nonbinary output symbols is immediate.

Definitions and bounds on efficiency. A random process $X = \{x_n,\ n \geqq 1\}$ is *acceptable* if it is stationary, takes values from an enumerable set A, and has marginal probabilities $Q_a = \Pr\{x_n = a\}$ with finite entropy

$$H(X^1) = -\textstyle\sum_{a \in A} Q_a \log_2 Q_a < \infty\,. \tag{3}$$

Let $\mathbf{x}^N = (x_1, x_2, \cdots, x_N)$. Let Z be the collection of sequence $(z_1, z_2, \cdots)$, where $z_i \in \{0, 1, \Lambda\}$ and if $z_i = \Lambda$ then $z_{i+1} = \Lambda$. Let B_N be a function on the range of $\mathbf{x}_N$ taking values in Z and (with some notational abuse) let $B_N(\mathbf{x}_N) = \mathbf{z} = (z_1, z_2, \cdots)$. Let $\mathbf{z}^m = (z_1, z_2, \cdots, z_m)$. B_N is *randomizing* for X iff for each m and each $\mathbf{z}^{m-1}$

$$0 \leqq \Pr\{z_m = 1 \mid \mathbf{z}^{m-1}\} = \Pr\{z_m = 0 \mid \mathbf{z}^{m-1}\} \leqq \tfrac{1}{2}\,. \tag{4}$$

Therefore in that subset of $z^m = \{\mathbf{z}^m\}$ which has no Λ's, all z_i take values 0 and 1 with equal probability and statistical independence.

For $\mathbf{z} \in Z$, let $t(z)$ be the number of coordinates $z_i \neq \Lambda$. The *efficiency* η_N of a randomizing B_N is defined as

$$\eta_N = (1/N)E(t(B_N(\mathbf{x}^N)))\,. \tag{5}$$

A randomizing procedure B is a sequence of randomizing functions $\{B_{N_i}\}$ defined for an increasing sequence of integers $N_1, N_2, \cdots$ and its efficiency η is the limit of η_N for $N \to \infty$. An example is C, the von Neumann procedure.

C_2 is the mapping (1) followed by Λ's. C_{2j} is the application of C_2 to j successive input pairs and the concatenation of the resulting non-null output digits. C is randomizing for any $X_v(p)$, p in (0, 1). Another is D, the Samuelson–Pratt procedure. D_{2j} maps j pairs of x by the mapping (2), and maps the resulting output pairs by C_2. D is randomizing for any $X_s(p_0, p_1)$, p_0, p_1 in (0, 1).

THEOREM 1. *If X is an acceptable random process and B is a randomizing procedure for X, then the efficiencies η_N of B_N and η of B are bounded above:*

$$\eta_N \leqq \frac{H(X^N)}{N}\,; \qquad \eta \leqq \lim_{N\to\infty} \frac{H(X^N)}{N}\,. \tag{6}$$

PROOF. We make use of standard information theory notation: for any two sets S and T with joint probabilities defined,

$$\begin{aligned} H(S) &= -\textstyle\sum_{s\in S} \Pr\{s\} \log_2 \Pr\{s\} \geqq 0 \\ H(S\,|\,T) &= -\textstyle\sum_{s\in S, t\in T} \Pr\{s, t\} \log_2 \Pr\{s\,|\,t\} \geqq 0\,. \end{aligned} \tag{7}$$

Then

$$\begin{aligned} H(X^N) &\geqq H(X^N) - H(X^N\,|\,Z) \\ &= H(Z) - H(Z\,|\,X^N) \\ &= H(Z) \\ &= \textstyle\sum_{m=1}^{\infty} H(Z^m\,|\,Z^{m-1})\,. \end{aligned} \tag{8}$$

The first line of (8) follows from the positivity of conditional entropy, the second from the equality of the average logarithm of the two factorings of a bivariate distribution into a univariate times a conditional, the third from the deterministic character of B_N which makes the conditional entropy of Z given X^N vanish, and the fourth from averaging the logarithm of the chain rule factoring of the joint distribution of $z_1, z_2, \cdots z_m \cdots$ into a product of conditionals: See e.g., Gallager [1], Chapter 2, for more detail.

