SI 646: Week 6
The Long Tail...

Today’s Agenda:
1. Long Tail Stuff
2. Case: Long Wait for Long Tail?
3. Review Assignment 1, etc.
4. Next Time: Final Exam

Note: these slides are based partly on Jeff Mackie-Mason’s 646 slides from Winter term 2008.
Rhapsody: More than 2 million tracks
40% of revenue from tracks not at Wal-Mart
The new growth market: Products you can't find anywhere but online

**Rhapsody**
Total inventory: 1.5 million tracks

**Netflix**
Total inventory: 55,000 DVD titles

**Amazon**
Total inventory: 3.7 million book titles

---

**Total sales**
- Products not available in offline retail stores

- **40%**
- **21%**
- **25%**
Visits to Top 100 Blogs

(based on data obtained from “the truth laid bear”, 11 Feb 2009)

Blog Daily Visits v. Ranking

Note: top 20 blogs responsible for 62% of traffic...
Citations to journal-year observations in economics

Citations in Year Y to journal-publication-year X

Source: McCabe and Snyder, 2008
Pareto distribution:
What is the probability that a person has wealth $x$?

- Vilfredo Pareto, Italian economist and social theorist, 1848-1923
\[ f(x; k, x_m) = k \cdot \frac{x_m^k}{x^{k+1}} \quad \text{for } x \geq x_m. \]
Pareto distribution:
Most people have low wealth, few have very high
Other Examples:

• Human settlement sizes
• File sizes transferred over Internet
• Sizes of oil fields
• Rates of return on corporate equities (stocks)
• Areas burnt in forest fires
Zipf’s law (1949):

“In a corpus of natural language utterances, the frequency of any word is roughly inversely proportional to its rank in the frequency table”
More generally, the size of the r'th largest occurrence of the event is inversely proportional to its rank:

\[ y = a r^{-b} \]

with \( b > 1 \), but close to unity
In the “Brown” corpus,

“the” accounts for 7% = .07/1
“of” for 3.5% = .07/2
“and” for 2.8% = .07/2.5

The first 135 words account for 50%

Based on a 1967 analysis of the
Brown University Standard Corpus of Present-Day American English
For more info go to the following Wikipedia link
Pareto & Zipf are both examples of a power law:

\[ y = a x^k \]

Take logs of both sides: \( \log y = \log a + k \log x \)
which is log-log linear: \( z = a + b v \)
A plot of word frequency in Wikipedia (November 27, 2006).

The plot is in log-log coordinates.

$x$ is rank of a word in the frequency table; $y$ is the total number of the word’s occurrences. Most popular words are “the”, “of” and “and”, as expected.

Zipf's law corresponds to the upper linear portion of the curve, roughly following the green line $\log y = 1.3 \times 10^7 - \log x$.

http://en.wikipedia.org/wiki/Zipf%27s_law
Sites visited by AOL users, December day 1997.
L. Adamic, “Zipf, Power-laws, and Pareto - a ranking tutorial”,

Note: this resembles the journal citation plot – not a standard long tail plot
80 - 20 Rule?
For many phenomena, 80% of the consequences stem from 20% of the causes (Pareto principle)

Doesn’t need to add to 100! These are percents of different things.

EXAMPLE: Inventory typically 10% of products account for 80% of inventory (or 20% of products account for 95% of inventory)
“Small fraction of X responsible for large fraction of Y”
3000 out of 55,000 titles: 5.5%
79% of sales
For Netflix, “80 – 6 rule”
What causes power laws?

• Variety
• Quality variation (drives differences in popularity)
• Network effects (arising from reputation effects)
• Scarcity (without scarce time/attention/resources would consume everything!)
Generally, goods power laws apply to subcategories (e.g., genres) as well...
• Network effects and recommender services are most effective at genre or subcategory level
Let’s do the basic economics of variety: demand for, supply of
Why the long tail now?

Didn’t consumers demand variety before?
Need to know about and find variety.

Internet can help with this in big way, increasing demand for variety.
Other big changes are in cost of providing variety...

e.g. costs of distrib, mfg, inventory, mktg, etc
Suppose there are distribution fixed costs

• $\pi = (p-mc)Q - F$

• Produce if expect $\pi > 0 \rightarrow Q > F/(p-mc)$

• If $F$ gets smaller, more goods offered

• Retail display, warehousing, some marketing are mostly fixed.

• e.g., need to rent 1 inch of CD storage per title regardless of how many are sold.
So, fixed distribution costs favor mass market hits. Lower fixed costs favor niche goods.

Does marketing to the Long Tail imply a shorter “head”?

- Availability of variety: people shift attention and consumption away from head
- Lower search costs of finding what you want
- Sampling to overcome risk of purchasing less familiar experience goods
- **EVIDENCE:** Brynjolfsson et al (2005), women’s clothing company, lower search costs for fixed inventory:
  - Bottom 80% of products yielded 15.7% of catalog sales
  - but 28.8% of online sales
Does marketing to the Long tail increase demand or just shift it?

• Time and attention are limited (did Amazon increase avg. hours of reading?)
• But, by providing more desirable products, we may shift time and attention from other activities.
• Also, better pre-selection may help us consume more efficiently: “I may not read any more words than I once did, but they’re more likely to be meaningful to me.”
Should prices be higher or lower for products down the tail?

• Depends on demand elasticity, which can go either way.

• For example, demand for niche books is often less elastic (supports higher prices) because of specific preferences. Discounting higher on best sellers (why?)

• Music demand, on the other hand, might be more elastic further down: willing to experiment, but only if price lower.
Anderson: Long Tail spawns two imperatives.

• 1. Make everything available
  - Driven by reduced distribution costs

• 2. Help me find it
  - Driven by reduced costs of finding and recommending
Anderson’s nine rules

• A good synthesis review of SI 646!
• But Didn’t need to discover Long Tail to learn them. No new ideas....
1. Move inventory way in...or way out
2. Let customers do the work
3. One distribution method doesn’t fit all
4. One product doesn’t fit all (unbundle/rebundle)
5. One price doesn’t fit all (price discrimination!)
6. Share information (lose control)
7. Think “and”, not “or” (variety!)
8. Trust the market to do your job (related to 2)
9. Understand the power of free (experience goods)