So far we've seen that 2 of our 4 universal properties can be explained by a simple, independent connection of nodes:

i.e. $G(n,p)$ exhibits
(i) a giant wcc
(ii) a small diameter

Also, clustering seems to just be a result of node correlations...

But, we don't have any explanation yet about what causes the heavy-tailed degree distributions.
(and actually, we don't know too much about heavy-tailed degree distributions yet.)

That's the focus of today:

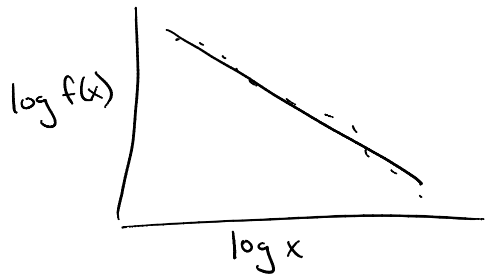
1) What do we mean by "heavy-tails"?

2) Why are "heavy-tails" so common?

(<u>Note</u>: We'll spend a lot of time on straight probability, without talking about networks, before we finally come back to a network model that leads to heavy-tails. We're doing this because heavy-tails show up in many places besides just networks (as you saw in lecture3) and many of the mechansims that lead to HTs outside of networks provide insight for why they occur in networks too.)

What are heavy-tailed distributions?

So far, to us heavy-tails means
"linear on a log-log plot"



$f(x) = \text{p.d.f.}$
$= c\,x^{-\alpha}$
$\Rightarrow \log f(x) = \underbrace{\log c}_{A} + (-\alpha)\log x$

Recall, if the cdf is linear on a log-log plot then it's a:

① Pareto Distribution

- Named after Vilfredo Pareto (1843-1923)
  → an economist who studied the income distribution
  "80% of wealth in 20% of population

Note: ccdf: $\bar{F}(x) = \Pr(X > x) = \int_x^\infty c\,t^{-\alpha}\,dt \qquad (\alpha > 1)$

$= \frac{c}{-\alpha+1}\left(t^{-\alpha+1}\right)\Big|_x^\infty$

$= c_1\,x^{-\alpha+1}$

$\Rightarrow$ still linear on a log-log scale.

this is important since $\bar{F}(x)$ is much less noisy than $f(x)$

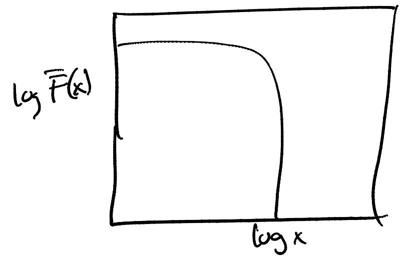
see ppt

Pareto (x_L, α) has $\bar{F}(x) = \left(\frac{x_L}{x}\right)^\alpha$ $\quad x \in [x_L, \infty)$

x_L: scale (min value); α: shape

Key 1: Polynomial tail is "heavier" than the exponential tail of the normal, Binomial, Poisson, etc.

→ when we plot those on log-log scale, the tail "dies"



see ppt

Key 2: Very highly variable

- $E[X^i] = \infty$ if $i > \alpha$

$\Rightarrow$ $Var[X] = \infty$ if $\alpha < 2$

(many real situations have $\alpha \approx 2+\varepsilon$)

To see this, recall

$$E[X^i] = \int_{x_L}^{\infty} t^i f(t)\, dt$$

$$= \int_{x_L}^{\infty} t^i\, c t^{-\alpha-1} dt$$

$$= c\int_{x_L}^{\infty} t^{i-\alpha-1}\, dt < \infty \iff i-\alpha-1 < -1$$

(recall $\int_c^{\infty} \frac{1}{t} = \infty$)

☆ Also called Zipf distn,
power law,
scale-free.

☆ Many things have been claimed to be Pareto based on log-log plots, but in reality are better modeled by Other Heavy-tailed distributions that are "nearly" linear in log-log scale.

See ppt

Other Heavy-tailed Distributions

Q: Who knows of other HT distns?

A: Others include:

A: Others include:

- ☆ the Weibull
- ☆ the log Normal
- Cauchy
- Student-t

② The Weibull(α, λ)

Named after Walodi Weibull (1887–1979)

$\bar{F}(x) = e^{-(x/\lambda)^\alpha}$ $\qquad x \in [0, \infty)$

$f(x) = \underbrace{\frac{\alpha}{\lambda}\left(\frac{x}{\lambda}\right)^{\alpha-1}}_{\text{polynomial component}} \underbrace{e^{-(x/\lambda)^\alpha}}_{\text{exponential component}}$

see ppt for plot

→ nearly linear on log log when α is small

→ very highly variable when α is small

$E[X^i] = \lambda^i \Gamma\left(1 + \frac{i}{\alpha}\right)$ . . .

