Lecture Notes: Quasi-reaction systems

Quasi-reaction systems are dynamical processes where a number of agents, particles or nodes interact in a stochastic way throughout reactions, generating different configurations for the studied system over time. These systems include applications such as protein dynamics, gene expression, ecological systems, kinetic systems, compartmental models, actor-oriented systems, among others.

In many of these applications, statistical methods for the inference of kinetic rates dynamics are crucial. Kinetic systems can be mathematically modeled as a Poisson processes, like the ones seen in lecture 6.

Quasi-reaction systems are dynamical processes where agents, particles, or nodes interact in a stochastic manner, leading to various system configurations over time. These systems are integral to numerous scientific and engineering fields.

- **Protein Dynamics:** Central to biochemistry, quasi-reaction systems model the intricate processes of protein synthesis and folding. Stochastic models are particularly useful in analyzing protein-ligand interactions, a critical aspect in drug development and molecular biology. Understanding these interactions aids in predicting drug efficacy and protein behavior.

- **Gene Expression:** These systems are instrumental in the study of gene regulation networks. Modeling the stochastic aspects of gene expression helps in deciphering complex cellular processes, contributing to advancements in genetic engineering and personalized medicine.

- **Ecological Systems:** Quasi-reaction systems find applications in ecology, particularly in modeling complex interactions within ecosystems. They help in understanding predator-prey dynamics, nutrient cycling, and population dynamics, offering insights into ecological stability and species survival strategies.

- **Kinetic Systems in Chemistry:** These systems are fundamental in chemical kinetics, where reaction rates are influenced by the stochastic interactions of molecules. Understanding these interactions is crucial for developing new chemical processes and materials.

- **Compartmental Models in Epidemiology:** In epidemiology, quasi-reaction systems model disease spread, with compartments representing different stages of disease progression. This modeling is vital for predicting disease trajectories, informing public health policies, and understanding epidemiological patterns.

In many applications, particularly in kinetic systems, statistical methods are employed to infer dynamic rates. When dealing with large numbers of species, these stochastic kinetic models can often be approximated by deterministic models, employing a mean field approach. This allows for a more tractable analysis of systems that operate on varying time scales.

Basic Combinatorics in Reaction Systems

Before delving into the complexities of reaction systems, it's essential to understand the basics of combinatorics, which plays a critical role in these systems. Combinatorics, the branch of mathematics dealing with combinations and arrangements of objects, is fundamental in calculating reaction rates and understanding reaction dynamics.

Consider a simple scenario where we have a set of distinct objects, and we want to determine how many different ways we can arrange or select these objects. The principles of combinatorics allow us to calculate these possibilities. For example, if we have n different objects and want to

Figure 1: Scientific Diagram Illustrating Population Dynamics in an Ecological Context: This visualization depicts different species populations as graphs or curves over time, showing fluctuations in population sizes. It includes elements such as predator-prey relationships, with arrows indicating the impact of one species on another, and reflects the complex interplay of birth rates, death rates, and environmental factors affecting these populations. The diagram serves as a comprehensive visual aid in understanding the dynamic and interconnected nature of ecological systems, emphasizing the critical role of various biological and environmental factors in shaping population trends.

choose r of them, the number of different combinations we can form is given by the binomial coefficient, denoted as $\binom{n}{r}$. This coefficient is calculated as:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

where $n!$ (n factorial) is the product of all positive integers up to n. This combinatorial calculation is crucial in reaction systems, particularly when determining reaction rates and the number of possible interactions between different molecular species.

This formula calculates the number of ways to choose r items from a set of n items, where the order of selection does not matter. Let us explore the rationale behind this formula:

The factorial of a number n, denoted as $n!$, is the product of all positive integers up to n. It represents the total number of ways to arrange n items in order:

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 1$$

Since combinations focus on selection where order does not matter, we adjust for overcounting caused by different arrangements:

- **Arrangements of Chosen Items:** For any selection of r items, there are $r!$ ways to arrange these items. To eliminate order consideration, we divide by $r!$:

$$\frac{n!}{r!}$$

- **Arrangements of Unchosen Items:** There are also $(n-r)!$ ways to arrange the remaining $n-r$ items. Dividing by $(n-r)!$ further removes the ordering of unchosen items:

$$\frac{n!}{r!(n-r)!}$$

By combining these adjustments, we derive the formula for combinations, which accurately counts the number of ways to choose r items from n without considering the order:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

This formula corrects for overcounting inherent in permutations and focuses on the unique selections where the order of items is irrelevant, a crucial aspect in many fields including probability, statistics, and biology.

