Correspondence

The Entropy of a Randomly Stopped Sequence

Laura Ekroot, *Student Member, IEEE*, and
Thomas M. Cover, *Fellow, IEEE*

Abstract—**A Wald-like equation is proved for the entropy of a randomly stopped sequence of independent identically distributed discrete random variables $X_1, X_2, \cdots$ with a nonanticipating stopping time N. Specifically, it is shown that**

$$H(X^N) = (EN)H(X_1) + H(N|X^\infty),$$

where X^N denotes the randomly stopped sequence. Thus, the randomness in the stopped sequence X^N is the expected number of "calls" for X times the entropy per call plus the residual randomness in the stopping time conditioned on the unstopped sequence X^∞.

Index Terms—**Entropy, stopping time, Wald's equation, stopped sequences.**

I. Introduction

The entropy of a randomly stopped sequence of independent identically distributed (i.i.d.) random variables depends not only on the texture of the sequence but on the length of the sequence as well. Let $X_1, X_2, \cdots$ be independent identically distributed random variables distributed according to the probability mass function $p(x)$, $x \in \mathscr{X}$, where $\mathscr{X}$ is a finite alphabet. We shall consider stopping times N which are (perhaps randomized) nonanticipating functions of the observed process $X_1, X_2, \cdots$. In particular, the stopping event $\{N = n\}$ is required to be independent of the future $\{X_{n+1}, X_{n+2}, \cdots\}$. We denote by X^N the finite length sequence $X_1 X_2 \cdots X_N \in \mathscr{X}^*$, where $\mathscr{X}^*$ denotes the set of all finite-length sequences with alphabet $\mathscr{X}$. Let X^0 denote the sequence with no elements. The stopping time N induces a probability mass function $q(x)$, $x \in \mathscr{X}^*$, where $q(x) = \Pr\{X^N = x\}$. We wish to evaluate the entropy $H(X^N) = -\sum_{x \in \mathscr{X}^*} q(x) \log q(x)$ of the stopped sequence X^N. Thus, $H(X^N)$ is the descriptive complexity of the randomly stopped sequence.

Consider, for example, a gambler who starts with one dollar and plays a game where at each time $n = 1, 2, \cdots$ his wealth increases one dollar with probability p or decreases one dollar with probability $q = 1 - p$. The game ends when the gambler's wealth is zero. The ups and downs of the gambler's wealth form an i.i.d. sequence. For $p < \frac{1}{2}$, it is certain the gambler will go broke. Although the outcome is certain, the path is random. We will show that the path entropy is $H(p)/(q-p)$, where $H(p) = -p \log p - q \log q$.

We first define a general stopping time and the associated stopped sequence, and then present the two main theorems for the entropy of a stopped sequence. The formal proofs of the lemmas necessary for the final step of the proof of the theorems are deferred until Section VI. Stopping times are discussed in [6] and [4]. Wald's equation, which is related to our work, is proved in [7] and discussed in [6]. An expression, due to Abramov, for the entropy rate of stopped sequences with deterministic stopping rules for stationary ergodic processes is given by Brown [8, p. 140]. Some related results on the asymptotic entropy rate for channels without synchronization are developed in [5] and [1]. Related work on the entropy of rooted trees is presented by Massey in [3]. Further work extending the entropy of stopped sequences to Markov trajectories may be found in [2].

Manuscript received June 18, 1990; revised April 12, 1991. This work was supported in part by the National Science Foundation under NSF Grant NCR-89-14538, DARPA Contract JFBI-91-218, and JSEP DAAL03-91-0010.
L. Ekroot is with the Department of Electrical Engineering, Durand 141A, Stanford University, Stanford, CA 94305.
T. Cover is with the Departments of Electrical Engineering and Statistics, Durand 121, Stanford University, Stanford, CA 94305.
IEEE Log Number 9102262.

II. Stopping Times

We first formalize the definition of a general stopping time N, which admits the possibility of subsidiary randomization.

