

# A Glimpse of Representing Stochastic Processes

Nathaniel Osgood

Using Modeling to Prepare for Changing  
Healthcare Needs

Duke-NUS

April 16, 2014

# Dynamic Uncertainty:

## Stochastic Processes

- Examples of things commonly stochastically approximated
  - Stock market
  - Rainfall
  - Oil prices
  - Economic growth
- What considered “stochastic” will depend on the scope of the model
  - Detailed model: Individual behaviour, transmission, differential severity of infection, etc.
  - A meteorological model may not consider rainfall stochastic

# Stochastic Processes in AnyLogic

- In AnyLogic, ABM and Discrete Event Models (“Network-Based Modeling”) are typically stochastic
  - Transitions between states
  - Event firing
  - Messages
    - (Frequent) timing of message send
    - Target of messages
  - Duration of a procedure
- As a result, there will be variation in the results from simulation to simulation

# Summarizing Variability

- To gain confidence in model results, typically need to run a “Monte Carlo” ensemble of realizations
  - Deal with means, standard deviations, and empirical fractiles
  - As is seen here, there are typically still broad regularities between most runs (e.g. rise & fall)
- Need to reason over a population of realizations
  - ⇒ statistics are very valuable
    - Fractile within which historic value falls
    - Mean difference in results between interventions

# Monte Carlo Methods in AnyLogic

- Monte Carlo methods draw repeated samples from distributions & stochastic processes of interest
- When running Monte Carlo method, we'd like to summarize the results of multiple runs
- One option would be to display each trajectory over time; downside: quickly gets messy
- AnyLogic's solution
  - Accumulate data regarding how many trajectories fall within given areas of value for a given interval of time using a "Histogram2D Data"
  - Display the Histogram2D Chart

# MonteCarlo2D Histogram

- Divides up time into user-specified # of intervals
  - This forms a set of divisions along the horizontal (time) axis
- Divides up value axis for quantity being displayed into a user-specified # of interval
  - This forms a set of divisions along the vertical (value) axis
- Together, the divisions define a uniform (2D) grid
  - For each cell on that grid, a “Histogram2D Data” object accumulates data regarding how many trajectories include a value within that cell
    - i.e. how many trajectories have hold a range of values during a given interval of time)



# Hands on Model Use Ahead



Load Sample Model:

**SIR Agent Based Calibration**

(Via “Sample Models” under “Help” Menu)

# Monte Carlo Analysis with Fixed Parameter Values

The screenshot displays the AnyLogic Professional software interface. The main workspace shows a simulation titled "Agent Based SIR Model - Monte Carlo Simulation". A button labeled "Run 100 replications" is visible. Below the title, a 2D histogram plot shows the distribution of simulation results, with the y-axis ranging from 6,000 to 8,000. The AnyLogic logo and the text "This model is © The AnyLogic Company. www.anylogic.com" are also present.

The left sidebar shows the "Projects" panel with a tree view containing:

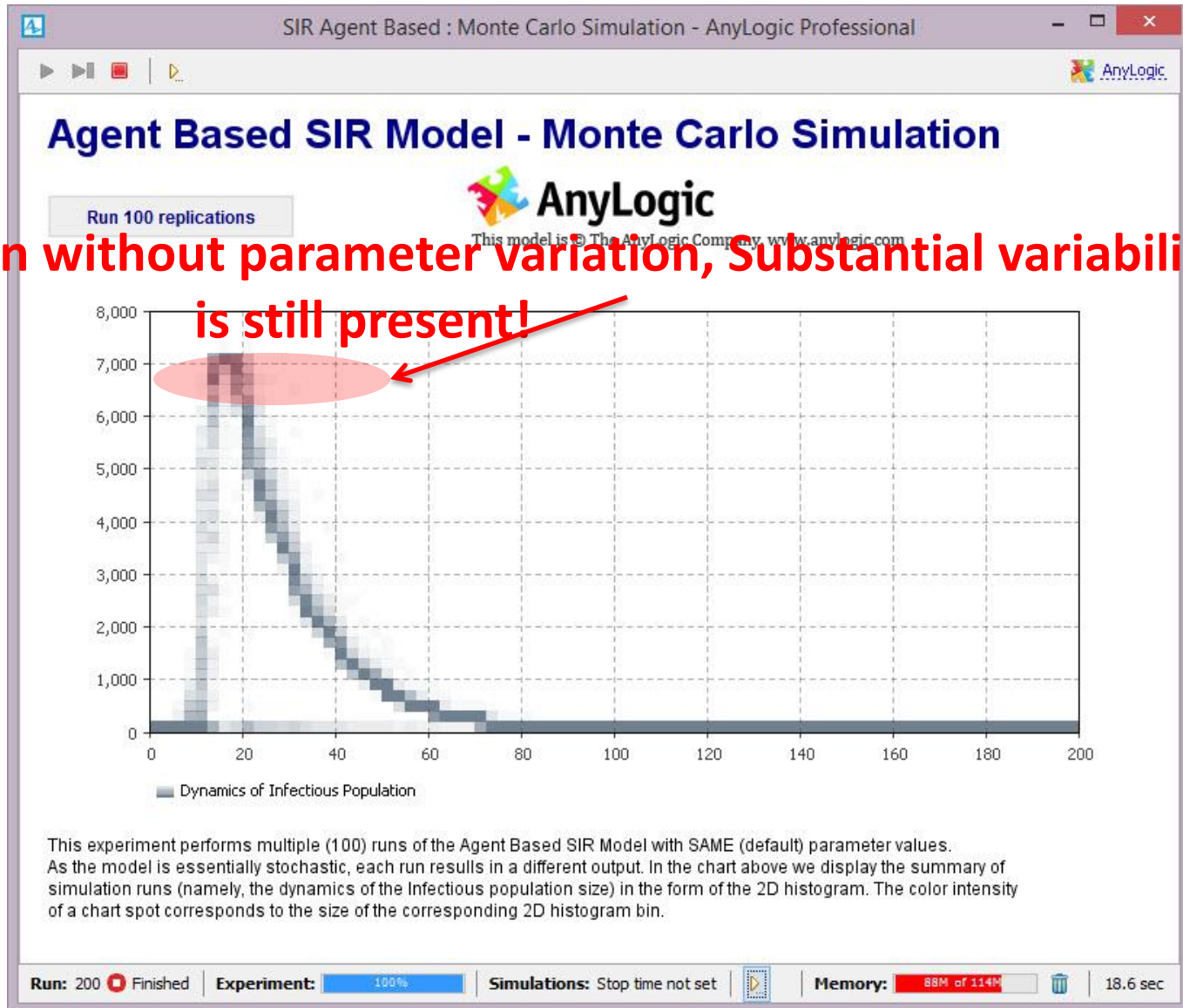
- SIR Agent Based Calibration
  - Main
  - Person
  - Calibration: Main
  - MonteCarlo2DHistogram: Main

The bottom panel shows the "Properties" section for "MonteCarlo2DHistogram - Parameter Variation Experiment". The "Name" field is set to "MonteCarlo2DHistogram" and the "Top-level agent" is set to "Main". The "Maximum available memory" is set to "128 Mb". The "Stop" condition is set to "Stop at specified time".

The status bar at the bottom indicates "Time units: days" and "X=...40".



# Results of Monte Carlo Simulation



# 2D Histogram Data

The screenshot displays the AnyLogic Professional interface. The main workspace shows a diagram of an "Agent Based SIR Model - Mo" with a "Run 100 replications" button and a "dataInfectious2D" data source. The left sidebar shows the project structure, including "SIR Agent Based Calibration", "Main", "Person", "Calibration: Main", and "MonteCarlo2DHistogram: Main". The bottom panel shows the "dataInfectious2D - Histogram2D Data" configuration, which includes settings for the horizontal and vertical axes, envelopes, and the number of intervals for the X and Y axes.