Stoichiometry of Reaction Systems

Stoichiometry in reaction systems examines the quantitative relationships between reactants and products in chemical reactions. Consider a system with r different reactants and p different products involved in a series of reactions.

Each reaction can be described by an equation. For the j-th reaction, the equation is:

$$\sum_{i=1}^{r} k_{ij} R_i \xrightarrow{\theta_j} \sum_{i=1}^{p} s_{ij} P_i. \tag{1}$$

Here, R_i represents the i-th reactant and P_i represents the i-th product. The coefficients k_{ij} and s_{ij} are the stoichiometric coefficients for reactants and products, respectively, in the j-th reaction.

Example 1 (Basic Predator-Prey Model)**.** *Consider a simplified ecological model with two types of organisms: predators (denoted as P) and prey (denoted as Q). The dynamics of this system are captured by the following reactions, each with its associated rate:*

$$Q \xrightarrow{\beta_{repro}} 2Q \quad \textit{(Prey reproduction at rate } \beta_{repro}\textit{)}, \tag{2}$$

$$P + Q \xrightarrow{\beta_{pred}} 2P \quad \textit{(Predation and predator reproduction at rate } \beta_{pred}\textit{)}, \tag{3}$$

$$P \xrightarrow{\beta_{death}} \emptyset \quad \textit{(Predator death at rate } \beta_{death}\textit{)}. \tag{4}$$

We define the reactants and products for each reaction as follows:

- *Reactants:* $R = \{Q, P, P+Q\}$,
- *Products:* $P = \{2Q, 2P, \emptyset\}$.

When does something happen:

$$\begin{aligned}\Delta T_{P\to\emptyset} &\sim Exp(\beta_{death}),\\ \Delta T_{P+Q\to 2P} &\sim Exp(\beta_{pred}),\\ \Delta T_{Q\to 2Q} &\sim Exp(\beta_{repro}).\end{aligned}$$

About reaction in Δt time:

- $P\to\emptyset$*: will have happened approx* $2\beta_{death}\Delta t$ *times,*
- $P+Q\to 2P$*:* $6\beta_{pred}\Delta t$*,*
- $Q\to 2Q$*:* $3\beta_{repro}\Delta t$*.*

We expect to see

$$\begin{aligned}E[Y_P(\Delta t)] &= Y_P(0)+6\beta_{pred}\Delta t-2\beta_{death}\Delta t,\\ E[Y_Q(\Delta t)] &= Y_Q(0)+3\beta_{repro}\Delta t-6\beta_{pred}\Delta t.\end{aligned}$$

We actually saw at $\Delta t=2$*:*

$$\begin{aligned}Y_P(2)&=5,\\ Y_Q(2)&=4.\end{aligned}$$

Now we want to know what we expect to happen at $2+\Delta t$*:*

$$\begin{aligned}R_1 &: 4\beta_{repro}\Delta t \text{ times},\\ R_2 &: 20\beta_{pred}\Delta t \text{ times},\\ R_3 &: 5\beta_{death}\Delta t \text{ times}.\end{aligned}$$

So we expect to see

$$\begin{aligned}E[Y_P(2+\Delta t)] &= 5+20\beta_{pred}\Delta t-5\beta_{death}\Delta t,\\ E[Y_Q(2+\Delta t)] &= 4+3\beta_{repro}\Delta t-20\beta_{pred}\Delta t.\end{aligned}$$

We actually see

$$\begin{aligned}Y_P(4)&=8,\\ Y_Q(4)&=1.\end{aligned}$$

What are the most likely values for β_{repro}*,* β_{pred}*,* β_{death}*?*
We saw 3×2 *values of the states:*
First interval:

$$\begin{aligned}5 &\approx 2+6\beta_{pred}\cdot 2-2\beta_{death}\cdot 2,\\ 4 &\approx 3+3\beta_{repro}\cdot 2-6\beta_{pred}\cdot 2.\end{aligned}$$