Definition 1 (Stopping time): Let $\{\mathscr{B}_n\}$ be an increasing sequence of σ-fields, and let $X_1, X_2, \cdots$ be i.i.d. random variables adapted to $\{\mathscr{B}_n\}$. Then a random variable N that takes values in $\{0, 1, 2, \cdots\}$ is a *stopping time* if for every $n = 0, 1, 2, \cdots$, the event $\{N = n\}$ is in $\mathscr{B}_n$, and $\{N = n\}$ is independent of $\{X_{n+1}, X_{n+2}, \cdots\}$.

Definition 2: A stopping time N is *bounded* if there is some nonnegative integer m such that $\Pr\{N \le m\} = 1$.

We note that if $\mathscr{B}_n$ is the σ-field $\sigma(X_1, X_2, \cdots, X_n)$ generated by $X_1, X_2, \cdots, X_n$ then N is determined by the sequence and there is no subsidiary randomization in the stopping time. A stopping time may depend entirely on the sequence, as with the rule "Stop after the first occurrence of a head" or "Stop at time 3." A stopping time that also has subsidiary randomness is illustrated by the stopping rule "Stop at any head, and stop at each tail with conditional probability α," or "Stop with probability α at each opportunity." These examples will be evaluated in Section V.

III. The Stopping Time Theorem

The following theorem gives the entropy of a randomly stopped sequence. Throughout, we assume $X_1, X_2, \cdots$ to be independent identically distributed discrete random variables. Let $X^N \in \mathscr{X}^*$ denote the randomly stopped sequence and let X^∞ denote the unstopped sequence $X_1, X_2, \cdots$.

Theorem 1: Let N be any stopping time for the i.i.d. $X_1, X_2, \cdots$. Then

$$H(X^N) = (EN)H(X_1) + H(N|X^\infty), \tag{1}$$

provided that it is not the case that both $EN = \infty$ and $H(X_1) = 0$.

Proof: The proof will be separated into two parts. First, the proof for a bounded stopping time will be presented. Then the boundedness condition will be removed.

Let N be a bounded stopping time with bound n, i.e., $\Pr\{N \le n\} = 1$. Let X_1^n denote $(X_1 X_2 \cdots X_n)$, and X_1^0 denote the sequence with no elements. Let X^∞ denote the unstopped sequence $X_1 X_2 \cdots$. Assume that $X_1, X_2, \cdots, X_n$ are i.i.d. random variables drawn according to probability mass function $p(x)$, $x \in \mathscr{X}$.

Since N is a function of X_1^N, and since $(X_{k+1}, X_{k+2}, \cdots, X_n)$ is conditionally independent of X_1^k given $\{N = k\}$, it follows that the sequence $X_1^N \to N \to X_{N+1}^n$ forms a Markov chain.

We now expand $H(X_1^n, N)$ in two ways. First, by the chain rule, and the fact that $X_1, X_2, \cdots, X_n$ are independent and identically distributed, we obtain

$$H(X_1^n, N) = H(X_1^n) + H(N|X_1^n) \tag{2}$$

$$= nH(X_1) + H(N|X_1^n). \tag{3}$$

Alternatively, since $N \le n$, we have the expansion

$$H(X_1^n, N) = H(X_1^N, X_{N+1}^n, N) \tag{4}$$

$$= H(X_1^N) + H(N|X_1^N) + H(X_{N+1}^n|X_1^N, N) \tag{5}$$

$$= H(X_1^N) + H(X_{N+1}^n|X_1^N, N) \tag{6}$$

$$= H(X_1^N) + H(X_{N+1}^n|N) \tag{7}$$

$$= H(X_1^N) + \sum_{k=0}^{n-1} \Pr\{N = k\} H(X_{N+1}^n|N = k) \tag{8}$$

$$= H(X_1^N) + \sum_{k=0}^{n-1} \Pr\{N = k\} H(X_{k+1}^n) \tag{9}$$

$$= H(X_1^N) + \sum_{k=0}^{n-1} \Pr\{N = k\}(n-k) H(X_1) \tag{10}$$

$$= H(X_1^N) + nH(X_1) - (EN)H(X_1), \tag{11}$$

where (5) is the chain rule for entropy, (6) follows from the fact that N is a deterministic function of the finite sequence X_1^N, and (7) follows from the Markovity of X_1^N, N, and X_{N+1}^n.

