diversity density

Alan Dix
http://www.hiraeth.com/alan/ebulletin/diversity-density/
alan@hcibook.com

abstract
If one looks through the supply chain from kitchen cupboard to factory, the diversity of goods reduces. Diversity density is a measure of this reduction and corresponds to a loss of information through the supply chain. This can be informally defined but also given a precise measure using conditional entropy. The loss of information is due to the fact that it is only the implicit information due to financial transactions that is available in traditional supply chains. In electronic business individual end-customer orders are transmitted via the internet or similar channels far up the supply chain with correspondingly greater information and higher diversity density.

keywords: ecommerce, ebusiness, eshopping, information theory, informational role of money, internet shopping

what is diversity density?

Open your food cupboard at home. What do you see? Something like this?

Now look at a supermarket shelf:

Whereas in the kitchen cupboard there are only a few items of each kind next to each other, on the supermarket shelf there are hundreds of cans of the same item next to each other. And not only are there hundreds of the same kind of thing (beans here!), but the nearby items are similar products (beans of different brands). Now imagine the shelves in the warehouse behind the supermarket, there are crates and crates of baked beans together. In the truck that delivered the beans, maybe 10 thousand cans all together and perhaps other products from the same manufacturer. In the warehouse the truck came from maybe hundreds of tons of beans and again only similar things from the same manufacturer. Finally, in the factory they were produced there may be tens of millions of cans of beans all together and perhaps nothing else.

Notice as we move up the supply chain there is less and less diversity. If you take a 100 items next to one another in your kitchen cupboard there may be 30 different kinds of things, in the supermarket may be only two or three kinds and in the truck or warehouse every item would be identical.

This difference is diversity density:

$$\text{diversity density} = \frac{\text{number of different kinds of item}}{\text{number of items sampled}}$$

or equivalently:

$$\frac{1}{\text{average number of items of the same kind together}}$$

It is possible to calculate this numerically for different places, perhaps 0.5 for the kitchen cupboard, 0.01 for the supermarket and 0.0001 for the truck. However it is really the concept that is most important.

diversity density as information

Diversity density is a measure of information or entropy - how much knowing about one item tells you about it's neighbours.

Imagine I have selected three items from my cupboard and show you the first item. It is a tin of baked beans. The second item is from the same shelf 20 cm to the left. Do you have any idea what it is? In fact it is a box of tea bags. The final item is from the shelf below 20 cm in the other direction. Again any idea what it may be? No - well in fact it is a packet of biltong.

item 1 item 2 item 3

kitchen cupboard items

Now imagine I have three items from the supermarket and tell you the first item is a tin of beans. Any idea what the item 20 cm along the same shelf is? You guessed it - a tin of beans Now 20 cm the other way on the shelf below - yes, another tin of beans, just a different brand!

item 1 item 2 item 3

supermarket cupboard items

It is clear that in the case of the kitchen cupboard knowing the first item told us very little about the second or third item. However, in the case of the supermarket the knowledge of the first item gave us lots of information.

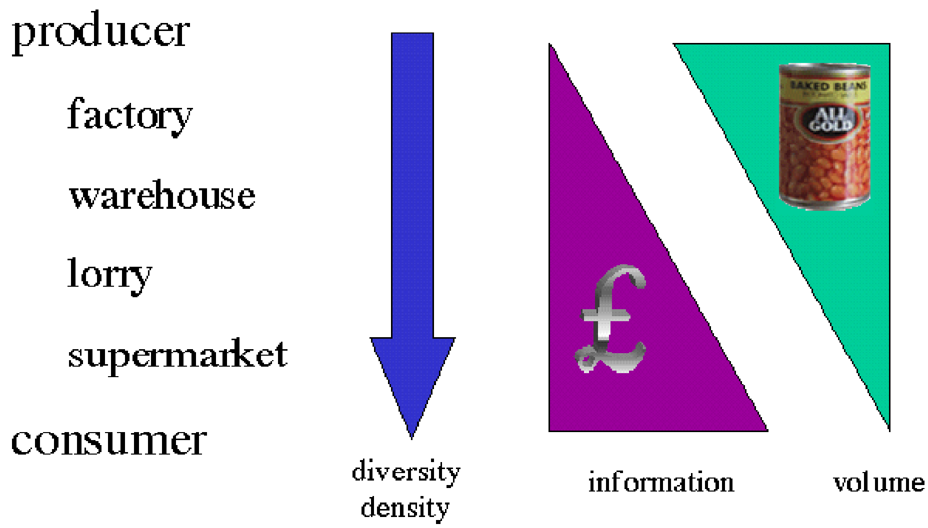
Again it is really the qualitative aspects of this information that are of primary interest. However, it is also possible to make this information precise using information theoretic measures of conditional entropy (see section 'precise calculation of diversity density' later).

