discrete_event_simulation

a practical example

dataroots
experts in data science

DSPT DAY 2019

contents

1. what?
2. why?
3. how?
4. does it fly?

dataroots
experts in data science

DSPT DAY 2019

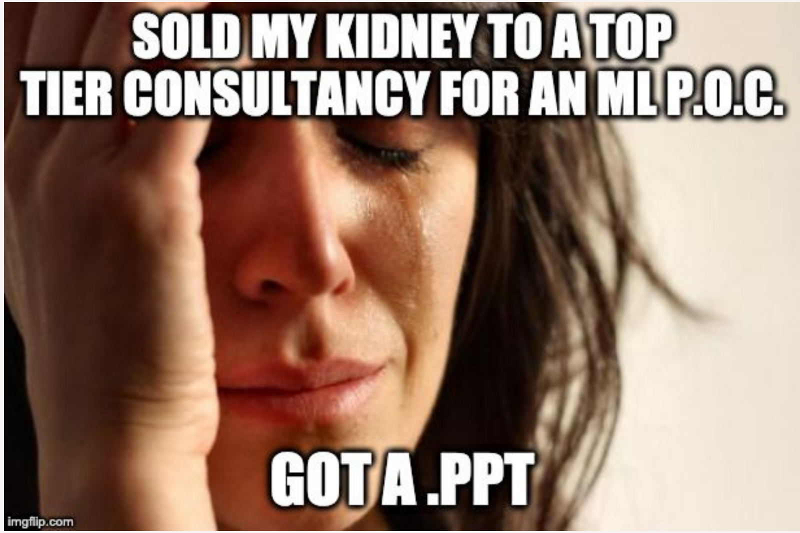
contents

dataroots
experts in data science

DSPT DAY 2019

we?

December 2016



"Hold my Belgian ale"

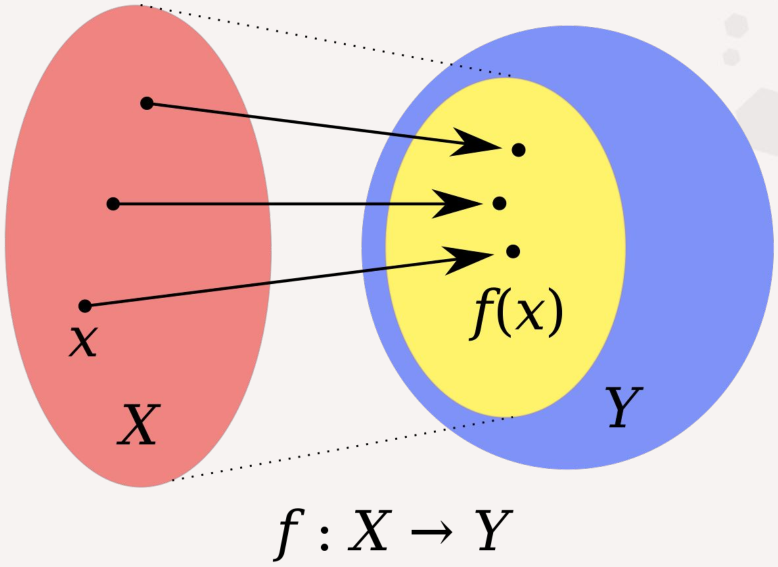
"When everything you have is a hammer, everything looks like a nail."

~ Unknown carpenter, 33 A.D.

dataroots
experts in data science

DSPT DAY 2019

95% of models used today are just “fancy mappings” (at best!)



Time?
States?
Entity interplay?
Resources?

Time?

Ingenious solution!

