Heavy tailed distributions with applications to broadband communication systems traffic

Alexander S. Rumyantsev (Petrozavodsk State University, Russia)
Evsey V. Morozov (Karelian Research Centre of Russian Academy of Science, Russia)
Michele Pagano (University of Pisa, Italy)

What

- Our work is concerning the following:
  - ✳ Definitions
  - ✳ Properties
  - ✳ Applications
- All this connected with Heavy Tails, Self Similarity and Long Range Dependence

Where

- You can find Heavy Tails, Self Similarity and Long Range Dependence here:
  - LAN, WAN traffic measurements.
  - Single Processor: times of service.
  - Variable Bit Rate video traffic.
  - TCP traffic.
  - Insurance: claim size.
  - Number of signs on this slides.
- Lots of activities connected with human.

When

- **Heavy Tails** may occur when you see
  - ✳ Enormous variance.
  - ✳ Huge values of r.v. with big probability.
  - ✳ Infinite spectral density at the origin.
  - ✳ Linear curve on the log-log plot for the tail.
- **Long Range Dependence** may be when
  - ✳ Hurst phenomenon: slow decay of R/S statistics.
  - ✳ Autocorrelations decay slowly.
- **Self Similarity** waits for you when
  - ✳ On high levels of aggregation process doesn’t “smoothen”. It still has “pikes”.

Definitions

- Tail of distribution
- Heavy Tail (HT)
- Long Tail (LT)
- Subexponential Tail (ST)
- Regularly Varying Tail (RV(α))
- Long Range Dependence (LRD)
- Self-Similar Process (SS)

- $\overline{F}(x) = 1 - F(x)$
- $\forall \varepsilon > 0,\ E\{e^{\varepsilon X}\} = \infty$
- $\frac{\overline{F}(x+y)}{\overline{F}(x)} \to 1,\ x \to \infty$
- $\frac{P(X_1+X_2>x)}{P(X>x)} \to 2,\ x \to \infty$
- $\forall c > 0,\ \frac{L(cx)}{L(x)} \to 1,\ x \to \infty$
- $\overline{F}(x) = x^{-\alpha} L(x),\ \alpha > 0$
- ? phase transition
- $\forall c > 0,\ (Y(ct),\ t \geq 0) \sim c^H (Y(t),\ t \geq 0)$

Properties: RV

- The most convenient type of Heavy Tailed random variable.
- RV(α): the less the index α is, the "heavier" the tail is, the bigger is the probability for the random variable to have huge value.
- If index $\alpha<2$, then the dispersion is infinite, if index $\alpha<1$, then mean is infinite.
- The distribution of unfinished work (or equilibrium distribution, residual lifetime distribution) has tail 1— α heavier than the distribution of whole work.

Properties: SS

- Self Similarity on the application level is not important to be noticed while Self Similarity on the network layer is to be of great importance.
- The most interesting result (I. Norros): storage model: fractional Brownian input with mean input rate m, Self-Similarity parameter H and output at constant rate C. Let the probability of high buffer level be P(V>x)=ε. What is the behavior of level x while increasing the utilization level m/C?
  - ✴ Classical case H=1/2 (Brownian motion, no dependence).

$$\frac{1-\rho}{\rho}\cdot x = const.$$

    - Reducing by half free capacity 1-ρ costs doubling needed level x.
  - ✴ Self-Similar case 1/2<H<1

$$x = x(\rho) = const\cdot\rho^{1/(2(1-H))}\cdot(1-\rho)^{-H/(1-H)}.$$

    - Ex.: H=0.9. Reducing by half free capacity 1-ρ costs 512 times increase of x.

Relations: examples

- LRD may follow from HT. For instance, On-Off process is LRD if and only if On and Off times have infinite variance.
- Superposition of On-Off processes with Pareto distributed On and Off times with infinite variance produce Self-Similar traffic.

Why

- We need to study HT, LRD and SS because:
  - ✳ Definitions differ. Misunderstanding between authors may occur.
  - ✳ Consequences of wrong analysis may be rather dramatic for the finances of the company.
- So let us explore some applications.

Consequences of wrong analysis: example

- A single server is connected to k=99 client machines and works at a constant rate r=100. What will happen if the service rate increases as r′=1.1 r?
- M/M/1 model. Let the intensity of the client be equal to λ=1. Then the intensity of the input as a superposition of Poisson processes has an intensity 99. Then the load is equal to 99/100 and the stationary queue size is 99. For the service rate r′=110 the load is 99/110 and the queue size is 9 that is **11 times less**! So the upgrade is essentially needed!
- On-Off model. The source alternates between two states: transmitting and waiting. The time for On and Off states have RV distributions, Off has a lighter tail. Let the average time spent in On state be 1 and in Off state be 100 (i.e. the client silently edits the document for a long time and then sends it for a short time). Then the intensity of transmission is a=101 from the equation $\lambda = \frac{\mu_{on}}{\mu_{on} + \mu_{off}} a$. The crucial parameter is $\kappa_0 = \left\lceil \frac{r}{a} \frac{\mu_{on} + \mu_{off}}{\mu_{off}} - k \frac{\mu_{on}}{\mu_{off}} \right\rceil$ related to the time to overflow in system with finite buffer. If it changes, then the performance also changes. But the parameter is equal to 1 in case r=100 and r′=110. So the upgrade **doesn't change anything**. Moreover, one can count the r″ needed to change the parameter, and can find r″>199 that is **increase of service rate in about two times!**

