G54SOD (Spring 2018)

Workshop 05

Representing Unpredictable Variability + Model Verification and Validation

Peer-Olaf Siebers

pos@cs.nott.ac.uk

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Representing Unpredictable Variability

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Recap: Data and Information

- Data Requirements
  - Preliminary or contextual data
  - Data for model realisation
  - Data for model validation

- Three categories of data
  - Category A:
    - Available
  - Category B:
    - Not available but collectable
  - Category C:
    - Not available and not collectable

Recap: Data and Information

- Data Requirements
  - Preliminary or contextual data
  - Data for model realisation
  - Data for model validation

- Three categories of data
  - Category A:
    - Available
  - Category B:
    - Not available but collectable
  - Category C:
    - Not available and not collectable

Representing Unpredictable Variability

- Modeller must determine how to present the variability that is present in each part of the model

- Three options
  - Traces (A or B)
  - Empirical distributions (A or B)
  - Statistical distributions (A or B or C)

Representing Unpredictable Variability

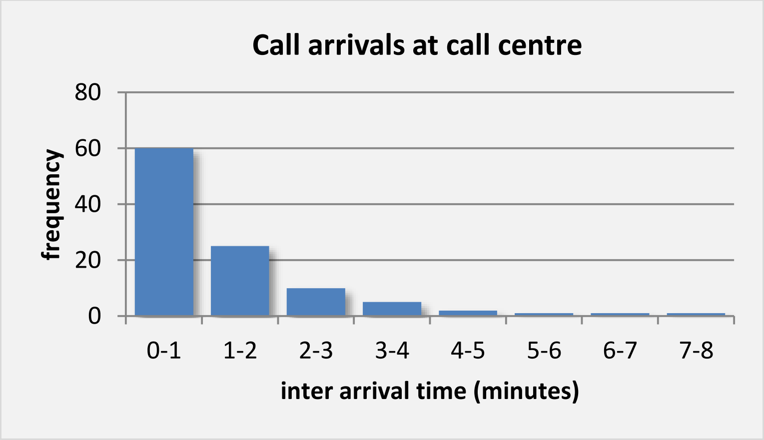
- Traces
  - Streams of data that describe a sequence of events
    - Data about the time the events occur (e.g. call arrival times)
    - Additional data about the event (e.g. call type)
  - Trace is read by the simulation as it runs and events are recreated in the model as described by the trace
  - Automatic monitoring systems are a common source of trace data

Call arrival time (minutes)	Call type
0.07	1
0.44	3
0.97	1
1.01	1
1.20	2
...	...

Representing Unpredictable Variability

- Empirical distributions
  - Show the frequency with which data values or ranges of values occur
  - Based on historical data, often formed by summarising trace data
  - As simulation runs values are sampled from the distribution
  - Most simulation software allows the user to enter empirical distribution data directly

