CSCC XIV Aug. 2015

Sand Pile Modeling of Economic Variables For Credit Risk Applications

João Pires da Cruz[1,2], Helena Cruz[1], Kanshukan Rajaratnam[3,4], Peter Beling[5], George A. Overstreet Jr.[6]

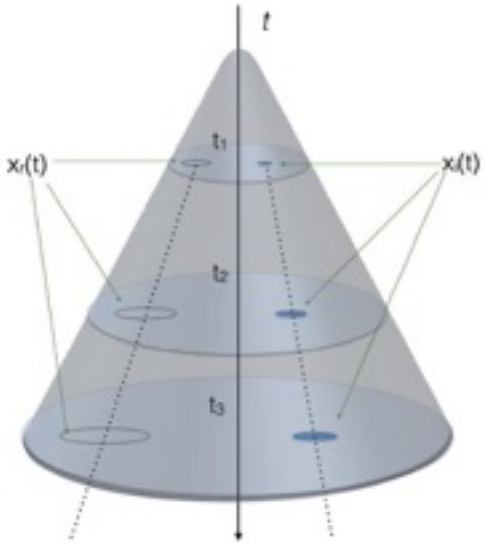
(1) Closer Consulting Ltd (UK)
(2) Center for Theoretical & Computational Physics, U. of Lisbon (Portugal)
(3) African Collaboration for Quant. Finance & Risk Research, U. of Cape Town (South Africa)
(4) Dep. of Finance & Tax, U. of Cape Town (South Africa)
(5) Department of Systems and Information Engineering, U. of Virginia(USA)
(6) McIntire School of Commerce, U. of Virginia(USA)

Outline

What is a Sand Pile Model?

What are the empirical evidences of it?

How it works on relevant variable reading



Non - Equilibrium

Suppose our system is changing the number of sides each instant we throw the dice.

We do not have a probability measure because our event space is changing.

“Equilibrium” is fundamental for any model.

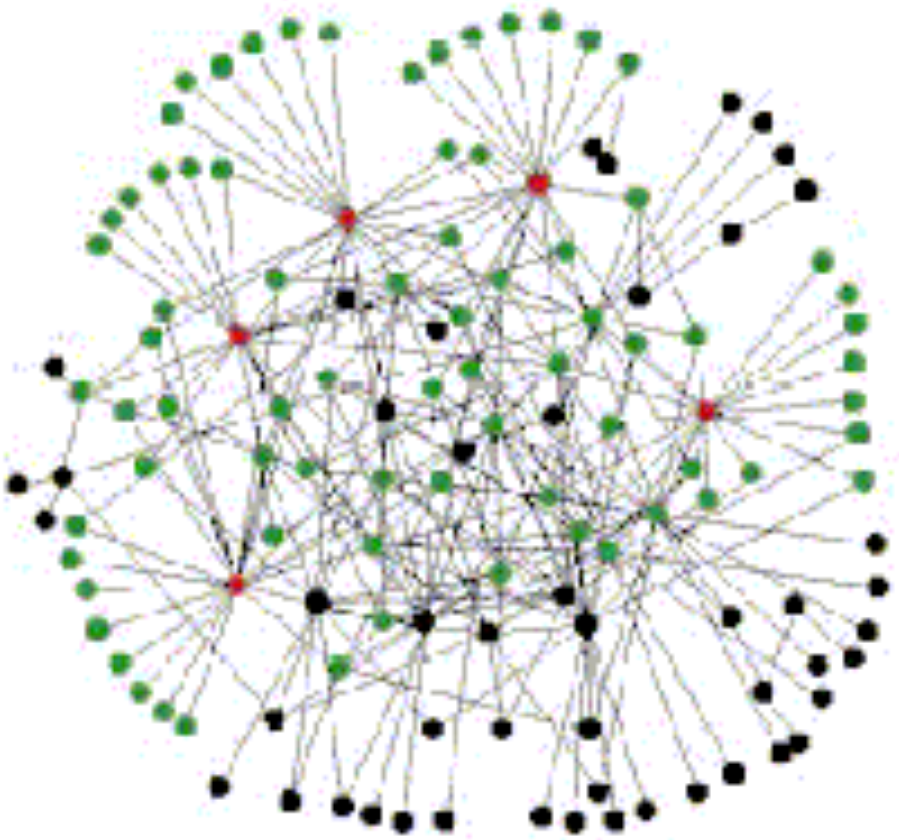
Source: http://justinandrewjohnson.com

Markovian

If the event space is changing then there is no basis for the definition of stochastic process.