Second interval:

$$\begin{aligned}8 &\approx 5+20\beta_{pred}\cdot 2-5\beta_{repro}\cdot 2,\\ 1 &\approx 4+5\beta_{repro}\cdot 2-20\beta_{pred}\cdot 2.\end{aligned}$$

This leads to the matrix equation:

$$\begin{bmatrix}3\\1\\3\\-3\end{bmatrix}=\begin{bmatrix}0&12&-4\\6&-12&0\\40&-10&0\\10&-40&0\end{bmatrix}\begin{bmatrix}\beta_{repro}\\\beta_{pred}\\\beta_{death}\end{bmatrix}+\begin{bmatrix}\epsilon_1\\\epsilon_2\\\epsilon_3\\\epsilon_4\end{bmatrix}$$

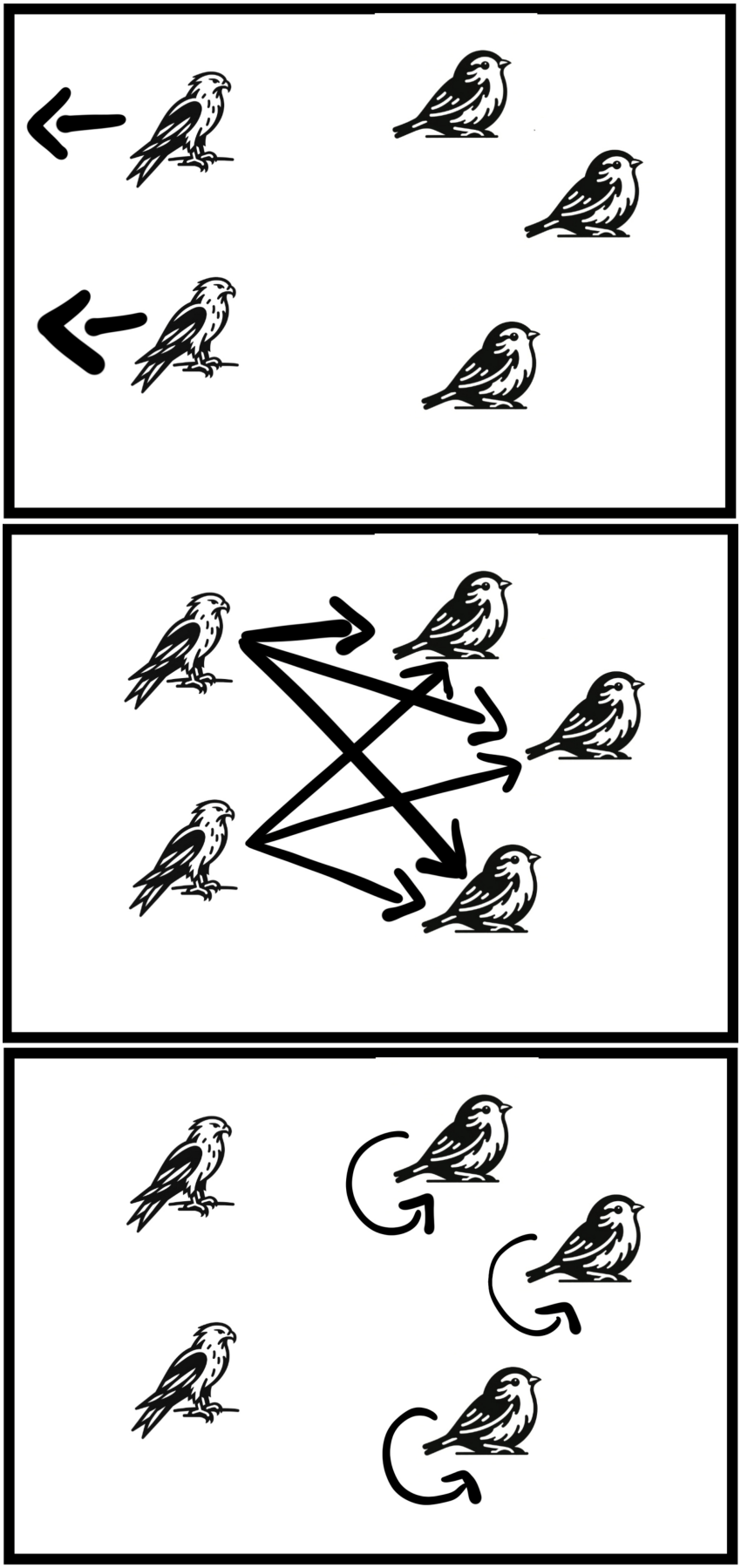


Figure 2: Diagrams showing the different number of combinations of reactions in the predator-prey system.

To estimate the model parameters (β_{repro}, β_{pred}, β_{death}), we use the observed values from two intervals, forming the following matrix equation:

$$\begin{bmatrix} 3 \\ 1 \\ 3 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 & 12 & -4 \\ 6 & -12 & 0 \\ 0 & 40 & -10 \\ 10 & -40 & 0 \end{bmatrix} \begin{bmatrix} \beta_{repro} \\ \beta_{pred} \\ \beta_{death} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$$

The least squares estimate $\hat{\beta}$ is calculated as $\hat{\beta} = (X^T X)^{-1} X^T Y$. Using this formula, the estimated values are:

$$\begin{aligned} \hat{\beta}_{repro} &\approx 0.36, \\ \hat{\beta}_{pred} &\approx 0.155, \\ \hat{\beta}_{death} &\approx 0.235. \end{aligned}$$