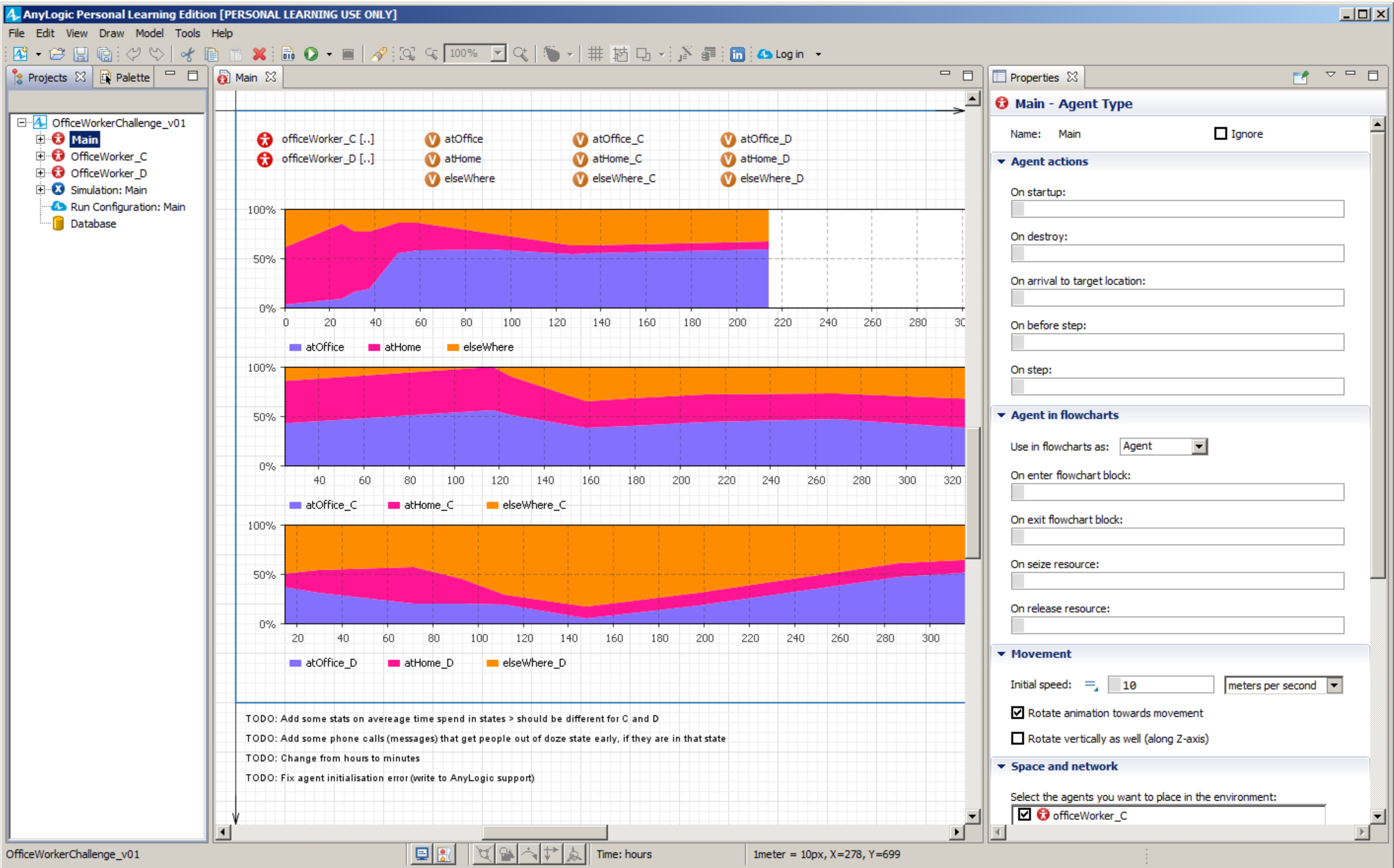
arrival time	inter-arrival time
0	-
1.5	1.5
2	0.5
4.1	2.1
4.2	0.1
7.5	3.3
7.8	0.3
8.2	0.4



