Oleksandr Romanko, Ph.D.

Senior Research Analyst, Risk Analytics, Watson Financial Services, IBM Canada Adjunct Professor, University of Toronto

MIE1624H – Introduction to Data Science and Analytics Lecture 9 – Simulation Modeling

University of Toronto November 20, 2018

Simulation Modeling

Sums of random variables

lacktriangle For any random variable $oldsymbol{x}$ and a constant w

$$\mathbb{E}[w \cdot \mathsf{x}] = w \cdot \mathbb{E}[\mathsf{x}]$$

Expectation of the sum of two random variables is equal to the sum of expectations $\mathbb{E}[x+y] = \mathbb{E}[x] + \mathbb{E}[y]$

and, therefore

$$\mathbb{E}[w_1 \cdot \mathbf{x} + w_2 \cdot \mathbf{y}] = w_1 \cdot \mathbb{E}[\mathbf{x}] + w_2 \cdot \mathbb{E}[\mathbf{y}]$$

Example: expected value of a portfolio

$$\mathbb{E}[0.4 \cdot r_1 + 0.6 \cdot r_2] = 0.4 \cdot \mathbb{E}[r_1] + 0.6 \cdot \mathbb{E}[r_2]$$

For the variance

$$var[w \cdot x] = w^2 \cdot var[x]$$

$$var[x + y] = var[x] + var[y] + 2 \cdot cov(x, y)$$

$$var[w_1 \cdot x + w_2 \cdot y] = w_1^2 \cdot var[x] + w_2^2 \cdot var[y]$$

$$+ 2 \cdot w_1 \cdot w_2 \cdot cov(x, y)$$

3

Sums of random variables

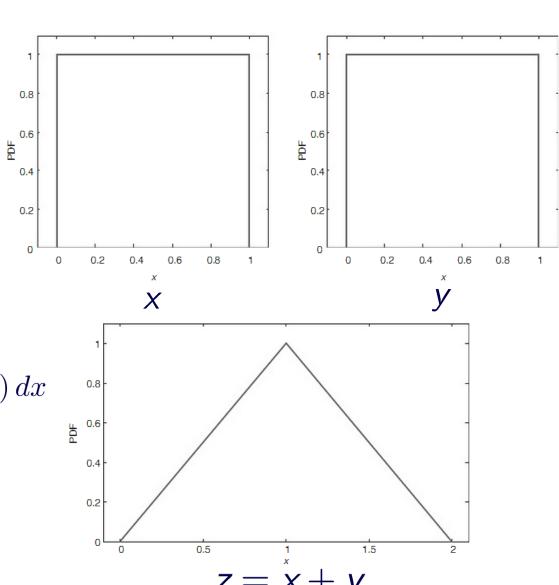
How to compute the probability distribution of the sum of random variables?

$$z = x + y$$

- We cannot add PDFs or PMFs
- The formula involves nontrivial integration and is known as convolution:

$$f_{z}(z) = \int_{-\infty}^{\infty} f_{y}(z - x) f_{x}(x) dx$$

 Use simulation to evaluate such complex integrals



Sums of random variables

$$f_{z}(z) = \int_{-\infty}^{\infty} f_{y}(z - x) f_{x}(x) dx$$

$$f_{\mathsf{x}}(x) = 1 \text{ only in } [0, 1]$$

$$f_{z}(z) = \int_{0}^{1} f_{y}(z - x) dx$$

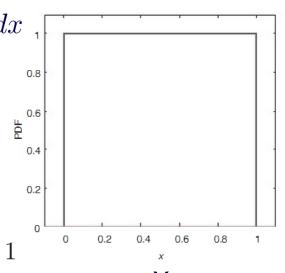
This is zero unless $0 \le z - x \le 1$ $(z - 1 \le x \le z)$

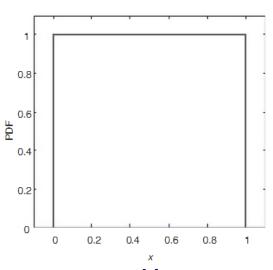
Case 1:
$$0 \le z \le 1$$

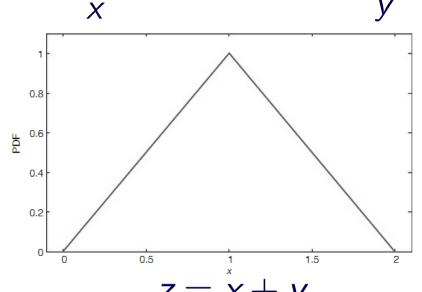
$$f_{\mathbf{z}}(z) = \int_0^z dx = z$$

Case 2: $1 \le z \le 2$

$$f_{z}(z) = \int_{z-1}^{1} dx = 2 - z$$







Simulation Modeling in Finance – Portfolio Selection

- lacktriangle We want to invest \$1000 in the US stock market for 1 year: $v_0=1000$
- Invest into the S&P 500 market index (index fund)
- Value of investment at the end of year 1: V_1
- Market return over the time period [0,1) is $r_{0,1}$

$$\mathbf{v}_1 = v_0 + \mathbf{r}_{0,1} \cdot v_0 = (1 + \mathbf{r}_{0,1})v_0$$

- lacktriangle Generate scenarios for the market return over the year and compute v_1
 - $lue{}$ decide on the number of scenarios and the set of scenarios for $r_{0,1}$
 - generate scenarios
 - ✓ use historic scenarios
 - √ draw randomly from historic scenarios (bootstrapping)
 - ✓ draw random numbers from the assumed distribution (Monte Carlo)
 - \square visualize and analyze the approximate probability distribution of V_1
- In our example we assume that the return of the market over the next year follow Normal distribution

- Between 1977 and 2007, S&P 500 returned 8.79% per year on average with a standard deviation of 14.65%
- Generate 100 scenarios for the market return over the next year (draw 100 random numbers from a Normal distribution with mean 8.79% and standard deviation of 14.65%):

 r01 = random.normal(0.0879, 0.1465, 100)

0.099278

-0.004262

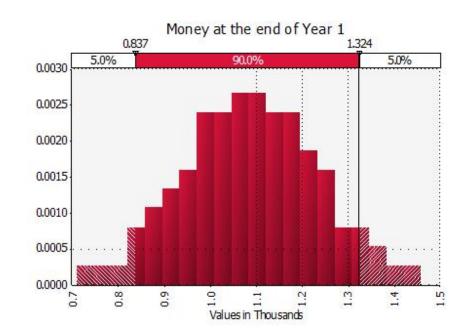
. .