**AnyLogic Professional**

File Edit View Draw Model Tools Help

Projects

- SIR Agent Based Calibration
  - Main
  - Person
  - Calibration: Main
  - MonteCarlo2DHistogram: Main

MonteCarlo2DHistogram

Agent Based SIR Model - Mo

Run 100 replications

dataInfectious2D

Any  
This model is © TI

Properties Progress

**dataInfectious2D - Histogram2D Data**

visible: ☐ no

Horizontal axis value:

Vertical axis value:

Envelopes:

**X-axis values range**

Number of intervals:

Range, from:

to:

**Y-axis values range**

Number of intervals:

Range, from:

to:

Time units: days

# Important Distinction (Declining Order of Aggregation)

- Experiment
  - Collection of simulations
- Simulation
  - Collection of replications that can yield findings across set of replications (e.g. mean value)
- Replication
  - One run of the model

# Flexibility Typically Ignored

- In most AnyLogic models, an Experiment is composed of a single Simulation, which is composed of a single Replication
- In most AnyLogic models which run “ensembles” of realizations, a simulation is composed of only a single realization

# Accumulating the Histogram2D dataset from other datasets

The screenshot displays the AnyLogic Professional interface for a model named "MonteCarlo2DHistogram". The main workspace shows a diagram with a "dataInfectious2D" component and a "Run 100 replications" button. The "Properties" panel on the left shows the "MonteCarlo2DHistogram" model selected. The "Java actions" section is expanded, showing the following configuration:

- Initial experiment setup:
- Before each experiment run:  
`dataInfectious2D.reset();`
- Before simulation run:
- After simulation run:  
`dataInfectious2D.add( root.InfectiousDS );`
- After iteration:
- After experiment:

Two annotations are present:

- A blue arrow points from the text "The accumulating Histogram2D dataset is in Experiment" to the "After simulation run" section.
- A red arrow points from the text "The source dataset is in Main" to the `root.InfectiousDS` variable in the "After simulation run" section.

The bottom status bar indicates "Time units: days".

# Monte Carlo Sensitivity Analyses in AnyLogic

The screenshot displays the AnyLogic Professional interface. The main workspace shows a 2D histogram titled "Agent Based SIR Model - Mo" with a "Run 100 replications" button. The histogram plots data from "dataInfectious2D". A red text overlay reads: "Choice between showing envelopes of empirical fractiles & showing counts in histogram bins". A red arrow points from this text to the "Show bins" radio button in the "chart - Histogram2D" properties panel. The properties panel also includes a "Name" field set to "chart", an "Ignore" checkbox, and a "Data" section with a "Title" field set to "Dynamics of Infectious Popu", a "Histogram" field set to "dataInfectious2D", and a "Color" dropdown set to "slateGray". The bottom status bar indicates "Time units: days".

AnyLogic Professional

File Edit View Draw Model Tools Help

Projects

- SIR Agent Based Calibration
  - Main
  - Person
  - Calibration: Main
  - MonteCarlo2DHistogram: Main

MonteCarlo2DHistogram

Agent Based SIR Model - Mo

Run 100 replications

Choice between showing envelopes of empirical fractiles & showing counts in histogram bins

chart - Histogram2D

Name: chart ☐ Ignore

☐ Show envelopes

☒ Show bins

Data

Title: Dynamics of Infectious Popu

Histogram: dataInfectious2D

Color: slateGray

Time units: days

# Difference Between Chart Options

## “Show envelopes”

- This option shows **envelopes of empirical fractiles**
  - These are associated with empirical fractiles defined in terms of percentages (e.g. “0.25” means boundary between lowest and 2<sup>nd</sup> lowest quartile; “0.50” means median)
  - e.g. These define envelopes of (contours) around the median within which data from different % of realizations fall
  - A “slice” through the output at a particular moment in time would be like an **extended boxplot** (showing fractiles)
- The empirical fractiles to use are themselves defined in the associated Histogram2D Data object



# Reminder: 2D Histogram Data

Note definition of envelopes to be used in The Histogram2D Chart if "Show envelopes" is selected.