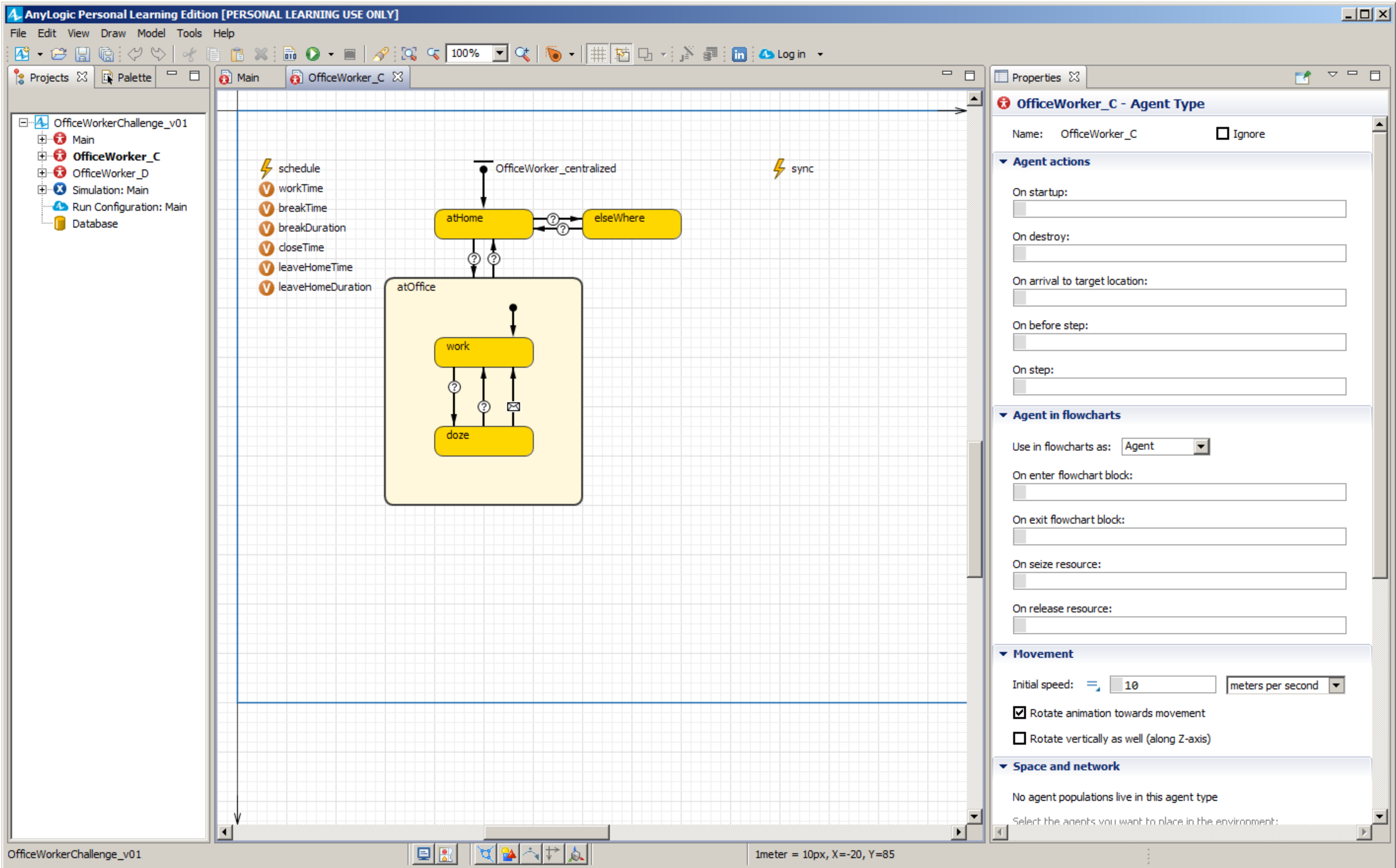
Challenge

- How can we create empirical distributions in AnyLogic?

