1. When monkeys type ...

(a) The probability that the monkey types a c-letter word.

$$P = \frac{1-q}{n} \cdot q = \frac{q(1-q)}{n}$$

(b) We assume rank r corresponds to word length $l(r)$, then, for words with length $l(r) = l$:

$$\sum_{i=0}^{l-1} n^i < r < \sum_{i=0}^{l} n^i$$

$$\frac{n^l - 1}{n-1} < r < \frac{n^{l+1} - 1}{n-1}$$

then we have,

$$\log_n(rn - r + 1) - 1 < l < \log_n(rn - r + 1)$$

we know that the monkey types a letter with l has probability.

$$\left(\frac{1-q}{n}\right)^l q$$

then $P_r = \left(\frac{1-q}{n}\right)^{l(r)} \cdot q \Rightarrow \log(P_r) = l(r) \log\left(\frac{1-q}{n}\right) + \log q$

$$< \left[\log_n(rn - r + 1) \cdot \log\left(\frac{1-q}{n}\right)\right] + \log q$$

$$\lim_{n \to \infty} \frac{\log P(r)}{\log(r)} < \lim_{n \to \infty} \frac{\log q + \log_n(rn - r + 1) \log \frac{1-q}{n}}{\log(r)}$$

$$= \lim_{n \to \infty} \frac{\cancel{\log(rn - r + 1)} \log \frac{1-q}{n}}{\cancel{\log(r)} \log(n)}$$

$$= \log_n\left(\frac{1-q}{n}\right)$$

On the other hand, $\log(P_r) = l(r) \log\left(\frac{1-q}{n}\right) + \log q$

$$> \left[\log_n(rn - r + 1) - 1\right] \cdot \log \frac{1-q}{n} + \log q$$

$$\lim_{n \to \infty} \frac{\log P(r)}{\log(r)} > \lim_{n \to \infty} \frac{\left[\log_n(rn - r + 1) - \cancel{1}\right] \log \frac{1-q}{n} \cancel{+ \log q}}{\log(r)}$$

$$= \lim_{n \to \infty} \frac{\cancel{\log(rn - r + 1)} \log \frac{1-q}{n}}{\cancel{\log(r)} \log(n)} = \log_n \frac{1-q}{n}$$

Therefore,

$$\log_n\left(\frac{1-q}{n}\right) \le \lim_{n\to\infty} \frac{\log(Pr)}{\log(r)} \le \log_n\left(\frac{1-q}{n}\right)$$

Then, we have,

$$\lim_{n\to\infty} \frac{\log P(r)}{\log(r)} = \log_n\left(\frac{1-q}{n}\right)$$

(c). We can say that the distribution of Pr is a Heavy tail distribution, Because it is linear on log-log scale ($\log_n\left(\frac{1-q}{n}\right)$ is a constant).

3. Exploiting the long tail

(a) The expected increase in market size for product i at time t:

$$\mathbb{E}(m_i(t) \mid m_i(t-1)) - m_i(t-1)$$

(i) if $i = t$, then $\mathbb{E}(m_i(t) \mid m_i(t-1)) - m_i(t-1) = 1$

(ii) if $i < t$, then this purchase is based on social behavior.

$$\mathbb{E}(m_i(t) \mid m_i(t-1)) - m_i(t-1) = \frac{m_i(t-1)}{\sum_{j=1}^{t-1} m_j(t-1)}$$

$$= \frac{m_i(t-1)}{2t-4}$$

(b) We can assume,

the rate of change of volumn of product i at time t:

$$\frac{dm_i(t)}{dt} = \frac{\mathbb{E}(m_i(t+1) \mid m_i(t)) - m_i(t)}{(t+1)-t} = \frac{m_i(t)}{2t-1} \quad (t > i)$$

(c) solve the above differential equation:

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

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

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

$$m_i(t) = C(2t-1)^{\frac{1}{2}} = \sqrt{\frac{t}{i}}$$

Using the initial condition, $m_i(i) = 1 \Rightarrow C = \frac{1}{\sqrt{2i}}$

(d) Consider 1^{st} product the most popular one, its market share:

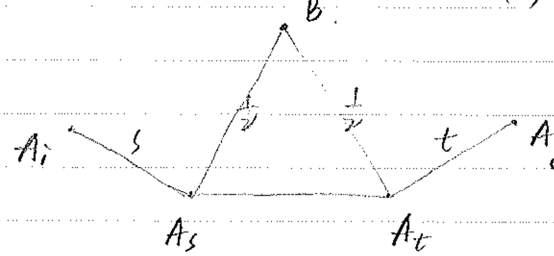
$$\frac{m_1(t)}{2t-1} \approx \frac{\sqrt{t}}{2t} \xrightarrow[t\to\infty]{} 0$$

And Consider last coming product (∂ of total)

$$\frac{\int_{(1-\partial)t}^{t} \sqrt{\frac{t}{i}}\, dt}{\int_0^t \sqrt{\frac{t}{i}}\, dt} = \frac{\sqrt{t} - \sqrt{(1-\partial)t}}{\sqrt{t}} = 1 - \sqrt{1-\partial}$$

This means that unpopular items still make up a large revenue source in the aggregate which leads Amazon to success.

