Density Estimation of Initial Conditions for Populations of Dynamical Systems

PASCAL 2008 Workshop on Approximate Inference in Stochastic Processes and Dynamical Systems

A. G. Busetto, B. Fischer, J. Buhmann

ETHZ – Swiss Federal Institute of Technology Zurich

May 27th, 2008
Cumberland Lodge, UK

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Modeling Cell Populations

Time Series from Experiments

In the biological sciences, time series can now be routinely collected from experiments. These data permit modeling, analysis and simulation.

Modeling Techniques

Many quantitative modeling techniques has been proposed. For

- continuous-valued
- continuous-time
- deterministic

systems, the traditional approach based on ODEs is still the most common (descriptive and analytical power!).

Measuring Single Cells

Single Cells and Populations

Dynamical modeling can be performed

- at the single cell level
  (e.g. fluorescent protein degradation) or
- averaged over a cell population
  (e.g. gene expression).

This depends on data availability and on the required detail.

Single Cells VS Populations!

What if we are interested in the dynamics of the single cell but only population measurements are available?!

Single-Cells VS Populations

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Single and Average Behaviors

The dynamical behaviors of single cells and populations can be significantly different!

Experimental Observations

For instance, in GFP degradation

- zero-order dynamics are measured *in vitro*,
- first-order dynamics are measured *in vivo*.

This distortion can be caused by the mentioned discrepancies[a].

[a] W. W. Wong *et al.* Single-cell zeroth-order protein degradation enhances the robustness of synthetic oscillator. *Mol Sys Biol*, 3, 2007.

Causes

Discrepancies!

What causes the observed discrepancies between single cells and populations?

Causes

The main causes of discrepancy are

- heterogeneously parametrized models,
- heterogeneous initial conditions for every cell,
- other reasons (incomplete modeling, ...).

Biologically Significant?

Scenario

Our scenario: recovering single cell behaviour (hidden variables) from measurements of a cell population.

We are interested in the behavior of a "generic" cell, not of a specific one. All the cells follow the same biological law.

Protein Degradation Example

Example

A possible model[a] for single-cell GFP protein degradation is

$$f(x, t, c, \delta, \gamma, K, V) = \underbrace{\frac{c}{\gamma}\exp(-\gamma t)}_{(1)} \underbrace{-\delta x}_{(2)} - \underbrace{\frac{Vx}{K + x}}_{(3)},$$

(1) transcription/translation,

(2) dilution,

(3) enzymatic decay.

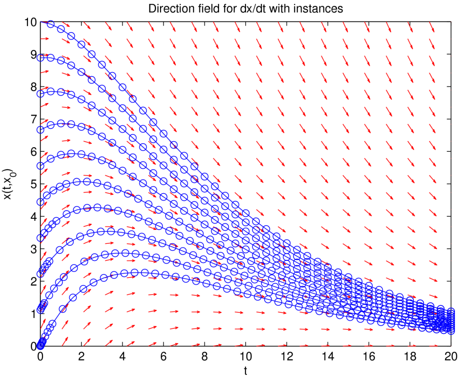
[a] C. Grilly *et al.*, A synthetic gene network for tuning protein degradation in Saccharomyces cerevisiae. *Mol Sys Biol*, 3, 2007.

Single Cells

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Example

In this example, time-dependent fluorescence [AU;AU] trajectories are plotted (for single cells with different initial conditions).

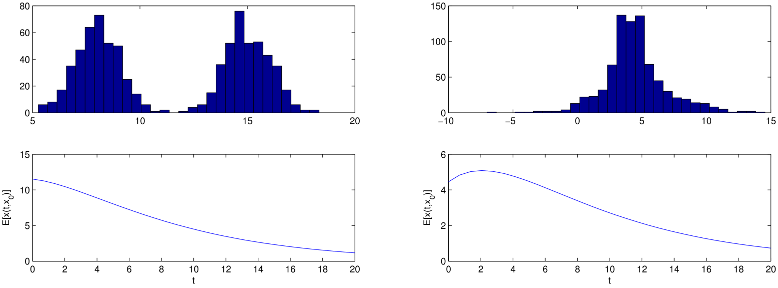


Masking Single Cell Behavior

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Example

Single-cell behavior can be masked by population averages.
Different density of initial conditions give quite different dynamical results!