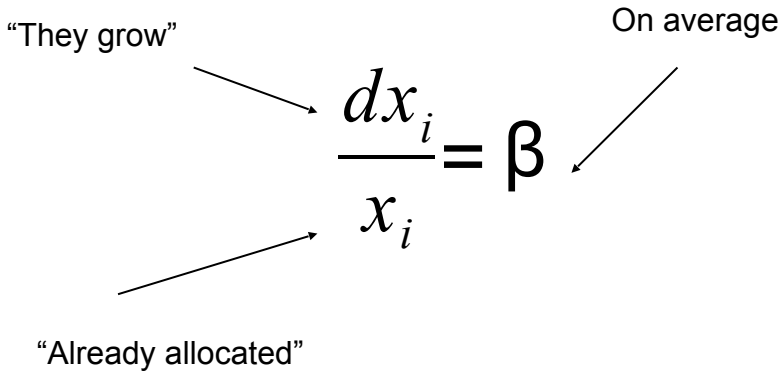
On each instant we have differen random variables / probability measures.

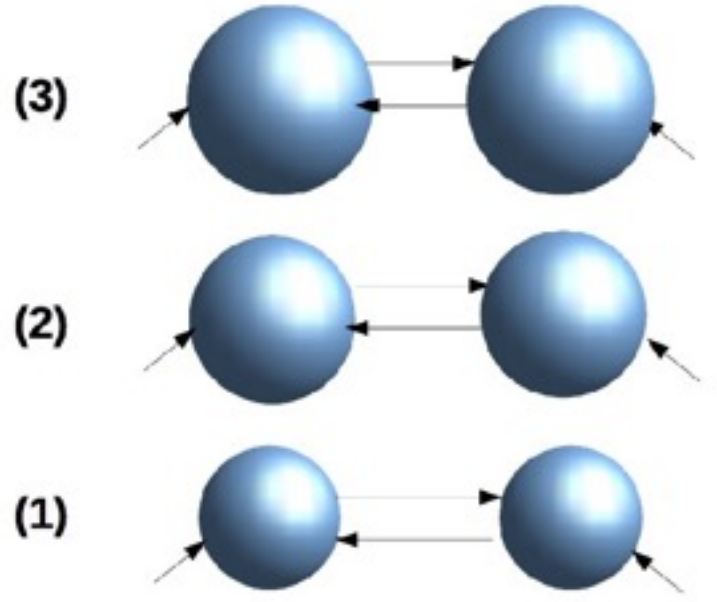
Is it an unsolvable problem?



Fundamentals

Economic agents status change by correlated growth. They grow with the resources already allocated to them.

