An Upper Bound on the Entropy Series

A. D. Wyner

Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey

An upper bound is established for the entropy corresponding to a positive integer valued random variable $X$ in terms of the expectation of certain functions of $X$. In particular, we show that the entropy is finite if $E \log X < \infty$. Further, if $\Pr(X = n)$ is nonincreasing in $n$ $(n = 1, 2,...)$, then the entropy is finite only if $E \log X < \infty$.

1. INTRODUCTION

Let $p = (p_1, p_2, \ldots)$ be a probability sequence, i.e., $p_n \geq 0$, $\sum_{n=1}^{\infty} p_n = 1$. The corresponding entropy is

$$H(p) = -\sum_{n=1}^{\infty} p_n \log p_n,$$

where $0 \log 0$ is taken as 0. Assume that the $p_n$ are nonincreasing in $n$ (reordering of the $p_n$ does not affect $H$). In a recent paper, Keilson (1971) showed that a sufficient condition for the series in (1) to be summable is that (i) $\sum_{n=1}^{\infty} p_n \log n < \infty$ and (ii) $p_{n+k} \sim 1/k p_n$ as $n \to \infty$ for all fixed $k = 1, 2, \ldots$.

It will follow from the results given here that condition (i) alone is sufficient as well as necessary.

Our main result is an upper bound on the entropy series (1) in terms of expectation of certain functions of the random variable $X$, where $X$ takes the value $n$ with probability $p_n$ $(n = 1, 2, \ldots)$.

2. STATEMENT OF RESULTS

We state our results here, leaving the proofs for Section 3.

THEOREM 1. If $p$ is such that $H(p) < \infty$, then

$$E \log X = \sum_{n=1}^{\infty} p_n \log n < \infty.$$
Next define the set \( \mathcal{A} \) as the set of nonnegative, nondecreasing sequences \( a_n, n = 1, 2, \ldots \), with the property that for some \( \lambda > 0 \),

\[
\Phi(\lambda) = \sum_{n=1}^{\infty} e^{-\lambda a_n} < \infty.
\]  

(2)

Note that \( a_n = n^e (e > 0) \) and \( a_n = \log n \), in particular, are sequences in \( \mathcal{A} \). Of course, if \( \{a_n\} \in \mathcal{A} \), then \( \lim_{n \to \infty} a_n = \infty \). Our main result is an upper bound on the entropy corresponding to a probability sequence \( \mathbf{p} \) in terms of \( E_{ax} = \sum_{1}^{\infty} a_n p_n \), for sequences \( \{a_n\}_{n=1}^{\infty} \) in \( \mathcal{A} \).

For \( \{a_n\} \in \mathcal{A} \), let \( \lambda_0 = \inf \{ \lambda : \sum_{1}^{\infty} e^{-\lambda a_n} < \infty \} \). Then \( \Phi(\lambda) < \infty \) for \( \lambda > \lambda_0 \).

Let us define

\[
\psi(\lambda) = \sum_{1}^{\infty} a_n e^{-\lambda a_n} / \Phi(\lambda), \quad \lambda_0 < \lambda < \infty.
\]  

(3)

We will establish

**Lemma 1.** (i) \( \psi(\lambda) < \infty \);

(ii) \( \psi(\lambda) \) is strictly monotone decreasing;

(iii) \( \lim_{\lambda \to \infty} \psi(\lambda) = a_1 \).

Thus setting \( \psi(\lambda_0) = \lim_{\lambda \to \lambda_0} \psi(\lambda) \), we can, for \( \beta \) satisfying \( a_1 \leq \beta < \psi(\lambda_0) \), unambiguously define \( \lambda^* = \lambda^*(\beta) \) as the solution of \( \psi(\lambda^*) = \beta \). We then can prove

**Theorem 2.** Let \( \mathbf{p} \) be such that for some \( \{a_n\} \in \mathcal{A} \), \( E_{ax} = \sum_{1}^{\infty} a_n p_n \leq \beta \), where \( a_1 \leq \beta < \psi(\lambda_0) \). Then

\[
H(\mathbf{p}) \leq \log \Phi(\lambda^*(\beta)) + \beta \lambda^*(\beta).
\]

Since, for \( a_n = \log n \), we have \( \psi(\lambda_0) = \infty \), Theorem 2 yields an upper bound on \( H(\mathbf{p}) \) for all \( \beta \). Thus, in particular, we have established a

**Corollary.** If \( \mathbf{p} \) is such that \( E \log X = \sum p_n \log n < \infty \), then \( H(\mathbf{p}) < \infty \).