gamma function is the continuous factorial

so all moments exist, but are very large if α is small

→ captures many other distns as special cases

- $\alpha = 1$ → exponential
- $\alpha = 2$ → Rayleigh
- $\alpha \sim 3,4$ → $\approx$ Normal
- $\alpha \to \infty$ → $\approx$ Determistic
- ☆ $\alpha < 1$ → heavy-tailed

③ The log Normal Distribution

$X \sim \text{logNormal}(\mu, \sigma^2)$

$\Rightarrow X \stackrel{d}{=} e^Y$ for $Y \sim \text{Normal}(\mu, \sigma^2)$

→ Again, nearly linear of log log plot if σ^2 is large

→ Again, all moments exist, but are very large:

e.g. $\mathrm{Var}[x] = (e^{\sigma^2} - 1)\, e^{2\mu + \sigma^2}$

see ppt for plot

We've seen 3 distributions that I've said are "heavy-tailed", but I still haven't defined what I mean by heavy-tailed really....

Q: What have these examples had in common?

A:
- nearly linear on a log-log scale
- tail that has a "polynomial" component

} "heavy tail"

- very large variance (and higher moments)

(Note that the 1st 2 imply the third.)

> **def:** A random variable is heavy-tailed iff
> $\forall s > 0 \quad \lim_{x \to \infty} e^{sx}\, \bar{F}(x) = \infty$

i.e. if the tail is "heavier" than any exponential.

If you want to learn more about heavy tails, take my course in the spring

147: Network Perf. Modeling

→ we'll cover many subclasses of HT distributions with important properties useful for system design

For this class (144), you should just take away 2 key insights for thinking about

distributions:

Heavy vs. Light tailed distributions:

1) LT → lot's of "near avg" samples
   HT → a few huge samples and many small.

   e.g. heights ⇒ ≈ Normal ⇒ LT
   ⇒ everyone has "nearly" the same height.

   incomes ⇒ ≈ Pareto ⇒ HT
   ⇒ a few really rich people and lots of people w/ < $50k/yr

2) LT → "conspiracy principle"
   lots of little things conspire to create a "bad" event
   HT → "catastrophe principle"
   a "bad" event is the result of 1 catastrophic event.

Formally:

$$\frac{\Pr(\max(X_1 \ldots X_n) > x)}{\Pr(X_1 + \ldots + X_n > x)} \to 1 \quad \text{as } x \to \infty.$$

← 1 bad event

↖ catastrophe

Where do heavy-tails come from?

This should seem a bit contradictory in light of the Central Limit theorem (CLT) which says that random sums converge to the Normal... and so we should expect the Normal distribution to be "universal"

... and it is very common:

heights,
weights,
IQ,
.... and many more.

Recall what the CLT says:

CLT As $n \to \infty$,

$$\frac{1}{\sqrt{n}}\left(X_1 + X_2 + \dots + X_n - n\mu\right) \xrightarrow{d} Z \sim \text{Normal}(0, \sigma^2)$$

when (i) $X_i = X$ are iid w/ $E[X_i] = \mu$
(ii) $E[X^2] < \infty$

<u>Aside</u>: Why do I write it this way?

Law of Large #s: $\frac{X_1 + \dots + X_n}{n} \to \mu$

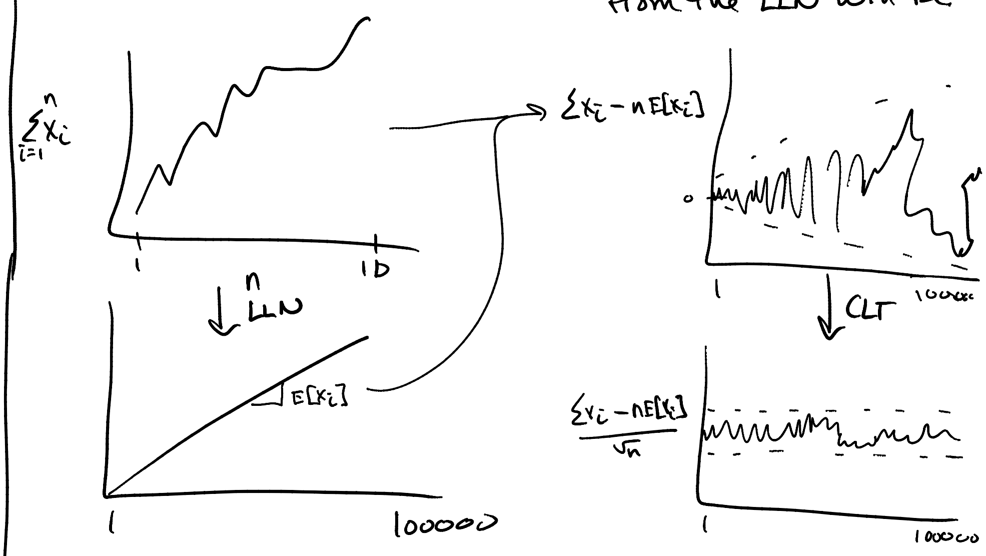
$\Rightarrow X_1 + \dots + X_n - n\mu = o(n)$

ie $X_1 + \dots + X_n = n\mu + o(n)$

☆ CLT gives the 2^{nd} order correction term

$X_1 + \dots + X_n = n\mu + \Theta(\sqrt{n})$ ← Normally dist

ie. tells how big deviations from the LLN will be



LLN → 1^{st} order approximation
CLT → 2^{nd} order correction for LLN

<u>Q</u>: What happens if we relax (i) & (ii)?

