

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 window shows a simulation titled "Agent Based SIR Model - Monte Carlo Simulation". A button labeled "Run 100 replications" is visible. Below the title, there is a 2D histogram plot showing the distribution of simulation results. The plot has a vertical axis ranging from 6,000 to 8,000 and a horizontal axis. The histogram consists of numerous small blue squares representing individual simulation runs, showing a distribution centered around 7,000.

The interface includes a menu bar (File, Edit, View, Draw, Model, Tools, Help), a toolbar with various icons, and a "Projects" panel on the left. The "Projects" panel shows a tree view with the following structure:

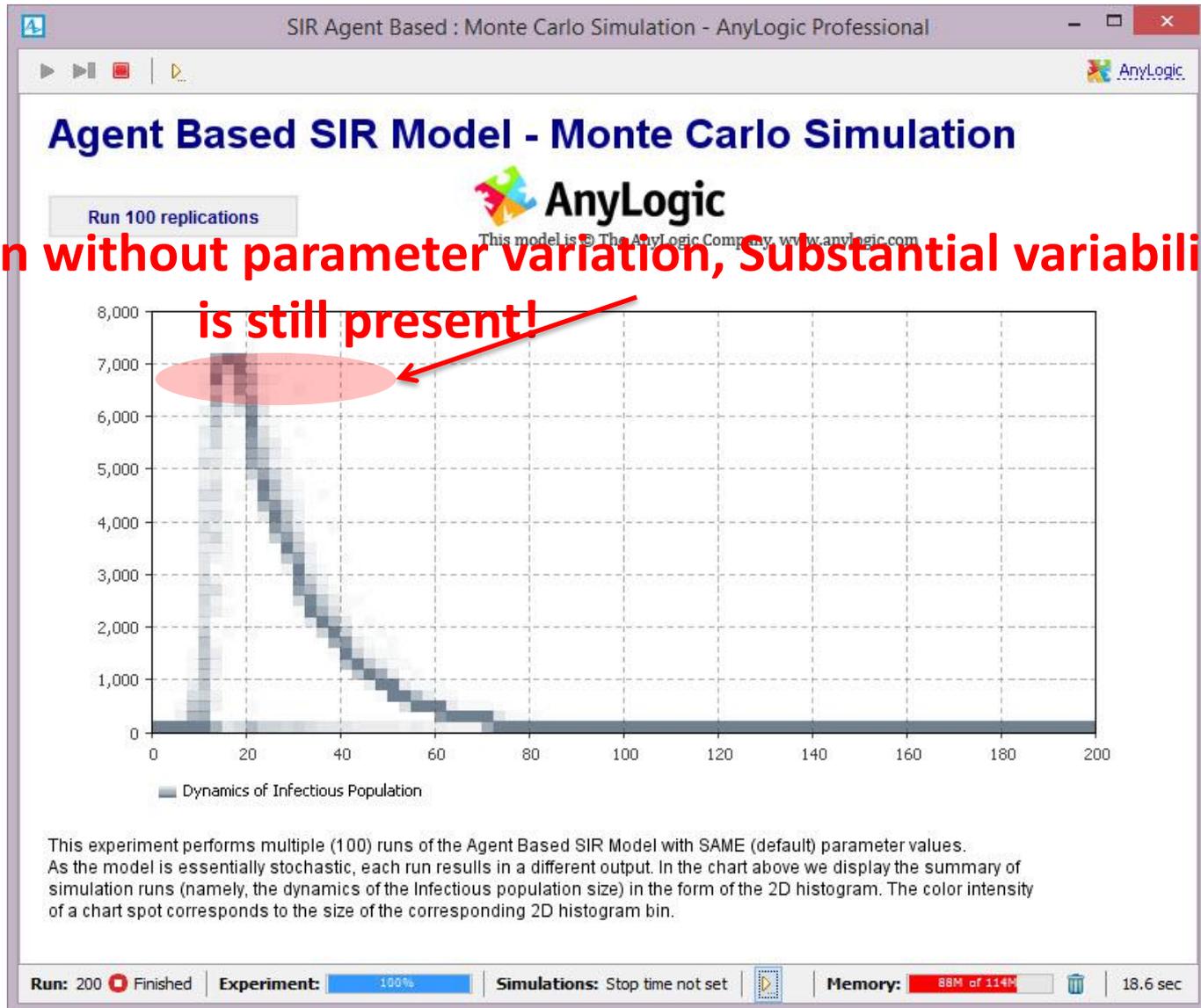
- SIR Agent Based Calibration
 - Main
 - Person
 - Calibration: Main
 - MonteCarlo2DHistogram: Main