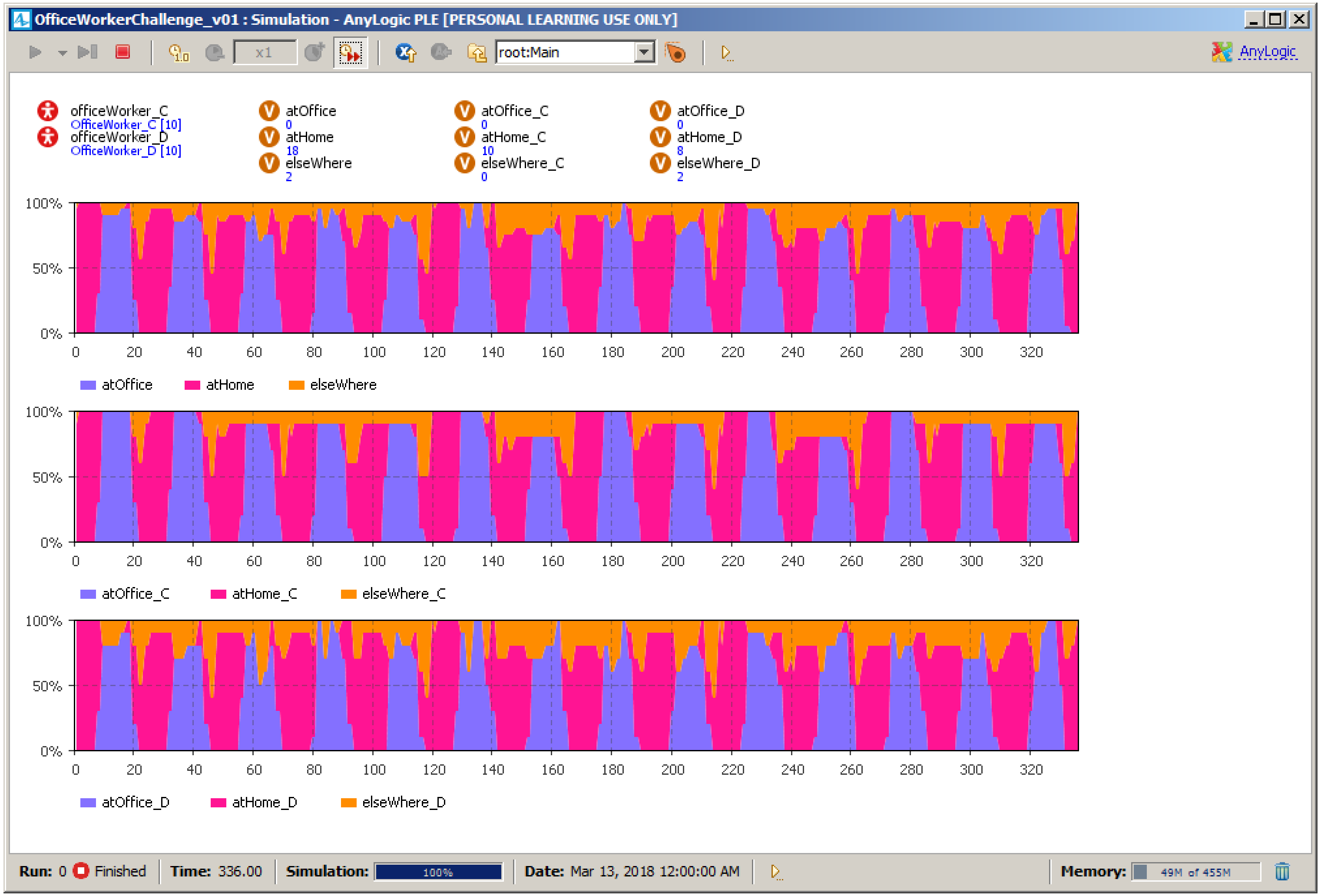
Previous Challenge



Previous Challenge



Previous Challenge



Representing Unpredictable Variability

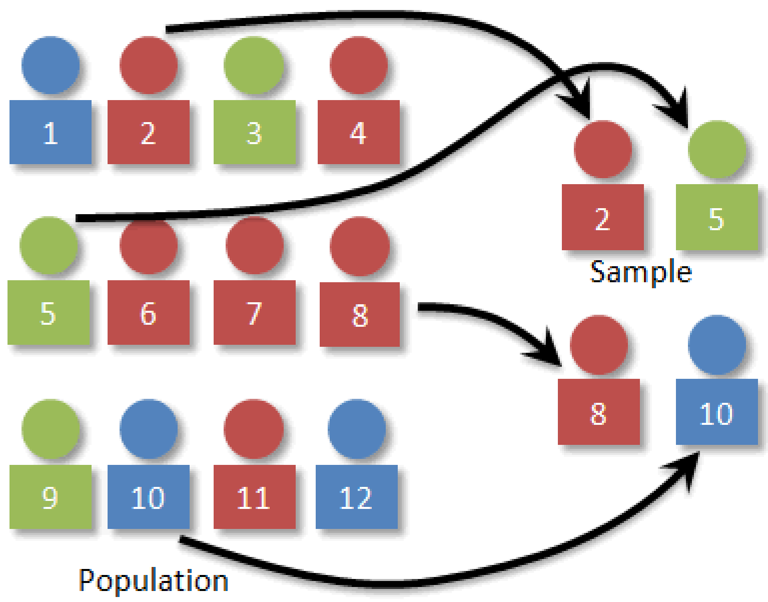
- Statistic distributions commonly used in simulation models
  - ***Continuous*** distributions Most Common?
    - For sampling data that can take any value across a range
      - Uniform distribution (continuous version)
      - Normal distribution
      - (Negative) exponential distribution
      - Erlang distribution
  - ***Discrete*** distributions Most common?
    - For sampling data that can take only specific values across a range, for instance only integer or non-numeric values
      - Uniform distribution (discrete version)
      - Binomial distribution
      - Poisson distribution

Representing Unpredictable Variability

- Statistic distributions commonly used in simulation models
  - ***Approximate*** distributions
    - Used in the absence of data
      - Discrete and continuous uniform distribution
      - Triangular distribution

For more details see Wikipedia or AnyLogic Help

Random Sampling



The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Random Sampling

- Sampling
  - Example: Booking clerk with one arrival process
    - 60% of customers: Personal enquirers (= type X)
    - 40% of customers: Phone callers (= type Y)
    - Top hat method:
      - 100 pieces of paper, 60 with X and 40 with Y
      - Every time a customer arrives draw one piece
      - Important to replace the paper to keep the ratio (60:40)
  - In computer simulation a similar principle is adopted based on random numbers

Random Sampling

- Random numbers:
  - Sequence of numbers that appear in a random order
  - Presented as integer (e.g. [0-99]) or as real numbers (e.g. [0-1[)
  - Top head method (replacement method)

  - Properties
    - ***Uniform:*** same probability of any number occurring at any point in the sequence

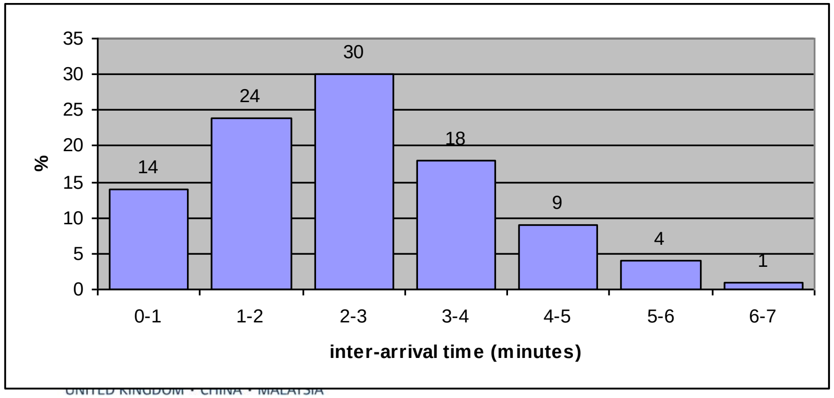
    - ***Independent:*** once a number has been chosen this does not effect the probability of it being chosen again or of another number being chosen

