UNIVERSIDADE DE SÃO PAULO INSTITUTO DE FÍSICA DE SÃO CARLOS

Guilherme David Araujo

A Bayesian framework of reaction networks for dynamical population models

São Carlos

2021

Guilherme David Araujo

A Bayesian framework of reaction networks for dynamical population models

Thesis presented to the Graduate Program in Physics at the Instituto de Física de São Carlos, Universidade de São Paulo to obtain the degree of Doctor of Science.

Área de concentração: Basic Physics

Supervisor: Prof. Dr. Leonardo Paulo Maia

Corrected version

(Original version available on the Program Unit)

São Carlos 2021 I AUTHORIZE THE REPRODUCTION AND DISSEMINATION OF TOTAL OR PARTIAL COPIES OF THIS DOCUMENT, BY CONVENTIONAL OR ELECTRONIC MEDIA FOR STUDY OR RESEARCH PURPOSE, SINCE IT IS REFERENCED.

> Araujo, Guilherme David A Bayesian framework of reaction networks for dynamical population models / Guilherme David Araujo; advisor Leonardo Paulo Maia - corrected version -- São Carlos 2021. 160 p.

> Thesis (Doctorate - Graduate Program in Theoretical and Experimental Physics) -- Instituto de Física de São Carlos, Universidade de São Paulo - Brasil , 2021.

1. Eco-evolution. 2. Bayesian probability. 3. Reaction networks. 4. Markov jump processes. 5. Individual specialization. I. Maia, Leonardo Paulo, advisor. II. Title.

I dedicate all my work to every being who has died in suffering.

ACKNOWLEDGEMENTS

My solitude was the main ground from which this work was unearthed. During these most afflicted and uncertain years of my life, my body almost snapped. But it didn't. And that is because I could make many right personal decisions, amidst an ocean of mistakes. However, all my right decisions were not actually mine; they came from a web of love that nurtured and yielded them with me, the same web that have always helped to protect my sanity and my needed solitude, from the brightest days to the darkest nights. This web is family, is passion, is community, is work. These are the beings I am most grateful to. Rosângela, you are my sun, without you I would never have met love itself, I would never be. Virgínia, Iago, Violeta, Daniel, you know all there is to know about me, and yet you love me; your existences shape my heart and my soul. Sergio, Leticia, Gesiel, Neiva, Fátima, Édina, Márcio, Marcelo, Cida, my dear family, my web since the beginning. Rafael, Carol, Lari, Guilherme da mecânica, Guilherme sô, Rúbia, Alex, Patrícia, Pedro da lê, Pedro primo, my closest friends during these vears. Miguel, Dani, Vivian, Felipe, my little partners in crime, my gang during these years. Rafa, you are the best thing that happened to my career so far and the reason I haven't sold myself yet, my newest friend. Léo, thank you for all the support, advice, faith, and respect; I have learned a lot from you and I am sincerely inspired by your spirit and your ideals. And I am also thankful to many others, guest stars in my life during these years.

And thank you **IFSC-USP** and **CNPq** for all the structure and money. I'm at your service. I'm also grateful to **Neusa**, for all the kindness in the correction of the formatting of this thesis. And I'm grateful to my examination board for the joyful moment that was my thesis defense; in special I thank **Diogo** for all the kind words and acknowledgement that deeply impacted me, and I thank **Alexandre** for the unmatched enthusiasm about my work.

Then I bow in gratitude to the **Moon**, the eternal witness, the ultimate source of yin, the shaper of my tides, whose silver light shines upon my path and symbolizes the entirety of my purpose. A priestess is nothing without her guiding light, especially a priestess of the Absurd. Above all the written symbols, the moon knows my real name...

"I will stand in the rain Hoping sun will come through, Then I'll see the colours of a misty rainbow. I'll stay up in the night Looking on shooting stars To tell you how magic Is the all universe. If you look inside your soul The world'll open to your eyes, You'll see..." ERA - Misere Mani

ABSTRACT

ARAUJO, G. D. A Bayesian framework of reaction networks for dynamical population models. 2021. 160p. Thesis (Doctor in Science) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2021.

The tradition of mathematical modeling in the biological sciences is yet to reach a mature state in many fields. The most pressing issues are the difficulty in first translating the complexities of life to quantitative modeling terms and the lack of robust frameworks providing structure and cohesion to the building and interpretation of models. In particular, the quantitative study of biological populations, as for example in behavioral ecology and evolutionary dynamics, is composed of a set of scattered methodologies that generate models without an anchored conceptual foundation. Modeling concepts are often ambiguous and do not directly translate to actual biological terms. Inspired by modeling advances in biochemistry, this thesis aims at the conceptualization and application of a general modeling framework for dynamical populations in biology. Combining a Bayesian probabilistic paradigm with the theory of reaction networks, I was able to structure a framework of relational interactions among populations, one that extends biochemical applications to all types of populations, unifying and generalizing existing methods in eco-evolutionary dynamics. The framework comprises both stochastic and deterministic models, and also their connection; it considers the connection with data through statistical model determination and brings a large emphasis on unambiguous design-informed dynamical equations. I validated the framework through applications to genetic regulation, parental investment, and ecological predator-prev dynamics.

Keywords: Eco-evolution. Bayesian probability. Reaction Networks. Markov jump processes. Individual specialization.

RESUMO

ARAUJO, G. D. Um framework Bayesiano de redes de reação para modelos dinâmicos de populações. 2021. 160p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2021.

A tradição de modelagem matemática nas ciências biológicas ainda não atingiu sua maturidade em diversas áreas. As maiores complicações enfrentadas são a dificuldade de traduzir a complexidade da vida em termos quantitativos de modelagem e também a falta de frameworks robustos que propiciam estrutura e coesão na criação e interpretação de modelos. Em particular, o estudo quantitativo de populações em biologia, como por exemplo em ecologia comportamental e dinâmica evolutiva, é composto por um conjunto de métodos difusos que geram modelos sem se ancorar em uma fundação conceitual. Conceitos de modelagem são muitas vezes ambíguos e não se traduzem a termos diretamente biológicos. Se inspirando em avanços de modelagem em bioquímica, essa tese tem como objetivo a conceitualização e aplicação de um framework geral de modelagem para dinâmica populacional em biologia. Combinando um paradigma Bayesiano de probabilidade e a teoria de redes de reação, eu pude estruturar um framework de interações relacionais entre populações, que extende aplicações bioquímicas a todo tipo de populações, assim unificando e generalizando métodos existentes em dinâmica eco-evolutiva. O framework consiste tanto de modelos estocásticos quanto determinísticos, e também da conexão entre eles; ele considera a conexão com dados através da determinação estatística de modelos e traz uma grande ênfase a equações não ambíguas e informadas por design. Eu validei o framework através de aplicações em regulação genética, investimento parental e dinâmicas ecológicas de presa-predador.

