Statistical model checking based analysis of biopathway models

P.S. Thiagarajan
Dept of Computational and Systems Biology
University of Pittsburgh
(Faeder Lab)
Classical model checking

• *Verify* if a *model* satisfies a specified *property*

• model
  - a *dynamical system*
    - hardware circuits
    - Programs
    - .....
Classical model checking

• Properties:
  - Assertions about the executions (trajectories) of the dynamical system.
    - At some time in the future the program will terminate ($\phi_1$)
    - Starting from now at every time it will be the case that $x_1 + x_2 = 100$ ($\phi_2$)

• $l_1$: Input $x_1$, $x_2$
• $l_2$: while $x_1 > 0$:
  - $x_1 := x_1 - 1$; $x_2 := x_2 + 1$
• $l_3$: stop
Classical model checking

• Model satisfies a property if *every run/execution* of the model satisfies the property.

• *At some time* in the future the program will terminate ($\phi_1$)

• Starting from now *at every time* it will be the case that $x_1 + x_2 = 100$ ($\phi_2$)

• $l_1$: Input $x_1$, $x_2$
• $l_2$: while $x_1 > 0$:
  
  $x_1 := x_1-1; x_2 := x_2+1$
• $l_3$: stop

$\phi_1$ is satisfied by the program

$\phi_2$ is *not* satisfied by the program
Classical model checking

• Properties:
  - specified as temporal logic formulas

• At some time in the future the program will terminate
  \( F(l3) \)

• Starting from now at every time it will be the case that \( x1 + x2 = 100 \)
  \( G(x1+x2 == 100) \)
Classical model checking

• Properties:
  - specified as *temporal logic formulas*
• future (F), always (G), until (U), next (X)
• and, or, not ....

• Precise
  - (machine readable) syntax
  - mathematical semantics

Amir Pnueli
(Turing Award 1996)
Classical model checking

• Properties:
  □ specified as *temporal logic formulas*
• future (F), always (G), until (U), next (X)
• and, or, not ....

• Precise
  □ (machine readable) syntax
  □ mathematical semantics
• The model checking problem can be solved *automatically!*

Ed Clarke

Alan Emerson

Turing award 2007

Joseph Sifakis
Probabilistic model checking

• *Verify* if a *model* satisfies a specified property *with a certain probability.*

• Models:
  - Stochastic dynamical systems
    - Discrete time Markov chain
    - Continuous time Markov chain

• Model satisfies a property with probability $p$ if:
  - The probability of a *randomly chosen run/execution* of the model satisfying the property is $p$.

• This is hard problem!
Statistical model checking

- Probabilistic model checking via:
  - *sequential hypothesis testing.*
- H0: $P(\varphi) \geq r$ (null hypothesis)
- H1: $P(\varphi) < r$ (alternative hypothesis)
- $r$ chosen by the user.
- User also fixes
  - $\alpha$ - false positives probability
  - $\beta$ - false negatives probability
- These parameters determine the thresholds $L$ and $U$
Statistical model checking

- $\varphi$ – “within two steps the state $F$ will be reached”
- $H_0$: $P(\varphi) \geq 0.8$
- $H_1$: $P(\varphi) < 0.8$
- $\alpha = \beta = 0.05$
- $L, U$

![Diagram](image-url)
Statistical model checking

• ϕ – “within two steps the state F will be reached”
• Suppose m sample trajectories have been drawn so far
• and the test ratio value $K_m$ lies between L and U
• Draw one more sample trajectory $σ$. 

