A note on Shannon's entropy as an index of product variety

Bas Straathof

2003-028

A note on Shannon's entropy as an index of product variety

Bas Straathof[*]

November 6, 2003

Abstract

This note shows that the antilog of Shannon's entropy is a suitable index of product variety for three reasons. First, for symmetric product types it is equal to the number of product types. Second, disaggregation of the underlying product set always leads to an increase in measured product variety. Third, the introduction or disappearance of a marginal type does not cause a discrete change in the variety index. These properties hold for a class of weights that includes, but is not limited to, frequencies.

Keywords: Product variety, entropy
JEL classification: C43, L11

---

[*]MERIT - Maastricht University, P.O. Box 616 - 6200 MD Maastricht - The Netherlands, tel. +31 43 3883879, b.straathof@merit.unimaas.nl.

1 Introduction

The point of departure of this note is that the number of types in a set of products is the perfect indicator of product variety *if* all the types are of equal importance. Usually, this is not the case and the need for a more sophisticated indicator of product variety arises. For example, if one would use the number of types as a variety index, then 'Crude oil' would have the same contribution to the index as 'Turkeys: whole: frozen, not cut in pieces' would have. It is argued here that entropy has properties that make it suitable as an index of variety when not all types are of equal importance.

Henry Theil (1965, 1967) introduced the concept of entropy to economics, taking it from information theory where Shannon (1948, theorem 2) had previously developed what turned out to be a particularly popular type of entropy. Theil (1967, pp. 91, 290) has suggested the use of entropy as a measure of income inequality and industry-concentration.

More recently, Alexander (1997) and Frenken, Saviotti and Trommetter (1999) have applied Shannon's entropy as an indicator of product variety. In these studies, product types have been weighted according to the frequency of their occurrence in a dataset. It is shown below that the properties of Shannon's entropy are preserved for a class of weights that includes, but is not limited to, frequencies.

2 Entropy and product variety

Suppose that we want to measure the variety of a set N that contains n different types of products. Each product type $i \in N$ is assigned a weight w_i that depends on how important the type is compared to the other types as indicated by a measure u. In particular, $w_i = u_i / \sum_{j \in N} u_j$. If not all types have the same weight, then the measurement of variety becomes non-trivial. In this situation, entropy (H) is a good index of product variety because it possesses the three properties stated below.

Property 1 *If $w_i = \frac{1}{n}$ for all i, then H is monotonically increasing in n.*

This property implies equivalence with the most basic variety-index possible, which is the number of types.

Property 2 $H(N) = H(G) + \sum_{j \in G} w_j H(N_j)$ *where N_j is a subset of N such that $\cup_{j \in G} N_j = N$ and $N_j \cap N_k = \emptyset$ for all $j \neq k$, and G is the collection of subsets.*
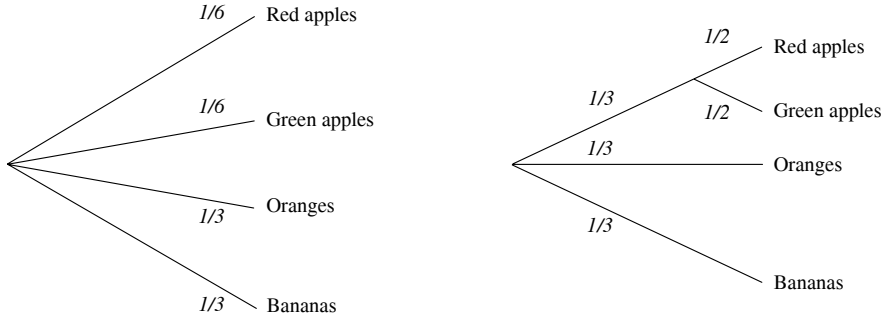
Figure 1: Entropy and product classification

In words, total variety is the variety of subsets plus the weighted sum of variety within each subset. This property ensures that the total entropy at a higher level of aggregation is lower than the total entropy at a more detailed level of aggregation, even when only one subset is disaggregated.