diversity density in the supply chain

> You may wonder why is diversity density is usually higher in your supermarket trolley than in your kitchen cupboard. When things go into the cupboard they are often sorted by kind and join items purchased on other days. That is the cupboard looses information about *when* you purchased things.

Typically, we see diversity density decrease as we move up the supply chain from customer to producer, with perhaps maximum diversity density in the supermarket trolley or in the car trunk on the way home from a shopping trip.

This corresponds to a gradual loss of information. The goods in your shopping trolley correspond to the things you want from that particular supermarket that particular week at the current price. Each supermarket trolley carries this information for different people. Now what happens when you get to the checkout. Alison has a tin of beans and a box of tea bags in her trolley and pays two pounds for them. Brian has two cans of beans and pays one pound for them. If the supermarket has no additional information systems to track purchases, all it knows is that three tins of beans and one box of tea bags were sold and it has three pounds. As different customer's money is mixed in the till it loses the information of what and who it was for.

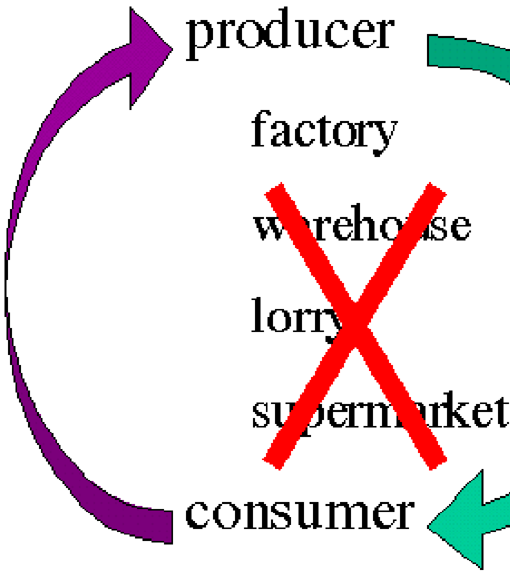


Notice also that the volume per order decreases as we move down the supply chain from producer (perhaps ship loads or lorry loads) to consumer (shopping basket full).

In general diversity density is lower higher up the supply chain because there is less information available. This is largely because information moves up the supply chain through financial transactions which looses information. For the purpose of value exchange it is important that every pound coin is the same – whether it came from Alison or Brian, whether it was used to buy beans or biltong. However, as a carrier of information this means money intrinsically looses information once it is pooled together.

changes for e-business

This picture of information loss and lower diversity density is completely changed when we look at e-businesses.



When you order online the information that you want a particular selection of goods travels much further up the supply chain. This information is transmitted directly in electronic form, rather than carried as a side effect of financial transactions. As the order is put together in the supplier's picking line it has high diversity density and small size (per order).

So, in the new economy not only is the value of money mediated electronically via credit cards and bank balances, but the informational role of money is gradually being replaced by direct electronic information.

precise calculation of diversity density

We can use standard information theory measures to give an exact measure of diversity density.

Suppose we choose an item at random from the kitchen shelf, supermarket shelf, or wherever we are considering the diversity density. If there are m different kinds of item all equally likely then the probability that an item will be of type i is 1/m. However. more generally these probabilities will differ and we'll write p_i for the probability that a randomly chosen item is of type i.

