MS&E 235: Lecture 8
Prof. Ashish Goel
Apr 23, 2009
Notes by Rio Goodman

Lecture 8: More Long Tail

In the previous lecture, we discussed the Long Tail, its applications to internet commerce, and some of the theory behind it. In this lecture we will see some statistics of long tailed distributions, and see how these statistics change as we change the shape of our distribution.

Distribution function

Last time we derived a sales volume function equation for a specific market:

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

This equation can be generalized by using the constants a, b, and e instead of $\sqrt{t}$, 0, and 1/2.

$$m_i \approx a(b+i)^{-e}$$

This equation defines a family of long tailed distributions. Notice that as $i \to \infty$, m_i decreases polynomially to zero. This slow rate of decrease (compared to exponential) gives the function its long tail.

Statistics of long tailed distributions

To find the **median** of a distribution, we look for the point, x, where half of the total volume is above x and half is below x. For our first distribution, $m_i = \sqrt{t/i}$, we have a total sales volume of approximately:

$$\int_1^t \frac{\sqrt{t}}{\sqrt{i}} di = \sqrt{t}\left[2i^{1/2}\right]_1^t = \sqrt{t}(2\sqrt{t}-2) = 2\sqrt{t}(\sqrt{t}-1) \approx 2t$$

Notice that we are approximating the total area under the curve by the area from 1 to ∞. To get an exact area, we would need to integrate from 0, but as our function is undefined at this point, it is easier to deal with the approximate area instead. To find the median, we then find the point, x, that cuts this volume in half:

$$\int_1^x \frac{\sqrt{t}}{\sqrt{i}} di = 2t/2$$
$$2\sqrt{t}(\sqrt{x}-1) = t$$
$$\sqrt{x} - 1 = \sqrt{t}/2$$
$$x \approx t/4$$

If you were the owner of a traditional brick and mortar store, you could interpret this as: in order to capture half the sales volume of the market, your store must stock 1/4 of the products in the market.

This median of $t/4$ assumes our original m_i function, $m_i(t) = \sqrt{t/i}$, where $e = 1/2$. Clearly as $t \to \infty$, $t/4 \to \infty$. In fact, this is true for $0 < e < 1$. Similar limits are true for other ranges of e:

$$0 < e < 1\text{: as } t \to \infty, \text{ median } \to \infty$$
$$1 < e < 2\text{: as } t \to \infty, \text{ E[Rank] } \to \infty$$
$$2 < e < 3\text{: as } t \to \infty, \text{ V[Rank] } \to \infty$$

As we continue to increase e, higher moments will approach infinity.

Zipfian distribution

If we have a market where, at time t, a consumer arrives and purchases $\begin{cases} \text{new product with probability } \frac{1}{t} \\ \text{old product with probability } \frac{t-1}{t} \end{cases}$, then we will get a *Zipfian* distribution, where $e = 1$.