Data Assimilation for Discrete-event Simulations

Peter W. Glynn

Societal Networks Workshop, Simons Institute, UC-Berkeley, March 27, 2018

What is “Date Assimilation”?

Weather prediction:

- How to initialize PDE-based circulation models?
- Data gathered from many different sources: ground stations, satellites, weather ships, etc.

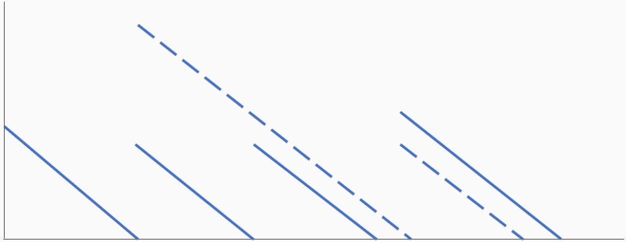
What are Discrete-event Simulations?

The type of dynamic simulation that arises in dealing with discrete entities:

- queues/customers
- inventory
- production systems
- ride-sharing platforms

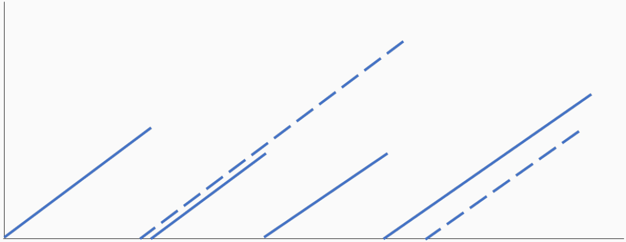
etc

G/G/1 Queue



“Residual Time” Representation

G/G/1 Queue



"Elapsed Time" Representation

Mathematical Formalism: Generalized Semi-Markov Processes (GSMP's)

Residual time Markov process:

$$(S(t), C(t))$$

Elapsed time Markov process:

$$(S(t), C(t))$$

East German probability school initiated study of GSMPs in 60s through 80s...

An Example: Container Fleet for Auto-maker

- multi-billion dollar inventory
- factory shuts down if it runs out of containers filled with incoming parts or containers to hold outgoing components
- very few containers have digital tags
- only know when containers pass through trans-shipment points

Decision:

- How to re-allocate containers to avoid shortages?

- GSMP Markov process:

$$X(t) = (S(t), C(t))$$

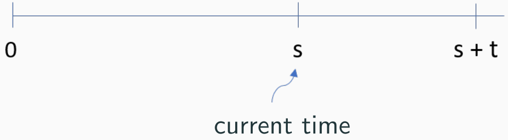
- Observables:

$$Z(t) = k(S(t)),$$

  where $k(\cdot)$ may significantly "collapse" $S(t)$

- We wish to compute:

$$\mathbb{E}[Y(s+t)|Z(u) : 0 \leq u \leq s]$$



$$
\begin{aligned}
&\mathbb{E}[Y(s+t)|Z(u) : 0 \leq u \leq s] \\
&= \int \mathbb{E}[Y(s+t)|X(s) = x] P(X(s) \in dx | Z(u) : 0 \leq u \leq s)
\end{aligned}
$$

If we can sample from

$$
P(X(s) \in \cdot | Z(u) : 0 \leq u \leq s),
$$

we can run simulation forward to time $s + t$

This is a filtering problem...

Evolution Equation for Filter for Elapsed Time Markov Process (G + Haas (2018))

Between jumps in $k(S(t))$:

$$
\begin{aligned}
&\frac{\partial}{\partial t} v(t, s', c) \\
&= - \sum_{e \in E(s')} \frac{\partial}{\partial c_e} v(t, s', c) - \sum_{e \in E(s')} v(t, s', c) r_e(c_e) \\
&\quad + v(t, s', c) \sum_{s'' \in k^{-1}(Z(t))} \sum_{e \in E(s')} \int v(t, s', c) r_e(c_e) dc_e \\
&\quad - v(t, s', c) \sum_{s'' \in k^{-1}(Z(t))} \sum_{s \in k^{-1}(Z(t))} \sum_{e' \in E(s)} \int v(t, s, c') r_{e'}(c''_e) p(t', s, e') dc' dc''
\end{aligned}
$$

A different update at the jump times of $Z(t) = k(S(t))\dots$

Another set of filtering equations for the “residual time” Markov process

Challenges:

- very high-dimensional
- Transition density intractable
- Particle filtering highly non-trivial

- Instead, we implement something simpler
- Sample from

  $$P(X(s) \in \cdot | Z(s))$$

  rather than

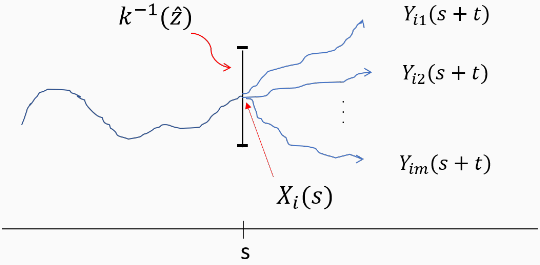
  $$P(X(s) \in c \cdot | Z(u) : 0 \leq u \leq s)$$

- Put $Z(s) = \hat{z}$; we want

  $$P(X(s) \in \cdot | Z(s) = \hat{z})$$

- If $P(Z(s) = \hat{z})$ is not too small, we can approximate via

$$\frac{\sum_{i=1}^{n} I(X_i(s) \in \cdot, Z_i(s) = \hat{z})}{\sum_{i=1}^{n} I(Z_i(s) = \hat{z})}$$



- To compute $\mathbb{E}[Y(s+t)|Z(s)=\hat{z}]$, we now run m simulations forward to time $s+t$

$$\frac{\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m} Y_{ij}(s+t)I(Z_i(s)=\hat{z})}{\sum_{i=1}^{n} I(Z_i(s)=\hat{z})}$$

Splitting...

What about if $P(Z(s) = \hat{z})$ is small?

e.g. container problem

Even though $k(\cdot)$ is typically discrete, we can apply density estimation ideas to an estimator with "local averaging":

$$\hat{P}_n(X(s) \in \cdot | Z(s) = \hat{z}) = \frac{\sum_{i=1}^n I(X_i(s) \in \cdot)\rho_n(Z_i(s), \hat{z})}{\sum_{i=1}^n \rho_n(Z_i(s), \hat{z})}$$

The smoothing kernel $\rho_n(\cdot)$ can be chosen using similar principles as in the standard (continuous) setting.

Justification: Imagine system as being embedded in a "fluid limit" environment

"discrete system can be approximated via continuous system"

Other Challenges/Issues

- Need to simultaneously estimate model parameters
- Possibility of assessing “model error” based on

$$Y(s+t) - \mathbb{E}[Y(s+t)|Z(s)]$$

Real Data

Simulation

Conclusions

- Discrete-simulation is likely to find increasing usage as a real-time decision-making tool
- Mis-match between "mathematical state" underlying simulation and what is observable leads to a very hard filtering problem with non-standard features
- Even initializing with $P(X(s) \in \cdot | Z(s) = \hat{z})$ leads to computational challenges