Combining (3) and (11), we obtain

$$nH(X_1) + H(N|X_1^n) = H(X_1^N) + nH(X_1) - (EN)H(X_1) \tag{12}$$

or

$$H(X_1^N) = (EN)H(X_1) + H(N|X_1^n). \tag{13}$$

Since the stopping time N is independent of the future and bounded by n, N is independent of the sequence after time n, and

$$H(N|X_1^n) = H(N|X^\infty). \tag{14}$$

Thus, combining (13) and (14) yields

$$H(X_1^N) = (EN)H(X_1) + H(N|X^\infty), \tag{15}$$

which establishes the theorem for bounded stopping times.

Now, to remove the boundedness condition on N, let N be unbounded, and consider the truncated stopping time N_n defined by

$$N_n = \min(n, N) \tag{16}$$

$$= \begin{cases} N, & \text{if } N < n, \\ n, & \text{if } N \ge n. \end{cases} \tag{17}$$

Since N_n is a bounded stopping time, application of (15) yields

$$H(X_1^{N_n}) = H(X_1)E(N_n) + H(N_n|X^\infty). \tag{18}$$

It remains to be shown that

$$H(X_1^{N_n}) \to H(X_1^N), \tag{19}$$

$$EN_n \to EN, \tag{20}$$

and

$$H(N_n|X^\infty) \to H(N|X^\infty), \tag{21}$$

in order to establish the theorem for unbounded stopping times. These will be proved in Lemmas 1–3 in Section VI, completing the proof that

$$H(X_1^N) = (EN)H(X_1) + H(N|X^\infty) \tag{22}$$

for arbitrary randomized stopping times. □

In general, the entropy $H(X^N)$ is greater than $(EN)H(X_1)$. We now show that the conditional entropy $H(X^N|N)$ is generally less than $(EN)H(X_1)$.

Theorem 2: Let N be any (possibly randomized) stopping time. Then

$$H(X^N|N) = (EN)H(X_1) - I(N; X^\infty), \tag{23}$$

where $X^N \in \mathscr{X}^*$ denotes the randomly stopped sequence.

Proof:

$$H(X^N|N) = H(X^N, N) - H(N) \tag{24}$$

$$= H(X^N) + H(N|X^N) - H(N) \tag{25}$$

$$= H(X^N) - H(N) \tag{26}$$

$$= (EN)H(X_1) + H(N|X^\infty) - H(N) \tag{27}$$

$$= (EN)H(X_1) - I(N; X^\infty), \tag{28}$$

where (26) follows from the fact that X_1^N reveals N, (27) follows from Theorem 1, and (28) follows from the definition of mutual information. □

Remark: Theorems 1 and 2 provide the upper and lower bounds on the expression $(EN)H(X_1)$ given by

$$H(X^N|N) = (EN)H(X_1) - I(N; X^\infty) \tag{29}$$

$$\le (EN)H(X_1) \tag{30}$$

$$\le (EN)H(X_1) + H(N|X^\infty) = H(X^N). \tag{31}$$

The next theorem is implicit in the theorem of Abramov [8, p. 140].

IV. Special Cases

We obtain simplified expressions for $H(X^N)$ under a variety of restrictions on the stopping time.

Theorem 3 (Determined Stopping Time): A stopping time N is said to be a *determined stopping time* if $\{N = n\} \in \sigma(X_1, X_2, \cdots, X_n)$ for all $n = 1, 2, \cdots$, where $\sigma(X_1, X_2, \cdots, X_n)$ is the σ-field generated by $X_1, X_2, \cdots, X_n$. Then, for a determined stopping time N,

$$H(X^N) = (EN)H(X_1), \tag{32}$$

where $X^N \in \mathscr{X}^*$ denotes the randomly stopped sequence.