<table>
<thead>
<tr>
<th>$σ$</th>
<th>A</th>
<th>B</th>
<th>F</th>
<th>$σ$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>A</td>
<td>F</td>
<td>F</td>
<td>Yes</td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>B</td>
<td>F</td>
<td>yes</td>
</tr>
</tbody>
</table>
Statistical model checking

• If $\sigma$ satisfies $\varphi$, increase $K_m$ to $K_{m+1}$
• else decrease $K_m$ to $K_{m+1}$
• If $K_{m+1} > U$ accept $H_0$ and stop
• If $K_{m+1} < L$ accept $H_1$ and stop
• Else draw one more sample and repeat.
Statistical model checking

• The hypothesis test is guaranteed to terminate with probability 1.
• Surprisingly few samples need to be drawn in practice
• Complexity depends on the hypothesis test parameters only
  ♣ Cost of drawing a sample will depend on the dimension of the system.
• Amenable to parallel implementation
• Scales well

Goal

- Apply the SMC method to analyze:
  - ODEs based models of biochemical networks.
- Parameter estimation
- Sensitivity analysis
- Model check (probabilistically, approximately) for properties.
- Assume a set (interval) of initial values:
  - For the variables
- Assume distributions over these sets of initial values.

\[
\begin{align*}
\frac{dS}{dt} &= -k_1 \cdot [E][S] + k_2 \cdot [ES] \\
\frac{dES}{dt} &= k_1 \cdot [E][S] - (k_2 + k_3) \cdot [ES] \\
\frac{dE}{dt} &= -k_1 \cdot [E][S] + (k_2 + k_3) \cdot [ES] \\
\frac{dP}{dt} &= k_3 \cdot [ES]
\end{align*}
\]
SMC for ODEs

• Assume a set (interval) of initial values:
  • For the variables
  • Assume for now all the rate constants are known

• Assume distributions over these sets of initial values.
  • Uniform
  • Normal
  • Log uniform
  • lognormal
\( \psi \) a BLTL formula

\[
TRJ_\psi = \{ \tau \mid \tau \text{ satisfies } \psi \}
\]

\[
P(\psi) = \frac{\#TRJ_\psi}{\#TRJ} = \frac{\mu(INIT_\psi)}{\mu(INIT)}
\]

\[
INIT_\psi = \{ \tau(0) \mid \tau \text{ satisfies } \psi \text{ and } \tau(0) \text{ in INIT} \}
\]
\[ T \text{R}J_\psi = \{ \tau \mid \tau \text{ satisfies } \psi \} \]

\[ P(\psi) = \frac{\#T \text{R}J_\psi}{\#T \text{R}J} = \frac{\mu(INIT_\psi)}{\mu(INIT)} \]

\[ INIT_\psi = \{ \tau(0) \mid \tau \text{ satisfies } \psi \text{ and } \tau(0) \text{ in INIT} \} \]

\( P(\psi) \) is well-defined because:
- The assumed continuity properties of the ODEs system
- BLTL semantics
- Basic measure theory
\[ \text{INIT}_\psi = \{ \tau(0) \mid \tau \text{ satisfies } \psi \} \]

\[ P(\psi) = \frac{\mu(\text{INIT}_\psi)}{\mu(\text{INIT})} \]
\[ \text{INIT}_\psi = \{ \tau(0) \mid \tau \text{ satisfies } \psi \} \]

\[ P(\psi) = \frac{\mu(\text{INIT}_\psi)}{\mu(\text{INIT})} = P(\text{INIT}_\psi) \]

- We can estimate \( P(\psi) \) by:
  - Estimating \( P(\text{INIT}_\psi) \)
  - Using the given distribution over \( \text{INIT} \)
For an ODEs system:
Given a distribution over the initial values sets
We can estimate/bound the probability of the system satisfying the property $\Psi$

\[
INIT_\psi = \{ \tau(0) \mid \tau \text{ satisfies } \psi \}
\]

\[
P(\psi) = \frac{\mu(INIT_\psi)}{\mu(INIT)} = P(INIT_\psi)
\]

- Use SMC to estimate $P(INIT_\psi)$
  - Sample a point $x_0$ from INIT
  - Generate a trajectory $\sigma$ starting from
  - Check if $\sigma$ satisfies $\psi$
  - .....
Parameter estimation

