Fuzzy summation of random processes

David Kruml[1]

SSAOS 2018, Špindlerův Mlýn

[1]Supported by project GAČR 18-06915S.

Industrial planning

Input: Orders, material, technologies, regimes, schedule, HR.

Output: Is the plan capable? Yes/no.

Traditional solutions: Discrete time simulation, Monte Carlo.

Our approach: Use regularity of mass production, direct fuzzy methods for probability.

Production flow

The flow is described in three modes:

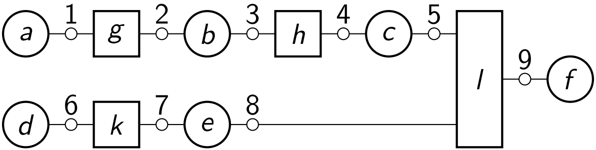
- space (production network),
- mass (material, products),
- time.

Each of the modes is organized to tree structures, we use a language of monoidal categories (composition, tensor).

The trees reflect also human reasoning — scalling between a global/rough root and local/precise leaves.

In mass production, data are compressed and computational speed is increased.

Network



a, b, c, d, e, f — stacks,

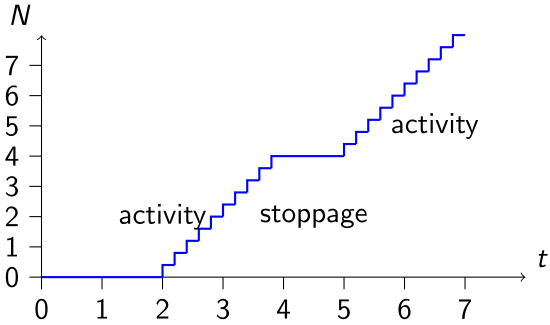
g, h, k, l — processes,

$1, \ldots, 9$ — states (nodes).

Signals

We assume that a plan is complete — *regimes* of processes are known and *jobs* are scheduled.

Then we know all inputs and outputs of processes. The flow is determined as a collection of *signals* at all states.

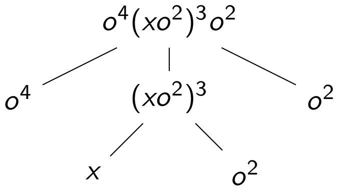


Signals as words

x — unit of mass (a product),

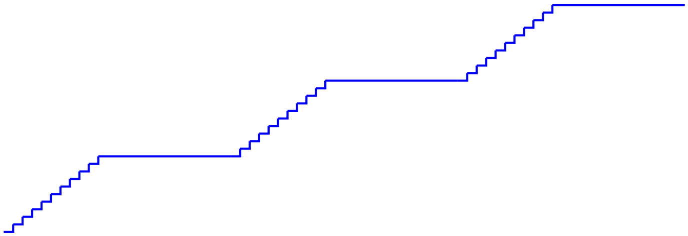
o — unit of time.

Subword tree:

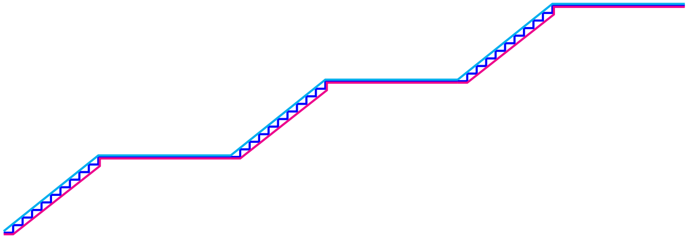


Data could be organized by markup languages. Each vertex can be equipped by additional information, calculated inductively from leaves.

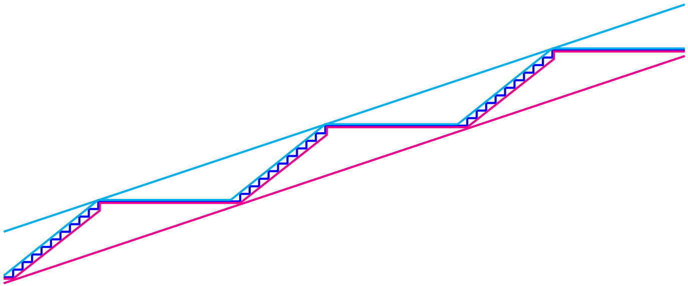
Sausage wrapping — start



Sausage wrapping — 1st step



Sausage wrapping — 2nd step



Why this?

The signal is approximated by linear segments.