The screenshot displays the AnyLogic Professional software interface. The main workspace shows a chart titled "Agent Based SIR Model - Monte Carlo 2D Histogram". A red arrow points from the text "Note definition of envelopes to be used in The Histogram2D Chart if 'Show envelopes' is selected." to the "Envelopes" field in the "dataInfectious2D - Histogram2D Data" properties panel. The "Envelopes" field is highlighted with a red oval and contains the values "0.25, 0.5, 0.75". The properties panel also shows fields for "Horizontal axis value", "Vertical axis value", "X-axis values range" (Number of intervals: 80, Range from: 0 to: 200), and "Y-axis values range" (Number of intervals: 40, Range from: 0 to: 8000). The status bar at the bottom indicates "Time units: days".

AnyLogic Professional

File Edit View Draw Model Tools Help

Projects

- SIR Agent Based Calibration
  - Main
  - Person
  - Calibration: Main
  - MonteCarlo2DHistogram: Main

MonteCarlo2DHistogram

dataInfectious2D

Run 100 replications

dataInfectious2D - Histogram2D Data

visible: no

Horizontal axis value:

Vertical axis value:

Envelopes: 0.25, 0.5, 0.75

X-axis values range

Number of intervals: 80

Range, from: 0

to: 200

Y-axis values range

Number of intervals: 40

Range, from: 0

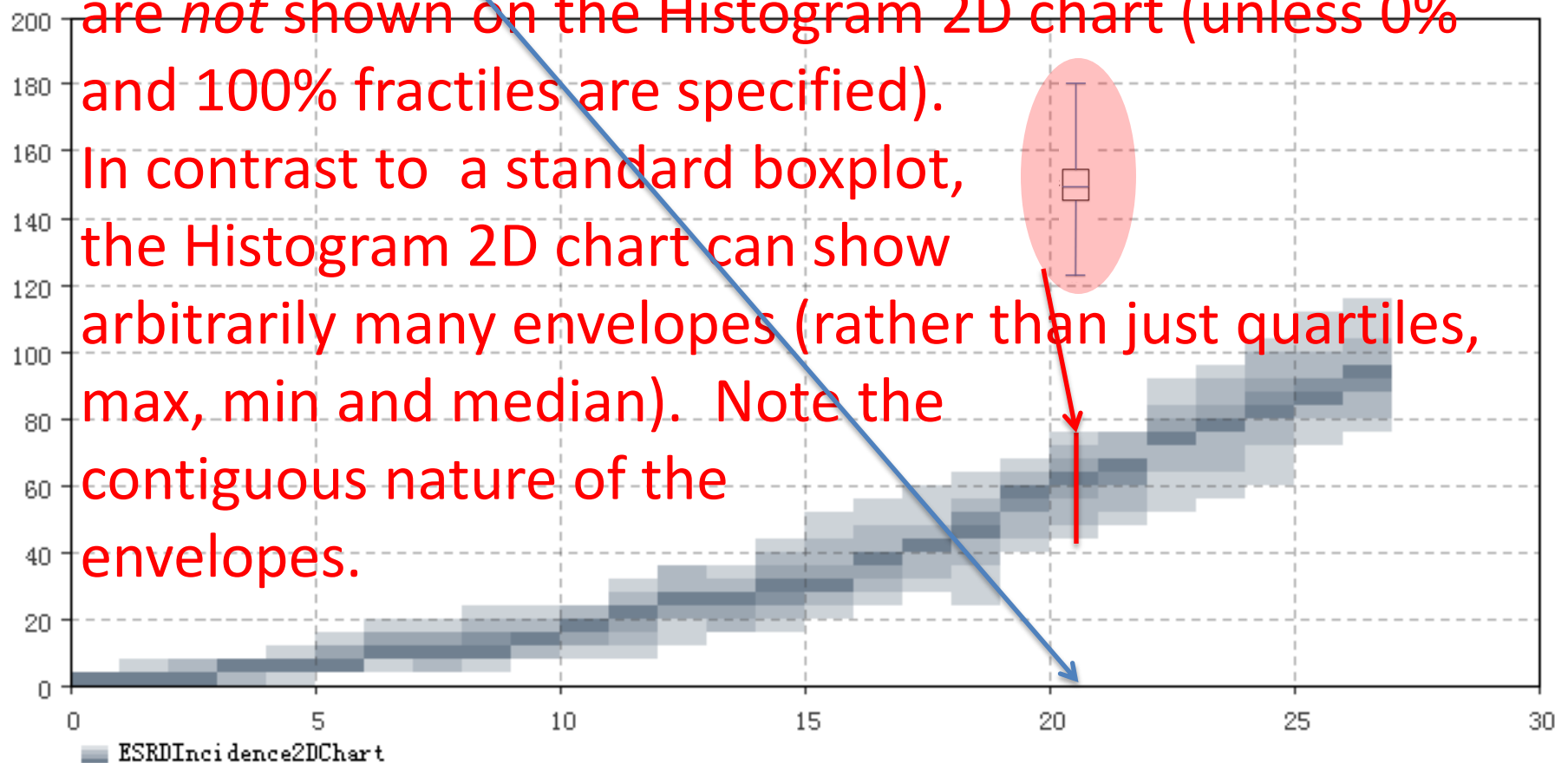
to: 8000

Time units: days