• Given an ODEs system:
  - Assume distributions over initial values sets
  - Assume distributions over intervals of values for unknown parameters
  - *Encode quantitative experimental data and known qualitative properties as a conjunction of BLTL formulas.*
  - Use SMC to evaluate the objective value of the current set of parameters
  - Use standard search techniques to traverse the parameter space.
Data encoding

• Quantitative experimental data
  - At time $t$ the value of the variable $x$ was observed to lie in the interval $[l, u]$
    - $F_t(l \leq x \text{ and } x \leq u)$
  - $\Psi_{\exp}$ – the conjunction of all such data point formulas.
Data encoding

• Known Qualitative trends
  
  - ERK concentration reaches a peak value and then drops off to a low value for good.
    
    - \( F([\text{ERK}] > 4.8 \text{ and } F(G([\text{ERK}] \leq 0.2)) \)

  - transient/sustained activation, oscillatory behavior, bistable, ...

• \( \Psi_{\text{qlt}} \) - the conjunction of all qualitative properties.

\[
\Pr_{_{\geq r}} (\Psi_{\text{exp}} \land \Psi_{\text{qlt}})
\]
SMC based Parameter Estimation

1. Guess $\mathcal{G}_l$
2. Verify $\psi_{\text{exp}} \land \psi_{\text{qly}}$ with the chosen strength
3. Compute $F(\mathcal{G}_l)$
4. Terminate or make a new guess (based on SRES) and repeat step 1

$$F(\theta) = J_{\text{qly}}(\theta) + \sum_{i \in O} \frac{J^i_{\text{exp}}}{J^i_{\text{exp}}}$$
MLC Phosphorylation Pathway

• Regulates the contraction of endothelial cells

• ODE model (Maeda et al 2006)
  • 105 species, 197 parameters (100 unknown parameters)

• Synthetic training data
  • Time serials: 10 species, 20 time points
  • Qualitative trend: 2 species

• Synthetic test data
  • 2 species, 12 time points

Maeda A¹ et al. Ca²⁺-independent phospholipase A2-dependent sustained Rho-kinase activation exhibits all-or-none response. Genes Cells. 2006 Sep;11(9):1071-83
MLC Phosphorylation Pathway

[Graphs of Rho.GTP, PKC.DAG, MYPT1_PPase₂, and MYPT1.Rho-kinase₂]
• TLR3 activation followed by TLR7 activation leads to synergistic production of cytokines
• Investigated the cross talk mechanism causing this synergy

Model Calibration using Training Data

112 ODEs

129 unknown parameters
Model Calibration and Validation

- Test data: [IL6mRNA], [IL12mRNA] at {0, 4, 8, 12, 16, 24, 28, 32, 40, 48 h}
The main findings

The JAK-STAT1/2 pathway is the main mechanism responsible for the induction of synergistic cytokine production.

The cytokine response is biphasic due to an incoherent type I feedforward loop.
QSP model of Sanofi’s bispecific antibody

- SAR440234 is a bispecific antibody
  - capable of co-engaging the CD3 receptor on T cells and
  - the CD123 receptor
    - highly expressed on AML blasts
- Two level model to capture:
  - PK dynamics
  - Synapse formation
  - Killing of Cd123+ cells (AML blasts)
  - Cytokines release
Going forward

- Add the SMC based method to the BioNetGen toolkit
- Current solutions: PTEMPEST, BioNetFit, SBML tools

JR Faeder, unpublished
Going forward

- Decompositions based parameter estimation
- Decompose the model into its maximal strongly connected components
- Use the resulting DAG to guide the parameter estimation procedure.
- Estimate the parameters of the upstream components first
- Complications:
  - Distribution of experimental data
  - Computing consistent global estimates from local ones.

  ➤ *Belief propagation*
Going forward

- Network based parameter estimation
- Estimate the parameters of the components individually
- Compute consistent global estimates from local ones.

> Belief propagation
Acknowledgements

Jim Faeder

Tim Lezon

Liu Bing

Benjamin Gyori (LSP, Harvard Medical School)