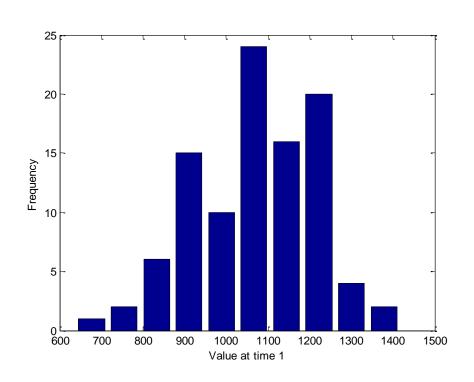
0.488364

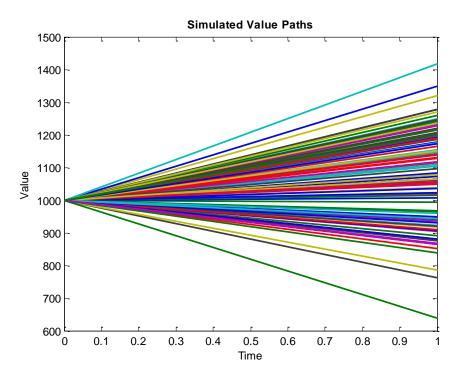
-0.119054

lacksquare Compute and plot $v_1 = (1 + r_{0,1})v_0$

Number of values	100
Mean	\$ 1,087.90
Std Deviation	\$ 146.15
Skewness	0.0034442
Kurtosis	2.871695
Mode	\$ 1,118.96
5% Perc	\$ 837.40
95% Perc	\$ 1,324.00
Minimum	\$ 708.81
Maximum	\$ 1,458.52







Why use simulation?

- Example 1 illustrates very basic Monte Carlo simulation system
- Simulation allows us to evaluate (approximately) a function of a random variable
 - f u in example 1 the function is simple ${\it v}_1=(1+{\it r}_{0,1})v_0$
 - given distribution of $r_{0,1}$, in some cases we can compute distribution of v_1 in closed form, e.g., if $r_{0,1}$ followed a Normal distribution, then v_1 also follows a Normal distribution with mean $(1+\mu_{0,1})v_0$ and standard deviation $\sigma_{0,1}v_0$
 - \Box if $r_{0,1}$ was not Normally distributed, or if the output variable v_1 were a more complex function of the input variable $r_{0,1}$, it would be difficult and practically impossible to derive the probability distribution of v_1 from the probability distribution of $r_{0,1}$
- Other advantages of simulation:
 - □ simulation enables visualizing probability distribution resulting from compounding probability distributions of multiple input variables (example 2)
 - □ simulation allows incorporating correlations between input variables (example 3)
 - □ simulation is a low-cost tool for checking the effect of changing a strategy on an output variable of interest (example 4)
- Next, we extend example 1 to illustrate such situations

- \blacksquare You are planning for retirement and decide to invest in the market for the next 30 years (instead of only the next year as in example 1). Your initial capital is still $v_0=1000$
- Assume that every year your investment returns from investing into the S&P 500 will follow a Normal distribution with the mean and standard deviation as in example 1.
- Value of investment after 30 years: V_{30}
- The return over 30 years will depend on the realization of 30 random variables

$$v_{30} = (1 + r_{0,1}) \cdot (1 + r_{1,2}) \cdot \dots \cdot (1 + r_{29,30}) \cdot v_0$$

$$r_{0,t} = (1 + r_{0,1})(1 + r_{1,2})\dots(1 + r_{t-1,t}) - 1$$

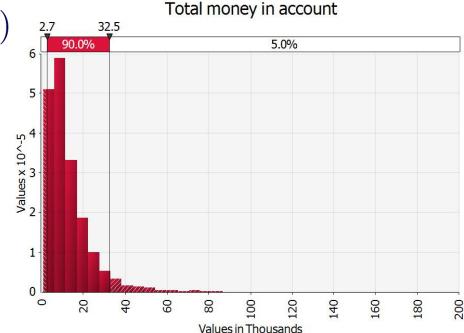
 $v_{0,t} = (1 + r_{0,t})v_0$

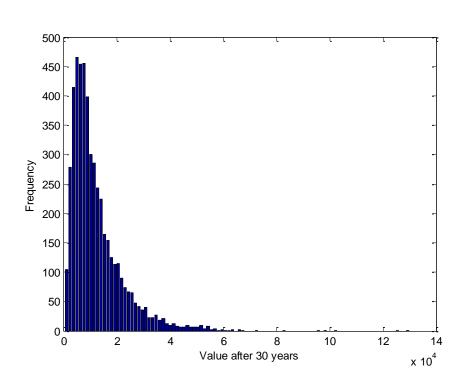
- Observations:
 - sum of Normal random variables is Normal
 - □ here we have multiplication of Normal random variables, is it Normal?

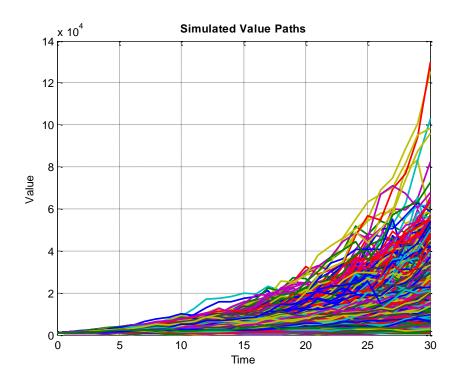
- Between 1977 and 2007, S&P 500 returned 8.79% per year on average with a standard deviation of 14.65%
- Simulate 30 columns of 100 observations each of single period returns:

lacksquare Compute and plot $extit{$
u_{30}=(1+ extit{$r_{0,30}$})$}$

Number of values	5000
Mean	\$ 12,587.62
Std Deviation	\$ 10,948.39
Skewness	3.349066
Kurtosis	28.24214
Mode	\$ 4,458.97
5% Perc	\$ 2,655.55
95% Perc	\$ 32,481.38
Minimum	\$ 609.75
Maximum	\$194,355.00