# Example of “Show Envelopes” Output (Different Model)

A slice at **this** point in time would yield a something like a **boxplot**. Note that the “whiskers” of the boxplot are *not* shown on the Histogram 2D chart (unless 0% and 100% fractiles are specified). In contrast to a standard boxplot, the Histogram 2D chart can show arbitrarily many envelopes (rather than just quartiles, max, min and median). Note the contiguous nature of the envelopes.



# Show Bins Option

The screenshot displays the AnyLogic Professional software interface. The main workspace shows a simulation titled "Agent Based SIR Model - Mo" with a "Run 100 replications" button and a 2D histogram plot. The histogram plot shows data points for "dataInfectious2D" with a y-axis ranging from 5,000 to 8,000. The "Properties" panel at the bottom is open, showing the "chart - Histogram2D" settings. In the "Properties" panel, the "Show bins" option is selected, and a red arrow points to it from the text "The 'show bins' option is here." on the left. The "Data" section of the properties panel shows the title "Dynamics of Infectious Popu", the histogram name "dataInfectious2D", and the color "slateGray". The "Presentation" palette on the right lists various shapes and objects.

AnyLogic Professional

File Edit View Draw Model Tools Help

Projects

- SIR Agent Based Calibration
  - Main
  - Person
  - Calibration: Main
  - MonteCarlo2DHistogram: Main