Proof: Since N is a deterministic function of the sequence, it follows that $H(N|X^\infty) = 0$ in Theorem 1. □

Theorem 4 (Independent Stopping Time): The stopping time N is said to be an *independent stopping time* if the event $\{N = n\}$ is independent of the process $\{X_i\}_{i=1}^\infty$. Then, for an independent stopping time N,

$$H(X^N) = (EN)H(X_1) + H(N). \tag{33}$$

Proof: Since N is independent of the sequence, $H(N|X^\infty) = H(N)$ in Theorem 1. □

V. Examples

We now analyze the examples introduced in Section II. Let $X_1, X_2, \cdots$ be independent identically distributed binary random variables with $\Pr\{X_i = 1\} = p$, $\Pr\{X_i = 0\} = q$, and $q = 1 - p$. Let $H(r) = -r\log r - (1-r)\log(1-r)$. Then $H(X_1) = H(p)$.

Example 1 (Determined Stopping Time):

N: Stop after the first occurrence of 0.

Then

$$EN = \frac{1}{q} \tag{34}$$

and

$$H(X_1^N) = (EN)H(X_1) = \frac{1}{q}H(p). \tag{35}$$

Example 2 (Fixed Stopping Time):

N: Stop at time 3.

Then

$$EN = 3 \tag{36}$$

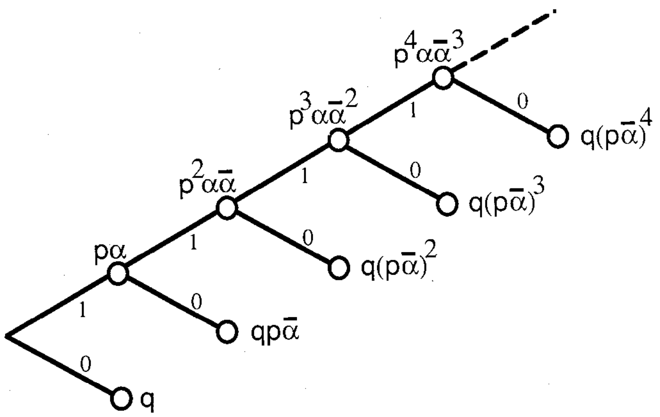
and

$$H(X_1^N) = (EN)H(X_1) = 3H(p). \tag{37}$$

Example 3 (Randomized Stopping Time):

N: Stop at any 0, and stop at each 1 with probability α.

Note that some of the stopped sequences correspond to interior nodes of the tree. In the diagram



the sequence $X^N = 11$ has probability $p^2\alpha\bar{\alpha}$.

We find

$$H(N|X^\infty) = \frac{p}{p\alpha + q}H(\alpha) \tag{38}$$

and

$$EN = \frac{1}{p\alpha + q}. \tag{39}$$

Thus,

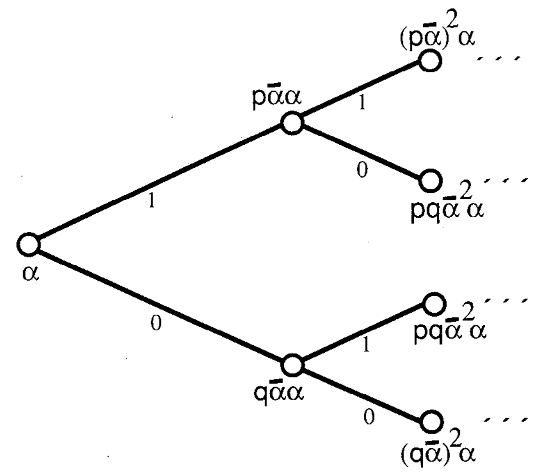
$$H(X_1^N) = (EN)H(X_1) + H(N|X^\infty) \tag{40}$$

$$= \frac{1}{p\alpha + q}H(P) + \frac{p}{p\alpha + q}H(\alpha). \tag{41}$$

Example 4 (Independent Stopping Time):

N: Stop with probability α at each opportunity.