Random Sampling

- Relating random numbers to variability in a simulation
  - Modelling proportions
    - Example: Booking clerk with one arrival process
      - For small sample sizes ratio might not be achieved (as the process is random)
      - For large sample sizes ratio will be achieved more or less

Random numbers	Customer type
00-59	X
60-99	Y

  - Modelling variability in times
    - To model continuous real variables (e.g. activity times) one could determine the range and then draw a second random number, divide it by 100 and add it to the range

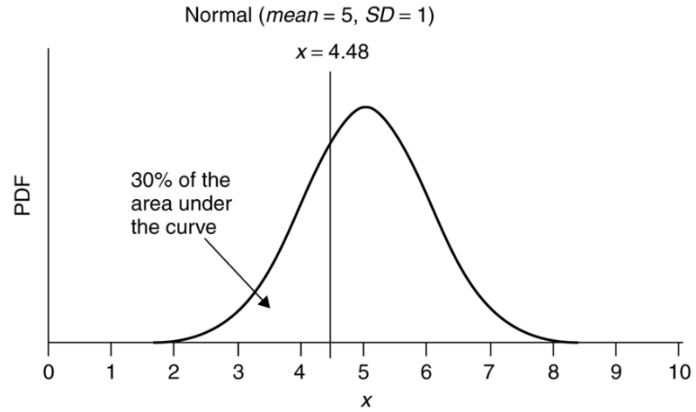


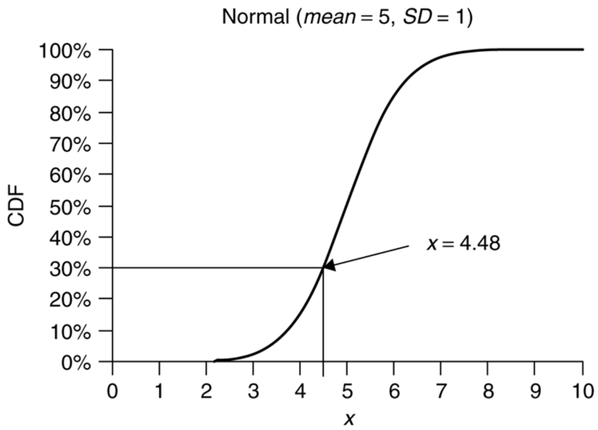
Random numbers	inter arrival time
00-13	0-1
14-37	1-2
38-67	2-3
68-85	3-4
86-94	4-5
95-98	5-6
99	6-7

Customer	rnd 1	range	rnd 2	inter arrival time
1	41	2-3	27	2.27
2	44	2-3	24	2.24
3	90	4-5	5	4.05

Random Sampling

- Random sampling (generating variates) from standard statistical distributions
  - To sample a value from the distribution the random number is taken to be the percentage of the area under the curve
    - Difficult to think in terms of identifying area under the curve
    - Instead of the probability density function (pdf) we use the cumulative distribution function (cdf)





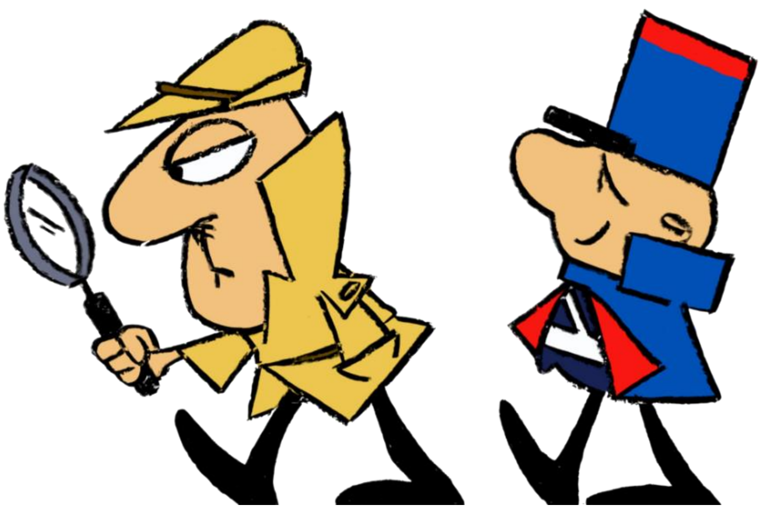
Random Sampling

- Computer generated random numbers
  - By nature computers do not behave in a random fashion
  - There are algorithms that give the appearance of producing random numbers although the results are completely predictable
    - Numbers produced by these algorithms (called pseudo random numbers) have the properties of uniformity and independence
  - Commonly used algorithm for generating random numbers:
    - $X_{i+1} = aX_i + c \pmod{m}$
      - X_i: Stream of random numbers (integer) on the interval (0, m-1)
      - a: Multiplier constant
      - c: Additive constant
      - m: modulus
      - X_0 = starting value for X = seed