Assumptions

Assumptions

The following mathematical formalization is based on the following assumptions:

- a cell population consists of a large but finite number of cells,
- every cell is independent,
- every cell is deterministic,
- cell models are heterogeneously parametrized,
- cell models exhibit heterogeneous initial conditions,
- the measurement noise is an additive stochastic process.

Modeling Single Cells

ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Modeling a Single Cell

Let x be a biological quantity (protein abundance, metabolite concentration, ...), the dynamics of a single cell with initial condition x_0 follows the initial value problem $\mathcal{U}_{x_0}$:

$$\mathcal{U}_{x_0} : \quad \begin{cases} \dfrac{dx(t, x_0)}{dt} = f(x(t, x_0), t, \theta) \\ x(t_0, x_0) = x_0, \end{cases}$$

restricted to the interval $[t_0, t_f]$, where $f : \mathbb{R} \times [t_0, t_f] \times \mathbb{T} \rightarrow \mathbb{R}$ and $\theta \in \mathbb{T}$ is a parameter vector[a].

[a] Assume also that the conditions of the Picard-Lindelöf (Cauchy-Lipschitz) theorem are satisfied.

Modeling Single Cells

ETH Eidgenössische Technische Hochschule Zürich Swiss Federal Institute of Technology Zurich

Density over the Initial Conditions

Let p be a probability density over the initial conditions x_0 of $\mathcal{U}_{x_0}$.

Random Initial Conditions

Let the continuous random variable $X_{0\mathcal{C}} \sim p$ determine the initial condition for the dynamics of the cell $\mathcal{C}$.
For a given realization with an initial value $x_{0\mathcal{C}}$, $\mathcal{C}$ follows the dynamical behavior $x(t, x_{0\mathcal{C}})$[a].

[a] From the Picard-Lindelhöf theorem, this trajectory exists and is unique.

Modeling Populations

Cell Populations

Consider a large but finite cell population consisting of s cells. Its dynamics is the average of the behaviors of the single components, whose initial conditions are the realization of a set of iid continuous random variables $X_{01}, X_{02}, \ldots, X_{0s}$.

Population Behavior

For a given realization $\mathbf{x}_0 = (x_{01}, x_{02}, \ldots, x_{0s})$, the population follows the dynamics given by the aggregation of $\mathcal{U}_{x_{01}}, \ldots, \mathcal{U}_{x_{0s}}$:

$$\mathcal{Z}_{\mathbf{x}_0} : \quad \begin{cases} \dfrac{dx(t, x_{0i})}{dt} = f(x(t, x_{0i}), t, \theta_i) \\ x(t_0, x_{0i}) = x_{0i} \end{cases} \qquad i = 1, 2, \ldots, s.$$

Observed Behavior

Averaged Behavior

The averaged behavior of the population is given by

$$z(t, \mathbf{x}_0) = \mathbb{E}[x(t, x_{0i})] = \frac{1}{s} \sum_{i=1}^{s} x(t, x_{0i}),$$

where $x(t, x_{0i})$ is the solution of $\mathcal{U}_{x_{0i}}$. Then, for $s \to \infty$, it tends to

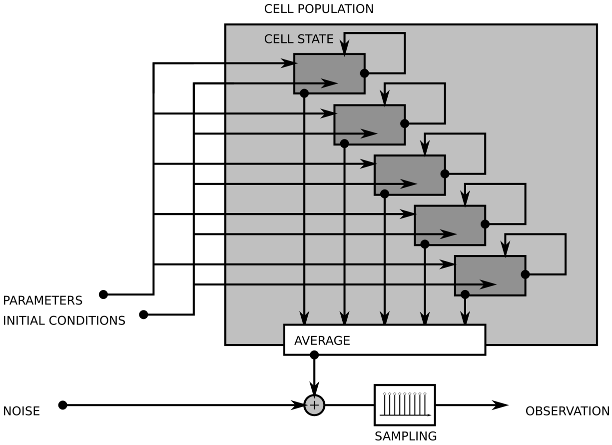
$$z_\infty(t) = \mathbb{E}[x(t, x_0)] = \int p(x_0) x(t, x_0) dx_0.$$

