Reprinted from:
Proceedings Beijing International Workshop on Information Theory,
Beijing, China, July 4-7, 1988.

On the Entropy of Integer-Valued Random Variables

James L. Massey
Professor für Digitaltechnik

Swiss Federal Institute of Technology
ETH - Zentrum
CH - 8092 Zürich

ABSTRACT

It is shown that the "discrete entropy" or uncertainty of an integer-valued random variable X with finite variance σ^2 satisfies the inequality

$$H(X) < \frac{1}{2} \log_2 \left[2\pi e(\sigma^2 + \frac{1}{12})\right] \text{ bits.}$$

This is proved by applying Shannon's upper bound on the "continuous entropy" of a continuous random variable with a given variance to the continuous random variable x whose probability density function is the piecewise constant function

$$p_x(\alpha) = p_i, \quad i - \frac{1}{2} < \alpha \leq i + \frac{1}{2}$$

for all integers i where $p_i = P(X = i)$. The discrete entropy of X then coincides with the continuous entropy of x, but the variance of x is greater by $\frac{1}{12}$ than the variance of X. It is further shown that the zero-mean integer-valued random variable that has maximum entropy for a given variance has a probability distribution proportional to the samples taken at integer arguments of a zero-mean Gaussian probability density function. This fact, which can be used numerically to find the best upper bound on H(X) for a given variance, is used to illustrate the relative tightness of the simple upper bound above when σ^2 is large.

I. FOREWORD

This paper presents an upper bound on the entropy of an integer-valued random variable in terms of its variance. We derived this bound just over 10 years ago, and it has been used with appropriate attribution in the literature [1, p. 1680], but the bound itself has not previously been published. We make this publication now not only to provide a literature citation for this bound, but also because its derivation may be of some independent interest.

II. DERIVATION OF THE BOUND AND NUMERICAL RESULTS

Let X be an integer-valued random variable and let $p_i = P(X=i)$. The discrete entropy of X is

$$H(X) = -\sum_{i:p_i \neq 0} p_i \log p_i . \tag{1}$$

Let x be a continuous random variable with probability density function $p(\alpha)$. The continuous entropy of x is

$$H(x) = -\int_{\alpha:p(\alpha)\neq 0} p(\alpha) \log[p(\alpha)] d\alpha . \tag{2}$$

In his 1948 paper, Shannon showed that [2, Section 20.5]

$$H(x) \leq \frac{1}{2} \log (2\pi e \sigma_x^2) \tag{3}$$

(where σ_x^2 is the variance of x) with equality if and only if x is Gaussian.

Now let X be an integer-valued random variable and let x be the continuous random variable whose probability density function is given by

$$p(\alpha) = p_i \quad , \qquad i - \frac{1}{2} < \alpha \leqslant i + \frac{1}{2} \tag{4}$$

for all integers i. It is easily checked that X and x have the same mean, but that the variance σ^2 of X is

$$\sigma^2 = \sigma_x^{\,2} - \frac{1}{12} \; . \tag{5}$$

Moreover, using (4) in (2) shows that

$$H(x) = H(X). \tag{6}$$

From (3), (5) and (6) it now follows that

$$H(X) < \frac{1}{2} \log \left[2\pi e (\sigma^2 + \frac{1}{12}) \right] \tag{7}$$

where the strictness of the inequality follows from the fact that x is not Gaussian.

Following Shannon's 1948 approach, we now seek the probability distribution that maximizes H(X) for an integer-valued random variable with second moment S by introducing two Lagrange multipliers to form the functional

$$L = - \sum_{i:p_i \neq 0} p_i \log p_i + \lambda_1 \sum_i p_i + \lambda_2 \sum_i (i)^2 p_i .$$

By standard calculus of variation arguments, one then finds the maximizing distribution to be

$$p_i = a\, e^{-bi^2} \tag{8}$$

for positive constants a and b chosen so that

$$\sum_i p_i = 1 \tag{9}$$

and

$$\sum_i (i)^2 \, p_i = S. \tag{10}$$

Equation (8) shows that $p_i = c\, p(i)$ where c is a positive constant and where $p(\alpha)$ is the Gaussian density with mean zero and with second moment (or, equivalently, variance) $\frac{1}{2b}$, i.e., the p_i's are the scaled samples of this Gaussian density taken at the integers. For a given b one can readily calculate numerically first a and S from (8), (9) and (10) and then H(X) from (1). This is the maximum entropy of any integer-valued random variable with variance σ^2 = S. That this entropy will be very close to the upper bound (7) for large σ^2 follows from the fact that the piecewise constant density of (4) then differs negligibly from a Gaussian density if the discrete distribution satisfies (8). This is confirmed by the numerical results given in Table I, which show that the bound is tight within 1% for $\sigma^2 > 2$ and within 0.1% for $\sigma^2 > 8$.

b	a	σ^2	max.H(X)	Bound (7)
∞	1	0	0 (bits)	0.2546 (bits)
1.813	0.7532	1/4	1.0630	1.2547
0.9980	0.5636	1/2	1.5472	1.6583
0.5000	0.3989	1	2.0471	2.1048
0.2500	0.2821	2	2.5471	2.5765
0.1250	0.1995	4	3.0471	3.0620
0.0625	0.1410	8	3.5471	3.5546
1/128	0.0499	64	5.0471	5.0480

TABLE I: Comparison of the Upper Bound (7) on H(X) to the maximum achievable H(X) for an integer-valued random variable with variance σ^2, whose distribution is given by (8).

REFERENCES

[1] J.F. Hayes and R.R. Boorstyn, "Delay and Overhead in the Encoding of Data Sources", IEEE Trans. Comm., vol. COM-29, pp. 1678-1683, November 1981.

[2] C.E. Shannon, "A Mathematical Theory of Communication", Bell System Tech. J., vol. 27, pp. 379-423 and pp. 623-656, July and October, 1948.