The "Properties" panel at the bottom shows the configuration for the "MonteCarlo2DHistogram - Parameter Variation Experiment":

- Name: MonteCarlo2DHistogram Ignore
- Top-level agent: Main
- Maximum available memory: 128 Mb
- Create default UI
- Parameters
 - Model time
 - Use calendar
 - Stop: 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 software interface. The main workspace shows a diagram of an "Agent Based SIR Model - Mo" with a "Run 100 replications" button and a "dataInfectious2D" component. The Properties panel is open for the "dataInfectious2D - Histogram2D Data" component, showing the following settings:

- 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:

The bottom status bar indicates "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. The main workspace shows a simulation model titled "Agent Based SIR Model - Mo" with a "Run 100 replications" button. A blue arrow points from the text "The accumulating Histogram2D dataset is in Experiment" to the "dataInfectious2D" component in the workspace. A red arrow points from the text "The source dataset is in Main" to the "InfectiousDS" component in the "After simulation run:" field of the "MonteCarlo2DHistogram - Parameter Variation Experiment" configuration panel. The configuration panel also shows "dataInfectious2D.add (root.InfectiousDS);" in the "After simulation run:" field.

The accumulating Histogram2D dataset is in Experiment

The source dataset is in Main

```
dataInfectious2D.add ( root.InfectiousDS );
```

Monte Carlo Sensitivity Analyses in AnyLogic

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 2nd 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