Using (7) in the last line of (8),

$$\begin{aligned} H(Z^m\,|\,Z^{m-1}) &= -\textstyle\sum_{\mathbf{z}^{m-1}\in z^{m-1}} \sum_{z_m\in\{0,1,\Lambda\}} \Pr\{\mathbf{z}^m\} \log_2 \Pr\{z_m\,|\,\mathbf{z}^{m-1}\} \\ &\geqq \Pr\{z_m = 0\} + \Pr\{z_m = 1\} = \Pr\{z_m \neq \Lambda\}\,, \end{aligned} \tag{9}$$

since for all $\mathbf{z}^{m-1}$, $\log_2 \Pr\{z_m = \Lambda\,|\,\mathbf{z}^{m-1}\} \leqq 0$ and by (4) $\log_2 \Pr\{z_m = 0\,|\,\mathbf{z}^{m-1}\} = \log_2 \Pr\{z_m = 1\,|\,\mathbf{z}^{m-1}\} \leqq -1$. Substituting (9) in (8), and using the definition of $t(z)$ and (5),

$$\begin{aligned} H(X^N) &\geqq \textstyle\sum_{m=1}^{\infty} \Pr\{z_m \neq \Lambda\} \\ &= \textstyle\sum_{m=1}^{\infty} \Pr\{t(z) \geqq m\} \\ &= E(t(B_N(x^N))) = N\eta_N\,. \end{aligned} \tag{10}$$

The existence of the limit in (6) is in e.g., Gallager, op. cit. □

The function E_N. We next construct a function E_N, for each integer $N \geqq 2$, with $E_2 = C_2$, the von Neumann mapping (1). E_N is randomizing for $X_v(p)$ and has efficiency near the bound of Theorem 1 for large N.

Given N, divide the set of 2^N possible input sequences into the $N+1$ composition classes S_k, $0 \leqq k \leqq N$, S_k containing the $\binom{N}{k}$ sequences of length N which have k ones and $N-k$ zeros.

Let

$$n_k = \lfloor \log_2 \binom{N}{k} \rfloor , \tag{11}$$
$$\binom{N}{k} = \alpha_n 2^n + \alpha_{n-1} 2^{n-1} + \cdots + \alpha_0 2^0$$

where $\lfloor y \rfloor$ is the largest integer $\leqq y$, so that $\alpha_n \alpha_{n-1} \cdots \alpha_0$ is the binary expansion of the integer $\binom{N}{k}$, with $\alpha_n = 1$, $\alpha_j = 0$ or 1, $n > j \geqq 0$. (We set $n_k = n$ for typographical convenience.)

For each non-vanishing α_j, $0 \leqq j \leqq n$, assign the 2^j possible output binary sequences of length j to 2^j distinct members of S_k which have not already been assigned. One member of S_k will be assigned to Λ if $\alpha_0 = 1$ so that S_k is odd. S_0 and S_N have only one member each, which is therefore assigned to Λ. Making such assignments in, say, binary number order for all S_k, $0 \leqq k \leqq N$, completes the definition of E_N.

To compute the efficiency $\eta_N(p)$ of E_N, we first compute $\bar{n}_k$, the average number of output digits per input digit given that the input sequence is in S_k. Since all of the input sequences in S_k have the same probability for any fixed input p, $\bar{n}_k$ is just the weighted count

$$\bar{n} = \bar{n}_k = \frac{\alpha_n 2^n n + \alpha_{n-1} 2^{n-1}(n-1) + \cdots + \alpha_0 2^0 (n-n)}{\alpha_n 2^n + \alpha_{n-1} 2^{n-1} + \cdots + \alpha_0 2^0} \tag{12}$$
$$= n_k - \frac{\alpha_{n-1} 2^{n-1} + 2\alpha_{n-2} 2^{n-2} + \cdots + n\alpha_0 2^0}{\alpha_n 2^n + \alpha_{n-1} 2^{n-1} + \cdots + \alpha_0 2^0} .$$