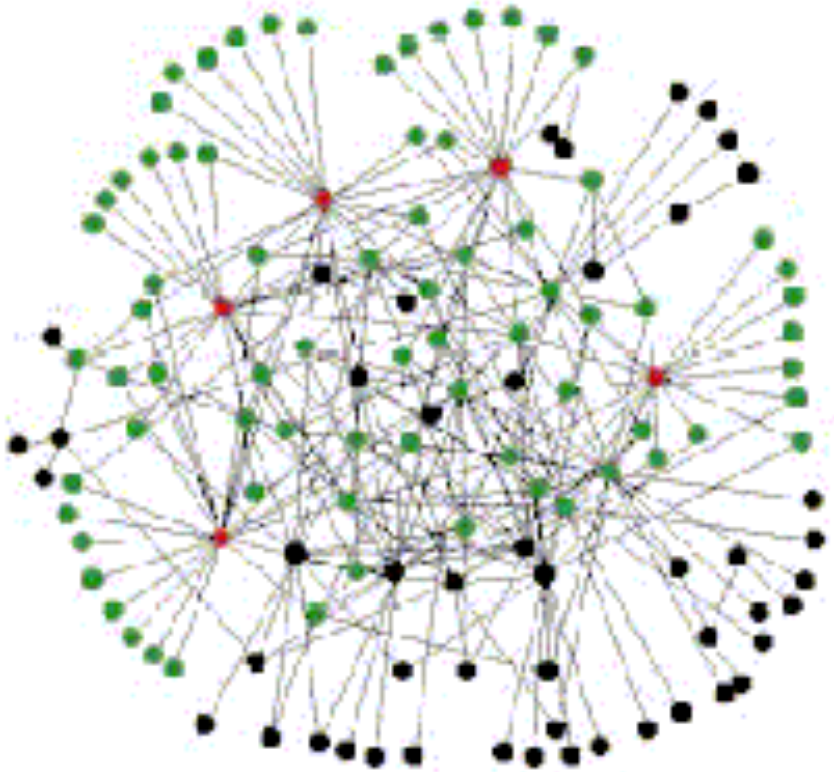




The name of the system is "Economy". And it has a geometry.

Why do we need a “geometry”?

We could represent an economy like a network. But there is no space, the length of the links do not make sense.

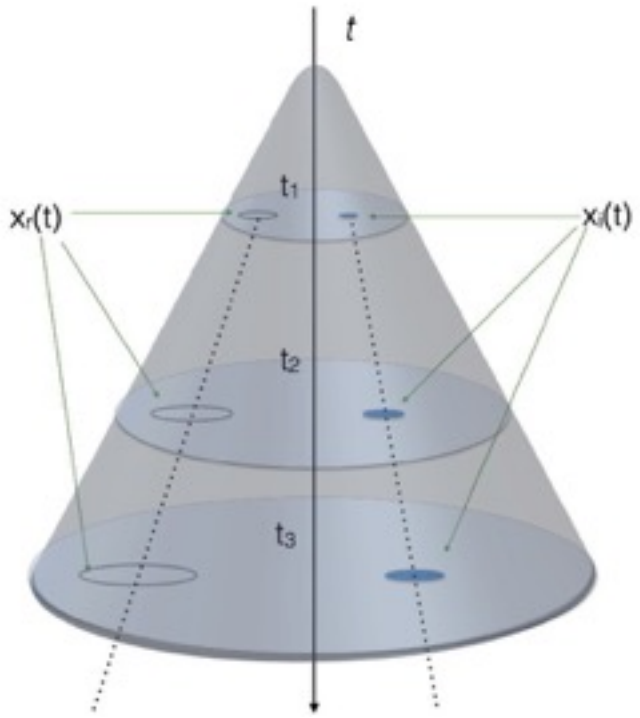


But there is a “geometry” in economic relations that is NOT something that we can look as Euclidean.

Geometry helps us to see if the mathematics we are using is adequate.

Sand Pile

Each economic link is characterized by point in a two dimensional space that inflates in time.



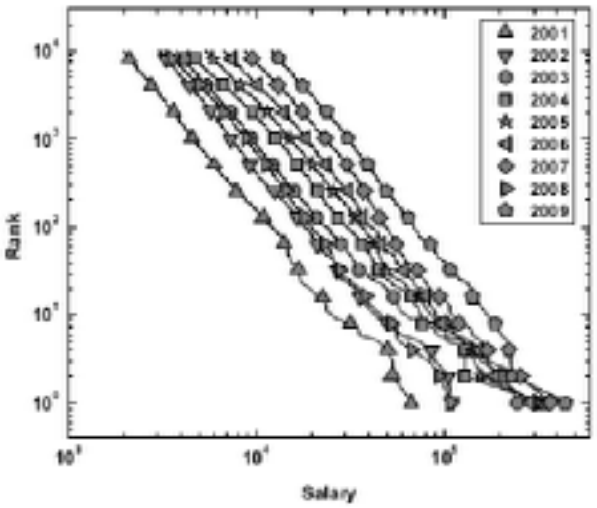
The economy fulfills the horizontal section of the pile as it is growing