Note that the null sequence X^0 occurs with probability α. Again, the stopped nodes do not form the leaves of a tree, as shown in the diagram:



Here,

$$H(N|X^\infty) = H(N) = \frac{1}{\alpha}H(\alpha) \tag{42}$$

and

$$EN = \bar{\alpha}/\alpha. \tag{43}$$

Consequently,

$$H(X_1^N) = (EN)H(X_1) + H(N) \tag{44}$$

$$= \frac{\bar{\alpha}}{\alpha}H(p) + \frac{1}{\alpha}H(\alpha). \tag{45}$$

Example 5 (Gambler's Ruin): Let a gambler have initial wealth $S_0 = 1$, and gamble 1 unit each time at even odds on independent gambles with success probability $p < 1/2$. The gambler stops at the fist time N that his wealth is zero. Thus, $S_k = \sum_{i=1}^k X_i$, where $X_1, X_2, \cdots$ are i.i.d. with $\Pr\{X_i = 1\} = p$, and $\Pr\{X_i = -1\} = q$. We are interested in the entropy $H(S^N)$ of the gambler's history $(S_1, S_2, \cdots, S_N)$. Note that the final wealth S_N is a rather dull random variable since the gambler is sure to go broke. Thus $H(S_N) = 0$. How exciting is the process by which he went broke? We observe that $\{S_i\}$ is not an i.i.d. process, but $\{X_i\}$ is i.i.d., and $(S_1, S_2, \cdots, S_n)$ is in one-to-one correspondence with $(S_0, X_1, X_2, \cdots, X_n)$. Thus, $H(S^N) = H(X^N)$ and Theorem 1 applies. We have

$$H(X_1) = H(p), \tag{46}$$

$$H(N|X^\infty) = 0, \tag{47}$$

and

$$EN = 1/(q-p). \tag{48}$$

Thus, the entropy of the gambler's ups and downs is

$$H(S^N) = H(X^N) \tag{49}$$

$$= (EN)H(X_1) = \frac{1}{q-p}H(p), \tag{50}$$

for $p < 1/2$.

We see, for example, that if $p = 0.49$ and the initial wealth S_0 is equal to 1, then the entropy of the gambler's story is 49.986 bits.

VI. Details

We now prove the three lemmas needed to complete the proof of Theorem 1. Throughout, let $N_n = \min(n, N)$ denote the truncated stopping time for N. As before, $\{X_i\}$ is an i.i.d. sequence of discrete random variables.

Lemma 1: Let N be a stopping time with $\Pr\{N<\infty\}=1$. Then

$$\lim_{n\to\infty} EN_n = EN. \tag{51}$$

Proof: We bound EN_n above by EN and then show that EN_n is greater than an expression that goes to EN in the limit. We use this method in the subsequent lemmas as well. Specifically,

$$EN = \sum_{k=0}^{n-1} k\Pr\{N=k\} + \sum_{k=n}^{\infty} k\Pr\{N=k\} \tag{52}$$

$$\geq \sum_{k=0}^{n-1} k\Pr\{N=k\} + n\Pr\{N\geq n\} \tag{53}$$

$$= EN_n \tag{54}$$

$$\geq \sum_{k=0}^{n-1} k\Pr\{N=k\} \tag{55}$$

$$\xrightarrow{n\to\infty} \sum_{k=0}^{\infty} k\Pr\{N=k\} \tag{56}$$

$$= EN.. \qquad \square \tag{57}$$

Lemma 2: Let N be a stopping time with $\Pr\{N<\infty\}=1$. Then

$$\lim_{n\to\infty} H(X^{N_n}) = H(X^N). \tag{58}$$

Proof: Let x denote a finite length sequence in $\mathscr{X}^*$, where $\mathscr{X}^*$ denotes the set of all finite length sequences with alphabet $\mathscr{X}$, and let $q(x)=\Pr\{X^N=x\}$. Then

$$H(X^N) = -\sum_{x\in\mathscr{X}^*} q(x)\log q(x). \tag{59}$$

Since we will be using the truncated stopping time N_n, we let $q_n(x)=\Pr\{X^{N_n}=x\}$ and note that

$$q_n(x) = \begin{cases} q(x), & \text{if } |x|<n, \\ \sum_{u\in\mathscr{X}^*} q(xu), & \text{if } |x|=n, \\ 0, & \text{if } |x|>n, \end{cases} \tag{60}$$

where $|x|$ denotes the length of x. For $|x|=n$, $q_n(x)$ is the lumped probability of all stopped sequences with prefix x.