MonteCarlo2DHistogram

dataInfectious2D

Run 100 replications

Agent Based SIR Model - Mo

This model is © T1

Properties Progress

chart - Histogram2D

Name: chart ☐ Ignore

☐ Show envelopes

☒ Show bins

Data

Title: Dynamics of Infectious Popu

Histogram: dataInfectious2D

Color: slateGray

Time units: days

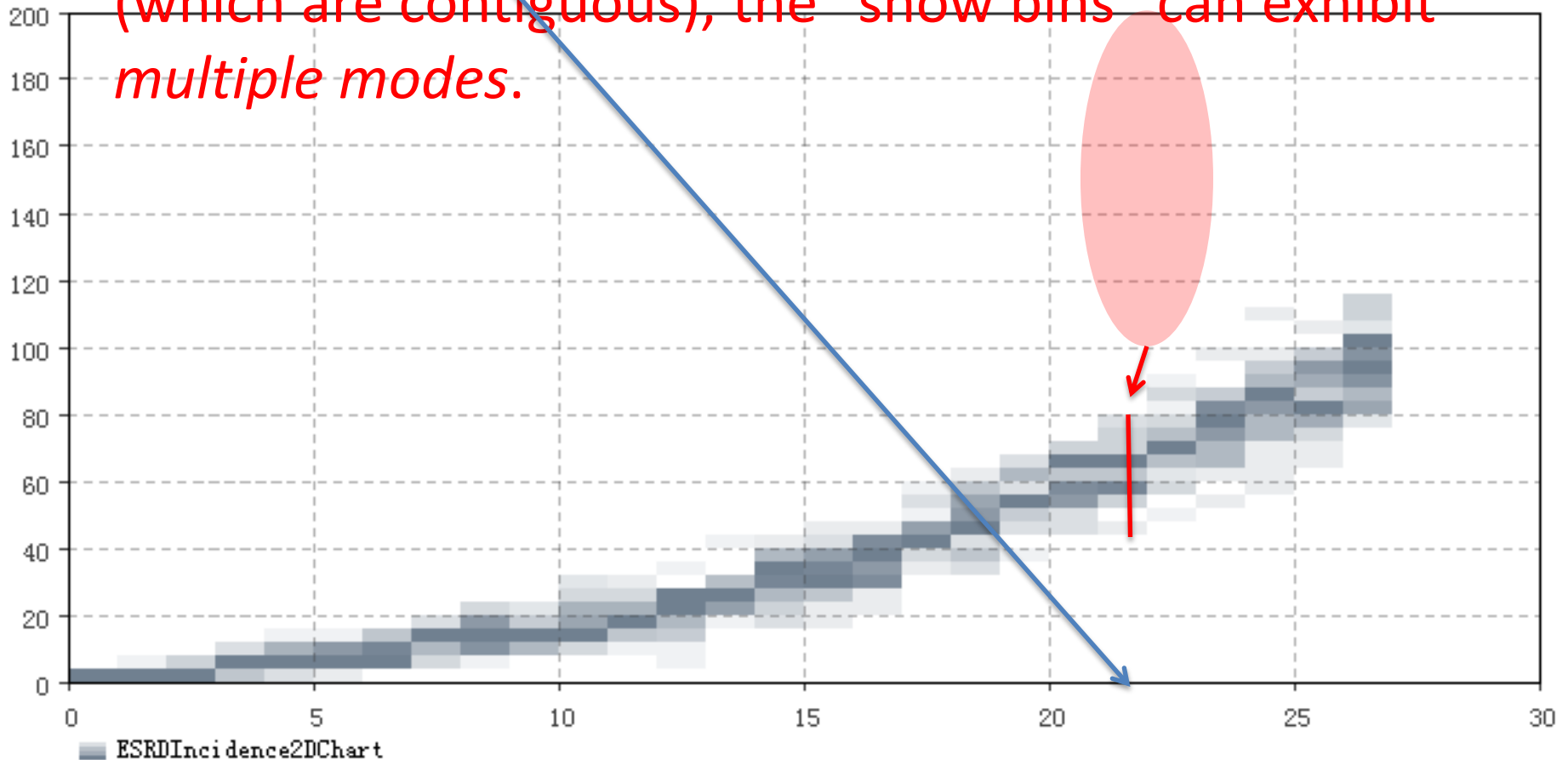
Palette

Presentation

- Rectangle
- Rounded Rect...
- Oval
- Arc
- Pixel
- Aa Text
- Image
- Group
- View Area
- 3D
  - 3D Window
  - 3D Object

# Example of “Show Bins” Output (Different Model)

A slice at **this** point in time would yield a *histogram*.  
Note: In contrast to the situation for the envelopes (which are contiguous), the “show bins” can exhibit *multiple modes*.