$$\frac{dx^i}{x^i} = \frac{1}{\alpha}\frac{d\Lambda}{\Lambda}$$

1- Independent

2- Completely Symmetric

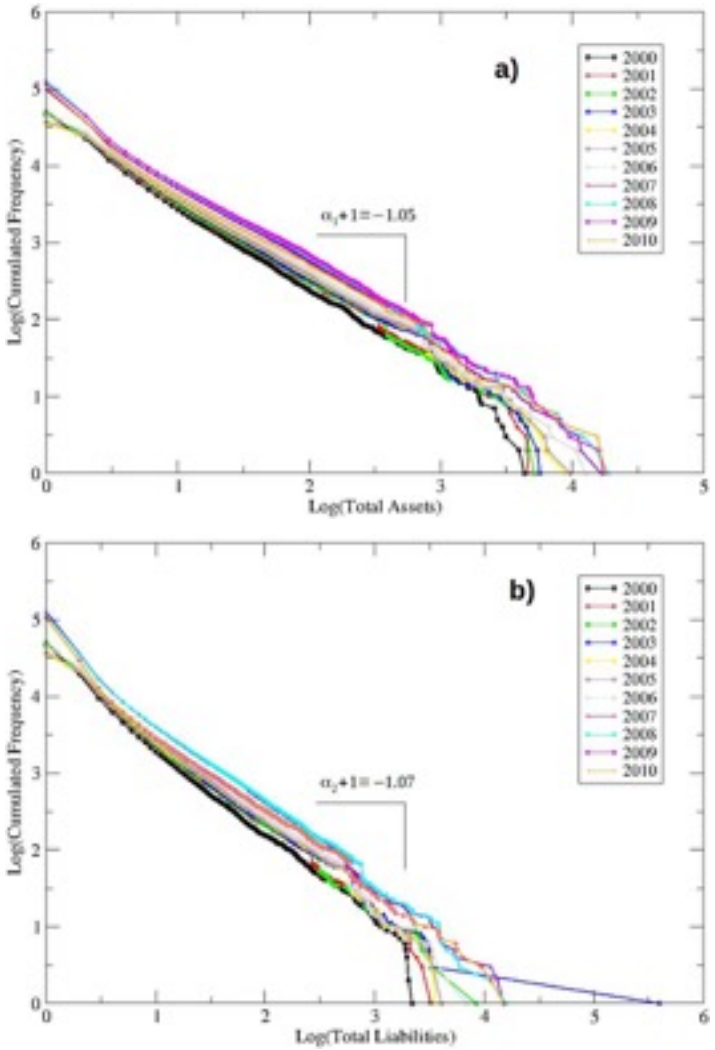
Empirical Evidences



N. Derzsy, Z. Néda,and MA Santos. Income distribution patterns from a complete social security database. Physica A: Statistical Mechanics and its Applications, 391(22): 5611–5619, 2012.

Individuals

Companies



$$p(x) = \alpha \frac{x_0^{\alpha}}{x^{\alpha+1}}$$

Check the paper for calcs.

Problem/Solution

The system is not an “equilibrium system”, the probability measure in one instant is not a probability measure in the the other

The variables have infinite variance! (the error of any learning process grows with the size of the sample)

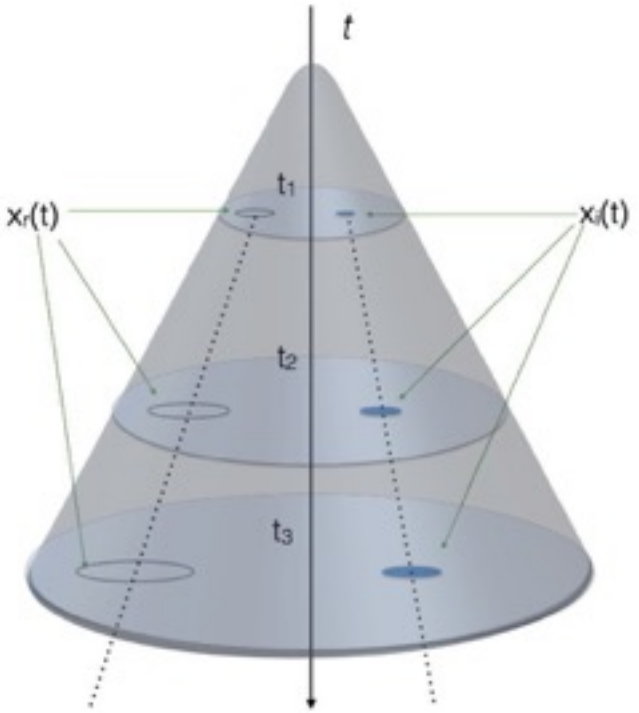
But now we have something to work with: the geometry is the invariant “quantity”