Random Sampling

- Computer generated random numbers (cont.)
  - Issues to think about when using random number generators
    - ***All generators will eventually cycle***, i.e. they return to their starting point and generate the same sequence of random numbers again.
    - ***Cycles should be easily divided into streams*** which are non-overlapping random-number sub-sequences that are themselves astronomically long, and which can be addressed easily by the modeller.
  - For more details on these issues see Kelton (2009)
  - For more details on Random Number Generators see L'Ecuyer (2006)

Model Testing (Verification and Validation)



Introductory Remark

- In this lecture we focus on DES verification & validation
  - Robinson (2004) Chapter 10 and 12
- More about SD verification & validation
  - Barlas (1996)
- More about ABS verification & validation
  - Midgley et al (2007)

Verification and Validation

- Model testing (verification and validation)
  - Required to place confidence in a study's results

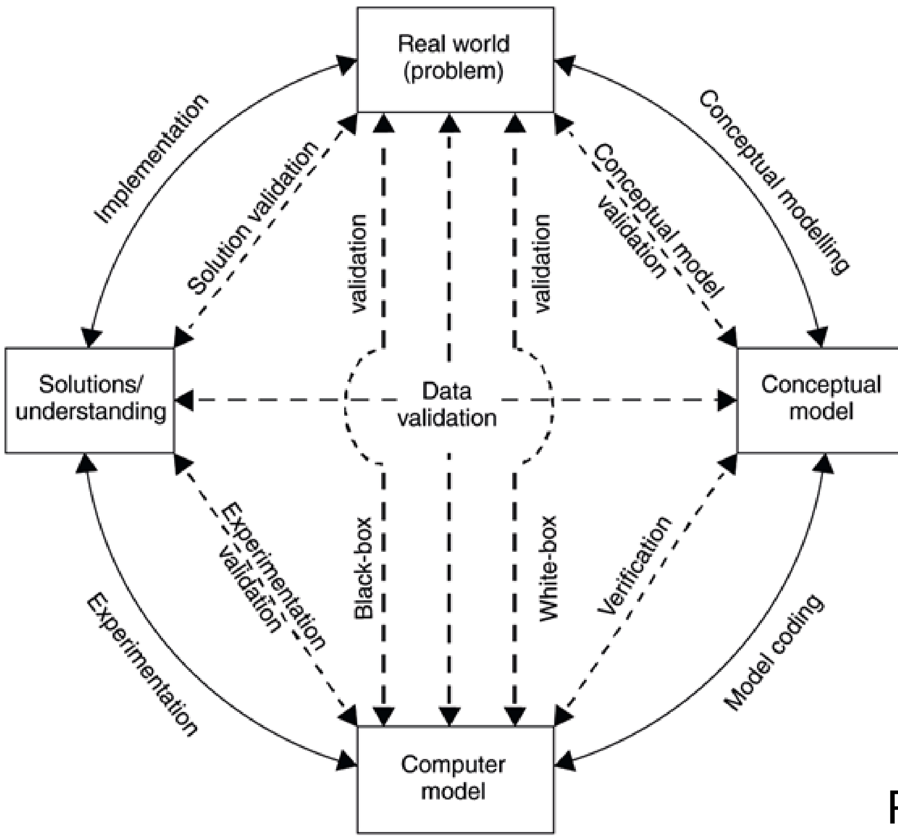
Model testing is not a process of trying to demonstrate that the model is correct but a process of trying to prove that the model is incorrect!

Verification and Validation

- Model verification and validation
  - **Model verification:** The process of ensuring that the model design has been transformed into a computer model with sufficient accuracy
  - **Model validation:** The process of ensuring that the model is sufficiently accurate for the purpose at hand
    - Models are not meant to be completely accurate
    - Models are supposed to be build for a specific purpose
- What is the difference between validity and accuracy?
  - Validity is a binary decision (conclusion: yes or no)
  - Accuracy is measured on a scale of 0 to 100%

Verification and Validation

- Conceptual model + data + white box + black box validation



Robinson (2004)

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Verification and Validation

- Conceptual Model Validation:
  - Determining that the content, assumptions and simplifications of the proposed model are sufficiently accurate for the purpose at hand.

