

A Glimpse of Representing Stochastic Processes

Nathaniel Osgood

CMPT 858

March 22, 2011

Recall: Project Guidelines

- Creating one or more simulation models.
- Placing data into the model to customize it to a particular context (e.g. to a particular region).
- Running a “baseline” scenario with the existing model parameters, and commenting on its plausibility.
- Running one or more “what if” scenarios with the model to explore different possible situations. These situations could reflect the results of implementing different policies, or different possible external conditions.
- Performing one or more “sensitivity analyses”, in which assumptions in the model (in the form of parameter values, or structural elements of the model) are changed.
- A well-structured written report describing the above, and your findings.
- Report Due date: April 24

What I'd like to See in the Report

- Introduction
- Background
- Comment on baseline scenario choice & plausibility of model results
- Sensitivity analyses (Parameters and/or structural)
- Investigation of “what if” scenarios
 - Potential policies
 - External conditions
- Model limitations & ideas for possible extensions
- Process learning
- Implications for our understanding of the world

Project Presentations

- Presentations will be 45 minutes in length
- Seeking 2 half days of presentations
- Encourage all students to attend
- Likely scheduling: Late in week of April 24 (just after report due)
 - afternoon of April 28 & 29?

Monte Carlo Analyses in AnyLogic

- When running Monte Carlo analysis, 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



Hands on Model Use Ahead



Load Sample Model:

SIR Agent Based Calibration

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

2D Histogram Data

The screenshot displays the AnyLogic software interface. The main workspace shows a 2D histogram titled "Agent Based SIR M" with a "Run 100 replicat..." button. The histogram's vertical axis ranges from 3,000 to 4,000. A data point is labeled "dataInfectious2D".

The Properties panel for "dataInfectious2D - Histogram 2D Data" is visible, showing the following settings:

- Name: dataInfectious2D
- Show Name
- Ignore
- Public
- Show At Runtime
- Horizontal axis value: []
- Vertical axis value: []
- Horizontal intervals: 80
- Range, from: 0 to: 200
- Vertical intervals: 40
- Range, from: 0 to: 4000
- Envelopes: 0.25, 0.5, 0.75
- Do not update automatically
- Auto update after every iteration

The left sidebar shows a project tree with folders for "SIR Agent Based Calibration*", "Influenza*", and "Family". The right sidebar contains a Palette with various modeling elements like Parameter, Flow Aux Variable, Stock Variable, Event, Dynamic Event, Plain Variable, Collection Variable, Function, Table Function, Port, Connector, Entry Point, State, Transition, Initial State Pointer, Branch, History State, Final State, and Environment. At the bottom right, there are buttons for Action, Analysis, Presentation, Connectivity, and Enterprise Library.

Important Distinction (Declining Order of Aggregation)

- Experiment
 - Collection of simulation
- 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

Monte Carlo Sensitivity Analyses in AnyLogic

Agent Based SIR Model - Monte Carlo Simulation

Run 100 replicat...

technologies
AnyLogic and this model is (c) XJ Technologies, www.anylogic.com. All rights reserved.

chart - Histogram2D

General Name: chart Show Name Ignore Public

Advanced

Dynamic Title: Dynamics of Infectious Populatic

Appearance Histogram: dataInfectious2D Color: slateGray

Description

Add Histogram Data

Show envelopes Show bins

Do not update automatically

Auto update after every iteration

- Parameter
- Flow Aux Variable
- Stock Variable
- Event
- Dynamic Event
- Plain Variable
- Collection Variable
- Function
- Table Function
- Port
- Connector
- Entry Point
- State
- Transition
- Initial State Pointer
- Branch
- History State
- Final State
- Environment

- Action
- Analysis
- Presentation
- Connectivity
- Enterprise Library
- More Libraries...

Monte Carlo Analyses in AnyLogic: Specifying Distributions for Parameters

The screenshot shows the AnyLogic Advanced [EDUCATIONAL USE ONLY] interface. The main window displays a simulation titled "Agent Based SIR Model - Monte Carlo Simulation". A button labeled "Run 100 replicat..." is visible. The simulation area shows a 2D histogram of "infectious2D" values, with the y-axis ranging from 2,000 to 4,000. The histogram shows a distribution of values across the simulation runs.

The Properties window is open, showing the configuration for the "MonteCarlo2DHistogram - Parameter Variation Experiment". The "General" tab is selected, and the "Advanced" section is expanded. The "Random number generation" section has "Random seed (unique simulation runs)" selected. The "Parameters" section has "Freeform" selected, and the "Number of runs" is set to 100.

