MTAT.03.231
Business Process Management

Lecture 6 – Quantitative Process Analysis II

Marlon Dumas

marlon.dumas ät ut . ee
Process Analysis

1. Process identification
   - Process architecture
   - Conformance and performance insights

2. Process discovery
   - As-is process model

3. Process monitoring and controlling
   - Executable process model
   - Insights on weaknesses and their impact

4. Process implementation
   - To-be process model
   - Process redesign
Process Analysis Techniques

**Qualitative analysis**
- Value-Added & Waste Analysis
- Root-Cause Analysis
- Pareto Analysis
- Issue Register

**Quantitative Analysis**
- Flow analysis
- Queuing analysis
- Simulation
1. Introduction
2. Process Identification
3. Essential Process Modeling
4. Advanced Process Modeling
5. Process Discovery
6. Qualitative Process Analysis
7. **Quantitative Process Analysis**
8. Process Redesign
9. Process Automation
10. Process Intelligence
Why flow analysis is not enough?

Flow analysis does not consider waiting times due to resource contention.

Queuing analysis and simulation address these limitations and have a broader applicability.
Queuing Analysis

• Capacity problems are common and a key driver of process redesign
  • Need to balance the cost of increased capacity against the gains of increased productivity and service
• Queuing and waiting time analysis is particularly important in service systems
  • Large costs of waiting and/or lost sales due to waiting

Prototype Example – ER at a Hospital
• Patients arrive by ambulance or by their own accord
• One doctor is always on duty
• More patients seeks help ⇒ longer waiting times

➤ Question: Should another MD position be instated?
Delay is Caused by Job Interference

If arrivals are regular or sufficiently spaced apart, no queuing delay occurs.

- **Deterministic traffic**
- **Variable but spaced apart traffic**

© Dimitri P. Bertsekas
Burstiness Causes Interference

Queuing results from variability in processing times and/or interarrival intervals
Job Size Variation Causes Interference

- Deterministic arrivals, variable job sizes

Queuing Delays
High Utilization Exacerbates Interference

- The queuing probability increases as the load increases
- Utilization close to 100% is unsustainable → too long queuing times

© Dimitri P. Bertsekas
The Poisson Process

• Common arrival assumption in many queuing and simulation models
• The times between arrivals are independent, identically distributed and exponential
  • $P(\text{arrival } < t) = 1 - e^{-\lambda t}$
• Key property: The fact that a certain event has not happened tells us nothing about how long it will take before it happens
  • e.g., $P(X > 40 | X \geq 30) = P(X > 10)$
Negative Exponential Distribution
Basic characteristics:

- $\lambda$ (mean arrival rate) = average number of arrivals per time unit
- $\mu$ (mean service rate) = average number of jobs that can be handled by one server per time unit:
- $c = \text{number of servers}$
Given $\lambda$, $\mu$ and $c$, we can calculate:

- **occupation rate**: $\rho$
- $W_q = \text{average time in queue}$
- $W = \text{average system in system (i.e. cycle time)}$
- $L_q = \text{average number in queue (i.e. length of queue)}$
- $L = \text{average number in system average (i.e. Work-in-Progress)}$
M/M/1 queue

Assumptions:
• time between arrivals and processing time follow a negative exponential distribution
• 1 server (c = 1)
• FIFO

\[
\rho = \frac{\text{Capacity Demand}}{\text{Available Capacity}} = \frac{\lambda}{\mu}
\]

\[
L = \frac{\rho}{1 - \rho} \\
W = \frac{L}{\lambda} = \frac{1}{\mu - \lambda}
\]

\[
L_q = \frac{\rho^2}{1 - \rho} = L - \rho \\
W_q = \frac{L_q}{\lambda} = \frac{\lambda}{\mu(\mu - \lambda)}
\]
M/M/c queue

- Now there are c servers in parallel, so the expected capacity per time unit is then $c^*\mu$

\[
\rho = \frac{\text{Capacity Demand}}{\text{Available Capacity}} = \frac{\lambda}{c^*\mu}
\]

**Little’s Formula**  $\Rightarrow$  $W_q = L_q / \lambda$

\[
W = W_q + (1/\mu)
\]

**Little’s Formula**  $\Rightarrow$  $L = \lambda W$
Tool Support

• For M/M/c systems, the exact computation of $L_q$ is rather complex...

\[
L_q = \sum_{n=c}^{\infty} (n-c)P_n = \ldots = \frac{(\lambda/\mu)^c \rho}{c!(1-\rho)^2} P_0
\]

\[
P_0 = \left( \sum_{n=0}^{c-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^c}{c!} \cdot \frac{1}{1-(\lambda/(c\mu))} \right)^{-1}
\]

• Consider using a tool, e.g.
  - \[\text{http://queueingtoolpak.org/}\] (for Excel)
  - \[\text{http://www.stat.auckland.ac.nz/~stats255/qsim/qsim.html}\]
Example – ER at County Hospital

- **Situation**
  - Patients arrive according to a Poisson process with intensity $\lambda$ ($\Leftrightarrow$ the time between arrivals is $\exp(\lambda)$ distributed).
  - The service time (the doctor’s examination and treatment time of a patient) follows an exponential distribution with mean $1/\mu$ ($=\exp(\mu)$ distributed)
    \[\Rightarrow \text{The ER can be modeled as an M/M/c system where } c = \text{the number of doctors}\]

- **Data gathering**
  \[\Rightarrow \lambda = 2 \text{ patients per hour}\]
  \[\Rightarrow \mu = 3 \text{ patients per hour}\]

- **Question**
  - Should the capacity be increased from 1 to 2 doctors?