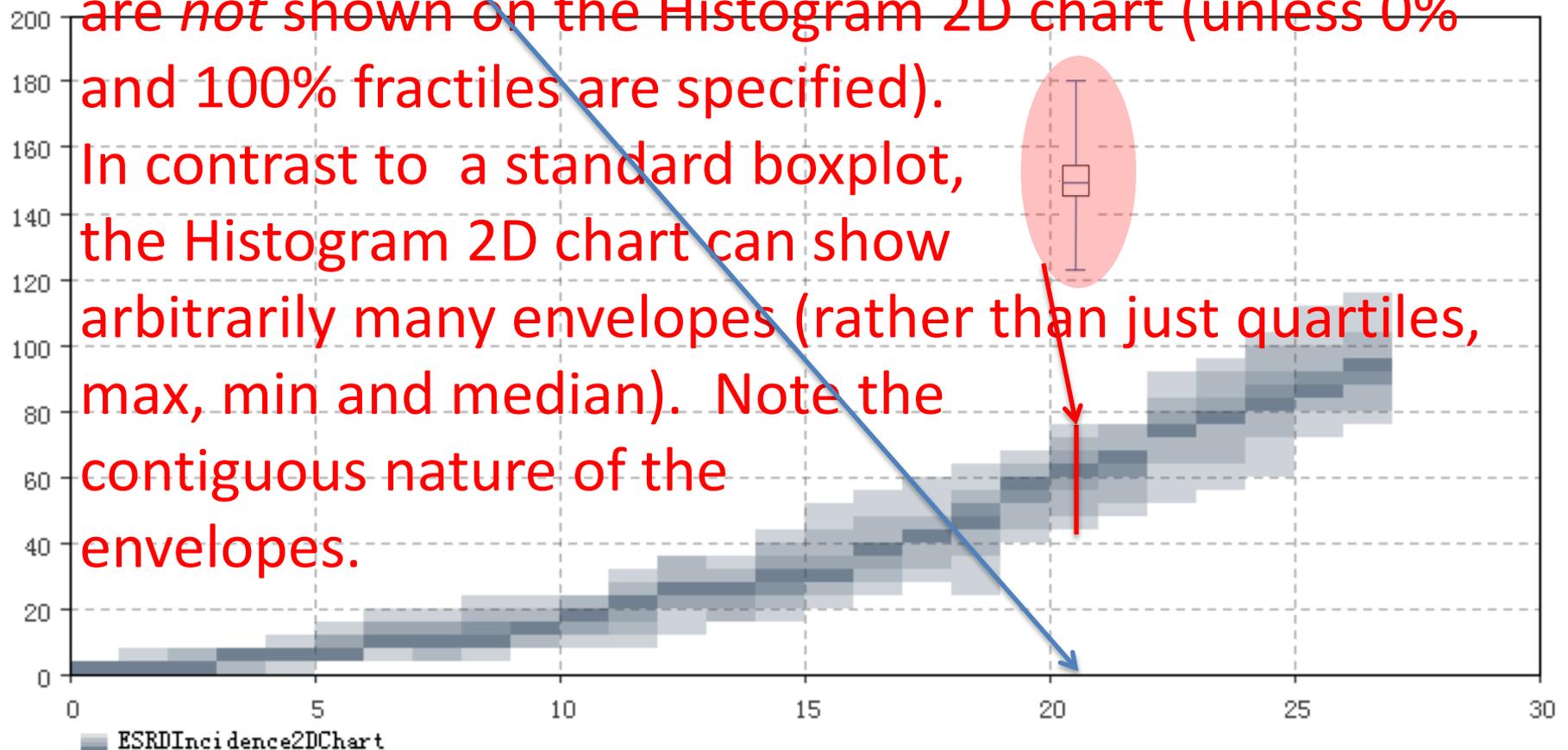
Reminder: 2D Histogram Data

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 simulation titled "Agent Based SIR Model - Monte Carlo 2D Histogram". A "Run 100 replications" button is visible. The "Properties" panel for the "dataInfectious2D - Histogram2D Data" component is open. The "Envelopes" field is highlighted with a red oval and contains the values "0.25, 0.5, 0.75". A red arrow points from the text "envelopes" in the overlay to this field. The "X-axis values range" section shows "Number of intervals: 80", "Range, from: 0", and "to: 200". The "Y-axis values range" section shows "Number of intervals: 40", "Range, from: 0", and "to: 8000". The status bar at the bottom indicates "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 interface. The main workspace shows a 2D histogram titled "Agent Based SIR Model - Mo" with a "Run 100 replications" button. The histogram data is plotted on a grid with a y-axis ranging from 5,000 to 8,000. The "Properties" panel at the bottom is open for the "chart - Histogram2D" object. In this panel, the "Show bins" radio button is selected and highlighted with a red circle and an arrow. Other options include "Show envelopes" (unselected) and "Ignore" (checkbox). The "Data" section shows the chart title as "Dynamics of Infectious Popu", the data source as "dataInfectious2D", and the color as "slateGray". The "Presentation" palette on the right lists various shapes and 3D objects.

The "Properties" panel for "chart - Histogram2D" includes the following options:

- Name: Ignore
- Show envelopes
- Show bins

The "Data" section includes:

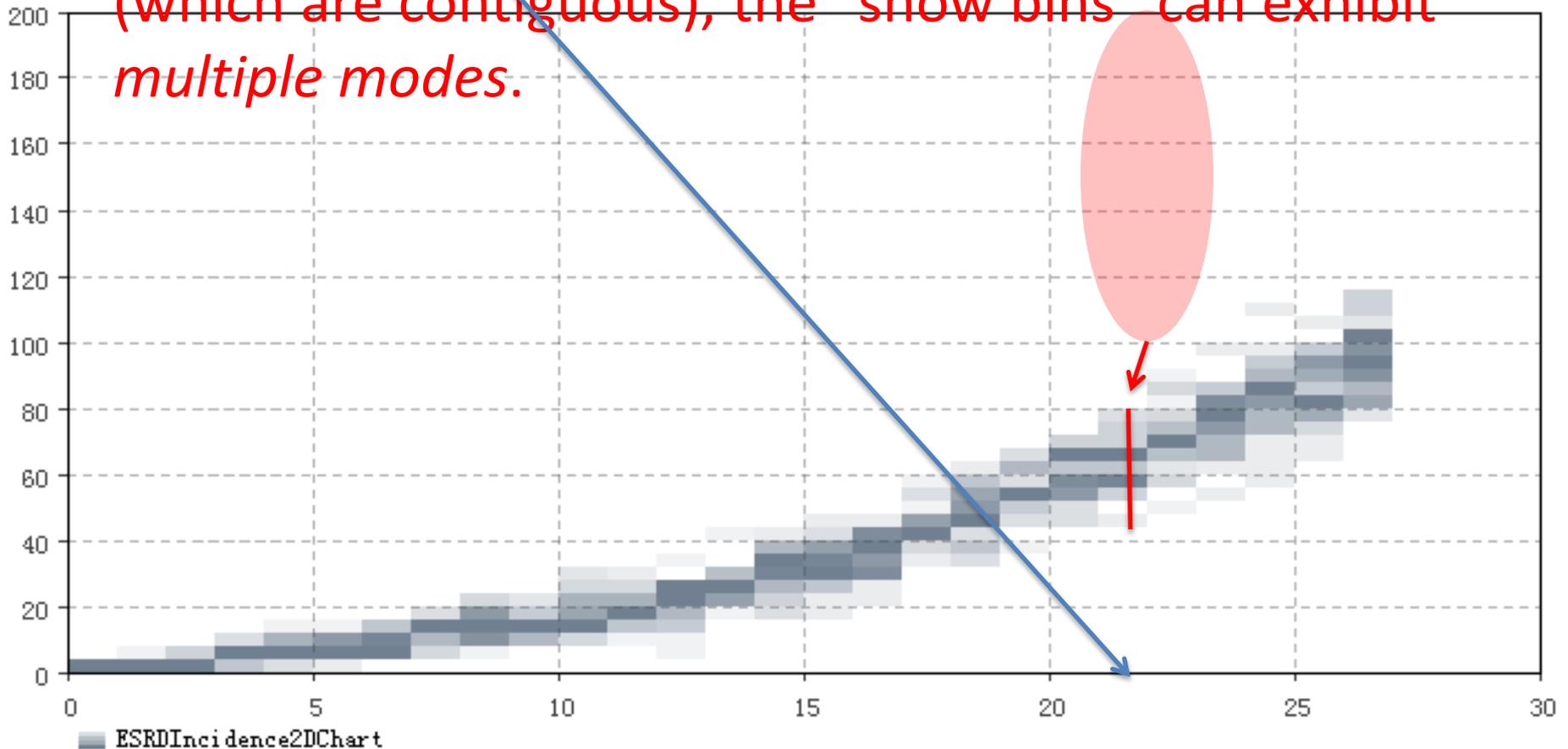
- Title:
- Histogram:
- Color:

Time units: days

The "show bins" option is here.

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*.



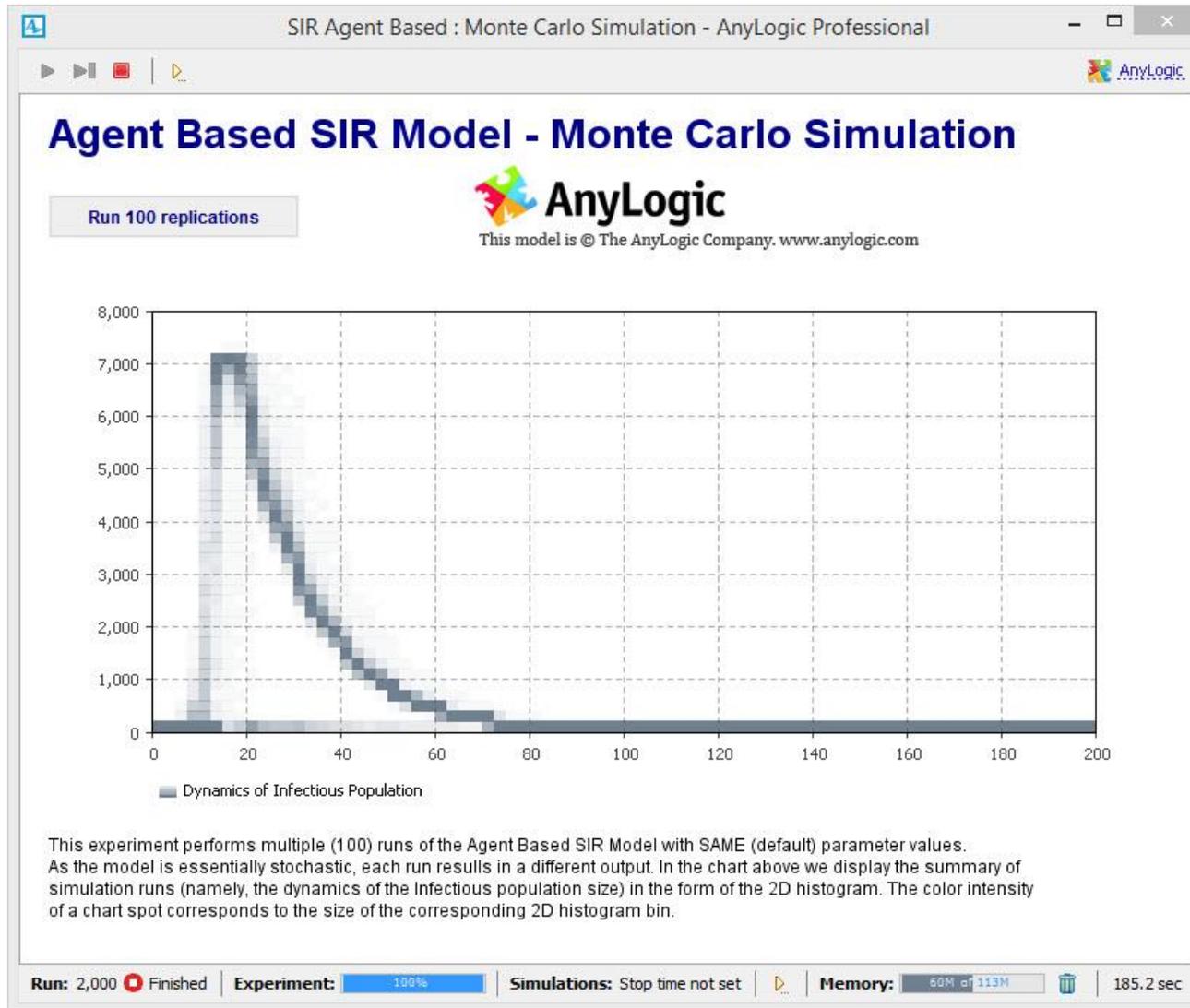
Automatic Throttling of Monte Carlo Analyses