Additive Noise

The measurement process assumes an additive stationary noise term:

$$z^\varepsilon(t) \simeq z_\infty(t) + \varepsilon(t).$$

System Diagram



Discretized Integral Equation

Approximated Integral Equation

With the introduced approximation,

$$
\begin{aligned}
z_\infty(t) &= \int p(x_0)x(t,x_0)dx_0 \\
&\simeq \int \widehat{p}_n(x_0,\mathbf{p})x(t,x_0)dx_0 \\
&= \sum_{i=1}^{n} p_i \underbrace{\int \frac{1}{h}K\left(\frac{x_0-\widehat{x}_{0i}}{h}\right)x(t,x_0)dx_0}_{\phi_i(t)}
\end{aligned}
$$

Approximate Subpopulations

Subpopulation Behavior

The averaged behavior of an approximate subpopulation is

$$\begin{aligned}\phi_i(t) &= \int w_i(x_0)x(t, x_0)dx_0 \\ &= \int \frac{1}{h} K\left(\frac{x_0 - \widehat{x}_{0i}}{h}\right) x(t, x_0)dx_0 \\ &= \mathbb{E}[x(t, x_0)],\end{aligned}$$

Dynamical Contributions

Therefore, before the sampling,

$$z^{\varepsilon}(t) \simeq \sum_{i=1}^{n} p_i \phi_i(t) + \varepsilon(t).$$

Sampling

Sampling

the sampled values are expressed in the following form

$$\forall j = 1, 2, \ldots, m \quad z^{\varepsilon}(t_j) = z_j, \quad \phi_i(t_j) = \phi_{ji} \quad i = 1, 2, \ldots, n.$$

Discretizing the Integral Equation

The integral equation that was introduced before can be rewritten as

$$j = 1, 2, \ldots, m \quad z_j \simeq \sum_{i=1}^{n} p_i \phi_{ji}.$$

that is, in matrix form,

$$\mathbf{z} \simeq \mathbf{\Phi p}.$$

Numerical Integration

Numerical Integration

Given x_0, $x(t, x_0)$ must be approximated by numerical integration[a], obtaining $\tilde{x}(t, x_0)$.

[a] Care must be taken, since the ODE can be stiff!

Numerically Integrated Subpopulation Dynamics

Assuming $x(t, x_0) \simeq \tilde{x}(t, x_0)$,

$$i = 1, 2, \ldots, n, \quad \phi_i(t) \simeq \int \frac{1}{h} K \left(\frac{x_0 - \widehat{x}_{0i}}{h} \right) \tilde{x}(t, x_0) dx_0 .$$

ML Estimation

Least Squares Problem

This can be stated as problem $\mathcal{P}$: find $\mathbf{p}^*$ such that

$$\mathbf{p}^* = \arg\min_{\mathbf{p}\in\mathbb{R}^n} \|\tilde{\mathbf{\Phi}}\mathbf{p} - \mathbf{z}\|_2^2,$$

subject to

$$\begin{cases} \sum_{i=1}^{n} p_i = 1, \\ 0 \leq p_i \leq 1 \quad i = 1, 2, \ldots, n. \end{cases}$$

Prior Information

Domain Knowledge

In systems biology, the simple processes are often understood quite well, but complex systems are still under investigation.

Prior Information

Domain knowledge is given under the form of priors over functions describing the dynamics of a cell. This is not possible with purely data-driven approaches and, when existing, must be exploited.

Robustness

Robustness

- Since prior domain information is often available,
- the existence of outliers cannot be denied and
- the least-squares approach by itself is not robust,

Bayesian regression with a mixture of regular observations and outliers can be employed[a].

[a] M. Kuss *et al.*, Approximate inference for robust Gaussian process regression. *Technical Report 136*, Tübingen, Germany, 2005.