- lacktriangle You are planning for retirement and decide to invest in the market for the next 30 years. Your initial capital is $v_0=1000$
- You have an opportunity to invest in stocks and Treasury bonds:
 - □ allocate 50% of your capital to the stock market (S&P 500 index fund) today
 - □ allocate 50% of your capital to bonds today
- Assume that every year your investment returns from investing into the S&P 500 and Treasury bonds will follow a Normal distribution with the mean and standard deviation as in example 2 (for S&P 500), mean 4% and standard deviation 7% for bonds. Assume correlation -0.2 between the stock market and the Treasury bond market.
- Covariance matrix:

$$\begin{pmatrix} 0.1465^2 & -0.2 \cdot 0.1465 \cdot 0.07 \\ -0.2 \cdot 0.1465 \cdot 0.07 & 0.07^2 \end{pmatrix} = \begin{pmatrix} 0.0215 & -0.0021 \\ -0.0021 & 0.0049 \end{pmatrix}$$

■ Value of investment after 30 years: *V*₃₀

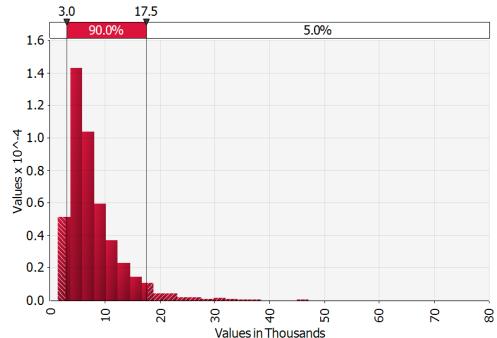
■ Simulate 30 years of 100 observations each of single period correlated returns:

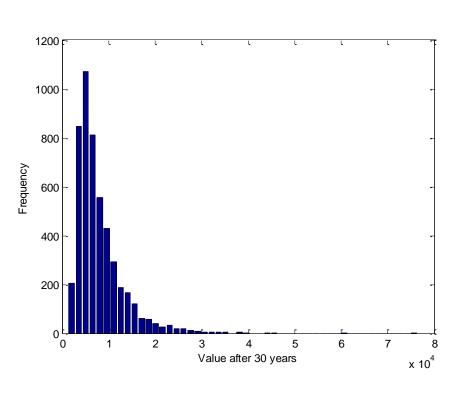
```
scenarios = random.multivariate_normal(mu, covMat, Ns)
for year in range(1, 31):
    scenarios = random.multivariate_normal(mu, covMat, Ns)
    stockRet *= (1 + scenarios[:,0])
    bondsRet *= (1 + scenarios[:,1])
```

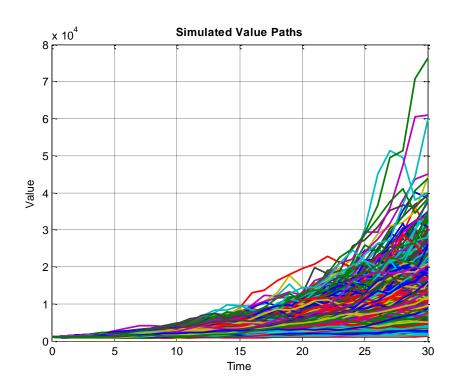
■ Compute and plot $\emph{v}_{30} = 0.5 v_0 (1 + \emph{r}_{0,30}^s) + 0.5 v_0 (1 + \emph{r}_{0,30}^b)$

Total amount in account

Number of values	5000
Mean	\$ 7,892.80
Std Deviation	\$ 5,233.10
Skewness	2.921482
Kurtosis	20.48869
Mode	\$ 5,050.96
5% Perc	\$ 2,951.82
95% Perc	\$17,457.43
Minimum	\$ 1,408.63
Maximum	\$79,729.34





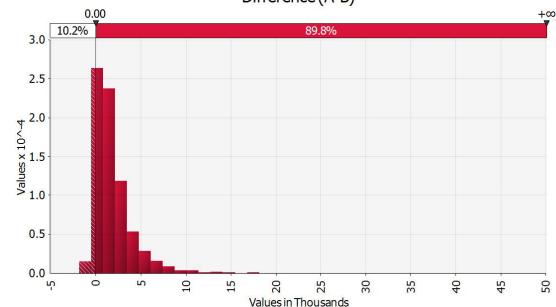


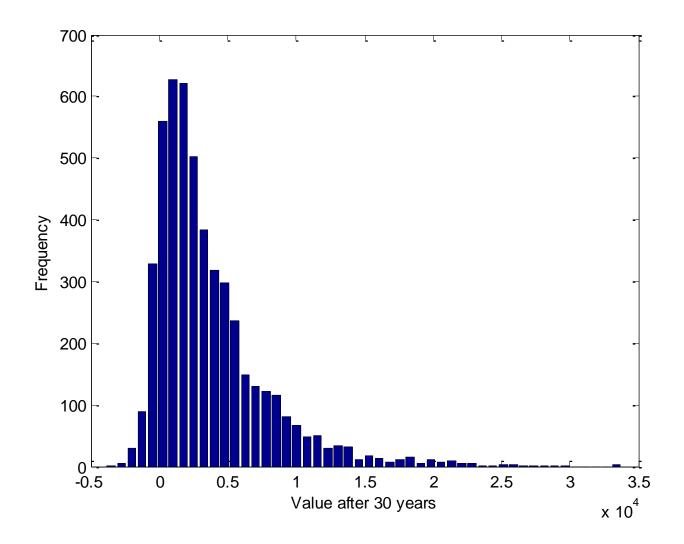
- Using scenario generation procedure from example 3 for decision-making
- Compare portfolios:
 - □ 50-50 portfolio allocation in stocks and bonds (Strategy A)
 - □ 30-70 portfolio allocation in stocks and bonds (Strategy B)

```
v30comp = []
for w in arange(0.2, 1.01, 0.2):
    v30comp += [w * v0 * stockRet + (1 - w) * v0 * bondsRet]
```

 \blacksquare Compute and plot $\emph{v}_{30}=w_sv_0(1+\emph{r}_{0,30}^s)+w_bv_0(1+\emph{r}_{0,30}^b)$