The screenshot displays the AnyLogic Professional software interface. The main workspace shows a simulation titled "Agent Based SIR Model - Monte Carlo Simulation" with the AnyLogic logo and the text "This model is © The AnyLogic Company. www.anylogic.com". A button labeled "Run 100 replications" is visible. Below the simulation area, the "Properties" panel is open, showing the "MonteCarlo2DHistogram - Parameter Variation Experiment" settings. The "Replications" section is expanded, showing the following configuration:

- 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

The "Window" section at the bottom of the Properties panel is also visible. The status bar at the bottom right 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 α , the standard deviation scales by the same factor of $\alpha \Rightarrow$ the variance scales by α^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 $1/n$

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

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

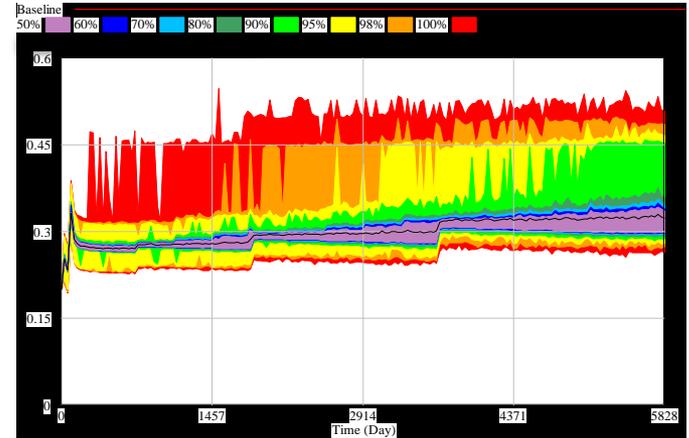
$$\text{StdDev}(m) = \sqrt{\text{Var}(m)} = \sqrt{\frac{\text{Var}(X)}{n}} = \sqrt{\frac{(\text{StdDev}(X))^2}{n}} = \frac{\text{StdDev}(X)}{\sqrt{n}}$$

- 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

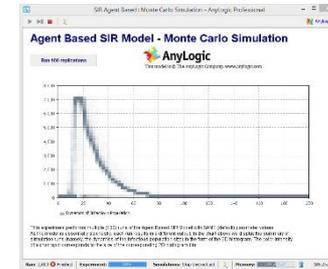
Closing Question: How can we best adapt our policies to deal with ongoing uncertainty?

- 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**