# Automatic Throttling of Monte Carlo Analyses

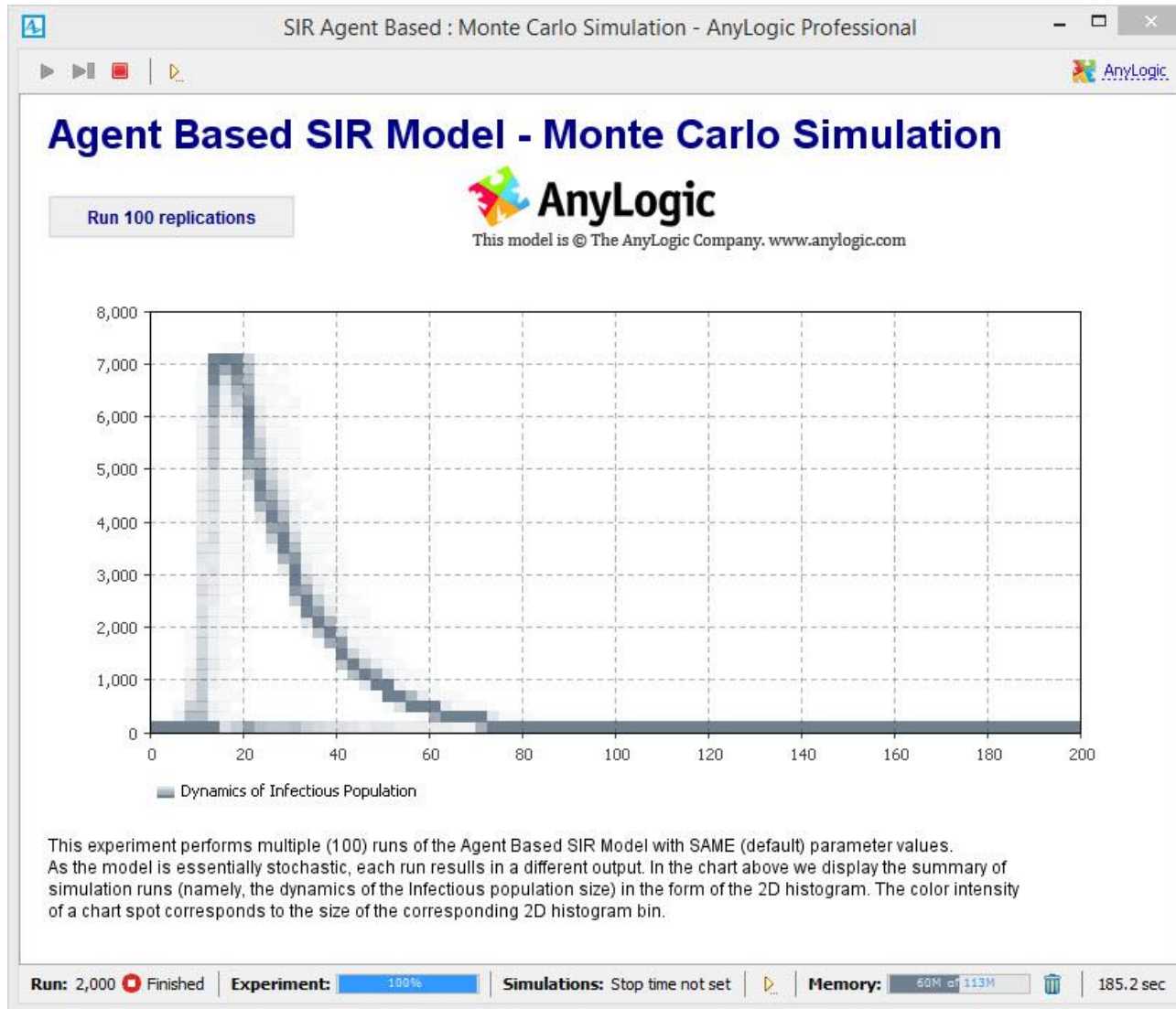
The screenshot displays the AnyLogic Professional software interface. The main workspace shows a simulation titled "Agent Based SIR Model - Monte Carlo Simulation". A button labeled "Run 100 replications" is visible. Below the title, a 2D histogram plot is shown, with the y-axis labeled "ctious2D" and values ranging from 7,000 to 8,000. The plot displays a distribution of data points.