To further assess the reliability of these estimates, a bootstrap analysis was performed. The table below provides the descriptive statistics for each bootstrapped beta coefficient:

Statistic	β_{repro}	β_{pred}	β_{death}
Mean	-0.344	-0.095	-0.809
Std	1.083	0.332	1.221
Min	-2.100	-0.450	-2.100
25%	-0.733	-0.450	-2.100
50%	0.360	0.155	-0.050
75%	0.633	0.233	0.235
Max	0.633	0.233	0.633

Table 1: Descriptive statistics for each bootstrapped beta coefficient.

The rate of each reaction depends on the combinations of reactant molecules. For reaction j:

$$h_j = \theta_j \times \prod_{i=1}^{r} \binom{Y_i}{k_{ij}}, \tag{5}$$

where Y_i is the number of molecules of the i-th reactant. The term $\binom{Y_i}{k_{ij}}$ represents the number of combinations of k_{ij} molecules from Y_i available molecules. The reaction rate θ_j is then the product of these combinations.

Consider the state of the system at a given time t, denoted Y_t. If the j-th reaction occurs first after starting with Y_0 molecules, the new state Y_t for each reactant is:

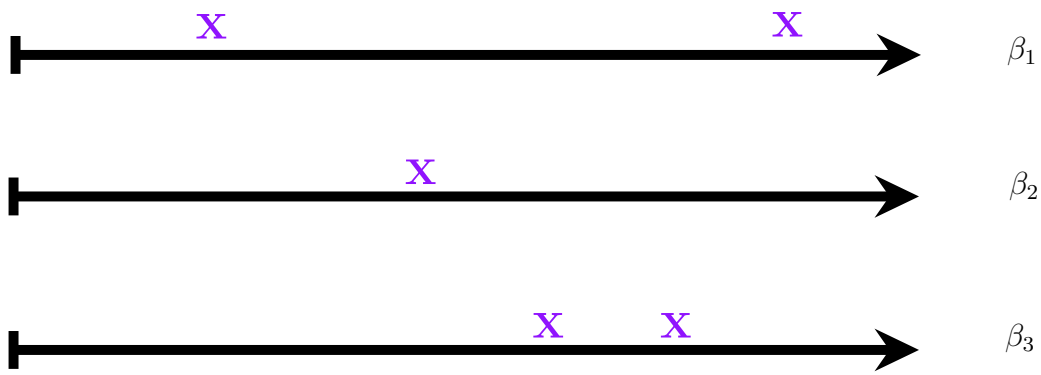
$$Y_{ti} = Y_{0i} + (s_{ij} - k_{ij}), \tag{6}$$

where $v_{ij} = s_{ij} - k_{ij}$ is the net change in the i-th reactant due to the j-th reaction.

