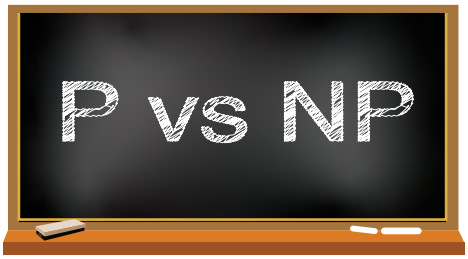
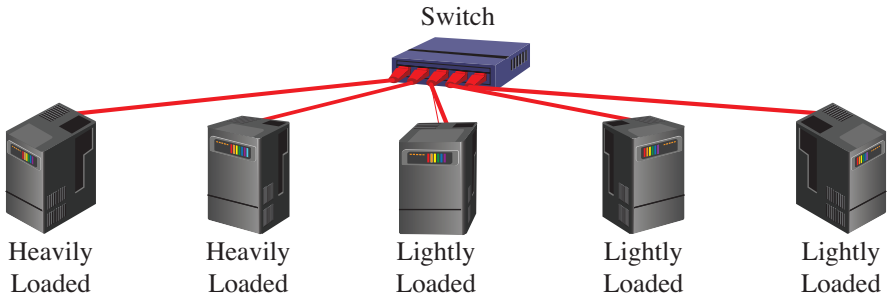
The Dark Ages of the 1990's ... Mor's PhD story





Definitions

A job's **size** refers to its total CPU requirement.
A job's **age** refers to its total CPU usage thus far.
A job's **remaining size** is its remaining CPU requirement.

Question: Which of these jobs likely has higher remaining size?

What we want to understand: $\mathbf{P}\{\text{Size} > x + a \mid \text{Size} > a\}$

Q: What's the answer if Size $\sim$ Exp(μ)?

Failure Rate – informally

$$\mathbf{P}\left\{\text{Size} > x + a \;\mid\; \text{Size} > a\right\}$$

Q: Examples?

Failure Rate more formally

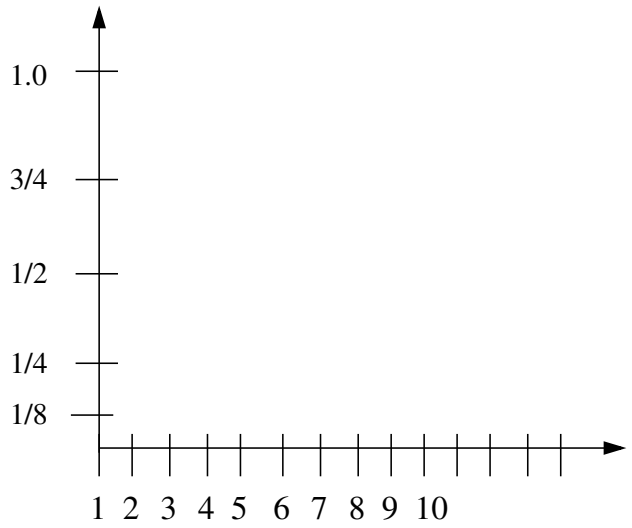
Definition: For continuous X with p.d.f $f_X(t)$ and tail $\overline{F}_X(t)$, the **failure rate function $\mathbf{r_X(t)}$** for X is:

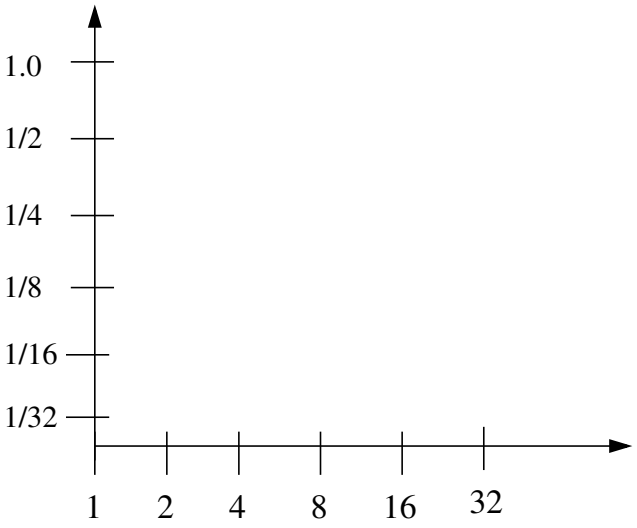
$$r_X(t) \equiv \frac{f_X(t)}{\overline{F}_X(t)}$$

- When $r_X(t)$ is strictly decreasing in t, we say: _____.
- When $r_X(t)$ is strictly increasing in t, we say: _____.

Question: What if $r_X(t)$ is constant?

To try to understand the failure rate, Mor measured millions of UNIX jobs ...