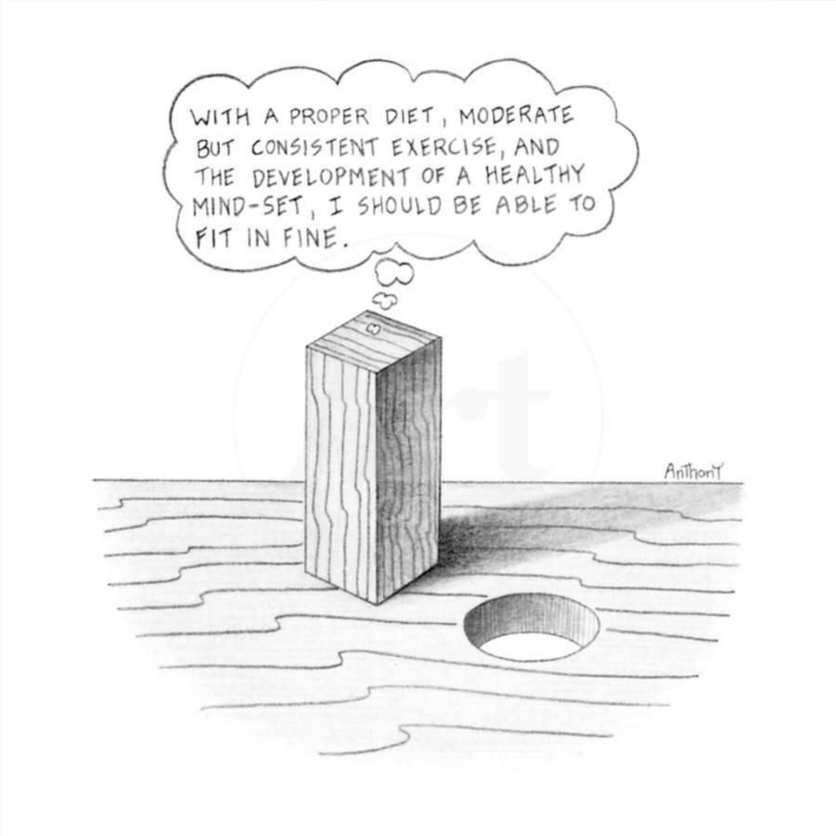
$X_t = [X_t, X_{t-1}, X_{t-2}]$

$\Rightarrow$ AR(I)MA(X) models

Still, different models for

$X_t \Rightarrow y_{t+1}, X_t \Rightarrow y_{t+5}, \ldots$

... can work, but a stretch



dataroots
experts in data science

DSPY 2019

Other issues

- How to encode prior/expert knowledge?
  - Forcing the model to learn something that can be parametrized in the process model is a waste.

- How to model complex interplay of entities and resources?
  - $X \Rightarrow y$ mappings are too naive for processes with multiple agents and resources.

- Model parameters “fused” to a specific process structure
  - How to test/predict outcomes of small structural changes in the process?

what if only...

... there existed a modelling approach

... and a library

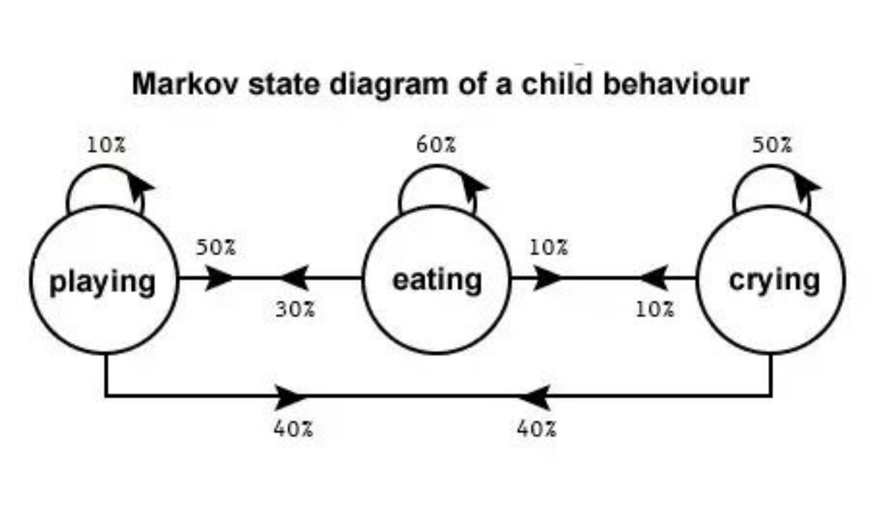
... that made all of this ~~simple~~ **less painful**?

dataroots
experts in data science

DSPT DAY 2019

enter: discrete event simulation (des)

