Lecture 7: Long Tail

In this lecture we will learn about impacts of the Long Tail[1] on internet commerce. Long tails are common in online markets, and have enabled innovative businesses, that might not have flourished outside of long tailed markets, to excel and become very successful.

Market distributions: long and short tails

Let us graph a market in the following way. We will order the products i by sales volume m_i, from highest to lowest, and graph the sales volume vs. the product. Note that this formulation requires that our graph be downward-sloping. Below is an example of such a graph.
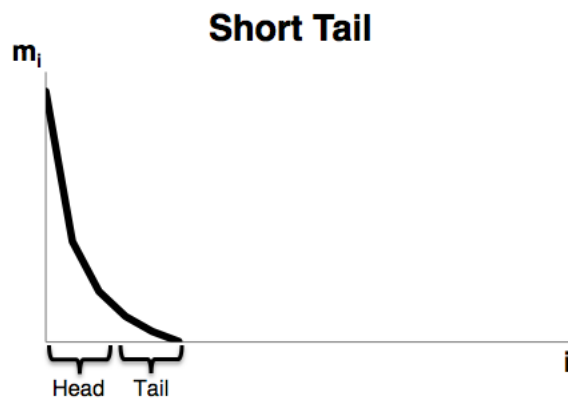

Figure 1: Short Tail

In the graph above, the market drops off fairly sharply to zero. This is an example of a "short tail," or "sharp tail." There are a small number of products with large sales volumes, and a small number of products with small sales volumes. This, however, will not accurately depict many online markets. One of the defining characteristics of online markets is that they look more like the graph in Figure 2.

In Figure 2 we can see that the tail is much longer, and as such is referred to as a "long tail" or "heavy tail." Such a distribution tells us that, in this market, there are a small number of products with high sales volumes, and a very large number of products with low sales volumes. Even though the products in the tail individually have low sales volumes, the total volume in the tail is relatively large because of the large number of products it encompasses.

---

[1] Although "long tail" has long been used to describe probability distributions, the proper noun "the Long Tail," was coined in 2004 by Wired magazine editor-in-chief Chris Anderson, who has since written a book on the subject.
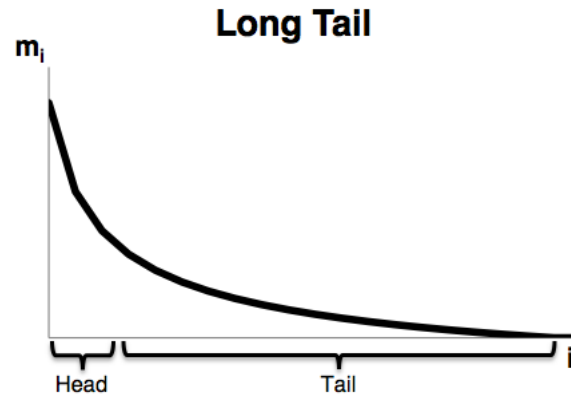
Figure 2: Long Tail

Impacts of the Long Tail

Tail length is a property of the market. In online markets, many different companies have found innovative ways to take advantage of long tails. Here are some examples.

Search engines

Search has a very long tail. There are a few keywords that are commonly searched for, but the vast majority of keywords are rarely searched for. In fact, folklore has it that $1/4$ of all searches are unique. As we discussed in previous lectures, this makes posted pricing for keyword advertising impractical, and has given rise to keyword auctions. It also means that there is very little statistical information available on most keywords, which leads to poor click-through-rate estimation, and a need for the Gittins Index.

Blinds.com

People do not buy blinds every day, and blinds themselves are not very expensive. Yet Blinds.com has been able to use internet advertising to monetize the Long Tail. Because of internet advertising, Blinds.com has a distinct advantage over brick and mortar blinds stores, since it does not have to maintain an inventory, and it can market to people all over the world.