Number of values	5000
Mean	\$ 1,865.13
Std Deviation	\$ 2,214.87
Skewness	3.506451
Kurtosis	40.18968
Mode	\$ 687.75
5% Perc	\$ -254.41
95% Perc	\$ 6,027.23
Minimum	\$-1,829.78
Maximum	\$45,972.08

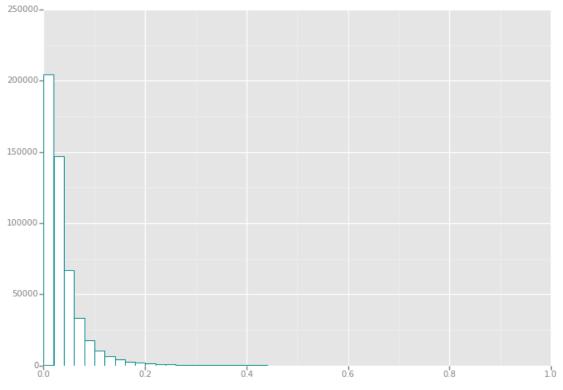




Simulation Modeling in Marketing – Marketing Campaign

Marketing campaign simulation modeling

- Case marketing campaign
- Data probability that client would buy a product
 - □ 500,000 clients
 - probability of buying a product for every client that is a result of another model
- Goal find target group parameters to maximize profit



Marketing campaign simulation modeling

Target function

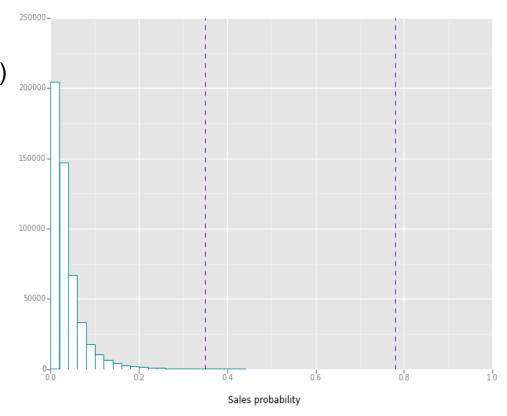
profit =
$$N_{\text{sales}} \cdot \text{avg}(\text{income}_{\text{sale}}) - N_{\text{contacts}} \cdot \text{avg}(\text{cost}_{\text{contact}})$$

- To compute profit we need information about sale income and contact cost
 - □ Sale income is \$10
 - □ Contact cost is \$2, it can be a distribution that we get from a more complex model for costs of contact, e.g., fixed cost plus variable cost based on duration of phone calls
- Our goal is
 - □ maximize sales, i.e., number of clients that buy the product
 - ☐ minimize cost of contacting each client, i.e., salary of client representatives
- Sales uplift we choose a simple model of "uplift", e.g., probability of sale will increase by 10% if a clients gets a call from the client representative
- Assume uniform distribution for the probability of sale
 - □ if random number generated from uniform(0,1) distribution is < prob of sale for that client -> client would buy the product
 - □ if random number generated from uniform(0,1) distribution is >= prob of sale for that client -> client would not buy the product

Marketing campaign simulation modeling

Algorithm

- Select target group parameters (min_probability, max_probability)
- 2. Compute contact uplift (add 0.1 to probability)
- 3. Simulate sales using obtained probabilities
- 4. Calculate profit function
- 5. Repeat 100 times to get average values
- Find the best target group parameters



- Final target group for contacting are clients with probabilities from 0.4 to 0.7
 - clients with low probability would not buy product even after being contacted
 - clients with high probability would buy the product without any additional stimulation

Simulation Modeling in Business – Restaurant Design

Business case study – optimal store design

Study environmental impact of restaurant operations

■ Restaurant

- order types and probabilities
- processing times (fixed portion and variable portion)
- design alternatives

■ Drive Through

- number of service windows
- queuing capacity

Parking Lot

- parking capacity
- customer prioritization

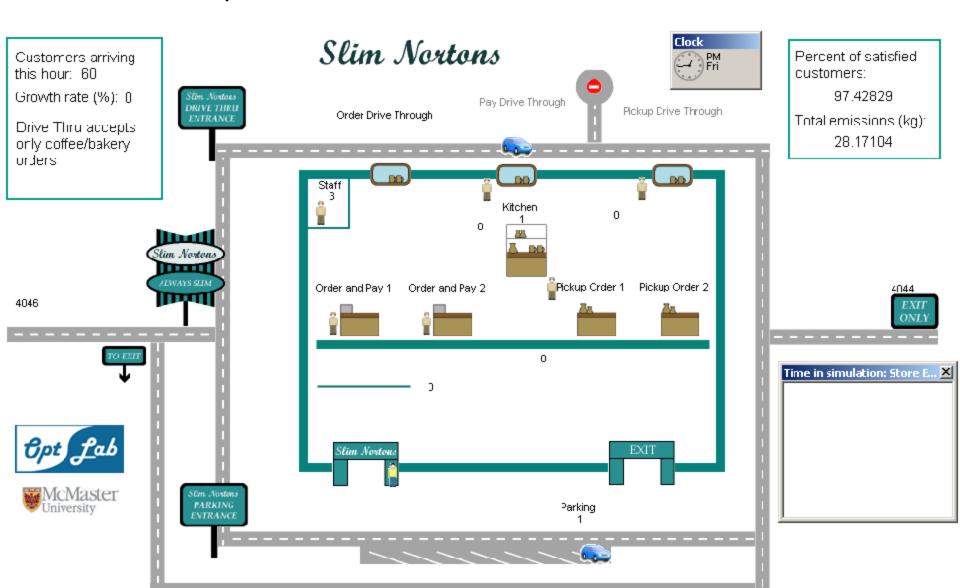
■ Goals:

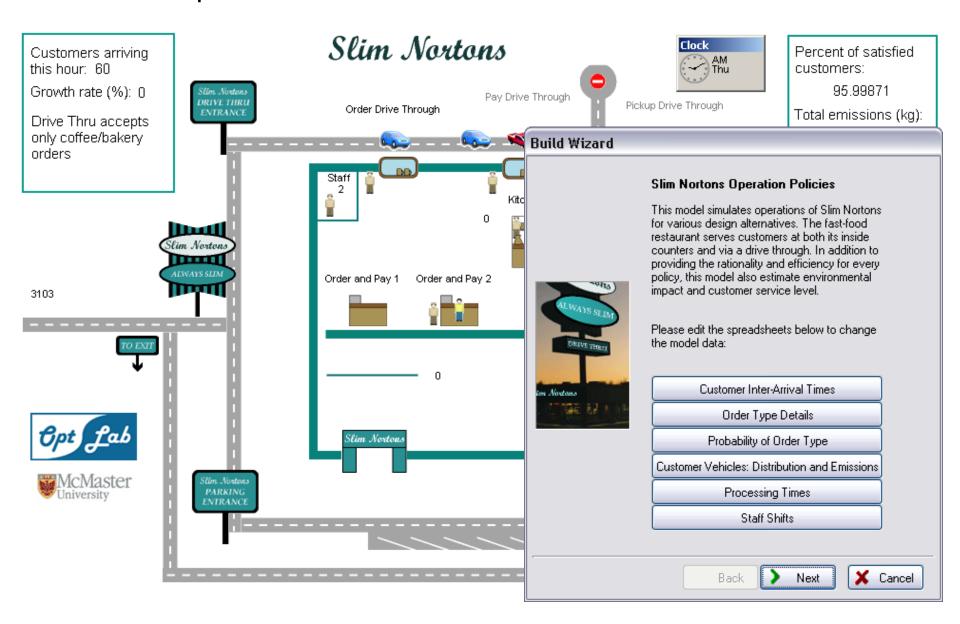
- maximize customer satisfaction (high customer service level)
- minimize environmental impact (quantity of emissions)

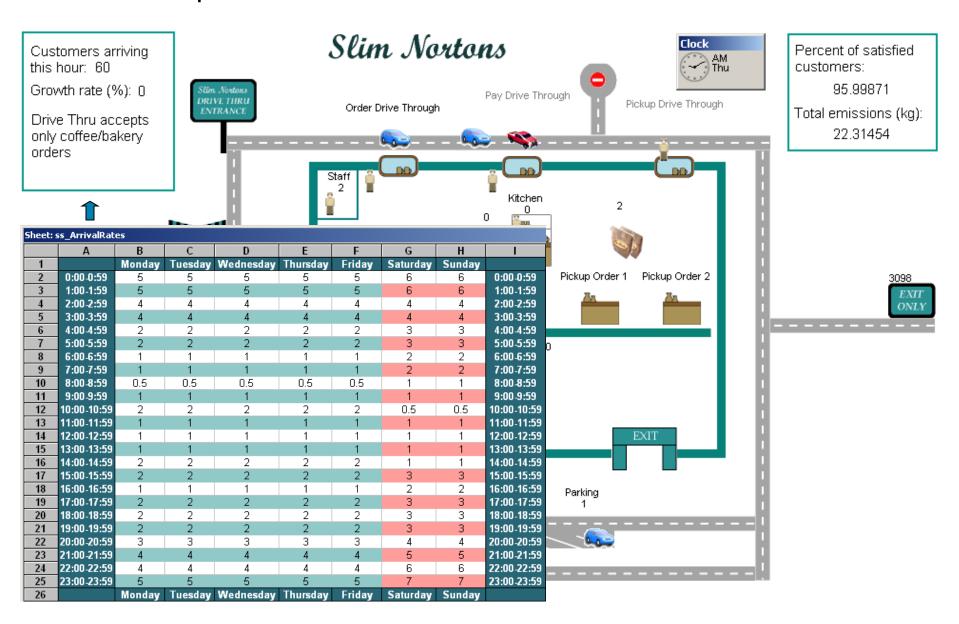
Business case study – outline

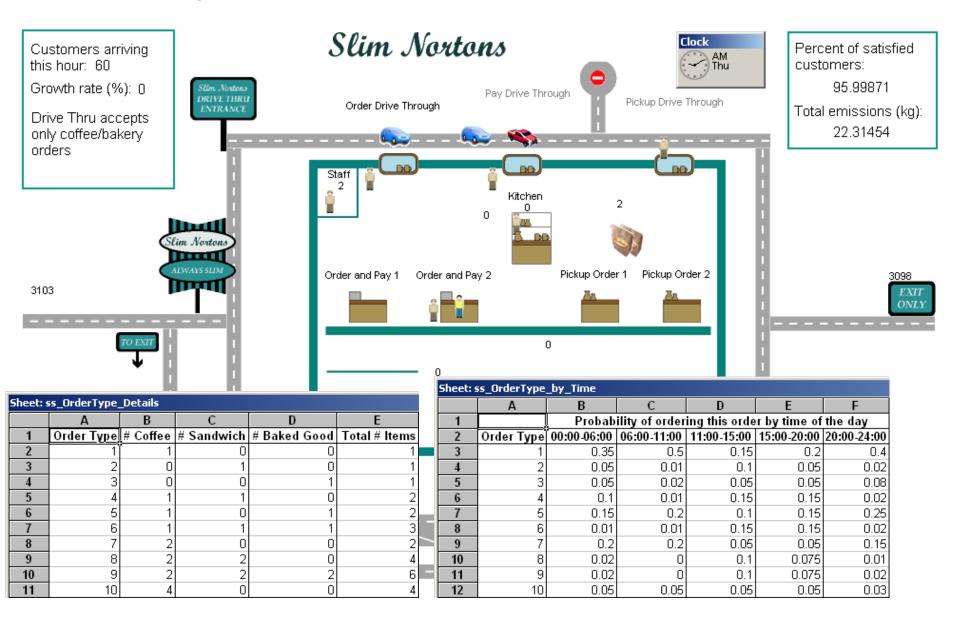
- Introduction
 - Problem description
 - Restaurant operations model
- Problems with the standard design
- Analysis of the alternative designs
- Additions to the optimal design solution
- Additional extensions and policies
- Conclusions and recommendations