- A **discrete-event simulation** (**DES**) models the operation of a system as a (discrete) sequence of events in time. Each event occurs at a particular instant in time and marks a change of state in the system



dataroots
experts in data science

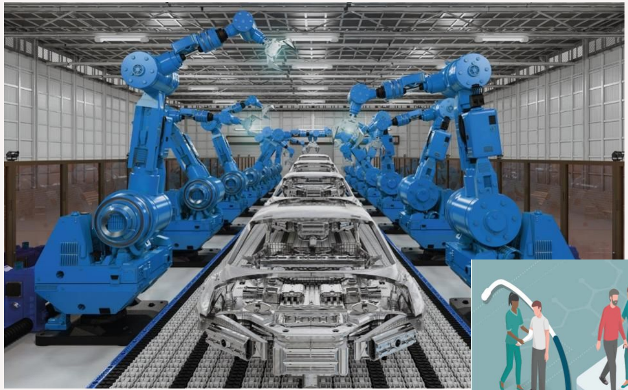
enter: discrete event simulation (des)

In other words: Replicating a real-world process in a virtual environment for the purpose of

- experimenting with different process inputs (e.g. working hours)
- better understanding the system behavior (e.g. bottlenecks)
- evaluating different scenarios to discover better processes (e.g. more resources)

dataroots
experts in data science

enter: discrete event simulation (des)



dataroots
experts in data science

how?

dataroots
experts in data science

process discovery

dataroots
experts in data science

Two options

1. Manual
   a. *Business analysis, interviews with stakeholders, data exploration*
2. Automated
   a. *Using Process Mining software*

Data based

General format is a table containing

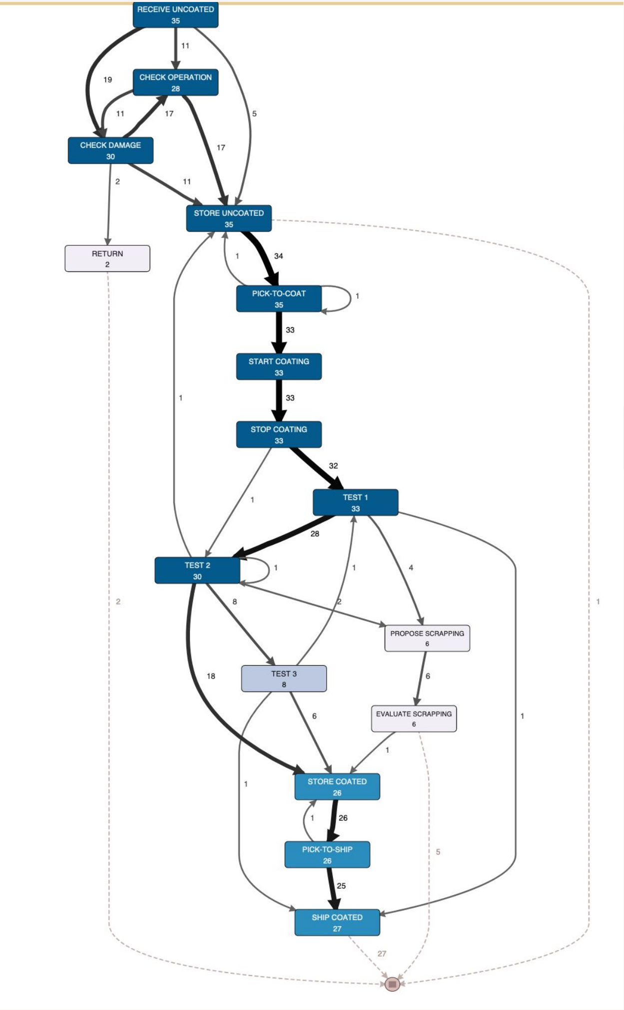
1. Entity of interest
2. Activity or resource occupied
3. Activity/resource start and end time

This allows us to reconstruct the trajectories.