The denominator of the fraction in the second line of (12) is bounded below by 2^n, and the numerator is bounded above by

$$2^{n-1} + 2 \cdot 2^{n-2} + \cdots + n \cdot 2^0 < 2^{n-1} \left(\frac{1}{1 - \frac{1}{2}} \right)^2 = 2^{n+1} . \tag{13}$$

Thus from the definition of n_k in (11), and (12) we have

$$\log_2 \binom{N}{k} \geqq n_k \geqq \bar{n}_k \geqq n_k - 2 \geqq \log_2 \binom{N}{k} - 3 . \tag{14}$$

The efficiency $\eta_N(p)$ of the mapping E_N is the average of the ratios $\bar{n}_k/N$, averaged with respect to the binomial probabilities of choosing an input sequence in S_k. It is therefore bounded by

$$\sum_{k=0}^{N} \binom{N}{k} p^k q^{N-k} \frac{\log_2 \binom{N}{k}}{N} \geqq \eta_N(p) \geqq \sum_{k=0}^{N} \binom{N}{k} p^k q^{N-k} \frac{\log_2 \binom{N}{k}}{N} - \frac{3}{N} . \tag{15}$$

Using the Stirling bounds on factorials gives, for fixed $\rho = k/N$,

$$\lim_{N \to \infty} \frac{1}{N} \log_2 \binom{N}{N\rho} = H_2(\rho) \tag{16}$$

where H_2 is the binary entropy function:

$$H_2(\rho) = -\rho \log_2 \rho - (1-\rho) \log_2 (1-\rho) . \tag{17}$$

By the weak law for the binomial distribution, given any $\varepsilon > 0$ and $\delta > 0$ there is an N_0 such that for $N > N_0$, all but at most δ of the probability in the binomial occurs in terms for which $|(k/N) - p| < \varepsilon$. This, together with the inequalities (15) and the fact that $\log_2 \binom{N}{k} \leqq N$ for $0 \leqq k \leqq N$ gives, for $N > N_0$,

$$\max_{|\alpha| \leqq \varepsilon} H_2(p + \alpha) + \delta \geqq \eta_N(p) \geqq (1 - \varepsilon) \min_{|\alpha| \leqq \varepsilon} H_2(p + \alpha) - \frac{3}{N} . \tag{18}$$

And (18), together with the continuity of H_2, proves that

$$\lim_{N \to \infty} \eta_N(p) = H_2(p) . \tag{19}$$

For any N, E_N is randomizing for $X_v(p)$, p in (0, 1), and may be used to define a randomizing procedure as C is defined via C_2. The sequence $x_1 x_2 \cdots$ is divided into N-tuples each of which is mapped by E_N and the non-null results concatenated. A more efficient admissible procedure E maps the first two input digits by E_2, the next four by E_4, the next six by $E_6 \cdots$ the next $2j$ by $E_{2j} \cdots$, and concatenates their outputs. It proves

THEOREM 2. *For the von Neumann ensemble $X(p)$, there is a randomizing procedure E which is independent of p, $p \varepsilon (0, 1)$, whose limiting efficiency $\eta(p)$ realizes the bound of Theorem* 1.

(After this paper was accepted for publication the author was informed that a result equivalent to Theorem 2 was independently obtained by J. A. Lechner at about the same time and by a similar method. J. Gill also informed the author at about the same time of similar results.)