We now use the fact that X^{N_n} is a function of X^N (since N_n is a function of N) to obtain the first inequality in the following chain:

$$H(X^N) \geq H(X^{N_n})$$

$$= -\sum_{\substack{x\in\mathscr{X}^* \\ |x|<n}} q(x)\log q(x) - \sum_{x\in\mathscr{X}^n} q_n(x)\log q_n(x) \tag{62}$$

$$\geq -\sum_{\substack{x\in\mathscr{X}^* \\ |x|<n}} q(x)\log q(x) \tag{63}$$

$$\xrightarrow{n\to\infty} -\sum_{x\in\mathscr{X}^*} q(x)\log q(x) \tag{64}$$

$$= H(X^N). \qquad \square \tag{65}$$

Lemma 3: Let N be a stopping time with $\Pr\{N<\infty\}=1$. Then

$$\lim_{n\to\infty} H(N_n|X^\infty) = H(N|X^\infty). \tag{66}$$

Proof: Since N_n is a function of N, we can bound $H(N|X^\infty)$ by

$$H(N|X^\infty) \geq H(N_n|X^\infty) \tag{67}$$

$$= E\Bigg(-\sum_{k=0}^{n-1} \Pr\{N=k|X^\infty\}\log\Pr\{N=k|X^\infty\}$$

$$-\Pr\{N\geq n|X^\infty\}\log\Pr\{N\geq n|X^\infty\}\Bigg) \tag{68}$$

$$\geq E\left(-\sum_{k=0}^{n-1} \Pr\{N=k|X^\infty\}\log\Pr\{N=k|X^\infty\}\right) \tag{69}$$

$$\xrightarrow{n\to\infty} E\left(-\sum_{k=0}^{\infty} \Pr\{N=k|X^\infty\}\log\Pr\{N=k|X^\infty\}\right) \tag{70}$$

$$= H(N|X^\infty), \tag{71}$$

where the expectation in (68) is over the random unstopped sequence X^∞. □

VII. Comments

Wald [7] considered randomly stopped sums $S_N=\sum_{i=1}^N X_i$ of i.i.d. random variables X_i, where N is a stopping time as defined in Section II. Wald's equation is

$$ES_N = (EN)(EX_1), \tag{72}$$

provided it is not the case that $EX_1=0$ and $EN=\infty$. In fact, if N is a determined stopping time, i.e., $\{N=n\}\in\sigma(X_1,X_2,\cdots,X_n)$, Wald's equation can be used to prove Theorem 3. The general equation for arbitrary randomized stopping times

$$H(X^N) = (EN)H(X_1) + H(N|X^\infty) \tag{73}$$

has a term due to subsidiary randomization that has no counterpart in Wald's equation.

Acknowledgment

The authors wish to thank J. A. Thomas, S. S. Shamai (Shitz), and L. B. Levitin for many helpful discussions.

References

[1] R. Ahlswede and J. Wolfowitz, "Channels without synchronization," *Adv. Appl. Prob.*, vol. 3, pp. 383–403, 1971.
[2] L. Ekroot and T. M. Cover, "The entropy of Markov trajectories," submitted to *IEEE Trans. Inform. Theory*, 1991.
[3] J. L. Massey, "The entropy of a rooted tree with probabilities," in *Abstract of Papers, IEEE Int. Symp. Inform. Theory*, San Jovite, Canada, Sept. 1983.
[4] P. Billingsley, *Probability and Measure*. New York: John Wiley, 1986.
[5] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors," *Probl. Peredach. Inform.*, vol. 3, no. 4, pp. 18–36, 1967.
[6] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*. San Diego, CA: Academic Press, 1975.
[7] Y. S. Chow, H. Robbins, and D. Siegmund, *Great Expectations: The Theory of Optimal Stopping*. Boston, MA: Houghton Mifflin Company, 1971.
[8] J. R. Brown, *Ergodic Theory and Topological Dynamics*. New York: Academic Press, 1976.