Data

Case ID	Activity description	Timestamp	Resource	Location
Phone 3651	RECEIVE UNCOATED	2/03/2016 8:10	Arthur	INBOUND AREA
Phone 3651	CHECK OPERATION	2/03/2016 9:13	Arthur	INBOUND AREA
Phone 3651	CHECK DAMAGE	2/03/2016 9:25	Arthur	INBOUND AREA
Phone 3651	STORE UNCOATED	2/03/2016 10:08	Arthur	WAREHOUSE UNCOATED
Phone 3651	PICK-TO-COAT	15/03/2016 10:51	Jerome	WAREHOUSE UNCOATED
Phone 3651	START COATING	16/03/2016 15:14	Alix	COATING ROOM
Phone 3651	STOP COATING	16/03/2016 15:34	Alix	COATING ROOM
Phone 3651	TEST 1	17/03/2016 16:34	Edward	TESTING ROOM
Phone 3651	TEST 2	18/03/2016 10:34	Edward	TESTING ROOM
Phone 3651	TEST 3	18/03/2016 14:34	Edward	TESTING ROOM
Phone 3651	STORE COATED	18/03/2016 16:04	Jerome	WAREHOUSE COATED
Phone 3651	PICK-TO-SHIP	24/03/2016 9:33	Jerome	WAREHOUSE COATED
Phone 3651	SHIP COATED	24/03/2016 15:33	Jerome	OUTBOUND AREA

https://www.horsum.be/sites/default/files/wysiwyg-uploads/Process%20mining%20explained%20by%20an%20example%20-%20Episode%202.pdf



dataroots
experts in data science

DSPT DAY 2019

Tools

Closed source

- DISCO
- Logpickr
- Blueprism
- Fluxicon

Open source

- PM4Py
- Apromore
- ProM Tools

des modelling in r

dataroots
experts in data science

simmer

simmer is a process-oriented and trajectory-based Discrete-Event Simulation (DES) package for R. Designed to be a generic framework like [SimPy]() or [SimJulia](), it leverages the power of [Rcpp]() to boost the performance and turning DES in R feasible.

Developers

Iñaki Ucar

Author, copyright holder, maintainer

Bart Smeets

Author, copyright holder

dataroots

experts in data science

Very (very) short tutorial

1. Define trajectory
2. Create resources
3. Create trajectory generators
4. Run the simulation N-times
5. Fetch and analyze results

```r
library(simmer)

set.seed(1234)

bank <- simmer()
```

```r
customer <-
  trajectory("Customer's path") %>%
  log_("Here I am") %>%
  set_attribute("start_time", function() {now(bank)}) %>%
  seize("counter") %>%
  log_(function() {paste("Waited: ", now(bank) - get_attribute(bank, "start_time"))}) %>%
  timeout(12) %>%
  release("counter") %>%
  log_(function() {paste("Finished: ", now(bank))})
```

```r
bank <-
  simmer("bank") %>%
  add_resource("counter") %>%
  add_generator("Customer", customer, function() {c(0, rexp(4, 1/10), -1)})

bank %>% run(until = 400)
```

dataroots experts in data science

DSPT DAY 2019

Basic grammar

```
seize(<resource>, <quantity>)
timeout(<n_intervals>)
release(<resource>, <quantity>)

set_attribute(<name>, <numerical_value>)
get_attribute(<name>)

leave()
rollback(<n_steps>)
branch(<condition>, <branch_1>, <branch_2>, ...)
```