Computational Costs

However, this is computationally expensive and feasible approaches must be approximated!

Undersampling

Undersampling

When undersampled, problem $\mathcal{P}$ is solved by the (constrained) linear subspace of $\mathbb{R}^n$ that satisfies

$$\tilde{\mathbf{\Phi}}\mathbf{p} - \mathbf{z} = 0.$$

Entropy Maximization

In order to maximize entropy, we solve $\mathcal{H}$: find $\mathbf{p}^*$ such that

$$\mathbf{p}^* = \arg \max_{\mathbf{p} \in \mathrm{Sol}(\mathcal{P})} H[\widehat{p}_n(x_0, \mathbf{p})],$$

where $H[p] = \int p(x) \log p(x) dx$ is the differential entropy.

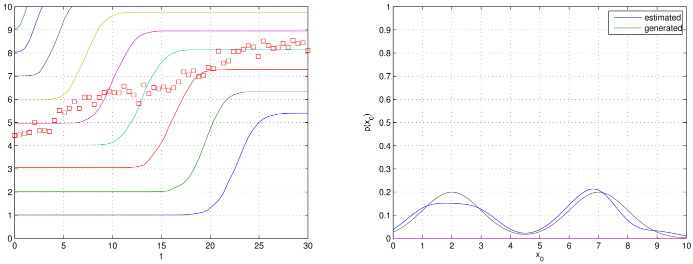
A Proof of Concept

Example

Consider the following function

$$f(x(t, x_0), t, \theta_1, \theta_2, \theta_3) = (\theta_1 t) \exp\{-(x(t, x_0) - \theta_2 + \theta_3 t)^2\},$$

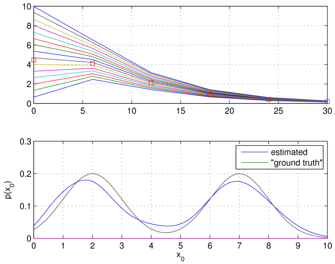
where $\sigma_\varepsilon = 0.2$, $m = 60$ and $n = 10$.



Undersampled Protein Degradation

Example

Now in the case of undersampling with $n = 15 > 6 = m$ (as before but with robust regression):



note that, for $m/n \to 0$ we have that $\mathbf{p}^*$ tends to the uniform distribution.

Outlook: Optimal Sampling

Sampling

Due to experimental costs, sample points are scarce. Whereas they are usually chosen uniformly spaced or according to heuristics, an optimal sampling is highly desirable.

Optimal Experiment Design

To maximize the average information gain, the optimal sampling minimizes the maximum entropy of the estimate. The result is

- the most informative of
- the least biased between the
- consistent with the observations.

Entropy Minimization

Entropy Minimization

We want to solve the problem $\mathcal{O}$: find $\mathbf{t}^*$ such that

$$\mathbf{t}^* = \arg\min_{\mathbf{t}} \left\{ \underbrace{\arg \max_{\mathbf{p_t} \in \mathrm{Sol}(\mathcal{P}_\mathbf{t})} H[\widehat{p}_{n\mathbf{t}}(x_0, \mathbf{p_t})]}_{\mathcal{H}_\mathbf{t}} \right\},$$

where $\mathcal{H}_\mathbf{t}$ is the entropy maximization problem subject to the sampling encoded by $\mathbf{t}$.

This is a constrained non-convex problem that is computationally expensive.

Considerations

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Considerations

1. In practice, taking prior information into account is strongly beneficial since it might reduce the effects of undersampling. Approximate inference permits a feasible approximation of the robust regression, extending the applicability of the whole approach.
2. The determination of the optimal experiment design is highly desirable for experimentalists and helps the improvement of the results, since it maximizes the information gain from the expensive measurement.

Open Questions

Open Questions

1

- The selection of a double model for outliers and regular observations seems promising, which model provides the best results? Which inference approximation technique provides the best results?
- Exact inference is intractable and must be approximated, but how? Which method provides the best tradeoff between quality and cost?

2

- How is it possible to speed up the non-convex experiment design optimization process?
- Which heuristics give the best results?