<u>A</u>:
(i) If X_i are dependent:

A:

(i) If X_i are dependent:

- If they "aren't too dependent", nothing changes
- If they are extremely correlated, you can get anything you want.

(ii) If $E[X^2]=\infty$, then we are assured the sum is heavy-tailed!

So, finite variance $\Rightarrow$ sum $\sim$ Normal

infinite variance $\Rightarrow$ sum $\sim$ heavytailed

There is some beautiful theory here...

Generalized CLT

As $n\to\infty$, for iid X_i

$$\frac{1}{n^{1/\alpha}}(X_1+X_2+\dots+X_n-d_n)\xrightarrow{d} Z$$

iff Z is a "α-stable" distribution w/ $\alpha\in(0,2]$

α-stable distribution are:

$$\begin{cases}\alpha=2 \rightarrow \text{Normal}\\ \alpha\in(0,2)\rightarrow \text{Pareto}(\alpha)\text{ tail}\end{cases}$$

body may be non-pareto.

☆ So the normal was just a special case and sums might just as easily lead to heavy-tails.

We just saw one process that leads to heavy-tails... but there are many others.

The CLT combines things "additively". But the world is not always additive.

Sometimes it is:

① Multiplicative $(Z = \prod_i X_i)$
② Extremal $(Z = \max_i (X_i))$

---

① Multiplicative processes:

In many cases we can think of something growing at a rate proportional to its size.

e.g. money → rich get richer
web sites → traffic, in degree
nature → forest fires

In such scenarios, a simple model is:

$$Y_0 = 1 \qquad (X_n > 0)$$
$$Y_n = X_n Y_{n-1} = \prod X_i$$

Q: What is Z such that
$Y_n \xrightarrow{d} Z$ as $n \to \infty$?

A: This is just the CLT in disguise!

$$\log Y_n = \sum \log X_i$$
$$\Rightarrow \log Y_n \to Z \sim \text{Normal}$$

(when $\log X_i$ has finite variance... which is almost always).

So $Y_n \Rightarrow$ LogNormal

☆ multiplicative processes lead to heavy-tails

(Note: We can also get Pareto tails if we take $Y_n = \max(X_n Y_{n-1}, \varepsilon)$ for some $\varepsilon > 0$.)

---

② Extremal Processes

Here the goal is to understand

$$Y_n = \max(X_1 \dots X_n)$$

This is a simplistic view of the processes that are important to

- insurance risk analysis
- evolution
- likelihood of large floods, fires, etc.

... and many other places.

We won't spend much time here, but I want to expose you to a result you probably haven't heard of...

This is the parallel of the CLT for "extreme value theory"

> Thm: As $n \to \infty$, for iid X_i,
>
> $$c_n(\max(X_1, \dots, X_n) - d_n) \to Z$$
>
> iff $Z = \begin{cases} \text{Type I} & \bar{F}(x) = e^{\left(-e^{-\left(\frac{x-\mu}{\sigma}\right)}\right)} & x \in \mathbb{R} \\ \text{Type II} & \bar{F}(x) = e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\delta}} & x \geq \mu \\ \text{Type III} & \bar{F}(x) = e^{-\left(\frac{\mu-x}{\sigma}\right)^{-\delta}} & x \leq \mu \end{cases}$

Type I is light-tailed (Gumbell), but

Types II & III are generalizations of the Weibull, so they can be heavy-tailed.

(Types II & III emerge when X_i have Pareto tails)

★ So, like additive processes, extremal processes can lead to heavy or light tails depending on X_i

---

This is a long lecture, so lets take stock of where we are:

(1) We saw 3 examples of heavy-tailed distributions

(2) We saw that HTs can emerge from additive, multiplicative, and extremal processes... but that additive & extremal processes take heavy-tailed x_i and "maintain" the heavy tail for the sum/max. On the other hand, multiplicative processes can take light-tailed X_i and "create" a heavy-tail.

Now, we're finally going to go back to networks, and try to come up with a simple "explanation" as to why degree distributions are heavy-tailed

---

Back to networks

Recall where we are:

We've seen that $G(n,p)$ has a giant component and a small diameter, but that it has a Binomial $\approx$ Normal degree distribution, which does not match the heavy-tailed degree distns we see in practice.