The overall molecular changes in the system can be encapsulated in the net change matrix $\mathbf{V}$:

$$\mathbf{V} = \mathbf{S} - \mathbf{K}, \tag{7}$$

where $\mathbf{S}$ and $\mathbf{K}$ are matrices of stoichiometric coefficients for products and reactants, respectively, across all reactions, with dimensions $p \times r$.



Example 2 (Circular Reaction System)**.** *Consider the following circular unitary reaction system,*

$$\begin{aligned} Y_1 &\xrightarrow{\theta_1} Y_2, \\ Y_2 &\xrightarrow{\theta_2} Y_3, \\ Y_3 &\xrightarrow{\theta_3} Y_1. \end{aligned}$$

In this system, each substance Y_i transforms into the next, forming a closed loop. The reactions are driven by the rates θ_1, θ_2, and θ_3.

To analyze this system, we define the stoichiometric matrices:

Reactants $R = \{Y_1, Y_2, Y_3\}$, Products $P = \{Y_2, Y_3, Y_1\}$.

The matrix of stoichiometric coefficients for reactants ($\mathbf{K}$) and products ($\mathbf{S}$) are:

$$\mathbf{K} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}. \tag{8}$$

The net change matrix ($\mathbf{V}$) is calculated as $\mathbf{V} = \mathbf{S} - \mathbf{K}$:

$$\mathbf{V} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix}. \tag{9}$$

This matrix $\mathbf{V}$ represents the net change in the quantities of Y_1, Y_2, and Y_3 due to each reaction in the system.

Inference of System Dynamics

The system under study is characterized by a set of reactions occurring over time from 0 to T. These reactions involve various compartments or states in the system, each with its distinct dynamics.

Each reaction in the system is associated with a waiting time T_i, which follows an exponential distribution:

$$T_i \sim \text{Exp}(\lambda_i) \quad \text{for } i = 1, 2, \ldots$$

Here, λ_i is the rate parameter of the i-th reaction. The count C indicates the number of occurrences of a particular event in a time interval Δt:

$$C = \#\text{ of occurrences in } \Delta t$$

For instance, the probability of no occurrences (i.e., $C = 0$) within Δt is:

$$P(C = 0) = 1 - (1 - e^{-\lambda \Delta t})$$

The probability of exactly one occurrence in Δt is:

$$P(C = 1) = e^{-\lambda \Delta t} \cdot \frac{(\lambda \Delta t)^1}{1!}$$

And for reaction i, the count C_i follows a Poisson distribution:

$$C_i \sim \text{Poisson}\,(\theta_i)$$

where θ_i is given by:

$$\theta_i = \lambda_i \prod_{j=1}^{p} \binom{Y(t)}{k_j} \Delta t$$

This equation models the frequency of each reaction, considering the current state $Y(t)$ of the system.

With the knowledge of C_i, we update the system states as follows:

$$\Delta Y(t) = Y(t + \Delta t) - Y(t) = V^T C$$

This expression calculates the change in state variables, $\Delta Y(t)$, based on the reactions that occurred.

Given observed data Y at specific time points, we aim to estimate the rate parameters λ that best describe the system dynamics. The dynamics can be represented as:

$$\Delta Y(t) = X(t)\lambda + \eta(t)$$

Here, $X(t)$ is the matrix capturing the influence of each reaction, and $\eta(t)$ represents the error term.

The residuals η are the difference between observed and predicted changes in state variables:

$$\eta = \Delta Y - X\lambda$$

The objective is to minimize the sum of squared residuals, $\Sigma\eta^2$, which is equivalent to minimizing $\eta^T\eta$. By differentiating with respect to λ and setting the derivative to zero, we find the optimal λ:

$$\frac{\partial}{\partial \lambda} \Sigma \eta^2 = 0$$

Solving this equation yields the estimated rate parameters $\hat{\lambda}$:

$$\hat{\lambda} = (X^T X)^{-1} X^T \Delta Y$$

This optimized $\hat{\lambda}$ provides the best-fit parameters for our model, describing the underlying dynamics of the system.