Now suppose that \( \psi(\lambda_0) < \infty \), then (from Lemma 1 (ii)) for \( \lambda > \lambda_0 \), and \( K > 0 \) arbitrary,

\[
\psi(\lambda_0) > \psi(\lambda) \geq \sum_{a_n \geq K} a_n e^{-\lambda a_n} / \Phi(\lambda) \geq K \sum_{a_n \geq K} e^{-\lambda a_n} / \Phi(\lambda).
\]  

(4)

Letting \( \Phi(\lambda_0) = \lim_{\lambda \to \lambda_0} \psi(\lambda) = \sum_{1}^{\infty} e^{-\lambda a_n} \), we have from (4) that \( \Phi(\lambda_0) < \infty \).
If this were not so, then for $K$ arbitrarily large the right member of (4) approaches $K$ as $\lambda \to \lambda_0$, yielding for all $K > 0$, $K \leq \psi(\lambda_0) < \infty$, a contradiction. We now state our final result:

**Theorem 3.** If $\{a_n\} \in \mathcal{C}$ is such that $\mathcal{V}(\lambda_0) < \infty$, and $p$ is such that $EaX = \sum_{i=1}^{\infty} a_n p_n \leq \beta < \infty$ where $\beta \geq \mathcal{V}(\lambda_0)$, then

$$H(p) \leq \log \Phi(\lambda_0) + \beta \lambda_0.$$

3. **Proofs**

**Proof of Theorem 1.** Since $p_n$ is nonincreasing,

$$1 = \sum_{i=1}^{\infty} p_i \geq \sum_{i=1}^{n} p_i \geq np_n,$$

and we have

$$\log (1/p_n) \geq \log n, \quad n = 1, 2, \ldots.$$

Thus if $H(p) < \infty$, then

$$E \log X = \sum_{i=1}^{\infty} p_n \log n \leq \sum_{n} p_n \log \frac{1}{p_n} = H(p) < \infty.$$

**Proof of Lemma 1.** (i) We must show that $\sum_{1}^{\infty} a_n e^{-\lambda a_n} < \infty (\lambda_0 < \lambda < \infty)$. Let $\lambda = \lambda_0 + 2\varepsilon$, and then let $K$ be sufficiently large so that $e^K \geq K$. Then

$$\sum_{1}^{\infty} a_n e^{-\lambda a_n} \leq \sum_{a_n < K} K e^{-\lambda a_n} + \sum_{a_n \geq K} e^{\varepsilon a_n} e^{-\lambda a_n} \leq K \Phi(\lambda) + \Phi(\lambda - \varepsilon) < \infty.$$

A similar proof will show that $\sum_{1}^{\infty} a_n^2 e^{-\lambda a_n} < \infty (\lambda < \lambda_0)$. From this we can conclude that the series $\sum_{n} a_n e^{-\lambda a_n}$ ($j = 0, 1$) are differentiable with respect to $\lambda$ term-by-term.

(ii) Taking derivatives,

$$\psi'(\lambda) = \frac{(\sum_{n} a_n e^{-\lambda a_n})^2 - \sum_{n} e^{-\lambda a_n} \sum_{n} a_n^2 e^{-\lambda a_n}}{(\Phi(\lambda))^2} < 0$$

by Schwarz's inequality (the strict inequality holding).
(iii) Let \( l = 2, 3, \ldots \), be the smallest index such that \( a_1 > a_i \). Then
\[
a_1 \leq \Psi(\lambda) \leq \frac{(l - 1)a_1 e^{-\lambda a_1} + \sum_{n=l}^{\infty} a_n e^{-\lambda a_n}}{(l - 1) e^{-\lambda a_1}} = a_1 \left(1 + \sum_{n=l}^{\infty} \frac{a_n}{a_1(l - 1)} e^{-\lambda(a_n - a_1)}\right) \rightarrow a_1, \quad \text{as} \ \lambda \rightarrow \infty.
\]

**Proof of Theorem 2.** Let \( \{a_n\} \subset (0, \infty) \) and \( \sum_n a_n p_n \leq \beta \), where \( a_1 \leq \beta < \Psi(\lambda_0) \). Let us define another probability sequence \( p^* = (p^*_1, p^*_2, \ldots) \) where
\[
p^*_n = e^{-\lambda^* a_n} \Phi(\lambda^*), \quad \text{and} \quad \lambda^* = \lambda^*(\beta).\]
Then
\[
H(p^*) = \sum_n p^*_n \log \frac{1}{p^*_n} = \log \Phi(\lambda^*) + \lambda^* \sum_n p^*_n a_n = \log \Phi(\lambda^*) + \lambda^* \Psi(\lambda^*)
\]
\[
= \log \Phi(\lambda^*) + \beta \lambda^* \geq \sum_n p_n \log \Phi(\lambda^*) + \lambda^* \sum_n p_n a_n
\]
\[
= -\sum_n p_n \log p_n^*.
\]
Thus,
\[
H(p) - H(p^*) \leq \sum_n p_n \log \frac{p_n^*}{p_n} \leq \sum_n \left[\frac{p_n^*}{p_n} - 1\right] = 0,
\]
which proves the theorem.