Palavras-chave: Eco-evolução. Probabilidade Bayesiana. Redes de reação. Processos de pulo Markovianos. Especialização individual.

CONTENTS

1	INTRODUCTION	17
1.1	Structure of the Thesis	19
2	PROBABILISTIC FOUNDATION	21
2.1	Bayesian Probability Framework	21
2.1.1	Probability Rules	23
2.1.2	Bayes Equation	24
2.2	Stochastic Processes and the Markov Property	25
2.2.1	Markovian Stochastic Processes	26
2.2.2	Bayesian Derivation of a Master Equation for Markov Jump Processes	27
2.2.2.1	Transitions	28
2.2.2.2	Master Equation	29
2.3	Parameter Estimation	31
2.3.1	Priors	32
2.3.2	Data Likelihood	32
2.3.3	Incomplete Measurements	35
3	REACTION NETWORKS	37
3.1	Reaction Networks	37
3.1.1	Definition	38
3.1.2	Stoichiometric Coefficients	39
3.1.3	Transition Rates	40
3.1.4	Physical Systems	41
3.1.5	The Continuous Limit	42
3.1.6	Common Systems Modeled as Reaction Networks	43
3.2	Deterministic System	47
3.2.1	The Event-function	48
3.2.2	Deterministic Equations	49
3.3	Stochastic System	49
3.3.1	Bayesian Probabilistic Assessment	50
3.3.2	Transition rates and the Markov Property	51
3.3.3	Connection to the Deterministic System	55
3.3.4	Simulation of the Limit	56
3.3.5	Final Remarks	59

4	ESTIMATION AND STOCHASTIC APPROXIMATION ON REAC-		
	TION NETWORKS	63	
4.1	Stochastic Simulation Algorithm	63	
4.2	System Size Expansion	64	
4.2.1	Linear Noise Approximation	66	
4.2.2	Relation to Other Approximations	67	
4.3	Parameter Estimation on Reaction Networks	68	
4.4	Introduction to Markov Chain Monte Carlo	69	
4.4.1	Monte Carlo	70	
4.4.2	Importance Sampling	71	
4.4.3	Rejection Sampling	71	
4.4.4	Markov Chain	72	
4.4.5	Metropolis-Hastings	74	
4.4.6	STAN	75	
4.5	Estimation Example: Lotka-Volterra	75	
5	STOCHASTIC ANALYSIS AND PARAMETER ESTIMATION OF		
	THE GOODWIN MODEL FOR GENETIC REGULATION	79	
5.1	Abstract	79	
5.2	Introduction	79	
5.2.1	Systems Biology	79	
5.2.2	Gene Expression	80	
5.2.3	Goodwin Oscillator	82	
5.3	Methods	82	
5.3.1	2D and 3D Models	82	
5.3.2	Deterministic Description	83	
5.3.3	Stochastic Description	84	
5.3.4	Noise Analysis	84	
5.3.5	Hill Feedback	85	
5.3.6	Parameter Estimation	86	
5.4	Results	87	
5.4.1	2D Model	87	
5.4.2	3D Model	88	
5.4.3	Parameter Estimation	92	
5.5	Discussion	94	
6	A UNIFYING MECHANISTIC FRAMEWORK OF EVOLUTION-		
	ARY DYNAMICS USING REACTION NETWORKS	97	
6.1	Abstract	97	
6.2	Introduction	97	

0.5	Methods
6.3.1	Reaction Networks for Evolutionary Dynamics
6.3.1.1	Stochastic Kinetics
6.3.1.2	Deterministic Kinetics
6.3.1.3	Modeling transition rates
6.4	Results
6.4.1	Reaction networks for evolutionary game theory
6.4.1.1	Evolutionary-level replicator dynamics
6.4.1.2	Ecological-level games
6.4.2	Stochastic Models
6.4.3	Stabilizing the Hawk-Dove game
6.4.4	Price Equation and Adaptive Dynamics
6.4.4.1	Games as Generations
6.4.5	Examples
6.4.5.1	Generalized Lotka Volterra
6.4.5.2	Stochastic Rock-Paper-Scissors Game
6.4.5.3	Predator-Prey with Hiding Preys
6.4.5.4	Model of Parental Care Evolution
6.5	Discussion
7	INDIVIDUAL SPECIALIZATION AND GENERALIZATION IN PREDATOR
7	INDIVIDUAL SPECIALIZATION AND GENERALIZATION IN PREDATOR PREY DYNAMICS
7 7.1	INDIVIDUAL SPECIALIZATION AND GENERALIZATION IN PREDATOR PREY DYNAMICS 131 Introduction 131
7 7.1 7.2	INDIVIDUAL SPECIALIZATION AND GENERALIZATION IN PREDATOR PREY DYNAMICS 131 Introduction 131 Methods 133
7 7.1 7.2 7.2.1	INDIVIDUAL SPECIALIZATION AND GENERALIZATION IN PREDATOR PREY DYNAMICS 131 Introduction 131 Methods 133 Lotka-Volterra Model 133
7 7.1 7.2 7.2.1 7.2.2	INDIVIDUAL SPECIALIZATION AND GENERALIZATION IN PREDATOR PREY DYNAMICS 131 Introduction 131 Methods 133 Lotka-Volterra Model 133 Specialization Model 134
7 7.1 7.2 7.2.1 7.2.2 7.2.3	INDIVIDUAL SPECIALIZATION AND GENERALIZATION IN PREDATORPREY DYNAMICS131Introduction131Methods133Lotka-Volterra Model133Specialization Model134N-prey extension135
 7 7.1 7.2 7.2.1 7.2.2 7.2.3 7.3 	INDIVIDUAL SPECIALIZATION AND GENERALIZATION IN PREDATORPREY DYNAMICS131Introduction131Methods133Lotka-Volterra Model133Specialization Model134N-prey extension135Results136
 7 7.1 7.2 7.2.1 7.2.2 7.2.3 7.3 7.3.1 	INDIVIDUAL SPECIALIZATION AND GENERALIZATION IN PREDATORPREY DYNAMICS131Introduction131Methods133Lotka-Volterra Model133Specialization Model134N-prey extension135Results136N-prey model137
 7.1 7.2 7.2.1 7.2.2 7.2.3 7.3 7.3.1 7.4 	INDIVIDUAL SPECIALIZATION AND GENERALIZATION IN PREDATORPREY DYNAMICS131Introduction131Methods133Lotka-Volterra Model133Specialization Model134N-prey extension135Results136N-prey model137Discussion138
 7 7.1 7.2 7.2.1 7.2.2 7.2.3 7.3 7.3.1 7.4 8 	INDIVIDUAL SPECIALIZATION AND GENERALIZATION IN PREDATORPREY DYNAMICS131Introduction131Methods133Lotka-Volterra Model133Specialization Model134N-prey extension135Results136N-prey model137Discussion138CONCLUSION143
 7 7.1 7.2 7.2.1 7.2.2 7.2.3 7.3 7.3.1 7.4 8 	INDIVIDUAL SPECIALIZATION AND GENERALIZATION IN PREDATORPREY DYNAMICS131Introduction131Methods133Lotka-Volterra Model133Specialization Model134N-prey extension135Results136N-prey model137Discussion138CONCLUSION143REFERENCES145
 7 7.1 7.2 7.2.1 7.2.2 7.2.3 7.3 7.3.1 7.4 8 	INDIVIDUAL SPECIALIZATION AND GENERALIZATION IN PREDATORPREY DYNAMICS131Introduction131Methods133Lotka-Volterra Model133Specialization Model134N-prey extension135Results136N-prey model137Discussion138CONCLUSION143REFERENCES145APPENDIX155
7 7.1 7.2 7.2.1 7.2.2 7.2.3 7.3 7.3 7.4 8	INDIVIDUAL SPECIALIZATION AND GENERALIZATION IN PREDATORPREY DYNAMICS131Introduction131Methods133Lotka-Volterra Model133Specialization Model134N-prey extension135Results136N-prey model137Discussion138CONCLUSION143REFERENCES145APPENDIX155APPENDIX157