4. It's a small world after all



(a)

(i) Between A_i, A_j, we have A_s, A_t, the nearest nodes next to A_i, A_j respectively, with distance s and t.

if $l < k$, (which means connect to B really helps)

$$P(l,k) = l \cdot p^2 (1-p)^{s+t}$$
$$= l \cdot p^2 (1-p)^{l-1}$$

explain: Shortest path is l, with connected to B, we have p^2, then left $l-1$ divided to 2 parts: s and t, has l combination, which not connected to B: $(1-p)^{l-1}$

if $l = k$, $P(l,k) = 1 - \sum_{i=1}^{k-1} P(l=i)$

$$= 1 - \sum_{i=1}^{k-1} i \cdot p^2 (1-p)^{i-1}$$
$$= 1 - p^2 \sum_{i=1}^{k-1} \frac{d}{dx}\left(\int i x^{i-1} dx\right) \qquad x = 1-p$$
$$= 1 - p^2 \frac{d}{dx} \sum_{i=1}^{k-1} x^i dx$$
$$= 1 - \cancel{p^2} \frac{1 - k(1-p)^{k-1} + (k-1)(1-p)^k}{\cancel{p^2}}$$
$$= k(1-p)^{k-1} - (k-1)(1-p)^k$$

explain: when $l = k$, which means from A_i to A_j we don't pass the node B and this is the worst case where our shortest path reachs maxima

(ii) The expected value of the shortest path from A_i to A_j is:

$$\mathbb{E}(\text{shortest path}) = \underbrace{\sum_{i=1}^{k-1} i^2 \cdot p^2 (1-p)^{i-1}}_{\text{part } \alpha} + \underbrace{k \cdot \left[k(1-p)^{k-1} - (k-1)(1-p)^k\right]}_{\text{part } \beta}$$

$$= \frac{2}{p} - 1 - (1-p)^k \left[\frac{2}{p} + k - 1\right]$$

Here, part α represents $l < k$, while part β represents $l = k$.

(b) The expected average shortest path length between the nodes on the ring of the graph:

$$\mathbb{E}[l_{avg}] = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{k=1}^{n-1} \mathbb{E}(\text{shortest path})$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{k=1}^{n-1} \left(\frac{2}{p} - 1 - (1-p)^k \left(\frac{2}{p} + k - 1\right)\right)$$

$$= \frac{1}{n-1} \sum_{k=1}^{n-1} \left[\left(\frac{2}{p} - 1\right) - \left(\frac{2}{p} + 1\right)(1-p)^k - k(1-p)^k\right]$$

$$= \frac{2}{p} - 1 + \frac{1}{n-1}\left(\frac{4}{p} - \frac{3}{p^2} - 1\right) + \frac{(1-p)^n}{n-1}\left(\frac{n-2}{p} + \frac{3}{p^2}\right)$$

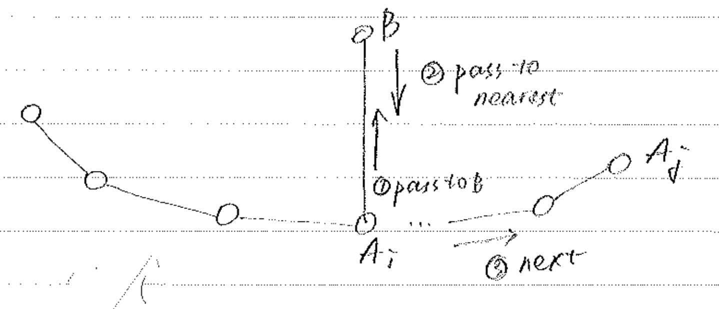
when $n \to \infty$, we have

$$\mathbb{E}[l_{avg}] \sim \frac{2}{p} - 1 + \frac{1}{n-1}\left(\frac{4}{p} - \frac{3}{p^2} - 1\right) \to \frac{2}{p} - 1$$

if we don't have central node B,
then every shortest path length from $A_i \to A_j$ (where $j = i+k$) is k,
then

$$\mathbb{E}[l_{avg}] = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{k=1}^{n-1} k = \frac{n}{2}$$

(c) This algorithm will always find the shortest path, except one special condition, when The closest node on the ring to destination B thinks is also A_i, then the route seems:



This condition can also be described as there is only one node along the way from A_i to A_j connect to B (except A_{j-1}, then A_{j-1} will not send packet to B, it will send it to A_j)

How close to the "shortest"?

This special case = "shortest" + 1