Solution

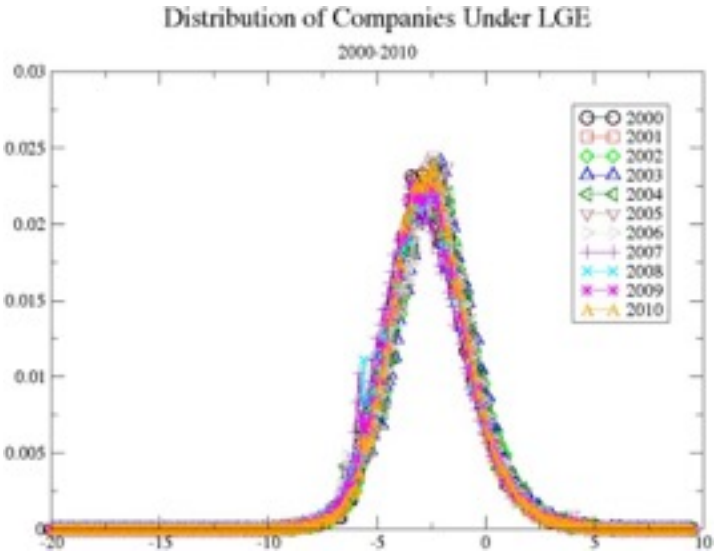
Not measuring with absolute gauges but with local ones! (LGE)

