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

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

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 $X^N$.

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 structure 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 $p(x), x \in \mathcal{X}$, where $\mathcal{X}$ is a finite alphabet. We shall consider stopping times $N$ which are (randomly permuted) 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 $X_1, X_2, \cdots X_N \in \mathcal{X}^N$, where $\mathcal{X}^N$ denotes the set of all finite-length sequences with alphabet $\mathcal{X}$. Let $X^0$ denote the sequence with no elements. The stopping time $N$ induces a probability mass function $q(n), n \in \mathcal{N}$, where $q(n) = \Pr(X^N = n)$. We wish to evaluate the entropy $H(X^N) = \sum_{x \in \mathcal{X}} -q(x) \log q(x)$ 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 a random walk. Although the outcome is certain, the path is random. We note that if $\mathcal{G}_n$ is the a-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 $\alpha$." 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 $X^N \in \mathcal{X}^N$ denote the randomly stopped sequence and let $X^n$ 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

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

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., $\Pr(N \leq n) = 1$. Let $X^n$ denote $(X_1, X_2, \cdots, X_n)$, and $X^0$ denote the sequence with no elements. Let $X^\infty$ denote the unstopped sequence $X_1, X_2, \cdots$. Assume that $X_1, X_2, \cdots, X_n$ are i.i.d. ran-
dom variables drawn according to probability mass function \( p(x) \), \( x \in \mathcal{X} \).

Since \( N \) is a function of \( X^N \), and since \((X_{k+1}, X_{k+2}, \ldots, X_n)\) is conditionally independent of \( X^k \) given \( \{N = k\} \), it follows that the sequence \( X^k \rightarrow N \rightarrow X^N_{k+1} \) forms a Markov chain.

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

\[
H(X^N_k, N) = H(X^N_k) + H(N | X^N_k) \tag{2}
\]

Alternatively, since \( N \leq n \), we have the expansion

\[
H(X^N_k, N) = H(X^N_k) + H(N | X^N_k) + H(X^N_{k+1} | N) \tag{3}
\]

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

\[
H(X^N_k) = (EN)H(X_k) + H(N | X^N_k) \tag{?}
\]

for arbitrary randomized stopping times. \( \square \)

In general, the entropy \( H(X^N) \) is greater than \( (EN)H(X_k) \).

We now show that the conditional entropy \( H(X^N | N) \) is generally less than \((EN)H(X_k)\).

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

\[
H(X^N | N) = (EN)H(X_k) + H(N | X^N_k) \tag{?}
\]

where \( X^N \in \mathcal{X}^* \) denotes the randomly stopped sequence.

**Proof:**

\[
\begin{align*}
H(X^N | N) &= H(X^N, N) - H(N) \\
&= H(X^N) - H(N) \\
&= (EN)H(X_k) + H(N | X^N_k) - H(N), \\
&= (EN)H(X_k) + H(N | X^N_k),
\end{align*}
\]

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

**Remark:** Theorems 1 and 2 provide the upper and lower bounds on the expression \((EN)H(X_k)\) given by

\[
H(X^N_k) \leq (EN)H(X_k) \leq H(X^N_k) \tag{?}
\]

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 \( \{N = n\} \in \mathcal{F}(X_1, X_2, \ldots, X_n) \) for all \( n = 1, 2, \ldots, \), where \( \mathcal{F}(X_1, X_2, \ldots, X_n) \) is the \( \sigma \)-field generated by \( X_1, X_2, \ldots, X_n \). Then, for a determined stopping time \( N \),

\[
H(X^N) = (EN)H(X_k) \tag{?}
\]

where \( X^N \in \mathcal{X}^* \) denotes the randomly stopped sequence.

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

**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 \( \{X_i\}_{i=1}^n \). Then, for an independent stopping time \( N \),

\[
H(X^N) = (EN)H(X_k) + H(N) \tag{?}
\]

where \( X^N \in \mathcal{X}^* \) denotes the randomly stopped sequence.

**Proof:** Since \( N \) is independent of the sequence, \( H(N | X^N) = H(N) \) in Theorem 1. \( \square \)
V. Examples

We now analyze the examples introduced in Section II. Let $X_1, X_2, \ldots$ be independent identically distributed binary random variables with $\Pr(X_1 = 1) = p$, $\Pr(X_1 = 0) = q$, 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

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

and

$$H(X_N^o) = (EN)H(X_1) = \frac{1}{q} H(p).$$

Example 2 (Fixed Stopping Time):

$N$: Stop at time 3.

Then

$$EN = 3$$

and

$$H(X_N^o) = (EN)H(X_1) = 3H(p).$$

Example 3 (Randomized Stopping Time):

$N$: Stop at any 0, and stop at each 1 with probability $\alpha$.

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

Here,

$$H(N | X^o) = H(N) = \frac{1}{\alpha} H(\alpha)$$

and

$$EN = \frac{\alpha}{\alpha}.$$

Consequently,

$$H(X^o) = (EN)H(X_1) + H(N)$$

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

Example 4 (Independent Stopping Time):

$N$: Stop with probability $\alpha$ at each opportunity.

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

$$H( N | X^o) = H(N) = \frac{1}{\alpha} H(\alpha)$$

and

$$EN = \frac{1}{q (q-p)}.$$

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

$$H(S^N) = H(X^N)$$

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

$$\lim_{n \to \infty} EN_n = EN.$$  \hfill (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,

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

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

$$= EN_n$$ \hfill (54)

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

$$\geq \sum_{k=0}^{n-1} k \Pr\{N = k\}$$ \hfill (56)

$$= EN_n.$$ \hfill (57)

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

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

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

$$H(X^N) = -\sum_{x \in \mathcal{Z}^*} q(x) \log q(x).$$ \hfill (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

$$q_n(x) = \begin{cases} q(x), & \text{if } |x| < n, \\ \sum_{u \in \mathcal{Z}^* : |x| = n} q(u), & \text{if } |x| = n, \\ 0, & \text{if } |x| > n, \end{cases}$$ \hfill (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:

$$H(X^N) \geq H(X^N_n)$$ \hfill (61)

$$= -\sum_{x \in \mathcal{Z}^* : |x| < n} q_n(x) \log q_n(x)$$ \hfill (62)

$$\geq -\sum_{x \in \mathcal{Z}^* : |x| < n} q(x) \log q(x)$$ \hfill (63)

$$\geq -\sum_{x \in \mathcal{Z}^* : |x| < n} q(x) \log q(x)$$ \hfill (64)

$$\Rightarrow \sum_{x \in \mathcal{Z}^*} q(x) \log q(x)$$ \hfill (65)

$$= H(X^N).$$ \hfill (66)

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

$$\lim_{n \to \infty} H(N_n|X^\infty) = H(N|X^\infty).$$ \hfill (67)

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

$$H(N_n|X^\infty) = H(N_n|X^\infty) \leq H(N_n|X^\infty)$$ \hfill (68)

$$\leq E \left( -\sum_{k=0}^{n-1} \Pr\{N = k\} \log \Pr\{N = k\} \right)$$ \hfill (69)

$$\leq E \left( -\sum_{k=0}^{n-1} \Pr\{N = k\} \log \Pr\{N = k\} \right)$$ \hfill (70)

$$= H(N|X^\infty).$$ \hfill (71)

where the expectation in (68) is over the random unstopped sequence $X^\infty$.

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

$$ES_N = (EN)(EX_1),$$ \hfill (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., $(N = n) \in \sigma(X_1, X_2, \ldots, X_n)$, Wald’s equation can be used to prove Theorem 3. The general equation for arbitrary randomized stopping times

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