**Proof of Theorem 3.** Say that we are given \( \{a_n\} \subset (0, \infty) \) for which \( \Psi(\lambda_0) < \infty \). For \( N = 1, 2, \ldots \), define
\[
\Psi_N(\lambda) = \sum_{n=1}^{N} a_n e^{-\lambda a_n} \Phi_N(\lambda), \quad 0 \leq \lambda < \infty,
\]
where
\[
\Phi_N(\lambda) = \sum_{n=1}^{N} e^{-\lambda a_n}.
\]
Now for \( \lambda < \lambda_0 \), \( \lim_{N \to \infty} \Phi_N(\lambda) = \infty \). Thus for \( \lambda < \lambda_0 \), we have
\[
\Psi_N(\lambda) = \sum_{n=1}^{N} a_n e^{-\lambda a_n} \Phi_N(\lambda) \geq K \sum_{1 \leq n \leq N} e^{-\lambda a_n} \Phi_N(\lambda) \geq K.
\]
Since \( K \) is arbitrary, we let \( K \to \infty \) to establish
\[
\lim_{N \to \infty} \Psi_N(\lambda) = \infty, \quad 0 \leq \lambda < \lambda_0.
\]
From (7) we have $\mathcal{W}_N(0) = 1/N \sum a_n \to \infty$. Thus if we are given $\beta(a_1 < \beta < \infty)$, we can find an $N$ sufficiently large so that $1/N \sum a_n \geq \beta$. With $N$ now held fixed, we can show, exactly as in Lemma 1, that there exists a unique $\lambda_N^*(\beta)$ for which $\mathcal{W}_N(\lambda_N^*) = \beta$. Now suppose we are given a $p$ such that $\sum p_n a_n \leq \beta$. Choose $N$ and $\lambda_N^* (\beta)$ as above, and define a new probability sequence $p_n^* = e^{-\lambda_N^* a_n} / q_n(\lambda_N^*)$, $n = 1, 2, \ldots, N$.

The corresponding entropy is, as in (5),

$$H(p^*) = \log \Phi_N(\lambda_N^*) + \beta \lambda_N^* \geq \sum_{n=1}^N p_n \log \Phi_N(\lambda_N^*) + \lambda_N^* \sum_{n=1}^N p_n a_n$$

$$= - \sum_{n=1}^N p_n \log p_n^*.$$

Let $H_N(p) = - \sum_{n=1}^N p_n \log p_n$, so that

$$H_N(p) - H(p^*) \leq - \sum_{n=1}^N p_n \log p_n + \sum_{n=1}^N p_n \log p_n^*$$

$$= \sum_{n=1}^N p_n \log \frac{p_n^*}{p_n} \leq \sum_{n=1}^N p_n \left( \frac{p_n^*}{p_n} - 1 \right)$$

$$= \sum_{n=1}^N p_n^* - \sum_{n=1}^N p_n = \sum_{n=N+1}^\infty p_n.$$

Thus

$$H_N(p) \leq \log \Phi_N(\lambda_N^*) + \beta \lambda_N^*(\beta) + \sum_{N+1}^\infty p_n. \quad (8)$$

Let us restrict our attention to $\beta \geq \Psi(\lambda_0)$. In this case, from (7), $\lambda_N^*(\beta) \to \lambda_0$. Further, since $\Psi_N(\lambda)$ is decreasing in $\lambda$,

$$\Psi_N(\lambda_N^*) = \beta \geq \Psi_N(\lambda), \quad \text{for } \lambda_N^* \geq \lambda \geq \lambda_0.$$

Thus

$$\beta(\lambda_N^* - \lambda_0) \geq \int_{\lambda_N^*}^{\lambda_0} \Psi_N(\lambda) \, d\lambda = \int_{\lambda_N^*}^{\lambda_0} \frac{-d}{d\lambda} \log \Phi_N(\lambda) \, d\lambda$$

$$= \log \Phi_N(\lambda_N^*) - \log \Phi_N(\lambda_0).$$

Thus letting $N \to \infty$,

$$\limsup_{N \to \infty} \log \Phi_N(\lambda_N^*) \leq \log \Phi(\lambda_0),$$
so that (8) yields (on letting $N \to \infty$)

$$H(p) \leq \log \Phi(\lambda_0) + \beta \lambda_0,$$

which is Theorem 3.

Received: July 1, 1971

**Reference**