Q: Given all the ways we've seen to create heavy-tails, how can we do it in a random graph model?

A: We should somehow use the multiplicative process idea...

One approach: "Rich get richer"

- create nodes in order $1 \ldots n$
- when a node is created it
  - chooses a node to link to uniformly at random w/prob p
  - chooses a node to link to proportionally to the in-degree of the node w/prob $1-p$

☆ all edges are directed

(This was introduced by [Price 1965, Albert & Barabasi 1999])

Called: "Preferential Attachment"

→ clearly this has a flavor of the multiplicative process since the "rate of connection" depends on the current degree.

So, we should "expect" a heavy-tail.

⇓

proving this rigorously is difficult, so we'll give a "heuristic proof" i.e. we'll prove that an approximation of preferential attachment has power-laws degrees.

Before doing the proof though,

Q: Is this a "realistic" model?

A: maybe?

Now to the proof:

$$\boxed{\text{Claim: "Pref attachment" leads to power law in-degrees}}$$

We will actually study a deterministic version of preferential attatchement based on the "expected change" to degrees at each step.

Let $X_j(t)$ be the indegree of node j at time t and let nodes "arrive" at times $t=1,2,\dots,n$

So $X_j(j)=0$.

Q: At time $t>j$, how does the degree of j change (in expectation)?

A: $$E[X_j(t+1)\,|\,X_j(t)] = X_j(t) + p\,(\Pr(\text{selected})\cdot 1) + (1-p)(\Pr(\text{selected})\cdot 1)$$

$$= X_j(t) + p\cdot\frac{1}{t} + (1-p)\cdot\frac{X_j(t)}{t}$$

# of total edges at time t.

This leads the the following (mean-value) deterministic approximation:

$$\frac{\partial X_j}{\partial t} = \frac{p}{t} + \frac{(1-p)X_j}{t}$$

We'll actually show that this approx of our model yields heavy-tails...

(Doing the full probabistic pf requires tools we don't have. I posted the paper on the web site though.)

"pf": We need to solve this differential equation: Dividing by $p+(1-p)X_j$ and integrating both sides gives:

$$\int \frac{1}{p+(1-p)X_j}\left(\frac{dX_j}{\partial t}\right)dt = \int \frac{1}{t}\,dt$$

$$\left(\frac{1}{1-p}\right)\log\left(p+(1-p)X_j\right) = \log t + c$$

for some constant c

$$\Rightarrow p+(1-p)X_j = A\,t^{1-p}$$

$$\Rightarrow \quad p + (1-p)x_j = \underset{e^c}{\underset{\uparrow}{A}}\, t^{1-p}$$

$$\& \quad x_j(t) = \frac{1}{1-p}\left(At^{1-p} - p\right)$$

solving for A using $x_j(j)=0$ gives $A = p/j^{1-p}$

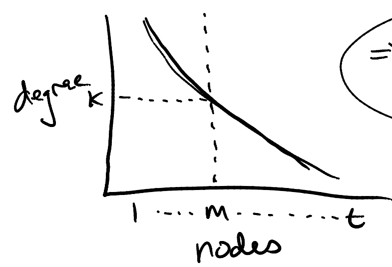
$$\Rightarrow \quad x_j(t) = \frac{p}{1-p}\left[\left(\frac{t}{j}\right)^{1-p} - 1\right]$$

Now, the last step is to understand the fraction of nodes with degree $\geq k$, i.e., to find the power law!

So, since $x_j(t)$ is deterministic, we know that all nodes that arrived early will have large degree. Specifically all j st:

$$x_j(t) = \frac{p}{1-p}\left(\left(\frac{t}{j}\right)^{1-p} - 1\right) \geq k$$

will have degree bigger than k



$\Rightarrow \Pr(\text{degree} > k) = \frac{m}{t}$

Rewriting this gives:

$$m = t\left[\frac{1-p}{p}k + 1\right]^{-1/1-p}$$

so all nodes with $j = 1, 2, \ldots \; t\left[\frac{1-p}{p}k+1\right]^{-1/1-p}$ have degree bigger than k.

which means the fraction of these nodes is

$$\frac{m}{t} = \frac{t\left[\frac{1-p}{p}k+1\right]^{-1/1-p}}{t}$$

← nodes w/ degree bigger than k

↖ total nodes

$$\approx \quad B\,k^{-1/1-p}$$ for some constant B

□ $\nwarrow$ power law!

Observation: $p \to 1 \Rightarrow$ power law disappears

(go back to light tails)

$p \uparrow \Rightarrow$ tail gets heavier.

So, that's one simple network mechanism that leads to heavy tails.

... there are many others too, so if you're interested talk to me and I'll point the way.

Next time: Why is it a small world after all?