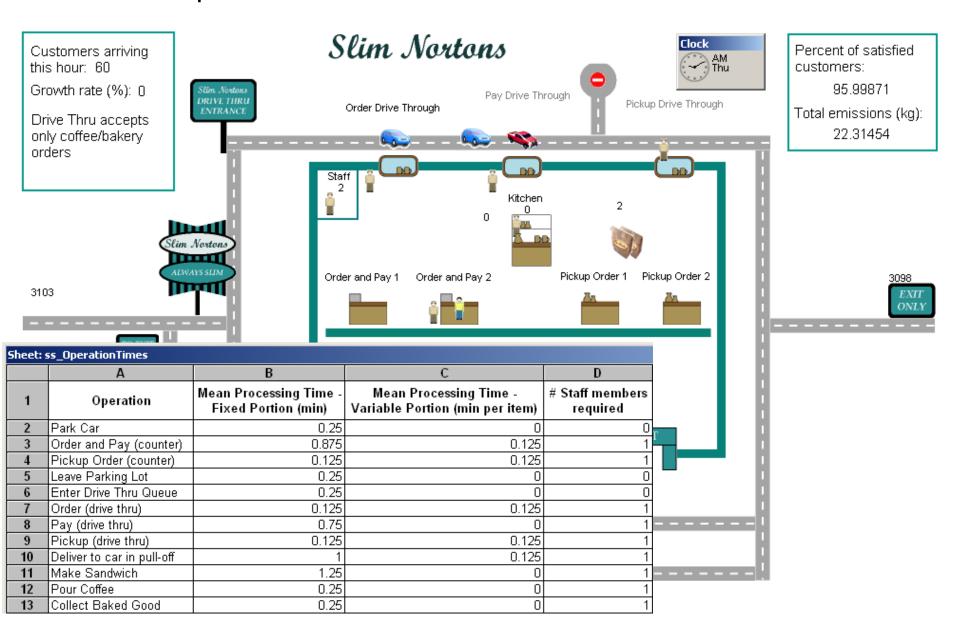
Problem description

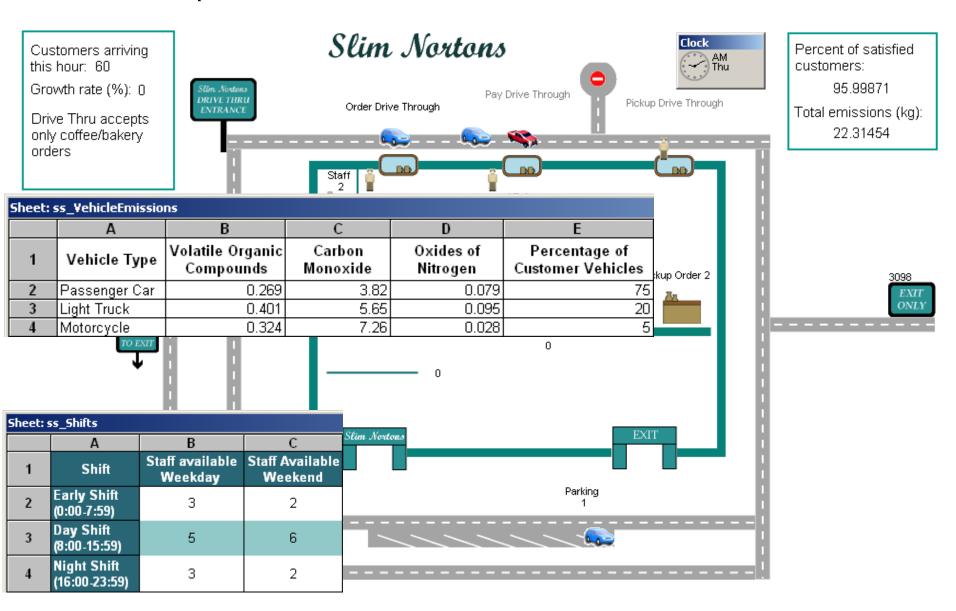


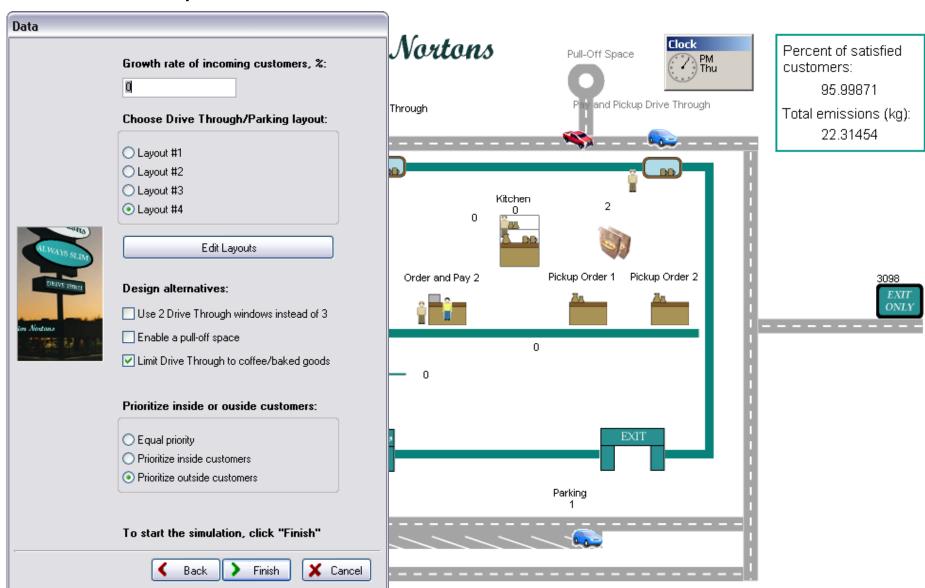




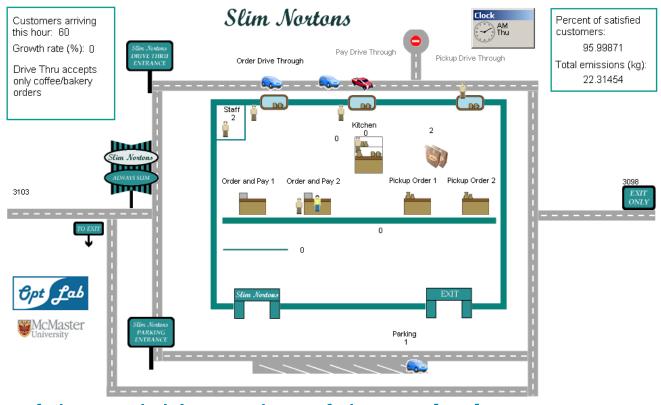








Problems with the standard design

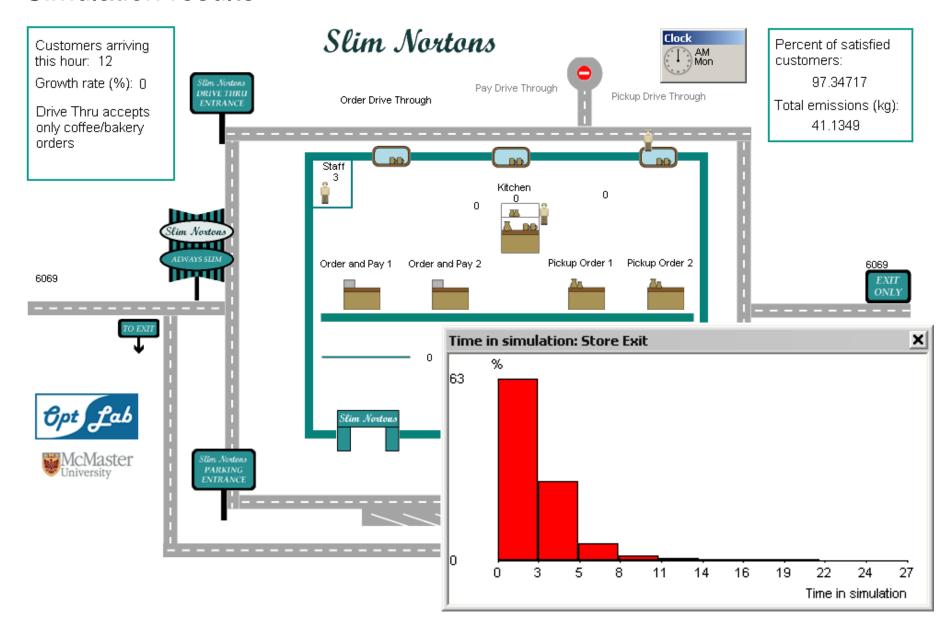