Now suppose we take two adjacent items. We'll write p_{ij} for the probability that the first item is of type i and the second item is of type j. Because the various probabilities need to add up, the following will always be true:

$$p_i = \sum_j p_{ij}$$
$$p_j = \sum_i p_{ij}$$

If the items were mixed completely randomly then p_{ij} would be exactly $p_i \times p_j$. However, because adjacent items are related one item gives us some better estimate of the probabilities for the next. This can be quantified using conditional entropy.

The entropy of an item (using the Shannon and Weaver measure of information) is:

(a) entropy single item $= - \sum_j p_i \log p_i$

This measures the information in bits we learn when we are told what the item is.

Similarly we can measure the entropy of the two items together:

(b) entropy two adjacent items $= - \sum_{i\,j} p_{ij} \log p_{ij}$

If the items were independent this would be exactly twice the information from a single item, but if there is any link it will be less than this.

The conditional entropy measures the extra information we get from knowing the second item:

entropy of second item knowing first = (a) – (b)

$$= - \left(\sum_{i\,j} p_{ij} \log p_{ij} - \sum_j p_i \log p_i \right)$$

To turn this into a diversity density measure we simply divide this by the entropy of the second item when we know nothing:

$$\text{diversity density} = \frac{\text{entropy of second item knowing first}}{\text{entropy single item}}$$

$$= \frac{-\left(\sum_{i\,j} p_{ij} \log p_{ij} - \sum_{j} p_i \log p_i\right)}{-\sum_{j} p_i \log p_i}$$

To see that this quotient corresponds to the more intuitive idea of diversity density, we consider the simple case when there are m different kinds of item and all are equally probable ($p_i = 1/m$). We assume that the items are all in a single shelf in groups of n identical items. Also we'll assume that there are very many different kinds of item – m is very large.

If we choose a group of N adjacent items from the shelf there will be typically around N/n different kinds of item. the intuitive diversity density is therefore:

$$\text{diversity density} = \frac{\text{number of different kinds of item}}{\text{number of items sampled}}$$
$$= 1/n$$

The probabilities for two identical items are:

$$p_{ii} = \frac{n-1}{n\ m}$$

$$p_{ij} = \frac{1}{n\ m\ (m-1)} \qquad (i \neq j)$$

(a) entropy single item $= -\sum_{j} p_i \log p_i$

$$= -m\ (1/m \log (1/m))$$
$$= \log m$$

(b) entropy two adjacent items $= -\sum_{i\,j} p_{ij} \log p_{ij}$

$$= -m \frac{n-1}{n\ m} \log \left(\frac{n-1}{n\ m}\right) - m(m-1) \frac{1}{n\ m\ (m-1)} \log\left(\frac{1}{n\ m\ (m-1)}\right)$$

$$= -\frac{n-1}{n} \log \left(\frac{n-1}{n\ m}\right) - \frac{1}{n} \log\left(\frac{1}{n\ m\ (m-1)}\right)$$

$$= \log nm - \frac{n-1}{n} \log (n-1) - \frac{1}{n} \log\left(\frac{1}{m-1}\right)$$

$$= \log nm - \frac{n-1}{n} \log (n-1) + \frac{1}{n} \log(m-1)$$

entropy of second item knowing first = (a) − (b)

$$= \log n - \frac{n-1}{n} \log (n-1) + \frac{1}{n} \log(m-1)$$

For large m the last term dominates and is approximately 1/n log m.

The entropy quotient is therefore equal to the intuitive diversity density:

$$\frac{\text{entropy of second item knowing first}}{\text{entropy single item}} = \frac{1/n \log m}{\log m}$$

$$= 1/n$$

$$= \text{diversity density}$$

This more precise measure of diversity density also takes into account 'similar' items on supermarket shelves such as cans of beans of different brands which are less diverse than totally different kinds of item.