A.2 3D Goodwin Model		58
----------------------	--	----

1 INTRODUCTION

Physical processes are naturally complex. They are messy with interconnected elements jointly acting through time in intricate dynamics. The historical approach in the study of physical sciences is to reduce nature in order to analyze its moving parts and experimentally control its behavior. From this tradition, with the aim to mechanistically understand, explain, and predict the many processes of nature, the idea of developing and studying models has been successful. A model of nature is designed so as to try and capture, in a sufficiently simple manner, the main factors responsible for the generation of observed processes. It is a method supported by the efficient assumption that a great part of that complexity of nature can be washed out as not determinant enough for the robust understanding of a given process. The main task in the design of models is to rightly sort out which are those determinant factors and then assemble their relationship in a theoretical framework that is able to capture the actual rules governing them. In order to rigorously lay out the rules of dynamical behaviors in a quantitative fashion, frameworks of modeling are better expressed in mathematical terms, and that has caused the physical sciences and mathematics to have a joint tradition in the development of their methods.

Biological processes, featuring a rich umbrella of autonomous systems built over the physical nature through billions of years of evolutionary accumulation, are much messier and much more complex in detailed behavior than physical systems. Because of this, the biological sciences traditionally faced a much harder challenge in terms of reducing life and its environments to quantitative mechanistic models.¹ This barrier has translated into a tradition of qualitative and verbal models, aiming mostly at description, classification, and hypothesis-driven assessment of biological processes and systems, that do much in paving the way for a comprehensive quantitative understanding of such processes. Then, only recently there has been a more robust movement towards quantitative mechanistic modeling in biology. Even then, there is much yet to be done in terms of laying down principled foundations that conceptualize behavioral rules in biology from the ground $up;^{2-4}$ this ground being understood as a mathematical-theoretical framework from where complex models can be derived from.

The main goal of this work is to research for the development and application of such a framework for biological processes, one aimed mainly at treating dynamical processes of biological populations. In this case, populations mean individuals grouped into collectives, that relate to their environments in the same ways, featuring simple kinds of local interactions that add up to complex global dynamics. We are particularly interested in applications to biochemical and eco-evolutionary processes, with individuals being molecules and lifeforms. Local interactions, for these specific applications, range from chemical reactions to the relationship between animals, as for example predator-prey relations.

The more complex a system is, and thus the more we need to reduce it to treatable and principal components, more we incur into the uncertainty accompanying this reduction. From the ignorance towards smaller or underlying processes and the impossibility to draw precise trajectories to the unavoidable measurement errors, we are left in need to consider the probabilistic components of modeling design. Thus, any framework that is proposed as a theoretical ground for these kinds of complex systems must provide a dynamical basis that is probabilistic. For example, we are not able to track the individual trajectories of animals in an ecological setting or molecules in a biochemical setting, so we must provide a probabilistic account of how these trajectories generate the global properties that we observe. This is akin to how statistical mechanics is a ground framework for thermodynamics.

We opt to use a Bayesian paradigm of probability to define and treat all probabilistic relations we use, the reason being that it is a strong perspective to tie together all layers in which we need to consider arguments that formalize uncertainty, from different kinds of modeling assumptions to model determination in the form of parameter estimation.⁵ We even base our use of stochastic calculus on the Bayesian formalism, all in a manner that is completely compatible with traditional derivations that do not bother to connect mathematical laws and results with modeling philosophy.⁶ In this context, the Bayesian framework is often considered an extension to propositional logic to encompass an uncertain assessment of nature.⁷ We highlight in particular the use of a Bayesian reasoning of uncertainty as a basis for the development of trajectories in reaction networks as probabilistic structures; this reasoning is the grounding over which we developed the ideas of recursive network modeling and nesting networks.

The main theory we use to sustain as a modeling framework is reaction network theory.⁸ Although being proven useful to model a wide variety of systems in chemistry and biology, reaction networks are yet not properly stressed in the literature as having the powers it actually has. Having been developed in the chemistry context, they are traditionally viewed as a strict framework for reaction-like events, and not as a general theory of relational interactions among populations of, in principle, abstract entities. They already caught the attention of researchers in the mathematical field of category theory,⁹ but their use as a general framework for biological entities, especially for ecoevolutionary processes, is yet to be fully grasped. For example, reaction networks allow for the generation of abstract quantities, such as payoffs, that do not directly influence the occurrence of reactions, but only indirectly (by affecting rates); however, this use has never been considered before. A recent work, for example, proposing reaction networks to build payoff matrices for game theory, had the trouble to redefine payoffs as proxy network elements and, in the process, restricting more than necessary their use in the context.¹⁰ Additionally, combining reaction networks with Bayesian arguments, we can properly extend the notion of what reaction rates and the law of mass action actually mean in an abstract form. This is able to help us formalize the proper way to build reaction network models outside the physical scope of elements being assumed to not express internal properties capable of driving their dynamics (such as payoffs or any evolutionary trait). Beyond this potential, reaction networks also provide us with a developed mathematical apparatus featuring a probabilistic basis that connects with deterministic models as an infinite system limit.