© Laguna & Marklund
Queuing Analysis – Hospital Scenario

- Interpretation
  - To be in the queue = to be in the waiting room
  - To be in the system = to be in the ER (waiting or under treatment)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>One doctor (c=1)</th>
<th>Two Doctors (c=2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>$2/3$</td>
<td>$1/3$</td>
</tr>
<tr>
<td>$L_q$</td>
<td>$4/3$ patients</td>
<td>$1/12$ patients</td>
</tr>
<tr>
<td>$L$</td>
<td>$2$ patients</td>
<td>$3/4$ patients</td>
</tr>
<tr>
<td>$W_q$</td>
<td>$2/3$ h = 40 minutes</td>
<td>$1/24$ h = 2.5 minutes</td>
</tr>
<tr>
<td>$W$</td>
<td>$1$ h</td>
<td>$3/8$ h = 22.5 minutes</td>
</tr>
</tbody>
</table>

- Is it warranted to hire a second doctor?
Your turn

• Textbook, Chapter 7, exercise 7.12 (page 249)
Simulation
Process Simulation

- Versatile quantitative analysis method for
  - As-is analysis
  - What-if analysis
- In a nutshell:
  - Run a large number of process instances
  - Gather performance data (cost, time, resource usage)
  - Calculate statistics from the collected data
Process Simulation

1. Model the process
2. Define a simulation scenario
3. Run the simulation
4. Analyze the simulation outputs
5. Repeat for alternative scenarios
Example
Example
Elements of a simulation scenario

1. Processing times of activities
   - Fixed value
   - Probability distribution
Exponential Distribution
Normal Distribution

\[ \mu = 0, \sigma^2 = 0.2 \]
\[ \mu = 0, \sigma^2 = 1.0 \]
\[ \mu = 0, \sigma^2 = 5.0 \]
\[ \mu = -2, \sigma^2 = 0.5 \]
Choice of probability distribution

• Fixed
  • Rare, can be used to approximate case where the activity processing time varies very little
  • Example: a task performed by a software application

• Normal
  • Repetitive activities
  • Example: “Check completeness of an application”

• Exponential
  • Complex activities that may involve analysis or decisions
  • Example: “Assess an application”
Simulation Example

Normal(10m, 2m)
- Check credit history
- Check income sources

Exp(20m)
- Assess application
- Notify rejection

Normal(10m, 2m)
- Make credit offer
- Receive customer feedback

Normal(20m, 4m)

0m
- Credit application processed
- Exit point
Elements of a simulation model

1. Processing times of activities
   • Fixed value
   • Probability distribution

2. Conditional branching probabilities

3. Arrival rate of process instances and probability distribution
   • Typically exponential distribution with a given mean inter-arrival time
   • Arrival calendar, e.g. Monday-Friday, 9am-5pm, or 24/7
Branching probability and arrival rate

Arrival rate = 2 applications per hour
Inter-arrival time = 0.5 hour
Negative exponential distribution
From Monday-Friday, 9am-5pm
Elements of a simulation model

1. Processing times of activities
   • Fixed value
   • Probability distribution

2. Conditional branching probabilities

3. Arrival rate of process instances
   • Typically exponential distribution with a given mean inter-arrival time
   • Arrival calendar, e.g. Monday-Friday, 9am-5pm, or 24/7

4. Resource pools
Resource pools

- Name
- Size of the resource pool
- Cost per time unit of a resource in the pool
- Availability of the pool (working calendar)

Examples
- Clerk: Credit Officer
- € 25 per hour: € 25 per hour
- Monday-Friday, 9am-5pm: Monday-Friday, 9am-5pm

In some tools, it is possible to define cost and calendar per resource, rather than for entire resource pool
Elements of a simulation model

1. Processing times of activities
   - Fixed value
   - Probability distribution

2. Conditional branching probabilities

3. Arrival rate of process instances and probability distribution
   - Typically exponential distribution with a given mean inter-arrival time
   - Arrival calendar, e.g. Monday-Friday, 9am-5pm, or 24/7

4. Resource pools

5. Assignment of tasks to resource pools
Resource pool assignment
Process Simulation

1. Model the process
2. Define a simulation scenario
3. Run the simulation
4. Analyze the simulation outputs
5. Repeat for alternative scenarios
Output: Performance measures & histograms
Process Simulation

Model the process

Define a simulation scenario

Run the simulation

Analyze the simulation outputs

Repeat for alternative scenarios
Tools for Process Simulation

- ARIS
- Bizagi Process Modeler
- ITP Commerce Process Modeler for Visio
- Logizian
- Oracle BPA
- Progress Savvion Process Modeler
- ProSim
- Signavio + BIMP
BIMP – bimp.cs.ut.ee

- Accepts standard BPMN 2.0 as input
- Simple form-based interface to enter simulation scenario
- Produces KPIs + simulation logs in MXML format
  - Simulation logs can be imported to the ProM process mining tool
BIMP Demo
Your turn

• Textbook, Chapter 7, exercise 7.8 (page 240)
Pitfalls of simulation

• Stochasticity
• Data quality pitfalls
• Simplifying assumptions
Stochasticity

• Simulation results may differ from one run to another
• Make the simulation timeframe long enough to cover weekly and seasonal variability, where applicable
• Use multiple simulation runs
• Average results of multiple runs, compute confidence intervals
Data quality pitfalls

• Simulation results are only as trustworthy as the input data
• Rely as little as possible on “guesstimates”
• Use input analysis
  • Deriver simulation scenario parameters from numbers in the scenario
  • Use statistical tools to check fit the probability distributions
• Simulate the “as is” scenario and cross-check results against actual observations
Simulation assumptions

- That the process model is always followed to the letter
  - No deviations
  - No workarounds
- That resources work constantly and non-stop
  - Every day is the same!
  - No tiredness effects
  - No distractions beyond “stochastic” ones
Next week