Most of the variable portion of the **emissions** are generated at the **drive through lane**

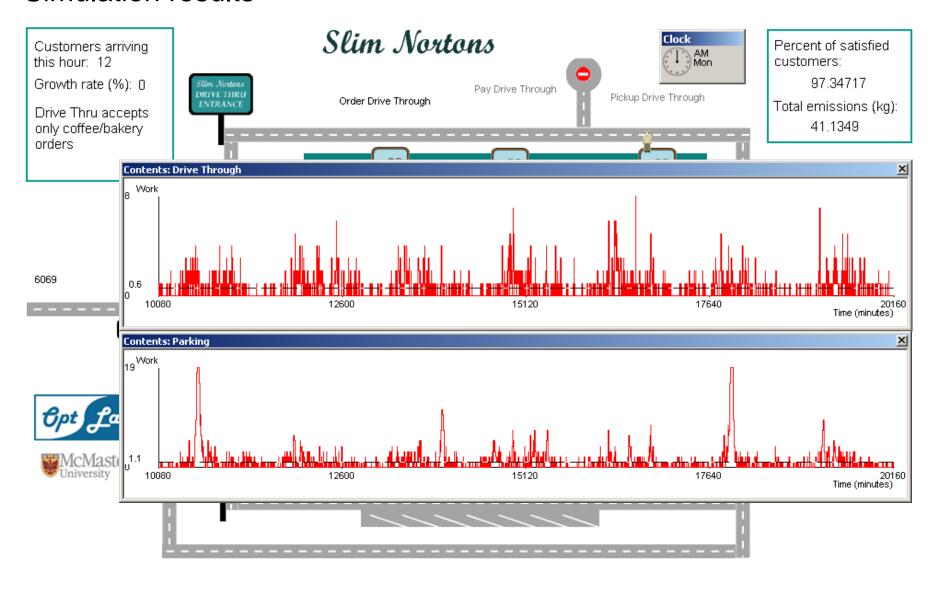
Customers should be encouraged to park their cars and enter the restaurant

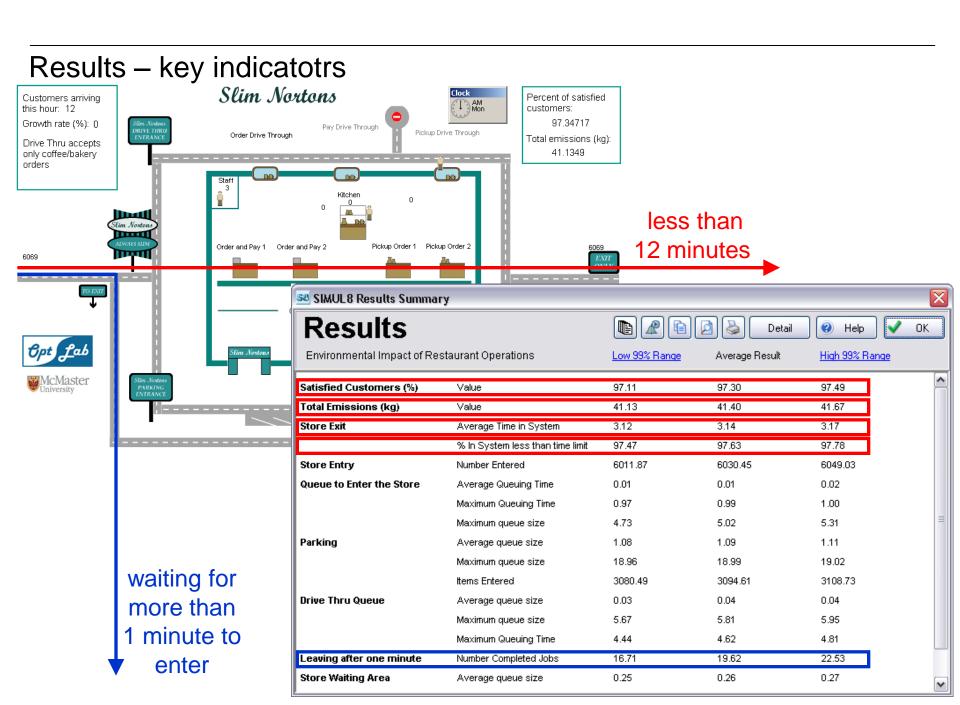
Drive through customers should be served as fast as possible