Figure (1) shows a graphical representation of the framework we develop in this thesis. We take the theory of reaction networks and embed its derivation and conceptualization into a Bayesian paradigm. From that, we generalize the types of models that can be designed with reaction networks, aiming at applications to eco-evolutionary population systems. Then, the theory yields two types of kinetics, one more fundamental and stochastic, and another that is deterministic and arises from a limit of infinite system. These kinetics allow us to dynamically analyze the models using standard methods. Then, developing a method for nesting networks, we can submit the generalized reaction networks to an ecologically-driven evolutionary scale dynamics, also providing a generalization of evolutionary game theory and adaptive dynamics. All these methods can then be subjected to a Bayesian parameter estimation pipeline as a basis for model determination and connection with experimental data. Our framework encompasses design-level relational modeling of populations, stochastic and deterministic dynamical analysis, statistical analysis through parameter estimation, and a specialized unifying paradigm to treat questions from an eco-evolutionary setting. Our original contributions comprehend a particular conceptualization of the theories, a new methodology, and applications to biochemical and ecological systems.

1.1 Structure of the Thesis

This work is structured in the following manner. Chapters 2, 3, and 4 develop the required theory, but with an original understanding and organization of concepts. Chapter 2 exposes a notion of the Bayesian interpretation of probability and parameter estimation methodology and, under this context, the derivation of the methods of stochastic calculus that are pertinent to our framework. Chapter 3 defines, in a principled manner, the theory of reaction networks along with the detailed derivation of the stochastic and deterministic dynamics, also with their connection. Chapter 4 works out stochastic approximation and statistical methods applied to reaction networks.

Chapters 5, 6, and 7 develop original works featuring applications and further development of the framework's methodology. Chapter 5 presents an application on a



Figure 1 – The Framework. Pink processes are of design-level modeling; yellow processes are mathematical description steps; blue processes are part of analysis and application. Red borders indicate processes with original contribution. The connection of all processes and the Bayesian basis is an original contribution. Source: By the author.

biochemical setting focused on the probabilistic and statistical branches of the framework, analyzing the stochastic properties of a seminal model for oscillations in the context of protein production and genetic regulation. Chapter 6 is the robust and important work that extends reaction networks and our framework to generalize, unify, and provide foundations to existing modeling methodology of eco-evolutionary dynamics. Lastly, chapter 7 brings a clean application of the eco-evolutionary branch of the framework that elucidates central mechanistic factors behind the scarcely quantified ecological question of specialist versus generalist diets and heterogeneity of populations in predator-prey interactions. This application stresses the functionality of the design-level emphasis of our framework, demonstrating the ease and efficiency in determining the interactions that are relevant to the system's composition. The works in these chapters are presented in chronological order.

2 PROBABILISTIC FOUNDATION

This chapter develops the basis of the Bayesian paradigm of probability, introducing the sum and product rules as the main tools to relate probabilities and build probabilistic models, and also the derivation of the Bayes equation, which is the backbone of statistical analysis. The Bayesian interpretation of probability is supposed to connect the philosophical meaning of uncertainty to the operational mathematical tools of probability spaces, resulting in a methodology that extends the binary propositional logic.

Under the Bayesian paradigm, we proceed to define stochastic processes and Markov jump processes, then we provide a fully Bayesian derivation of the master equation for the time-evolution of probability densities of jump processes. The master equation will be the mathematical basis for the stochastic treatment of our modeling framework. There is nothing new in our Bayesian derivation of Markov jump processes apart from the accommodation of these methods into a Bayesian type of reasoning.

Finally, we show the basics of the parameter estimation procedure of Bayesian statistics, comprised of the search for a posterior distribution of parameters that arises from prior knowledge and a data likelihood that encodes the data generating process defined by the model.

2.1 Bayesian Probability Framework

The Bayesian interpretation of probability assumes that the understanding of nature is always attached to a point of view, and that the picture of the world given by data is fundamentally uncertain.^{5,7} From that, probability is defined as a measure of uncertainty, from a point of view of the models themselves when trying to produce statements about nature. Under this view, other forms of interpretations, such as frequencies or propensities, are understood as models of probability; models of assessment and control of uncertainty.

Following that interpretation, we can understand the Bayesian framework as an extension to propositional deductive logic, where, instead of binary 0 or 1 true/false values of propositions, we have a [0, 1] interval of possible values of certainty about a proposition. In this view, the objects of attention are propositions, statements about information we have, and we use the tools of probability spaces as a mathematical treatment for measuring the degree of belief of a given proposition. Probability is then an operator over statements outputting their plausibility, given a model. Thus, models and parameters are treated as subjects of propositions, and therefore have their values and definitions attached to probability statements. This is done in contrast with the standard approach of having random variables as arguments of probabilities, instead of uncertain propositions about

fixed variables.

We can define a logical sum and a logical product operations over propositions in order to recursively combine them into compound propositions. We define A, B as a product operation, meaning A and B; it is true only if both A and B are true. And A + Bas a sum operation, meaning A or B; it is true when at least one of A or B is true. Then, for example, we can talk about the proposition C defined as C = A + B, A or B. We saw that this means that C is true if A is true, regardless of B, or if B is true, regardless of A; so C is only false if A is false and B is false, which is equivalent to say that $\overline{C} = \overline{A}, \overline{B}$, that reads as: not C is equal to not A and not B, by the use of a negation operator. Figure (2) shows the and and or operations as Venn diagrams. This reasoning gives us a tool to transform between sum and product operations through the use of the negation operation:

$$\overline{A,B} = \overline{A} + \overline{B},\tag{2.1}$$

$$\overline{A+B} = \overline{A}, \overline{B}. \tag{2.2}$$

We can understand these as logical rules relating propositions (note that we can consider every equality between propositions also as a proposition, that is trivially true if it's a rule). Now, treating propositions as varying in degree of belief, the same as varying in plausibility in the interval [0, 1], we need to define a good operator to rigorously convey the notion of plausibility. These are probability distributions.



Figure 2 – Venn Diagrams. The diagrams show a visual representation of the four most important logical expressions: or operation, and operation, conditional operation, and negation. Blue plus red regions represent the space of all possible outcomes, only blue shows the outcomes being considered by the composed propositions. Source: By the author.

2.1.1 Probability Rules

Standard probability distributions, the mathematical objects that model the concept of probability, are linked with the conceptual measure of uncertainty by being compatible with its goals. We express these goals in the form of three *desiderata*, asserting rules to be followed by the plausibility of every proposition:

I) Plausibility is represented by real numbers.

II) Plausibility must increase continuously and monotonically with the addition of information supporting the truth of propositions, as well as respect deductive limits. This is a desiderate of qualitative agreement with rational consideration of data.