Parameter	Expression
Average...ration*	max(0,normal(5,15))
ContactRate	1.0
Infection...bability	0.8
AreaSide	100
TotalPopulation	10000

Automatic Throttling of Monte Carlo Analyses

The screenshot displays the AnyLogic Advanced software interface for an educational use. The main window is titled "Agent Based SIR Model - Monte Carlo Simulation" and shows a 2D histogram of "infectious2D" data. A button labeled "Run 100 replicat..." is visible. The bottom panel shows the "Replications" settings for the "MonteCarlo2DHistogram" component, including options for "Use replications", "Fixed number of replications", and "Varying number of replications".

MonteCarlo2DHistogram - Parameter Variation Experiment

General

- Use replications

Advanced

- Fixed number of replications
 - Replications per iteration:
- Varying number of replications (Stop replications after minimum replications, when confidence level is reached)
 - Minimum replications:
 - Maximum replications:
 - Confidence level: of expression:
 - Error percent:

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, we have

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

- This means that standard deviation for the sample mean of n samples varies as

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

- 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

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, etc.
 - A meteorological model may not consider rainfall stochastic

Making a Vensim Flow Stochastic

Editing equation for - New Recoveries

New Recoveries Add Eq

= `RANDOM BINOMIAL(0, Infective, (1/Mean Time to Recover) * TIME STEP, Infective, 0, 1, Noise Seed) / TIME STEP`