- How can we do this?
  - Modeller should circulate model specification
  - Modeller and client should assess the assumptions and simplifications jointly

Verification and Validation

- Data Validation:
  - Determining that the contextual data and the data required for model realisation and validation are sufficiently accurate for the purpose at hand.

- How can we do this?
  - Modeller should investigate the source of data to determine their reliability

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Verification and Validation

- White-Box Validation:
  - Determining that the constituent parts of the computer model represent the corresponding real world elements with sufficient accuracy for the purpose at hand (micro check)

- How can we do this?
  - Checking the code
  - Visual checks; the following aspects should be checked:
    - Timings
    - Control of elements
    - Control of flows
  - Inspecting output reports

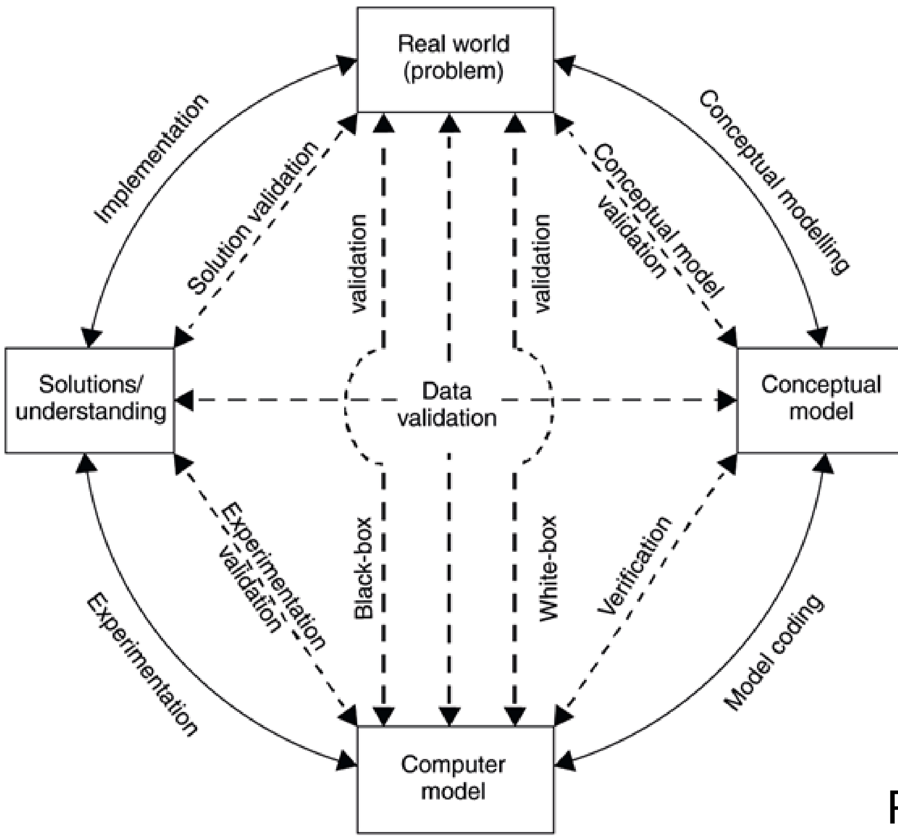
Verification and Validation

- Black-Box Validation:
  - Determining that the overall model represents the real world with sufficient accuracy for the purpose at hand (macro check)

- How can we do this?
  - Comparison with the real system
  - Comparison with other (usually simpler) models

Verification and Validation

- Experimentation + solution validation



Robinson (2004)

Verification and Validation

- Experimentation Validation:
  - Determining that the experimental procedures adopted are providing results that are sufficiently accurate for the purpose at hand.

- How can we do this?
  - Graphical or statistical methods for determining warm-up period, run length and replications (to obtain accurate results)
  - Sensitivity analysis (to improve the understanding of the model)