dataroots
experts in data science

DSPT DAY 2019

a real-world example

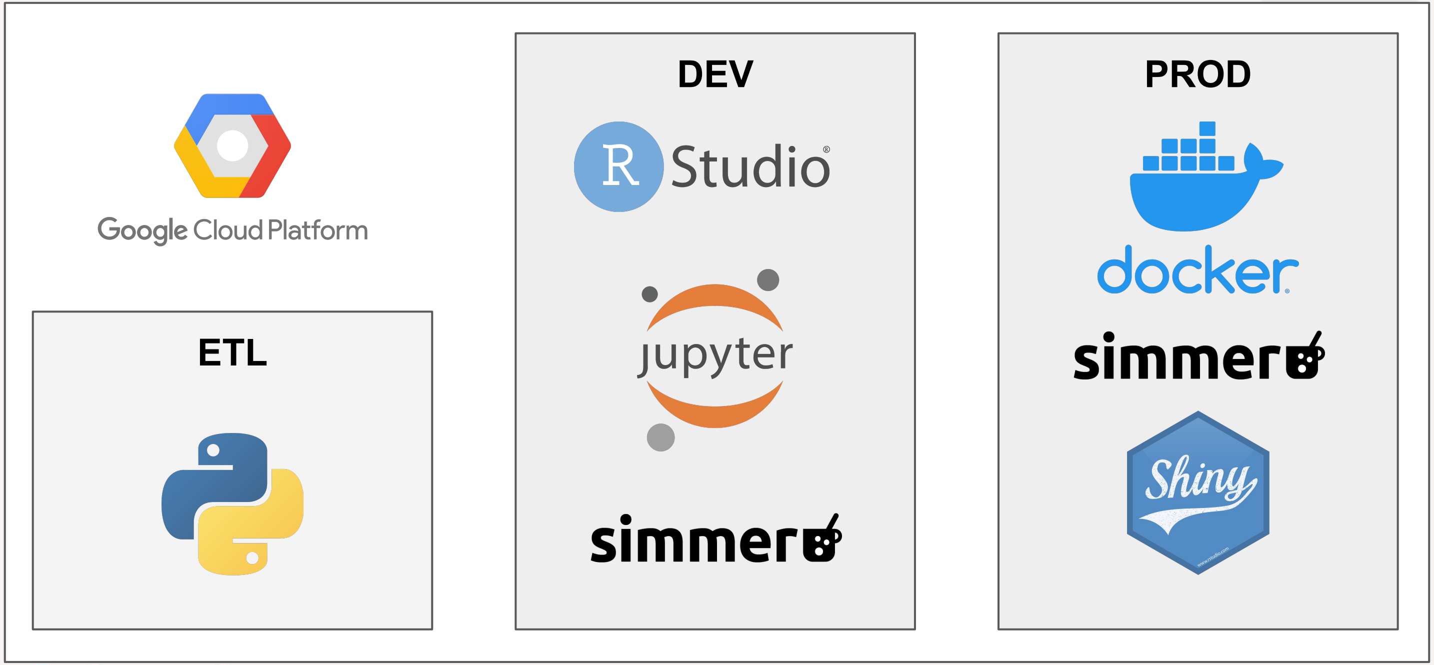
dataroots
experts in data science

The problem

- Our client is a the logistics department of a major brewery.
- Warehousing costs are very volatile and hard to predict using traditional BI approaches.
- Drivers of warehousing costs are
  - Volume of beer produced
  - Beer arrivals in and departures from warehouses
  - Total days all beers spent in each warehouse
- Movement profiles vary greatly by SKU