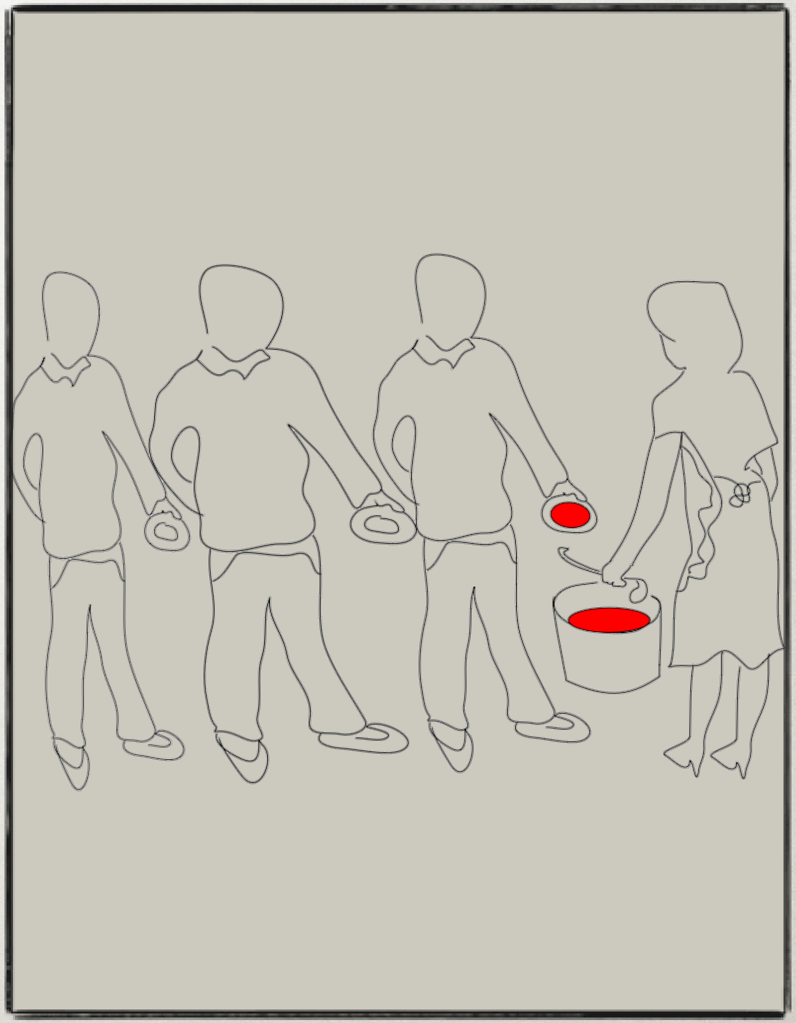
What's the difference?

- M/G/1 model: compare three service disciplines: First Come First Serve, Last Come First Serve, Processor Sharing. The event of large sojourn time is caused in different ways.
- k different sources. Let the service time distributions be Regularly Varying with indexes $\nu_1 < \nu_2 < \ldots < \nu_k$

First Come First Served (FIFO)

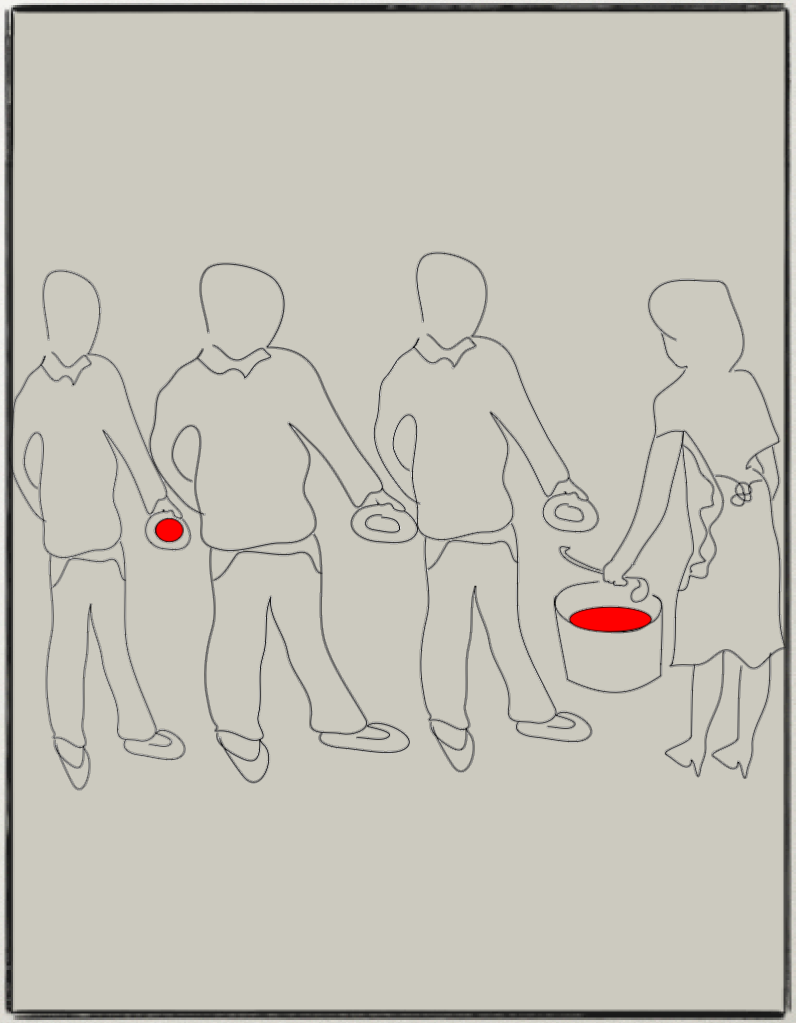
- RV(ν_i) service causes sojourn time RV($1 - \nu_1$) even heavier than the heaviest service tail.