Simulation results



Simulation results

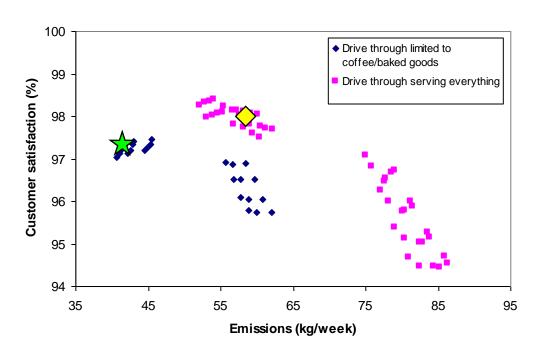




Results - alternatives

- Comparing 72 alternatives:
 - Limiting drive through to coffee/bakery orders
 - Pull-off space for large drive through orders
 - 2 or 3 service windows in drive through
 - Customer prioritization: inside, outside or equal
 - Varying queuing/parking capacity

Drive Through Food Variety



yes no 3 outside layout #4 (6/19)

Results – optimal staffing

	Original setup			Op	timal staf	fing
	Early	Day	Night	Early	Day	Night
	shift	shift	shift	shift	shift	shift
	0:00-	8:00-	16:00-	0:00-	8:00-	16:00-
	7:59	15:59	23:59	7:59	15:59	23:59
Weekday # staff available	3	5	3	3	5	3
Weekend # staff available	2	6	2	2	5	1

- The staffing pattern has significant impact on the overall throughput
- □ Staff utilization is around 38%

Introduction of more flexible shifts would result in 10-20% salary savings and virtually unchanged customer satisfaction and emission levels

Results – optimal staffing

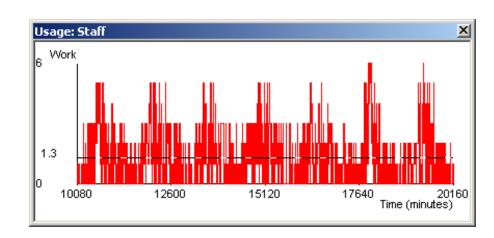
	Original setup			Optimal staffing		
	Early	Day	Night	Early	Day	Night
	shift	shift	shift	shift	shift	shift
	0:00-	8:00-	16:00-	0:00-	8:00-	16:00-
	7:59	15:59	23:59	7:59	15:59	23:59
Weekday # staff available	3	5	3	3	5	3
Weekend # staff available	2	6	2	2	5	1

Customer satisfaction: 97.31%

Emissions (kg/week): 41.39

Customer satisfaction: 96.55%

Emissions (kg/week): 42.44



Results – parking capacities

Reducing the parking lot or drive through queue capacity may be required in case of constructing the restaurant on a small piece of property

Drive through queue size	Parking lot size	Customer satisfaction	Emissions
6	19	97.31%	41.39
4	19	97.29%	41.02
3	19	97.27%	40.74
4	17	97.43%	41.04
2	17	97.41%	40.34
2	7	98.82%	40.42
2	5	98.83%	40.43
2	4	98.63%	40.46
2	2	95.43%	40.58

- ☐ The number of parking spots can be reduced from the original 19 to around 7-9
- The number of spots in the drive through queue can be reduced to around 2-3 from the original 6
- Current arrival rates do not justify large parking capacities

Results – store layout

Kitchen size:

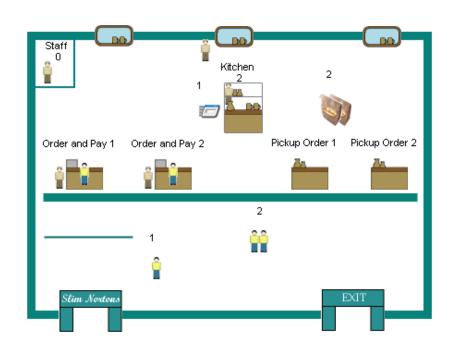


Kitchen size	Customer satisfaction	Emissions (kg/week)
5	97.31%	41.39
4	97.30%	41.03
3	97.20%	41.05
2	95.99%	41.21

Required: the space where at least 3-4 orders can be prepared simultaneously

Store waiting area size:

Required: space for at least 5-7 customers waiting to be served inside the restaurant



Results – effects of increased demand

- □ 10 kg/week reduction in greenhouse gas emissions translates into 5% increase in the number of customers
- □ Reduction for the proposed design is **17 kg/week**

Increase in customer base	Customer satisfaction (%)	Total emissions (kg/week)	Increase in emissions over projected (%)
0%	97.31	41.39	
2%	97.15	42.51	0.7%
4%	97.06	43.69	1.5%
6%	96.81	44.93	2.4%
8%	96.54	46.02	3.0%
10%	96.42	47.40	4.1%



- ☐ The store will be able to handle the increased demand while maintaining high customer service levels
- ☐ Gradually increasing staffing makes the proposed solution feasible over a long period of time

Additional extensions and policies

The "green" policy of the restaurant:

- make drive through more efficient or
- encourage customers to use parking lot instead
- Make orders more expensive for the drive through customers
 - equivalent of introducing the emission sales tax and can be justified from the environmental point of view
- Provide customers with the information about expected waiting times and greenhouse gas emissions per vehicle for the drive through lane and for using the parking lot
 - this information can be displayed on the illuminated indicator board (lighting panel) outside the restaurant

Recommendations

- We recommend implementing the following design:
 - Drive through limited to coffee and baked goods
 - No pull-off space
 - Separate pay and pickup windows at the drive through (3 service windows)
 - Priority given to drive through customers (or equal priority if any difficulties are expected with prioritizing the outside customers)
 - Any reasonable parking lot/drive through design would work (it depend more on the physical restrictions on the available space for the newly planned locations than on the other factors)
- Implement our additional recommendations about the staffing patterns and waiting area size as well as "green" policies

Optimal store design – conclusions

☐ The optimal restaurant design:

- allows reducing greenhouse gas emissions by 30% while keeping customer satisfaction level virtually unchanged
- indirectly enforces "green" customer behavior
- is sustainable in the long run
- additional policies can enforce "green" customer behavior directly

Problems:

- customers do not understand the problem of relatively high emissions while using the drive through as compared to parking
- legal and financial restrictions may prevent implementing optimal "green" policy of the restaurant
- the staffing patterns are not 100% efficient and do not follow well changes in the customer arrival rates

Simulation Software

Simulation software

What is AnyLogic?