III) Plausibility must be consistent: different ways of obtaining a result must give the same result; all given relevant information must be considered and equivalent propositions must be represented with equivalent plausibility.

With these three desiderata, we choose probability spaces as a good mathematical measure of plausibility. Thus, given a model, P(A) is the plausibility of proposition A from that model, a number between 0 and 1. P(A) = 0 when A is certainly false and P(A) = 1 when A is certainly true. This identification also provides the transformation of sum and product operations, given as the probability rules:

$$ProductRule: P(A, B) = P(A|B)P(B) = P(B|A)P(A)$$
(2.3)

$$SumRule: P(A + B) = P(A) + P(B) - P(A, B)$$
(2.4)

The proposition A|B is the conditional proposition of A under B, read as A given B; it means the proposition A when we know that B is true (note that this does not imply temporal order between A and B, this is to stress that we are not talking about causal connections, but logical connections). With these rules, we build the negation of a proposition, \overline{A} , as

$$P(\overline{A}) = 1 - P(A). \tag{2.5}$$

Figure (2) shows Venn representations of a conditional proposition and of a negation of a proposition. A proposition and its negation form an exhaustive $(P(A + \overline{A}) = 1)$ and mutually exclusive $(P(A, \overline{A}) = 0)$ set, composed of two propositions. When N propositions are exhaustive and mutually exclusive, they are called a partition of the event space. Consider a partition represented as the set $\{E_i\}$, with i = 0, 1, 2...N. By applying the product and sum rules, we can write the probability of any proposition A using the conditionals over the $\{E_i\}$. This is called the law of total probability, and is useful for designing probability terms of conditioned propositions that we know something about:

$$P(A) = P((\sum_{i} E_{i}), A) = P(\sum_{i} E_{i}, A) = \sum_{i} P(E_{i}, A) = \sum_{i} P(A|E_{i})P(E_{i}).$$
(2.6)

The first equality comes from the fact that $\{E_i\}$ are exhaustive, the second equality is a distributive property of propositional sum. The third is the sum rule with $\{E_i\}$ being mutually exclusive, and then the product rule. As an example, suppose we have a set of dice, but with different numbers of sides. Then we randomly pick one to throw. We don't know directly the probability of any outcome, but we know the probability of each outcome given the die has n faces, so we can build the probability of an outcome using the law of total probability. In that case, the partition would be of propositions E_i = The die has i faces.

Another useful related operation is the marginalization of a joint probability, that is just a form of the law of total probability. Note what happens here,

$$P(A) = \sum_{i} P(E_i, A).$$
(2.7)

If we have the propositions A and E_i and their joint probability $P(E_i, A)$, we may use a partition built over E_i to remove E_i with a sum and obtain P(A). This is called a marginalization of $P(E_i, A)$ over E_i . This is particularly relevant in the case where E_i asserts that some variable has a given value. Then we may consider as the partition the set with propositions for every value in that variable's domain; we marginalize over the domain of that variable. If that variable is continuous, we have $E_i = \{$ The variable x is in the range between x_i and $x_i + dx \}$. Then we may write the marginalization process as

$$P(A) = \int dx_i P(E_i, A).$$
(2.8)

2.1.2 Bayes Equation

Statistical methods are mainly concerned with two major types of problems: 1) model selection, that uses data to establish criteria to choose between models for describing systems of interest; and 2) parameter estimation, that, given a model, uses data to infer parameter values of the model. Both interests are centered around the Bayes equation, which is simply derived from the product rule of probability:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$
(2.9)

It has this quality of inverting the conditional, making it possible for us to update our knowledge about proposition A by the use of information acquired about B (because they're related, information is shared between them). This can translate into update of our theories in light of new data. Suppose we have a set of hypotheses $\{H_i\}$ and a proposition representing data, D. Then, with I representing our prior information on the matter, we have

$$P(H_i|D,I) = \frac{P(D|H_i,I)P(H_i|I)}{P(D|I)} = \frac{P(D|H_i)P(H_i|I)}{\sum_i P(D|H_i,I)P(H_i|I)},$$
(2.10)

where the last equality is a law of total probability for D over the set of hypotheses (it's assumed that they form a partition, because they are naturally mutually exclusive, and, if they aren't exhaustive, it is rational to include the hypothesis that is the negation of the sum of every other, then making it exhaustive). This is the base of model selection and parameter estimation, as the set of hypotheses can represent alternate models or alternate versions of a model with different parameter values. This is also valid in continuous form, for a set of hypotheses parameterized by continuous values.

In this context of estimation by the use of data, we call $P(H_i|D, I)$ the posterior probability of H_i , that is the probability of H_i given that D is true, so given that we know the data. The term $P(D|H_i)$ is called likelihood of the data D over H_i , meaning the probability of the data D given that H_i is true. The term $P(H_i|I)$ is the prior probability of H_i , what we know before considering the data. The term in the denominator is of less importance and is usually regarded as a normalization constant, with the estimation problem represented as

$$P(H_i|D,I) \propto P(D|H_i,I)P(H_i|I), \qquad (2.11)$$

with the product of prior and likelihood acting as a kernel for the posterior, called the odds of that hypothesis in light of data. The prior information I is a formalization of our previous knowledge about the hypotheses. Normally, prior information becomes increasingly irrelevant as we accumulate more data.

The likelihood is where the probabilistic model of the system comes in. For example, we assume the validity of our dynamical model, that gives us the probabilities of the system being in every possibility of states, and with that we have probabilities of the system being where it was seen in the data. And by maximizing the likelihood, we can arrive at point estimations for the parameters, but a more robust treatment is made by obtaining the posterior distribution and summarizing it in all desired manners in order to obtain estimations concerning H_i . Model selection can be performed by calculating the odds ratio between hypotheses, that give the relative values of their posteriors.

This concludes our sketch of the general theory of the Bayesian framework considered in this work. We now turn to the Bayesian treatment of Markov jump processes, the basis of our stochastic modeling approach.

2.2 Stochastic Processes and the Markov Property

We start by defining a notion of a stochastic process. Our interest is in describing dynamical systems, which are systems changing over time. When we model these systems as stochastic, the model evolution is not completely known to us. The model dynamics is described by probabilistic trajectories over their states. At first, we could think of propositions concerning the entirety of a dynamical process, but then we would restrict ourselves to probabilities about whole trajectories. So, following our interest in knowing probabilities dealing with each moment in time, we mostly consider propositions that are concerned with what is happening to the system at each instant. The definition of a stochastic process aims at pairing propositions with instants in time and chaining them in order to represent the whole dynamical process.