Type: Auxiliary Undo {()} 7 8 9 + Variables Functions More

Normal Choose Variable... Inputs

Supplementary 1 2 3 *

0 E . /

() . ^

Infected
Mean Time to Recover
Noise Seed
TIME STEP

Units: ▼

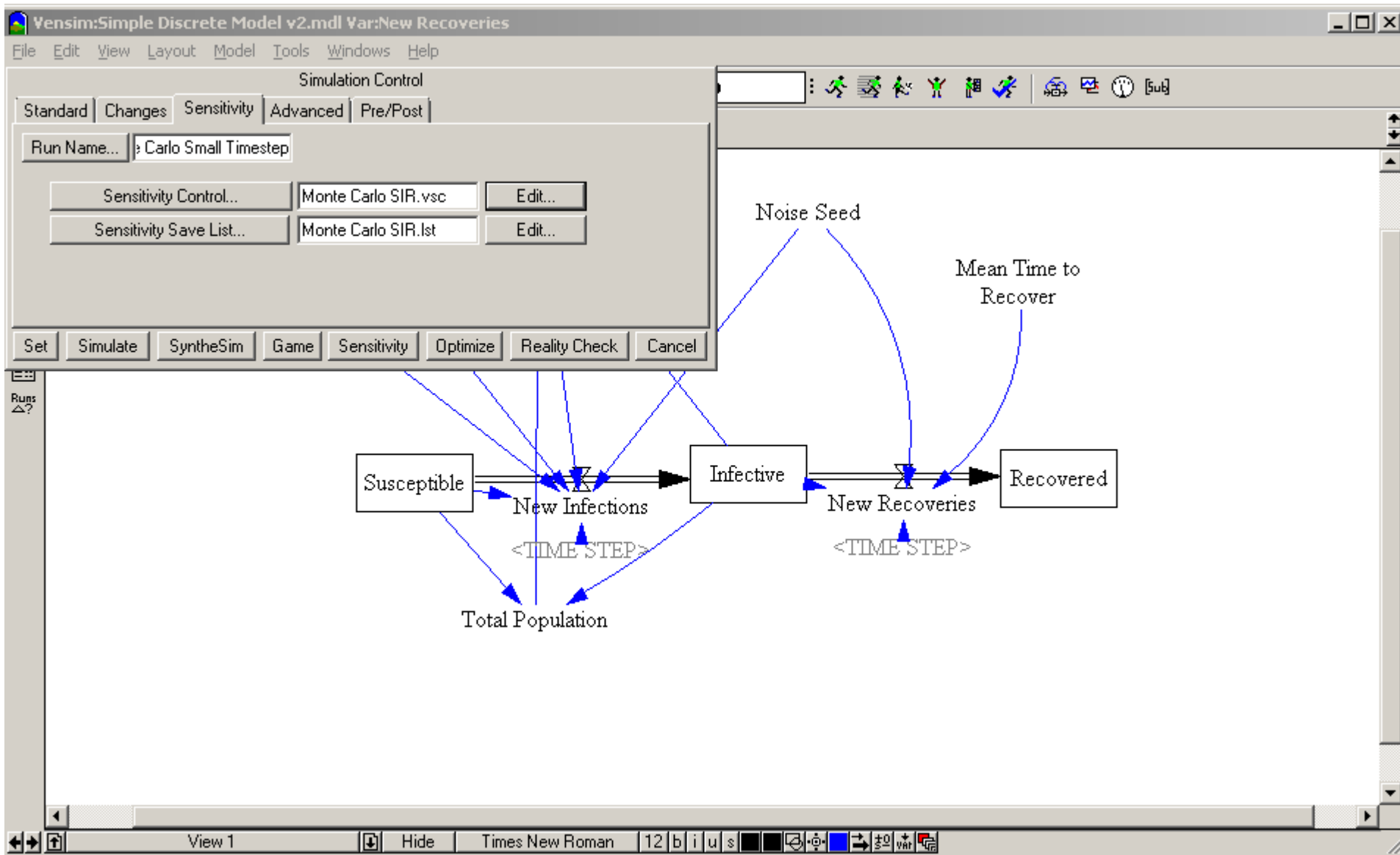
Comment: ▼

Group: .Simple discr ▼ Range: Go To: Prev Next ← Hilite Sel... New

Errors: Equation Modified ▼

OK Check Syntax Check Model Delete Variable Cancel

Treat as a Sensitivity Analysis



Setting the “Random Seed” to Differ between Simulations

Vensim: Simple Discrete Model v2.mdl Var: New Recoveries

File Edit View Layout Model Tools Windows Help

Sensitivity Control. Edit the filename to save changes to a different control file

Filename: Choose New File... Clear Settings

Number of simulations: Noise Seed:

Multivariate Univariate
 Latin Hypercube Latin Grid
 File Sel...

Display warning messages

Currently active parameters (drag to reorder)
Noise Seed=RANDOM_UNIFORM(0,10000)

Delete Selected
Modify Selected
Add Editing

Distribution

Parameter	Minimum Value	Maximum Value
<input type="text" value="RANDOM_UNIFORM"/>	<input type="text"/>	<input type="text"/>

OK Cancel

Noise Seed

Mean Time to Recover

Infected

Recovered

New Recoveries

<TIME STEP>

Total Population

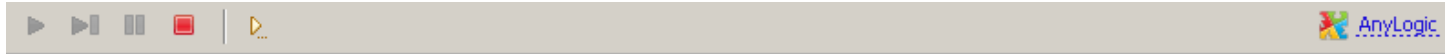
Monte Carlo Analysis with Fixed Parameter Values

The screenshot shows the AnyLogic Advanced software interface. The main window displays a 2D histogram titled "Agent Based SIR Model - Monte Carlo Simulation". The histogram shows the distribution of "infectious2D" values, with a y-axis ranging from 2,000 to 4,000. A button labeled "Run 100 replicat..." is visible above the histogram. The software interface includes a menu bar (File, Edit, View, Model, Window, Help), a toolbar, and a project browser on the left. The project browser shows a hierarchy for "SIR Agent Based Calibration*", including "Main", "Parameters", "Plain Variables", "Environments", "Embedded Objects", "Analysis Data", "dsInfectious", "Presentation", "Person", "Calibration: Main", "Functions", "Analysis Data", "datasetCurrentObjective", "datasetBestFeasibleObjective", "dsInfectiousHistoric", "dsInfectiousCurrent", "dsInfectiousBest", "Presentation", and "MonteCarlo2DHistogram: Main". The "MonteCarlo2DHistogram: Main" sub-project is expanded, showing "Analysis Data" and "Presentation". The "Properties" panel at the bottom is open, showing the "General" tab for the "MonteCarlo2DHistogram - Parameter Variation Experiment". The "General" tab includes fields for "Name" (MonteCarlo2DHistogram), "Main active object class (root)" (Main), and "Ignore" (unchecked). The "Random number generation" section has "Random seed (unique simulation runs)" selected. The "Fixed seed (reproducible simulation runs)" option is also visible, with a "Seed Value" field set to 1. The "Parameters" section has "Freeform" selected, and the "Number of runs" is set to 100. A table lists parameters and their expressions:

Parameter	Expression
AverageI...uration	15
ContactRate	1.0
Infection...bability	0.8
AreaSide	100
TotalPopulation	10000

The right side of the interface features a "Palette" with various analysis and visualization options, including "Data Set", "Statistics", "Histogram Data", "Histogram2D Data", "Bar Chart", "Stack Chart", "Pie Chart", "Plot", "Time Plot", "Time Stack Chart", "Time Color Chart", "Histogram", and "Histogram2D". The bottom right corner contains buttons for "Presentation", "Connectivity", "Enterprise Library", and "More Libraries...".