In order to clarify this point consider the following example. Suppose the set of products 'fruit' contains 'oranges', 'apples', and 'bananas'. Each type of fruit gets the weight $1/3$. The subset 'apples' in turn consists of 'red apples' and 'green apples', each type of apple being weighted equally. Property 2 allows for the decomposition of entropy of the entire set of products into the entropies of subsets (see figure 1). If the subset 'apple' would not be disaggregated, total variety would be $H(1/3, 1/3, 1/3)$. The difference between total variety with two types of apples and total variety without different types of apples equals the contribution to total variety of variety in 'apples':

$$H(1/6, 1/6, 1/3, 1/3) - H(1/3, 1/3, 1/3) = \frac{1}{3}H(1/2, 1/2).$$

Property 3 *H is continuous in all w_i.*

This property ensures that when an infinitesimal amount of a product is produced that did not exist previously, the variety index will not make a discrete jump but will change marginally. By this property, the introduction or disappearance of an unimportant type does not disrupt the variety index.

Theorem 4 [1] *The only function H that possesses all three properties stated above is*

$$H\left(N\right) = -K\sum_{i \in N} w_i \ln w_i \tag{1}$$

$$w_i \equiv \frac{u_i}{\sum_{j \in N} u_j} \tag{2}$$

where K is an arbitrary constant, and u is a measure along which the elements of N (the product types) can be compared such that

$$u_N = \sum_{j \in N} u_j. \tag{3}$$

Proof. Define $m_i \equiv u_i/\bar{u}$ and choose $\bar{u}$ such that m_i is an integer for all $i \in N$. By equation 3, m_N is the number of units of $\bar{u}$ possessed by the set of product types N. We will start by showing that $H\left(M\right)$ possesses properties 1 and 2, where M is the set of all m units of $\bar{u}$.

By definition, all units of $\bar{u}$ are equally weighted: $w_i = 1/m$. In this symmetric case, H depends only on m and the notation can be simplified into $H\left(M\right) = A\left(m\right)$. Suppose that the units are classified into s broad subsets and that all these subsets are again subdivided into s subsets and so on until the number of times sets are subdivided is p. After p steps, the total number of subsets at the most detailed level of disaggregation will be s^p.

By property 2 this implies that $A\left(s^p\right)$ should satisfy

$$A\left(s^p\right) = A\left(s\right) + s\frac{1}{s}A\left(s\right) + s^2\frac{1}{s^2}A\left(s\right) + .. + s^{p-1}\frac{1}{s^{p-1}}A\left(s\right)$$
$$= pA\left(s\right).$$

In the same way, if the units are q times subdivided into t subsets per set, the total number of subsets at the most detailed level is t^q. From property 2 follows again that $A\left(t^q\right) = qA\left(t\right)$.

For any value of q, s, and t, an p can be found such that

$$s^p \leq t^q < s^{p+1}.$$

After taking logarithms and dividing by $q\ln s$ we get

$$\frac{p}{q} \leq \frac{\ln t}{\ln s} < \frac{p}{q} + \frac{1}{q}.$$

---

[1]The theorem and its proof are more general than the similar theorem by Shannon (1948, theorem 2) as it shows that the properties stated above hold for a class of weights that includes, but is not limited to, frequencies.

Rearranging yields

$$\begin{array}{l} 0 \le \frac{\ln t}{\ln s} - \frac{p}{q} < \frac{1}{q} \\ \frac{p}{q} - \frac{\ln t}{\ln s} \le 0 < \frac{1}{q} \end{array} \Rightarrow \left| \frac{\ln t}{\ln s} - \frac{p}{q} \right| < \varepsilon.$$

As q can take on any value, ε can be taken to be arbitrarily close to zero.[2]
Property 1 implies that

$$A\left(s^p\right) \le A\left(t^q\right) < A\left(s^{p+1}\right)$$
$$pA\left(s\right) \le qA\left(t\right) < \left(p+1\right)A\left(s\right)$$
$$\frac{p}{q} \le \frac{A\left(t\right)}{A\left(s\right)} < \frac{p}{q} + \frac{1}{q}$$
$$\left| \frac{p}{q} - \frac{A\left(t\right)}{A\left(s\right)} \right| < \varepsilon,$$

and thus

$$\begin{array}{l} \frac{A(t)}{A(s)} - \frac{p}{q} + \frac{p}{q} - \frac{\ln t}{\ln s} < 2\varepsilon \\ \frac{p}{q} - \frac{A(t)}{A(s)} + \frac{\ln t}{\ln s} - \frac{p}{q} < 2\varepsilon \end{array} \to \left| \frac{A\left(t\right)}{A\left(s\right)} - \frac{\ln t}{\ln s} \right| < 2\varepsilon.$$

For ε arbitrarily close to zero, $A\left(t\right)$ should be of the form $K \ln t$. The constant K should be positive, otherwise H would not be increasing in q. The function $H\left(M\right) = K \ln m$ possesses all three properties.