- Long sojourn time due to **unfinished service** of some task with the heaviest tail that comes before.



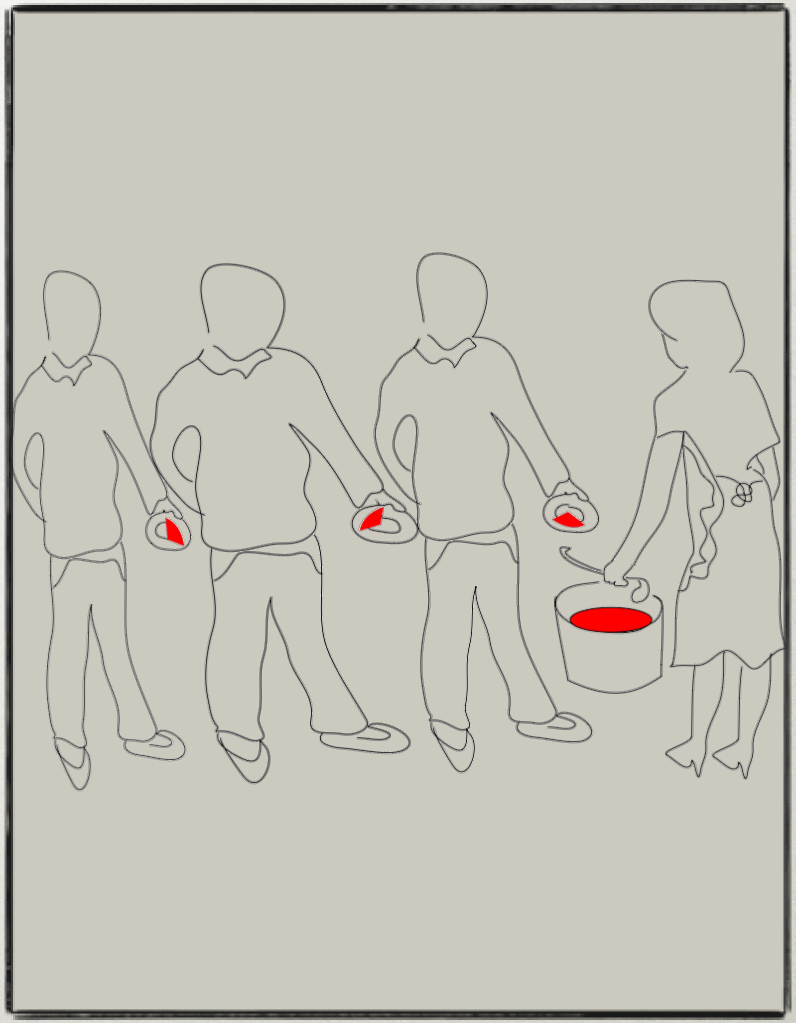
Last Come First Served (LIFO)

- RV(ν_i) service causes sojourn time RV(ν_1) equally heavy to the heaviest service tail.
- Long sojourn time because of later **service** of some task with the heaviest tail, that comes **after** this.



Processor Sharing

- RV(ν_i) service causes equally heavy RV(ν_i) sojourn time.
- Long sojourn time because of it's **own long service** time. Huge tasks don't affect the performance too much.



Upgrade?

- 10 Mbit Ethernet => 1.3 Mbps average service level
- 100 Mbit ATM net => 0.3 Mbps
- Fast network: TCP manages to send more bits before changing window size, so the collision becomes only worse.
- So why upgrade?
- Networks are to be deeply investigated.

Bibliography

- G. Samorodnitsky. Long range dependence, Heavy Tails, Rare Events.
- A. Zwart. Queueing Systems with Heavy Tails.
- Leland, Taqqu, Willinger. On the Self-Similar Nature of Ethernet Traffic.
- Crovella, Bestavos. Performance Characteristics of WWW Informational Systems.
- Goldie, Kluppelberg. Subexponential Distributions.
- I. Norros. A Storage Model with Self-Similar Input.