Results of Monte Carlo Simulation



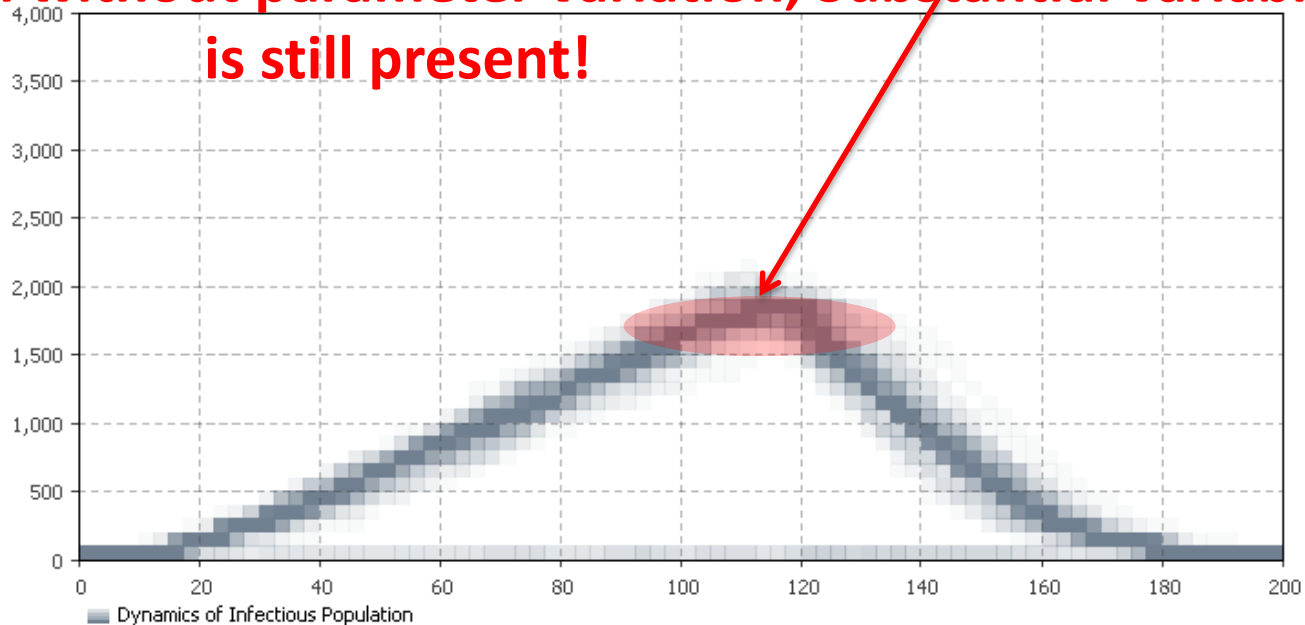
Agent Based SIR Model - Monte Carlo Simulation

Run 100 replications



AnyLogic and this model is (c) XJ Technologies, www.anylogic.com. All rights reserved.

Even without parameter variation, Substantial variability is still present!



This experiment performs multiple (100) runs of the Agent Based SIR Model with SAME (default) parameter values. As the model is essentially stochastic, each run results in a different output. In the chart above we display the summary of simulation runs (namely, the dynamics of the Infectious population size) in the form of the 2D histogram. The color intensity of a chart spot corresponds to the size of the corresponding 2D histogram bin.

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

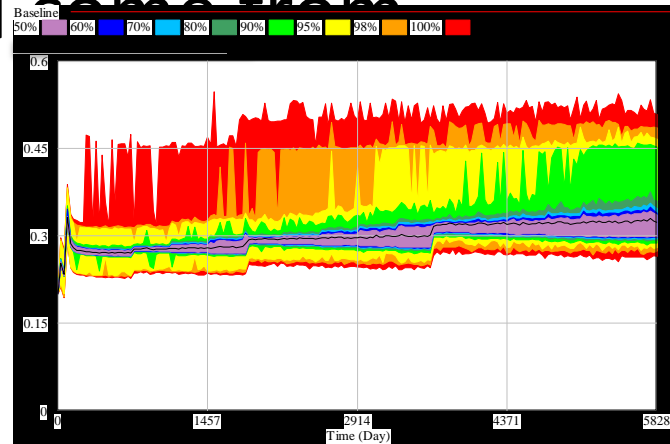
- To gain confidence in model results, typically need to run an ensemble of realizations
 - Deal with means, standard deviations, and empirical fractiles
 - As is seen here, there are typically still broad regularities between most runs (e.g. rise & fall)
- Need to reason over a population of realizations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

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

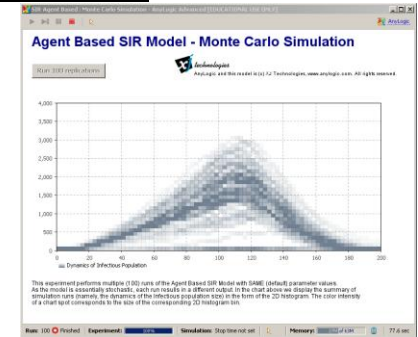
- We are dealing here with making decisions in an environment that changes over time

- This uncertainty could come from

- Stochastic variability



- Uncertainty regarding parameter value



- There is an incredibly vast # of possible policies

Stochastic Processes in Vensim