The bottom panel, titled "MonteCarlo2DHistogram - Parameter Variation Experiment", contains the following settings:

- Selection mode for simultaneous events: LIFO (Last In, First Out; in the reverse order of scheduling)
- Replications:
  - ☒ Use replications
  - ☒ Fixed number of replications
    - Replications per iteration: 10
  - ☐ Varying number of replications (Stop after minimum replications, when confidence level is reached)
    - Minimum replications: 2
    - Maximum replications: 10
    - Confidence level: 80% of expression: 0
    - Error percent: 0.5
- Window

The status bar at the bottom indicates "Time units: days".

# General Variety of Output



# Reminder: Statistical Scaling

- Consider Taking the sample mean of  $n$  samples that vary independently around a mean
- If two samples  $x$  and  $y$  are independent samples of random variables  $X$  and  $Y$ , then  $\text{Var}[x+y]=\text{Var}[X]+\text{Var}[Y]$ 
  - So if we have  $n$  indep. samples  $x_i$  from distribution  $X$ 
$$\text{Var}\left(\sum_{i=1}^n x_i\right) = n\text{Var}(X)$$
- If we scale a random variable by a factor  $\alpha$ , the standard deviation scales by the same factor of  $\alpha \Rightarrow$  the variance scales by  $\alpha^2$ 
  - i.e.  $\text{StdDev}[\alpha X] = \alpha \text{StdDev}[X]$ ,  $\text{Var}[\alpha X] = \alpha^2 \text{Var}[X]$

# Statistics of Sample Mean

- Recall: Sample Mean:  $m = \frac{\sum_{i=1}^n x_i}{n}$
- From the preceding, variance drops as  $1/n$

$$Var(m) = Var\left(\frac{\sum_{i=1}^n x_i}{n}\right) = \frac{Var\left(\sum_{i=1}^n x_i\right)}{n^2} = \frac{nVar(X)}{n^2} = \frac{Var(X)}{n}$$

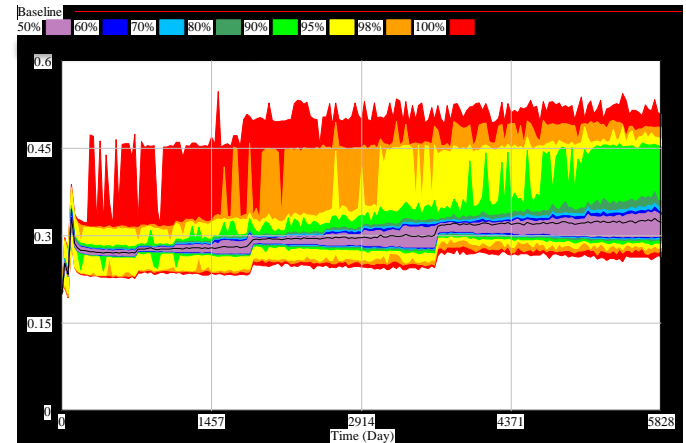
- This means that standard deviation for the sample mean of  $n$  samples drops as  $1/\sqrt{n}$

$$StdDev(m) = \sqrt{Var(m)} = \sqrt{\frac{Var(X)}{n}} = \sqrt{\frac{(StdDev(X))^2}{n}} = \frac{StdDev(X)}{\sqrt{n}}$$

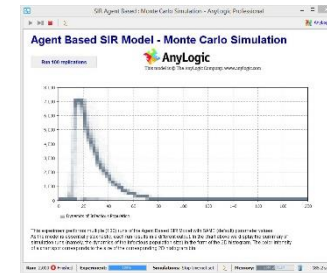
- So if we wish to divide the standard deviation of the sample mean by a factor of 2, we need to take 4x the number of Monte Carlo samples

# Closing Question: How can we best adapt our policies to deal with ongoing uncertainty?

- We are dealing here with making decisions in an environment that changes over time
- This uncertainty could come from
  - Stochastic variability



- Uncertainty regarding parameter values



- There is an incredibly vast # of possible policies
- **Reminder: Can successfully integrate decision analysis & simulation to neatly handle such cases**