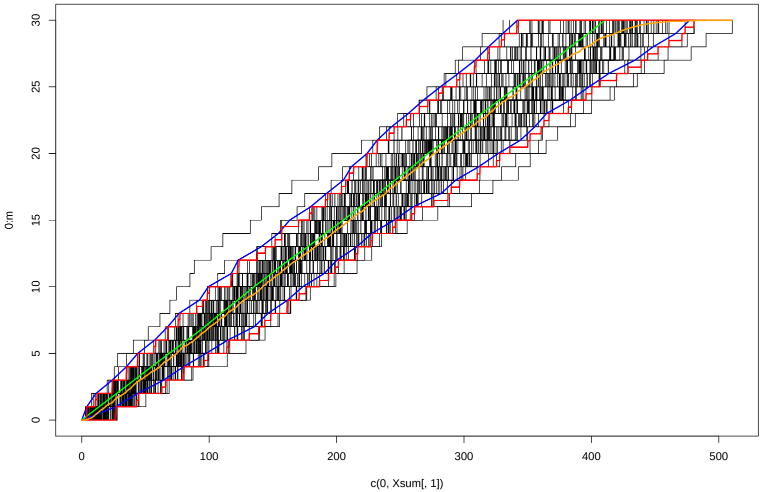
Good for answering questions about:

- maximum of the signal,
- time when some concrete value is exceeded, etc.

Time complexity: linear (search) $\rightarrow$ logarithmic (improved estimations).

Particular problem: summation of signals (of simultaneous processes). The signals could operate in a different “rhytm” and there is no good word description for the sum.

The real world signal



Random effects

We deal with signals like $o^4(xo^Y)^{50}$ where Y is a random variable, e.g.

- $Y \sim N(2, 0.3)$, or
- "Y takes values between 1.8 and 2.3, typically 2".

Signals are a sort of random walks, could be treated as "fuzzy clouds".

Precise theoretic results are scarce, need high mathematics, and computationally difficult.

We cheat with a fast fuzzy calculus.

Moments

There are three prominent characteristics:

- *mean* $\mu = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x$ (1st moment)
- *variance* $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\mathrm{d}x$ (2nd central moment)
- *3th central moment* $m_3 = \int_{-\infty}^{\infty} (x-\mu)^3 f(x)\mathrm{d}x$

Additivity: When $Y = Y_1 + Y_2$ is a sum of independent random variables, then the three moments of Y are sums of moments of Y_1 and Y_2 (regardless of the distribution). In particular, "power variable" $n * Y = Y + \cdots + Y$ has moments $n\mu, n\sigma^2, nm_3$.

Derived characteristics:

- *standard deviation* σ,
- *skewness* $\gamma = m_3/\sigma^3$.

3-point estimation

We assume that the variable is given by three values:

- "low" l — lower bound, or a small quantile,
- "typical" m — median,
- "high" h — upper bound, or a large quantile.

Inspired by a double-triangular probability distribution we set:

$$\mu = \frac{1}{6}l + \frac{2}{3}m + \frac{1}{6}h, \qquad l = \mu - \sigma + \gamma,$$

$$\sigma = -\frac{1}{2}l + \frac{1}{2}h, \qquad m = \mu + \frac{1}{3}\gamma,$$

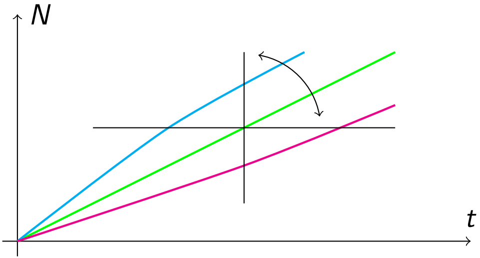
$$\gamma = \frac{3}{4}l - \frac{3}{2}m + \frac{3}{4} \qquad h = \mu + \sigma + \gamma.$$

Idea of the calculus

$$\begin{tikzcd}
\text{3pts input} \arrow[r, "\text{lin. transf.}"] & \text{moments input} \arrow[d, "\text{additivity}"'] \\
\text{3pts output} & \text{moments output} \arrow[l, "\text{lin. transf.}"]
\end{tikzcd}$$

Turn of modes

Summation of random variables works for subsequent events ("serial summation"). But we also need "parallel summation" for simultaneous signals. This would be the same when we are able to turn marginal distribution "time for mass N to a marginal distribution "mass for time t":



Fortunatelly, both the distributions "share quantiles", hence we can find easily the three-point estimations.

Conclusion

$-$:

- It is wrong.
- It is very wrong for exotic distrubutions, e. g. multimodal ones.

$+$:

- It is fast, simple, and stable.
- The engineers like the 3-point estimation and this is often the only knoledge about the process.
- The model works for discrete, continuous, or combined processes.