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Verification and Validation

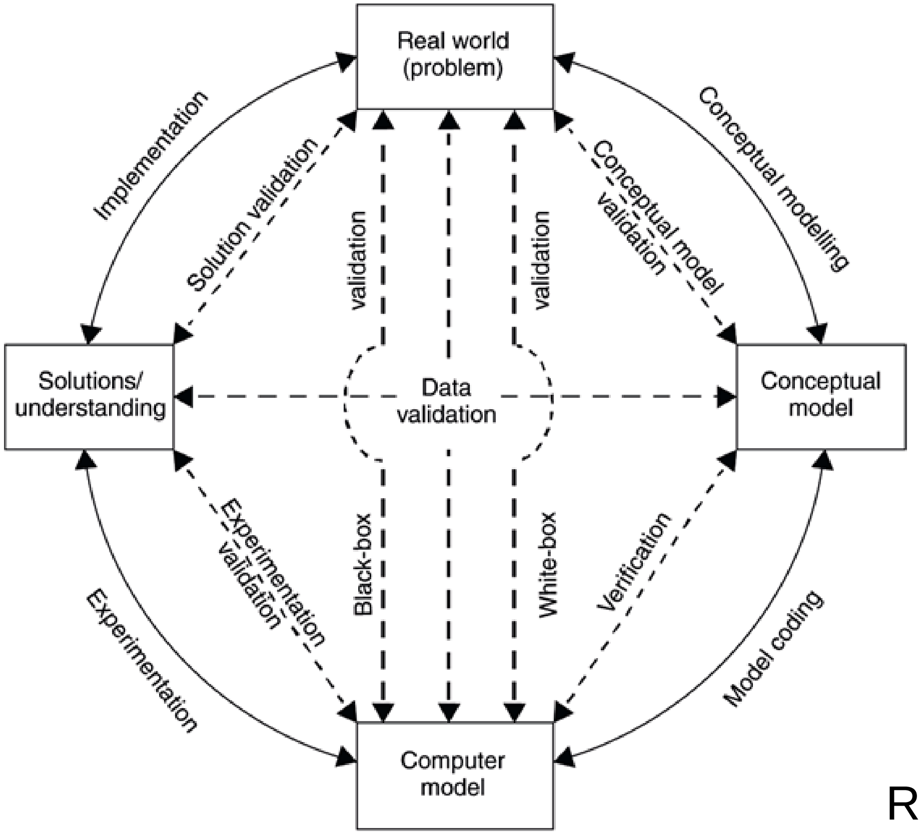
- Solution Validation:
  - Determining that the results obtained from the model of the proposed solution are sufficiently accurate for the purpose at hand

- How does this differ from Black Box Validation?
  - Solution validation compares the model of the proposed solution to the implemented solution while black-box validation compares the base model to the real world

- How can we do this?
  - Once implemented it should be possible to validate the implemented solution against the model results

Verification and Validation

- Verification



Robinson (2004)

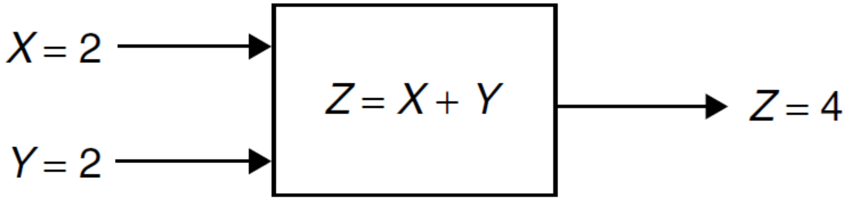
Verification and Validation

- Verification:
  - Testing the fidelity with which the conceptual model is converted into the computer model. Verification is done to ensure that the model is programmed correctly, the algorithms have been implemented properly, and the model does not contain errors, oversights, or bugs.

- How can we do this?
  - Same methods as for white-box validation (checking the code, visual checks, inspecting output reports) but …
    - Verification compares the content of the model to the conceptual model while white-box validation compares the content of the model to the real world

Verification and Validation

- Pitfalls:
  - Black-box validation: Is this a valid model?
    - Data has been obtained for two inputs (2,2) and one output (4)
    - A simple model is proposed: Z = X + Y



  - Relying on black-box validity alone can lead to the temptation to calibrate the model; in isolation it can lead to a simulation that is unrepresentative of the system it is trying to model

It is important to test both, white box and black box validity!

For more details see Robinson (2004)

Verification and Validation

- Difficulties of verification and validation
  - There is no such thing as general validity
    - A model is only valid with respect to its purpose
  - There may be no real world to compare against
  - Which real world?
    - Different people have different interpretations of the real world
  - Often real world data are inaccurate:
    - If the data are not accurate it is difficult to determine if the model's results are correct
    - Even if the data is accurate, the real world data are only a sample, which in itself creates inaccuracy
  - There is not enough time to verify and validate every aspect of a model

Verification and Validation

- Some final remarks:
  - V&V is a continuous and iterative process that is performed throughout the life cycle of a simulation study.
    - Example: If the conceptual model is revised as the project progresses it needs to be re-validated
  - V&V work together by removing barriers and objections to model use and hence establishing credibility.

- Conclusion:
  - Although, in theory, a model is either valid or not, proving this in practice is a very different matter. It is better to think in terms of confidence that can be placed in a model!

Questions / Comments

References

- Barlas (1996). Formal aspects of model validity and validation in system dynamics. System Dynamics Review 12(3), 183-210
- Midgley et al (2007). Building and assurance of agent-based models: An example and challenge to the field. Journal of Business Research 60(8), 884–893
- Robinson (2004). Simulation: The practice of model development and use. John Wiley & Sons: Chichester, UK