For us, stochastic process will be the set of propositions $\{X_{s,t}\}$, for $t \in T$, with T being the relevant set of time instants, for $s \in S$, with S being the relevant set of possible system states, where each $X_{s,t}$ is, in a general form, read as

 $X_{s,t} = \{$ The system is in state s at time $t\}.$

Then, we may talk about the probability of the system being in state s at time t, $P(X_{s,t})$, or the probability of the system being at state s' at time t' if we know it's in state s at time t, $P(X_{s',t'}|X_{s,t})$. The models we study in this work are stochastic processes of a certain kind, they are Markovian models.

2.2.1 Markovian Stochastic Processes

Markovian processes are stochastic processes for which the probabilities associated with the system in future times depend only on knowledge about its current state. It means that the model doesn't hold any memory of previous states, and past knowledge has no bearing in its future. In practice, any conditioning on previous times actually is dependent only on the closest previous time. So, if $t_1 < t_2 < t_3$, we have that $P(X_{s_3,t_3}|X_{s_1,t_1}, X_{s_2,t_2}) =$ $P(X_{s_3,t_3}|X_{s_2,t_2})$. This means that a transition between states is characterized only by the current state, regardless of the past. With that, we can talk about transition probabilities $P(X_{s',t_n}|X_{s,t_{n-1}})$, meaning a probability of the system moving to a state s' at a time t' from the state s at a time t. Knowledge about some initial state and transition probabilities is sufficient to build the whole chain probability, just using the product rule together with the Markov property. For ease of notation, let's consider $X_{s_{t_n},t_n} = X_n$, with $t_n > t_m$ for n > m:

$$P(X_0, X_1, X_2) = P(X_2 | X_0, X_1) P(X_0, X_1) = P(X_2 | X_1) P(X_1 | X_0) P(X_0).$$
(2.12)

Like that, we can build a chain for up to an arbitrary proposition X_n . This means that a Markovian model is completely characterized by an initial state and its transition probabilities (these transition probabilities are often presented as a transition matrix, and for a continuous time set T we often talk about *transition rates*).

The theory of Markovian processes spreads in four possibilities as we consider the nature of the states and times sets, S and T. Both can have continuous or discrete indexes

for its elements. For discrete states and discrete times, we have the theory of discrete Markov chains, the system jumps over a network of states through discrete iterations. For continuous states and discrete times, we simply have the case of a continuous Markov chain, a case we'll briefly visit through the method of Markov chain Monte Carlo. For continuous states and continuous times, we have the Markov continuous stochastic processes, such as diffusion processes. Finally, for discrete states and continuous times, we have the Markov jump processes, the branch we are mainly concerned with in this work. Here, by a continuous passage of time, the system jumps between different states at random times.

Markov jump processes are governed by a Chapman-Kolmogorov differential equation that we call a master equation. The master equation is very often not solvable, and we can approximate it to continuous-states differential equations for continuous stochastic processes or even to deterministic differential equations with a limit of infinite system.

2.2.2 Bayesian Derivation of a Master Equation for Markov Jump Processes

This section presents a derivation of the master equation using a Bayesian reasoning, and we emphasize our reliance on the simple probability rules defined in the previous section. We are interested in models that are Markovian processes of continuous time and discrete state space, Markov jump processes. Consider a system Γ of this type, defined by the following assumptions:

1. Γ exists in a discrete state space, with states that can be uniquely determined by a set of numbers, each describing a component of Γ (usually translated to integer count numbers of each type of component). So, if Γ is a system determined by two components, two species N_1 and N_2 with counts n_1 and n_2 , then at a given time it's determined by the pair (n_1, n_2) contained in the set of possible states. We denote the state of the system with the vector \boldsymbol{n} with dimension equal to the number of system's components. In the example, $\boldsymbol{n} = (n_1, n_2)$.

2. Γ evolves by changing states along a continuous passage of time. So, Γ has a continuous set of time instants and is a "jump process".

3. Γ obeys the Markovian property and we know the transition rates for the system. We'll rewrite them in terms of the transition probabilities. Also, the jumps to the many different states are independent events.

4. We can divide the time set into defined intervals dt for which we can consider $\mathcal{O}(dt^2)/dt \to 0$ for $dt \to 0$ and that we can assure transition rates to be approximately constant during dt.

So, if we know that Γ is in a state n_1 at a time t_1 , it can jump to any other state n_2 at a posterior time t_2 with a probability $P(\Gamma_{n_2,t_2}/\Gamma_{n_1,t_1}, Z_{t_1,t_2}) = Tr(n_1, t_1 \to n_2, t_2)$, with $Z_{t_1,t_2} = \{$ There are no jumps during the interval $t_2 - t_1 \}$ and $n_1 \neq n_2$. Since Γ is

Markovian, the transition probability does not depend on states before t_1 . Remember that $\Gamma_{n,t} = {\Gamma \text{ is in state } n \text{ at time } t}$. In order to completely specify the system, we must connect the transition probabilities to the known transition rates. They are defined as follows:

$$Tr(n_1, t_1 \to n_2, t_2) = W_{n_1, t_1 \to n_2, t_2} dt,$$
 (2.13)

as long as $t_2 - t_1 = dt$. But we have problems. When the system jumps, this probability breaks; how many times can we expect it to jump during a time interval dt? Also, what is the probability of the system remaining in the same state after dt, $\overline{\sum_{n_j} Tr(n, t \to n_j, t + dt)} = 1 - \sum_{n_j} Tr(n, t \to n_j, t + dt)$? Can we know it?

2.2.2.1 Transitions

We are interested in the limit $dt \to 0$, so we can solve our problems by proving the following statement: During a passage of time dt starting at time t, the system can jump once, from state \mathbf{n} to any different state \mathbf{n}_i with probability $W_{\mathbf{n},t\to\mathbf{n}_i,t+dt}dt$. Also, the system can jump more than once with probability $\mathcal{O}(dt^2)$ and remain in state \mathbf{n} with probability $1 - \sum_{n_i} W_{\mathbf{n},t\to\mathbf{n}_2,t+dt}dt + \mathcal{O}(dt^2)$.