Properties of Power-law Distribution

Suppose that

$$\overline{F}_X(t) = \frac{1}{t}, \qquad t \geq 1$$

1. Is the above a valid distribution?

2. What is $\mathbf{E}[X]$?

3. What is $r_X(t)$, $t \geq 1$? What kind of failure rate is this?

4. Derive $\mathbf{P}\{X > 2a \mid X > a\}$

The Pareto(α) Distribution

Definition:

Let $0 < \alpha < 2$.

$$X \sim \text{Pareto}(\alpha) \qquad \text{if} \qquad \overline{F}(x) = x^{-\alpha}, \quad x \geq 1$$

Let $X \sim \text{Pareto}(\alpha)$:

1. If $0 < \alpha \leq 1$, classify the moments of X as finite or infinite.

2. If $1 < \alpha < 2$, classify the moments of X as finite or infinite.

Three Properties of Pareto(α) Distribution:

1. DFR
2. Infinite Variance
3. "Heavy-tailed property"

Q: What do the above properties tell us about migration?

Measured distribution is BoundedPareto(k, p, α)

End of the story ...

Today's Distributions – 2024

Question: How have workloads changed between 1996 and today?

See: "Borg: the next generation" by Tirmazi et al., *EUROSYS*, 2020.

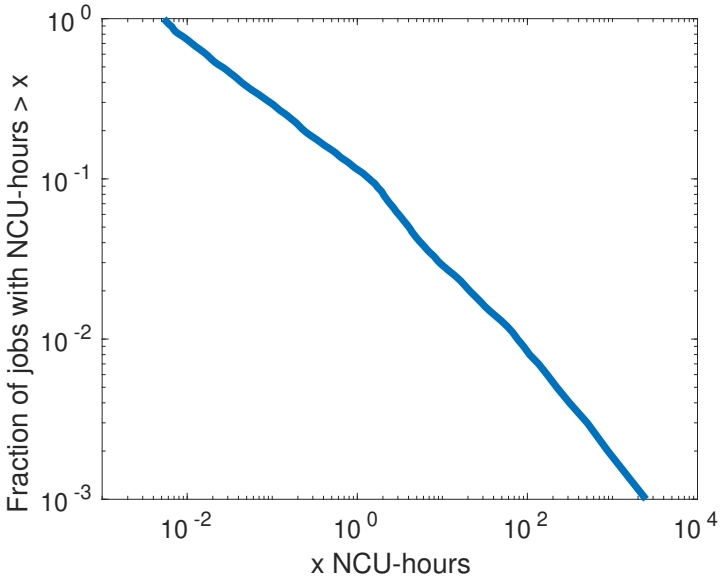
COMPUTE CONSUMPTION AT GOOGLE 2020:

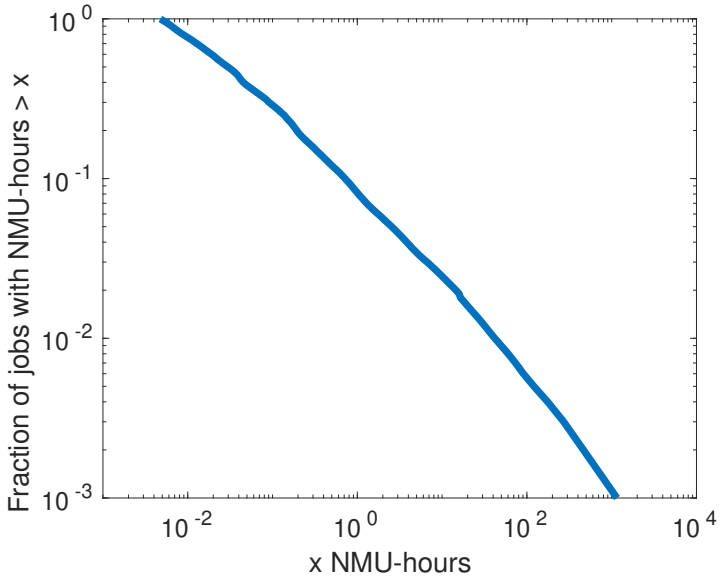
- Units: _____
- Distribution: _____
- C^2 = _____
- Span in job sizes: _____
- Top 1% jobs consume ____ % of total load

MEMORY CONSUMPTION AT GOOGLE 2020:

- Units:_____
- Distribution: _____
- C^2 = _____
- Span in memory consumption: _____
- Top 1% jobs consume ____ % of total load

Plots from the EUROSYS 2020 paper





Pareto Distributions are Everywhere

- Compute consumption across jobs
- Memory consumption across jobs
- Web file sizes
- Internet node degrees
- IP flow durations
- Wireless session times
- Phone call durations
- Wealth
- Natural disasters

WHY?