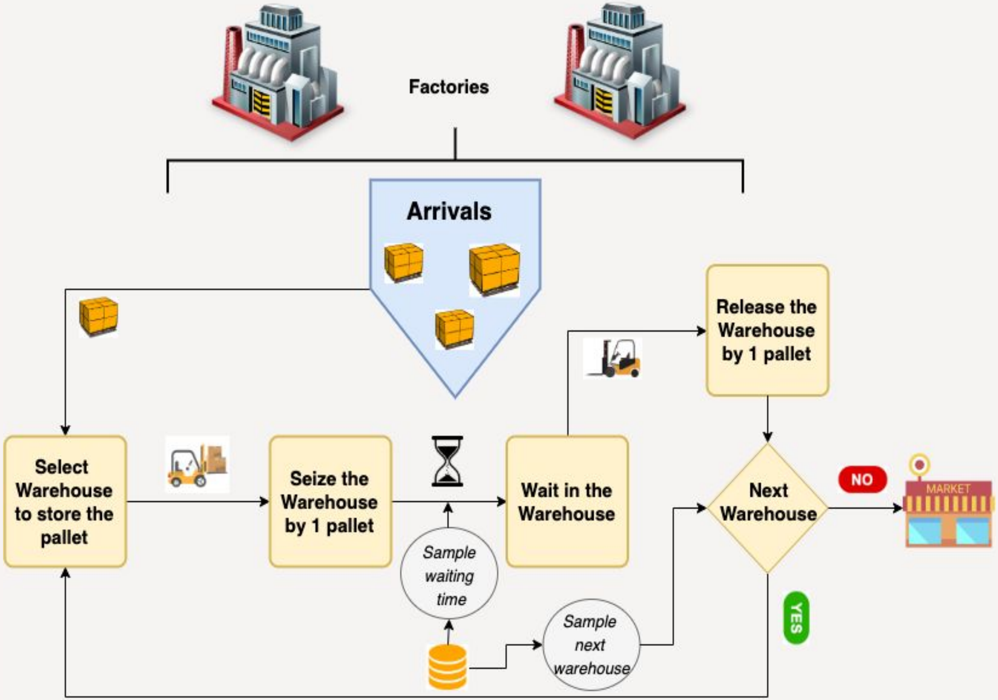
dataroots
experts in data science

PoC architecture



dataroots
experts in data science

Process diagram



Resource

- Warehouses

Attribute

- The capacity of warehouse

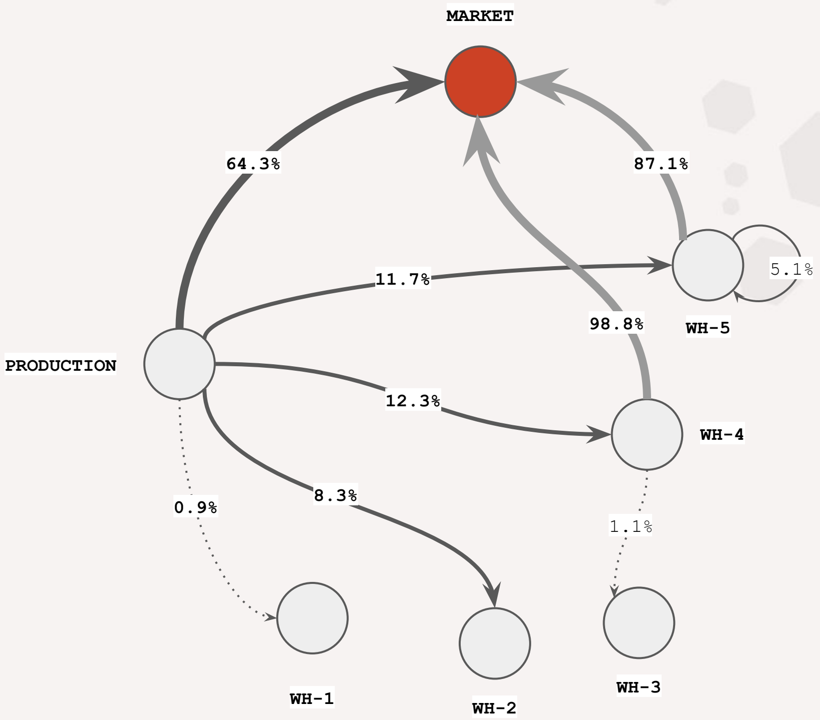
Variable

- Total stock in each warehouse

States

- Arrivals
- Selecting a warehouse
- Seizing the warehouse
- Waiting in the warehouse
- Releasing the warehouse

Parametrization: movement profile for SKU 123



dataroots
experts in data science

Performance

Key evaluation metric: 1 month cumulatives of:

- Moves into the warehouse (UNLOADS)
- Moves out of the warehouse (LOADS)
- Days spent in the warehouse (STORAGE DAYS)

Results:

- LOADS: 9% MAE
- UNLOADS: 8% MAE
- STORAGE DAYS: 4% MAE

dataroots
experts in data science

Conclusions

- DES allowed us to model the process in its “natural form”, leveraging the client’s business knowledge.
- Running N simulations in parallel allowed us to estimate worst and best case scenarios for a set of defined inputs (normally not available with classical models)

Improvement points

- Time-varying movement profiles (beer is very seasonal)

dataroots
experts in data science

recap

takeaways

- white/grey(ish) process modelling approach.
- requires process understanding.
- time consuming development, but granular output.
- ONE OF alternative approaches, not THE alternative -- be wise.

DISCLAIMER

- can be time consuming/overkill
    - Model design, documentation, validation, verification
- finite number of states / fixed graph
    - Not suitable for high-degree-of-freedom reinforcement learning

dataroots
experts in data science

thank you!

dataroots
experts in data science