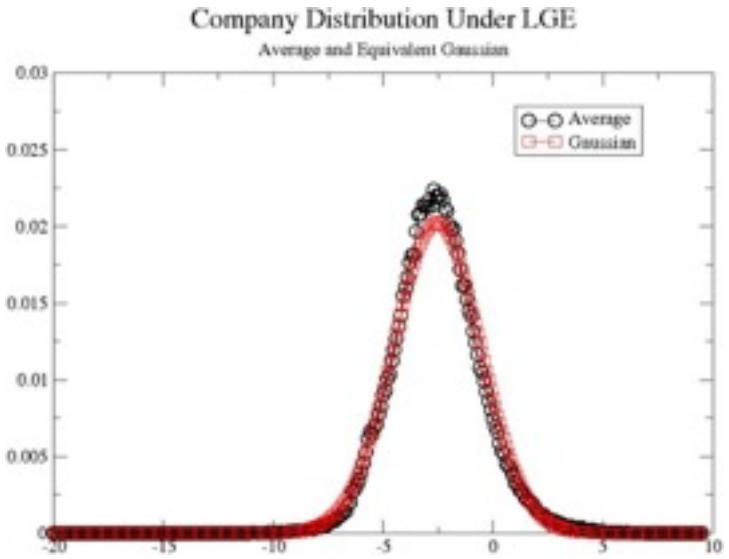
$$y = \frac{x^i}{x^r}$$

$$\frac{dy}{y} = 0$$



Converging Distributions

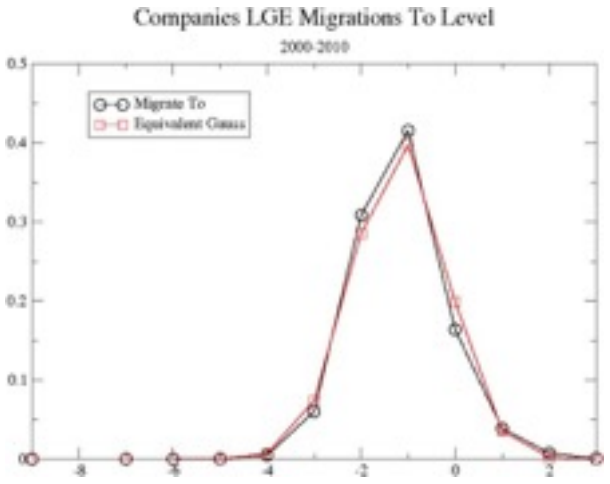


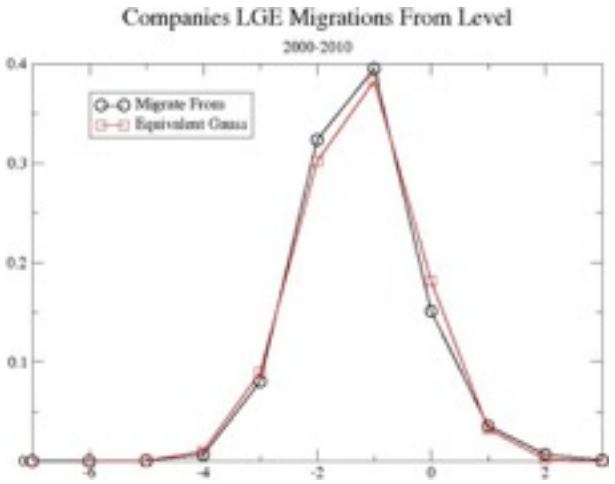


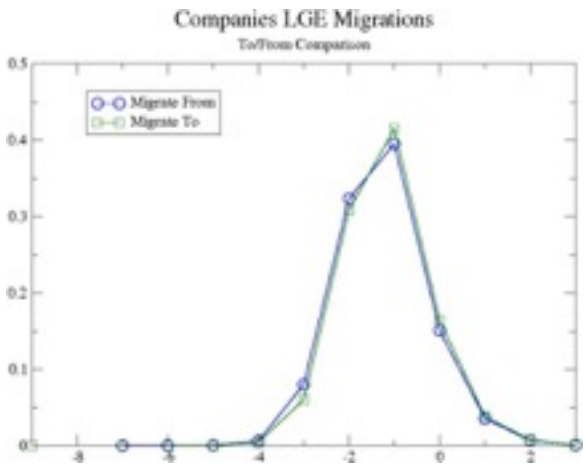
Same distributions we saw before

Under LGE transformation, distributions have finite variance and converge to Gaussian(CLT)

Gaussian Reversible Migrations







Migration matrices for ratings $R(x^1,x^2)$ based on the two dimensional sand pile become Gaussian and reversible.

$$dS = \mu dt + \sigma dW$$

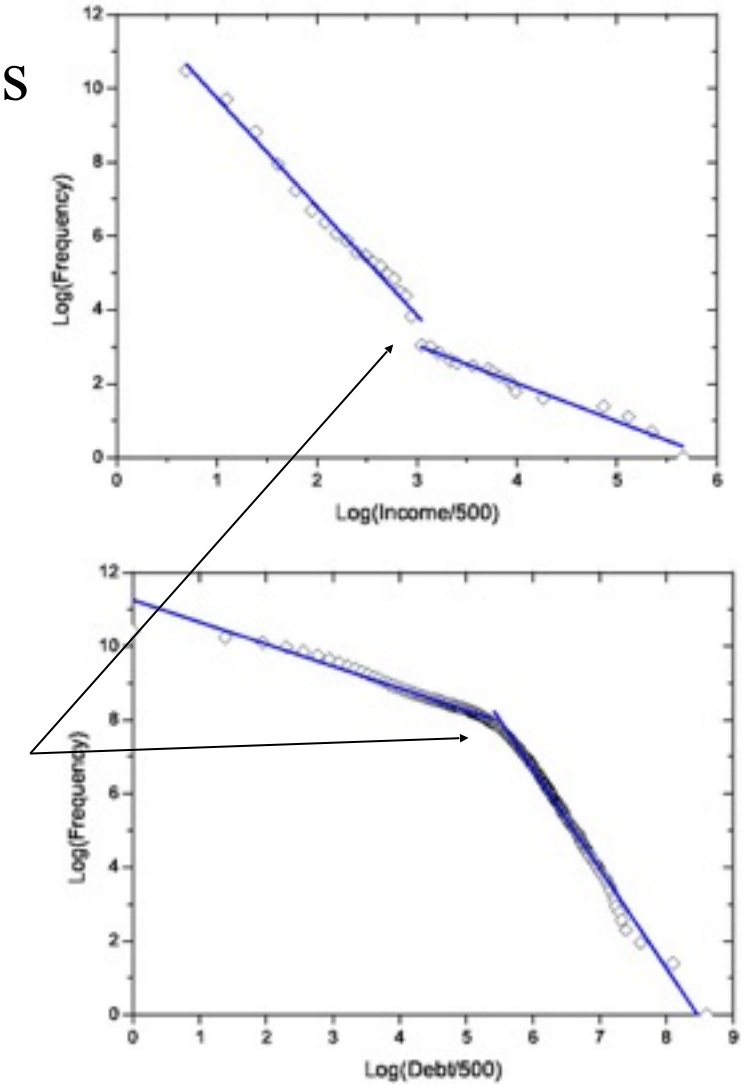
Conditional Probabilities

The problem started with 167 variables to identify the defaulting applicants.