The elements of M had equal weights. The form of H for heterogenous weights still has to be found. Fortunately, this general form of H can be found by the use of property 2. Allocate the m units of M over the product types in the set N. Property 2 then implies

$$A\left(m\right) = H\left(N\right) + \sum_N w_i A\left(m_i\right)$$

$$H\left(N\right) = \sum_N w_i \left(A\left(m\right) - A\left(m_i\right)\right).$$

Substituting for $A\left(m\right) = K \ln m$ yields

$$H\left(N\right) = \sum_N w_i \left(K \ln \frac{m}{m_i}\right).$$

By $w_i = m_i \bar{u} / \left(\sum_{j \in N} m_j \bar{u}\right) = m_i / m$ the expression for $H\left(N\right)$ can be written as

$$H\left(N\right) = -K \sum_N w_i \ln w_i.$$

---

[2]$\bar{u}$ can be chosen to be arbitrarily small, implying that m can be arbitrarily large.

That property 3 holds for $w_i > 0$ is trivial. That H is continuous in w_i when $w_i \to 0$ is demonstrated below.

$$\lim_{w_i \downarrow 0} H = -K \sum_{j \in N/i} w_j \ln w_j - K \lim_{w_i \downarrow 0} (w_i \ln w_i)$$

L'Hôpital's rule implies

$$\lim_{w_i \downarrow 0} (w_i \ln w_i) = \lim_{w_i \downarrow 0} \frac{\mathrm{d} \left(w_i^2 \ln w_i\right)}{\mathrm{d} w_i}$$
$$= \lim_{w_i \downarrow 0} (2 w_i \ln w_i) + \lim_{w_i \downarrow 0} w_i.$$

Substraction of $2 \lim_{w_i \downarrow 0} (w_i \ln w_i)$ yields $-\lim_{w_i \downarrow 0} (w_i \ln w_i) = \lim_{w_i \downarrow 0} w_i = 0$, which proves continuity of H in w_i. ∎

It is possible to normalize H in order to resemble more closely the number of product types as a measure of product variety. If we set $K = 1$ and take the exponential form of H we get

$$V(N) = \prod_N w_i^{-w_i}.$$

The properties of V that correspond to the three properties of H are:

Property 5 *If $w_i = \frac{1}{n}$ for all i, then $V = n$.*

Property 6 $V(N) = V(G) \prod_{j \in G} V(N_j)^{w_j}$

Property 7 *V is continuous in all w_i.*

The special case of V where u is chosen to be expenditure (such that the weights become expenditure shares), has direct relevance to the theory of index numbers (Straathof 2003).

3   Concluding remarks

The antilog of Shannon's entropy possesses three properties that make it a suitable index of product variety. First, for symmetric product types it equals the number of product types. Second, disaggregation of the underlying product set always leads to an increase in measured product variety. Third, the introduction or disappearance of a marginal type does not cause a discrete change in the variety index. It has been shown that these properties are preserved for weights based on any measure u for which holds that $u_N = \sum_{j \in N} u_j$. This result is a modest generalization of Shannon's second theorem.

References

ALEXANDER, P. J. (1997): "Product Variety and Market Structure: A New Measure and a Simple Test," *Journal of Economic Behavior and Organization*, 32(2), 207–14.

FRENKEN, K., P. P. SAVIOTTI, AND M. TROMMETTER (1999): "Variety and Niche Creation in Aircraft, Helicopters, Motorcycles and Microcomputers," *Research Policy*, 28(5), 469–88.

SHANNON, C. E. (1948): "A mathematical theory of communication," *The Bell System Technical Journal*, 27, 379–423, 623–656.

STRAATHOF, B. (2003): *Variety-robust axiomatic indices*. MERIT Research Memorandum 2003-25.

THEIL, H. (1965): "The information approach to demand analysis," *Econometrica*, 33(1), 67–87.

——— (1967): *Economics and information theory*. North Holland, Amsterdam.