Amazon.com

Amazon is one of the most amazing examples of a company profiting from the Long Tail, because it has used the Long Tail in many different ways. First of all, it has been able to sell rare books. This is extremely hard for a brick and mortar bookstore to do because if they choose any specific rare book, there are very few people that want to buy it. The chances of such a person coming into the store are slim. However, there is such a large volume of rare books that there is a significant amount of money to be made. Amazon can take advantage of this because it does not have to maintain an inventory, and can offer *all* rare books. The probability of a customer coming to amazon.com who wants to buy *a* rare book is high, so Amazon has been able to profit from this.

Another way Amazon has used the Long Tail is in its advertising. Since Amazon sells many products, it is hard to decide which keywords to advertise on, and where to advertise. To get around this difficulty, Amazon has subcontracted their advertising to everyone by creating Amazon Associates. Anyone can become an Amazon Associate by posting an Amazon ad. If this ad leads to a sale, the person who posted it gets 4% of the sale.

A third way that Amazon has capitalized on the Long Tail is their recommendation engine. In order to advertise low-demand items, Amazon has developed a recommendation engine that singles out specific items that a user has a higher probability of buying. They base the recommendations on previous purchases and similarity to other users.

Netflix

Like Amazon, Netflix uses a recommendation engine to advertise products in the Long Tail. Take an obscure movie that not many people have heard of. It is unlikely that someone will come into a movie rental house and rent this movie. Netflix uses its recommendation engine to target specific users that might be interested in this movie, based on their previous rentals. Netflix relies on their recommendation engine because they must pay royalties for the movies they rent out. The royalties on blockbuster movies that most people have heard of are much higher than the royalties on obscure movies that most people have not heard of.

Formulation of a long tailed market

As humans, we are both social and individual. We are more likely to do something that the people around us are doing, but we also look for new things, that the people around us are not doing. It is the tension between these two human behaviors that causes long tails.

Consider a market where at each time, t, a consumer arrives and makes two purchases: one based on individualistic behavior and one based on social behavior. We will define $m_i(t)$ to be the sales volume of product i at time t. The individual purchase is a product that no one has ever purchased before, and the social purchase is a product that many consumers have purchased. The likelihood that the consumer chooses a given product for his social purchase is proportional to the current volume, $m_i(t)$, of the product.

Note that in this market, the following identities hold:
At any time, t, there will be a total market size of $M(t) = 2t$.
The probability of product i being chosen at time t by social behavior is $\frac{m_i(t-1)}{M(t-1)}$
If product i is chosen, then $m_i(t) = m_i(t-1) + 1$
If product i is not chosen, then $m_i(t) = m_i(t-1)$
Thus, the expected increase in market size of product i at time t is $\frac{m_i(t-1)}{2(t-1)}$. For now, we will assume that this is the actual increase.
This leads to the following relationship:

$$m_i(t) - m_i(t-1) = \frac{m_i(t-1)}{2(t-1)}$$

Where $m_t(t) = 1$ and $m_{t+1}(t) = 0$. If we convert this to a continuous differential equation, we get:

$$\frac{dm_i(t)}{dt} = \frac{m_i(t)}{2t} \text{ for } t > i$$

We also know that $m_i(i) = 1$, so we have an initial condition, and we can solve this ordinary differential equation by separation of variables:

$$\int \frac{1}{m_i(t)} dm_i(t) = \int \frac{1}{2t} dt$$

$$\log(m_i(t)) = \frac{1}{2} \log(t) + C$$

$$m_i(t) = Ct^{1/2}$$

Using the initial condition, we solve for C:

$$m_i(i) = Ci^{1/2} = 1$$

$$C = i^{-1/2}$$

Plugging this back into our solution, we get:

$$m_i(t) = \sqrt{t/i}$$

This equation defines the classic heavy-tailed distribution.

Notice that $\frac{m_i(t)}{M(t)} \approx \frac{1}{\sqrt{t}}$, which means the first product is the most popular, but has a vanishingly small market share.

In the next lecture we will see some statistics (mean, variance, etc.) of distributions with different tails.