For that, consider the propositions, using the notation with implicit dependency of time $W_{n,n_j} = W_{n,t \to n_j,t+dt}$:

 $K_k = \{ \text{With } \Gamma \text{ being in state } \boldsymbol{n} \text{ at time } t, \text{ exactly } k > 0 \text{ transitions occur during the next interval } dt, k_j \text{ from } \boldsymbol{n} \text{ to } \boldsymbol{n}_j \neq \boldsymbol{n} \text{ with constant probability } W_{n,n_j} dt \text{ and the constraint } \sum_{n_j} k_j = k. \}$

With constant independent transitions, $P(K_k)$ follows a multinomial distribution with k trials and a number of possible outcomes equal to the number of possible states. One of its possible outcomes never happens in any trial, representing the system jumping to nowhere in that trial, so:

$$P(K_k) = \sum_{\sum k_j = k} \frac{k!}{\prod_j k_j!} \prod_{n_j \neq n} (W_{n,n_j} dt)^{k_j} (1 - \sum_{n_j \neq n} W_{n,n_j} dt)^0 = (\sum_{\sum k_j = k} \frac{k!}{\prod_j k_j!} \prod_{n_j \neq n} W_{n,n_j}^{k_j}) dt^k.$$
(2.14)

We can see that this probability is proportional to dt^k , so we have $P(K_k) = \mathcal{O}(dt^k)$. In particular,

$$P(K_1) = \sum_{n_j \neq n} W_{n,n_j} dt.$$
 (2.15)

For no transitions, we have the proposition K_0 , defined as $K_0 = \overline{\sum_k K_k} = \overline{\sum_{n_j} Tr(n, t \to n_j, t + dt)}$. Noting that the K_k s are mutually exclusive, using the sum rule, we have

$$P(K_0) = P(\sum_k \overline{K_k}) = 1 - P(\sum_k K_k) = 1 - \sum_k P(K_k) = 1 - \sum_k P(K_k) = 1 - P(K_1) - \sum_{k>1} P(K_k) = 1 - \sum_{n_j \neq n} W_{n,n_j} dt + \mathcal{O}(dt).$$
(2.16)

This ends our justification and solves our problems. We can now make sure that at most one transition occurs during dt in the limit.

2.2.2.2 Master Equation

Finally, we turn to the task of building the master equation. Let's give some easier names to our relevant propositions:

 $X_0 = \Gamma_{\boldsymbol{n}_0, t_0} = \{ \Gamma \text{ starts in an initial state } \boldsymbol{n}_0 \text{ at time } t_0 \}.$

 $X = \Gamma_{n,t} = \{ \Gamma \text{ is in state } \boldsymbol{n} \text{ at time } t > t_0 \}.$

And for each possible state n_i :

$$Y_i = \Gamma_{n_i, t} = \{ \Gamma \text{ is in state } n_i \text{ at time } t' < t \text{ with } t' > t_0 \}$$

The goal now is to assign a probability to proposition X using the Y_i s. Let's look at the proposition $\sum_i Y_i$; it means Y_1 , or Y_2 , or Y_3 , etc. It essentially means that Γ is in any possible state at time t', and this is always true, the set $\{Y_i\}$ is exhaustive. Also note that Y_i s are mutually exclusive, because at the same time Γ can only be in one state. So the set $\{Y_i\}$ is a partition of the event space at time t', a set of mutually exclusive events covering the whole space. Since the sum of Y_i s is always true, and using product properties, we can write

$$X = (\sum_{i} Y_{i}), X = \sum_{i} X, Y_{i}.$$
(2.17)

Let's begin assigning probabilities to porpositions. Note that the products X, Y_i are also mutually exclusive, so we have, using the sum rule

$$P(X|X_0) = P(\sum_i X, Y_i|X_0) = \sum_i P(X, Y_i|X_0).$$
(2.18)

Now we use the product rule

$$P(X|X_0) = \sum_{i} P(X|Y_i, X_0) P(Y_i|X_0)$$
(2.19)

and then the Markovian property, that says $P(X|Y_i, X_0) = P(X|Y_i)$,

$$P(X|X_0) = \sum_{i} P(X|Y_i) P(Y_i|X_0).$$
(2.20)

See that all this is just the law of total probability applied to X with the partition $\{Y_i\}$. Now, why is it relevant to rewrite $P(X|X_0)$ in terms of the Y_i s? It is because, with our specification of Γ , we have knowledge about local transition probabilities, but the known initial state X_0 may be as far as we wish from the arbitrary state X we want to describe. Using the Y_i s as bridges, we can make t' "adjacent" to t and smuggle the known transition probabilities into our derivation. With adjacent meaning distant by an interval dt.

We need to specify a t' of Y_i that is adjacent to the t of X: t' = t - dt. If this is true, we have the probabilities $P(X|Y_i)$ in terms of the transition rates. There are two

cases; 1) $n_i = n$ and it means that no transitions occur, and 2) $n_i \neq n$ and it means that some transition with rate $W_{n_i,t-dt\to n,t}$ occurs. So we separate the sum in these two possibilities

$$P(X|X_0) = \sum_{n_i \neq n} P(X|Y_i) P(Y_i|X_0) + P(X|Y_n) P(Y_n|X_0), \qquad (2.21)$$

with Y_n defined as Y_i for the case of $n_i = n$. The transition probabilities are, using the same implicit time-dependency notation as above,

$$P(X|Y_i) = W_{n_i,n}dt + \mathcal{O}(dt^2)$$
(2.22)

because we are going from n_i to n. The probability of no transition is

$$P(X|Y_n) = 1 - \sum_{n_i \neq n} W_{n,n_i} dt + \mathcal{O}(dt^2)$$

$$(2.23)$$

because we are going from n to all other n_i s. Note the exchange in the indexes of W. Putting more clearly, in case 1 the system is jumping from n_i to n, and in case 2 the system already is in n and we consider the negation of it going to any other possible n_i .

Inserting in the equation for $P(X|X_0)$, we have

$$P(X|X_0) = \sum_{n_i \neq n} \left(W_{n_i,n} dt + \mathcal{O}(dt^2) \right) P(Y_i|X_0) + \left(1 - \sum_{n_i \neq n} W_{n,n_i} dt + \mathcal{O}(dt^2) \right) P(Y_n|X_0).$$
(2.24)

Just reorganizing the equation, we arrive at

$$\frac{P(X|X_0) - P(Y_n|X_0)}{dt} = \sum_{n_i \neq n} \left(W_{\mathbf{n}_i, n} P(Y_i|X_0) - W_{n, n_i} P(Y_n|X_0) \right) + \frac{\mathcal{O}(dt^2)}{dt}.$$
 (2.25)

Finally, we perform the limit $dt \to 0$. With this, the left side of the equation becomes the derivative of $P(X|X_0)$ in relation to time and $\frac{\mathcal{O}(dt^2)}{dt} \to 0$. $P(Y_n|X_0)$ on the right side becomes $P(X|X_0)$ as $t' \to t$ (note that the Y_i s now represent Γ in time t with the limit imposing $t' \to t$). We have the Master Equation:

$$\frac{dP(X|X_0)}{dt} = \sum_{n_i \neq \mathbf{n}} \left(W_{n_i,n} P(Y_i | \{t' = t\}, X_0) - W_{n,n_i} P(X|X_0) \right).$$
(2.26)