Best result with Random Forest, but the almost everything was used.

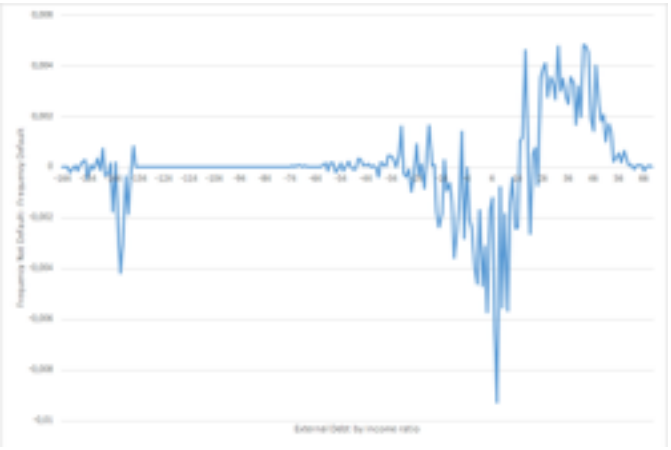
How a machine learning technique handles a conditional distribution?

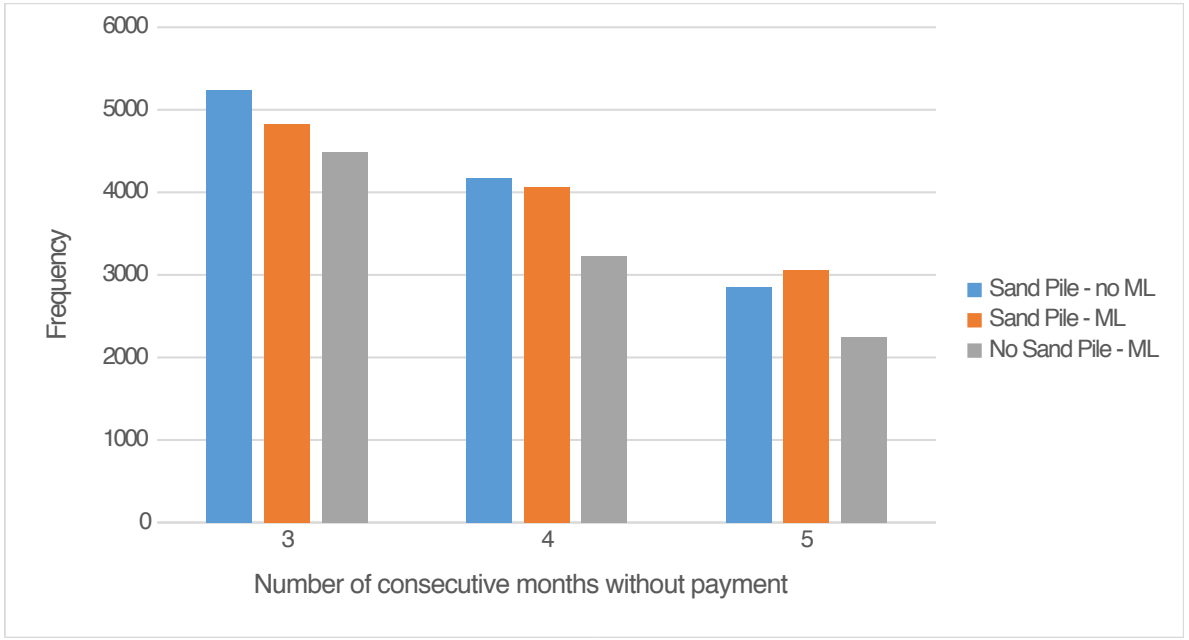
And the changes in the conditions? The infinite variance?



Credit Decision

Applying the model using just the two variables to reference the applicants in the economic space provides a separation better than the machine learning techniques.





“rich”

“poor”

Conclusions

- Stochastic, at last
- Stable, reversible, gaussian
- Learnable
- Reproducible