The finite-state case. We next consider Samuelson's binary Markov process $X_s(p_0, p_1)$ with $p_0 = \Pr\{x_m = 1 \mid x_{m-1} = 0\}$ and $p_1 = \Pr\{x_m = 1 \mid x_{m-1} = 1\}$ fixed but arbitrary, both in (0, 1). The equilibrium equation for the two state probabilities $P = \Pr\{x_n = 1\}$ and $Q = 1 - P$ gives

$$P = Qp_0 + Pp_1 = (1 - P)p_0 + Pp_1 , \tag{20}$$

$$P = \frac{p_0}{1 + p_0 - p_1}$$

Given a sequence $x_1 x_2 \cdots$ generated by $X_s(p_0, p_1)$, we use x_1 only to determine the state of the process when x_2 is generated, and decompose $x_2 x_3 \cdots$ into two sequences. S_0 consists of all x_m for which $x_{m-1} = 0$, concatenated in increasing index order, produced by the mapping

$$x_m \to x_m \quad \text{if} \quad x_{m-1} = 0 ; \qquad x_m \to \Lambda \quad \text{otherwise} \tag{21}$$

and S_1 consists of all x_m for which $x_{m-1} = 1$, similarly concatenated.

S_0 and S_1 are processes which generate independent binary symbols, of the type $X_v(p)$ of Theorem 2, with parameters $p = p_0$ and $p = p_1$ respectively. The process $X_s(p_0, p_1)$ is ergodic, so there exists an ε_M which $\to 0$ with increasing M such that when $X_s(p_0, p_1)$ generates M symbols, with probability $> 1 - \varepsilon_M$ S_0

generates $> M(Q - \varepsilon_M)$ symbols and S_1 generates $> M(P - \varepsilon_M)$ symbols. Applying procedure E independently to S_0 and to S_1, and concatenating the output sequences, the expected number of total output symbols for an input M-tuple from $X(p_0, p_1)$ is bounded below by

$$(22) \qquad (1 - \varepsilon_M)M(Q - \varepsilon_M)\eta_{M(Q-\varepsilon_M)}(p_0) + (1 - \varepsilon_M)M(P - \varepsilon_M)\eta_{M(Q-\varepsilon_M)}(p_1) .$$

Dividing by M, taking the limit as $M \to \infty$, using Theorem 2 and noting that (see e.g., Gallager, *op. cit.* or Shannon [4])

$$(23) \qquad H(X_s^N) = H_2(P) + (N - 1)[QH_2(p_0) + PH_2(p_1)]$$

proves

THEOREM 3. *Given an acceptable two-state Markov source $X_s(p_0, p_1)$, there is a randomizing procedure which is independent of p_0 and p_1 and has a limiting efficiency $\eta(p_0, p_1)$ which attains the bound of Theorem* 1.

Extensions. The results all extend trivially to the mapping of an independent or Markov input process X with finite alphabet of $a \geqq 2$ letters onto an independent equiprobable b- letter output process Z with $z_i \in \{0, 1, \cdots, b - 1, \Lambda\}$. $\log_b$ replaces $\log_2$ in all entropy calculations. E_N is constructed using b-ary enumeration of the integral parts of the $\log_b$ of multinomial coefficients in (11)—(16) to prove Theorem 2. The Markov process has a states and decomposes into a processes to prove Theorem 3. A kth order Markov process whose state is determined by k preceding letters can also be decomposed (into a^k processes), as can one of Shannon's finite-state sources with state known to the receiver ([1], [4]), and Theorem 3 holds.

REFERENCES

[1] GALLAGER, R. G. (1968). *Information Theory and Reliable Communication.* Wiley, New York.

[2] HOEFFDING, W. and SIMONS, G. (1970). Unbiased coin tossing with a biased coin. *Ann. Math. Statist.* **41** 341–352.

[3] SAMUELSON, P. A. (1968). Constructing an unbiased random sequence. *J. Amer. Statist. Assoc.* **63** 1526–1527.

[4] SHANNON, C. E. and WEAVER, W. (1949). *The Mathematical Theory of Communication.* Univ. of Illinois Press.

[5] VON NEUMANN, J. (1951). Various technique used in connection with random digits. Monte Carlo Method, Applied Mathematics Series, No. 12, U.S. National Bureau of Standards, Washington D.C. 36–38.