We can now change the probabilities to the more explicit distributions notation. The distribution that $P(X|X_0)$ follows has as variables the state vector \boldsymbol{n} and the time t. If we define the probability of no transitions occurring as $W_{\boldsymbol{n},\boldsymbol{n}}$, we can sum over all states of Γ without altering the equation (note that the additional term $n_i = n$ ends up being zero). Calling the distribution $P(X|X_0) = \Pi(\boldsymbol{n}, t)$, we have

$$\frac{d\Pi(\boldsymbol{n},t)}{dt} = \sum_{\boldsymbol{n_i}} \left(W_{\boldsymbol{n_i},\boldsymbol{n}} \Pi(\boldsymbol{n_i},t) - W_{\boldsymbol{n,n_i}} \Pi(\boldsymbol{n},t) \right),$$

$$\Pi(\boldsymbol{n}, t_0) = \delta(\boldsymbol{n}, \boldsymbol{n_0}). \tag{2.27}$$

Note that we can generalize the proposition X_0 into a set of propositions to mean that the state of Γ in t_0 is uncertain, with different probabilities of being in different states. We don't need to know the exact initial state for the equation to be valid. For systems with a finite number of states, we can even know nothing about the initial state, assigning to the set of X_0 a uniform probability distribution over the sates.

The solution of this equation gives the probability of proposition X happening once that X_0 happened, that means Γ has transitioned to state n after an arbitrary number of jumps during an arbitrary time interval $t - t_0$.

We can interpret the Master Equation in terms of gains and losses in probability; it means that the right side is viewed as a net gain in probability at time t, the first term being the gain from transitions into n and the second term being the loss from transitions away from n.

The Master Equation is the differential form of the Chapman-Kolmogorov equation. In this work, we'll consider only time-independent transition rates (homogeneous Markovian systems), so the abridged notation, $W_{n_i,n}$, will always be used.

2.3 Parameter Estimation

In this section, we'll connect model to data by building the parameter estimation process.¹¹ The input of the process is the data, measured from the physical systems of interest (this work is concerned specifically with reaction network systems and we'll test the estimation models using simulated data generated from stochastic simulations). The output of the process is a posterior probability function, the probability density of the estimated parameters under the model.

We start with the Bayes equation, that gives us the parameters' posterior distribution,

$$P(H_k|D,I) = \frac{P(D|H_k,I)P(H_k|I)}{P(D|I)}.$$
(2.28)

D is a proposition asserting the data, we'll specify it later. The set of hypotheses H_k will mean the following:

 $H_k = \{ \text{The model } m_k \text{ with parameter values } \boldsymbol{\theta}_i \text{ is true} \}.$

And we'll write $H_k = M_k, \Theta_i$, with these new propositions meaning:

 $M_k = \{ \text{The model } m_k \text{ is true} \}.$

 $\Theta_i = \{\text{The vector of parameters}, \boldsymbol{\theta}, \text{ for the given model is between } \boldsymbol{\theta}_i \text{ and } \boldsymbol{\theta}_i + d\boldsymbol{\theta}\}.$

We are interested in parameter estimation and will work with a fixed model, so we may omit the proposition M_k as always true for our model. We also omit the data probability, since it doesn't involve Θ_i . Then, we'll work only with the estimation kernel, on the form

$$P(\Theta_i|D,I) \propto P(D|\Theta_i,I)P(\Theta_i|I).$$
(2.29)

There are two elements to deal with for the estimation, 1) the data likelihood given the parameter values, $P(D|\Theta_i, I)$, and 2) the parameter's prior information, $P(\Theta_i|I)$.

2.3.1 Priors

The choice of prior depends on everything we know about the model and the parameters. It can reflect the form of our likelihood, we may choose them to be conjugate, so the form of the posterior doesn't change by addition of new data to the estimation problem. It also depends on the nature of the parameters. We'll consider here only continuous parameters; then we may deal with two kinds of parameters: space parameters, that can be negative and depend on a choice of origin; and scale parameters, that are only positive and express our chosen units. The prior's domain must contain all known possible values for the parameters and can't contain values we know to be impossible (see that the posterior is zero for values outside this domain). It has to, at least in order of magnitude, represent our state of knowledge about probabilities for different parameter values. When we have no useful probabilistic information to use, we need to use non-informative priors; this can be tricky, because the specifications of the problem may result in a nontrivial way to invoke uniformity over parameter spaces.

We may consider the posterior as a prior for inclusion of more data in the future, so the prior also loads information from possible previous measurements, and this also enables the notion of sequential data processing. Now, the specifics of prior choices is an experimental analysis and it depends on details of the problems, so we leave further developments to the applications when in need of them.

2.3.2 Data Likelihood

Before inputting the data into the likelihood, let's build the functional form of the likelihood. Consider a variable \boldsymbol{y} to represent the data point at a time t. Then, assuming a continuous data variable just for convenience, the probability of having a data point valued \boldsymbol{y} at this time is the same as the one for the proposition: $Y = \{\text{The data is in the range } [\boldsymbol{y}, \boldsymbol{y} + d\boldsymbol{y}] \text{ at time } t\}$. This analysis can be readily adapted to discrete variables. Now, consider what the model gives us. For each possible state of the model, \boldsymbol{x} , we have a probability for the system to be in that state at any time, given an initial state. The probability density for a continuous \boldsymbol{x} is the same as the probability of the proposition $X_t|X'_0$ (remembering to consider an initial state), with: $X_t = \{\text{The state of the model is in the model is in the model is in the state of the model is in the probability of the model is in the state of the model is in the probability of the model is in the state of the model is in the probability of the model is in the state of the model is in the probability of the model is in the state of the model is in the probability of the model is in the state of the model is in the probability of the model is in the state of the model is in the probability of the model is in the state of the$

the range $[\boldsymbol{x}, \boldsymbol{x} + \boldsymbol{dx}]$ at the time t}. If we consider a random error for each measurement, we have a variable \boldsymbol{e} representing the possible values for the error, following a probability density equal to the one for the proposition: $E = \{\text{The measurement error is in the range} [\boldsymbol{e}, \boldsymbol{e} + \boldsymbol{de}]\}$. Now we can write the data \boldsymbol{y} in terms of the variables representing the model and the error. If we define the variables \boldsymbol{x} and \boldsymbol{e} with the propositions $X_t|X_0'$ and E, we have that

$$\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{e}. \tag{2.30}$$

This equation represents a measurement model. The function $f(\mathbf{x})$ is any transformation of the physical quantities that matches their relation to the measured quantity. The simplest form of the measurement model is one where we consider no measurement errors and the measurements are exactly the variables of our physical model, so $\mathbf{y} = \mathbf{x}$. From now on, we'll consider this simple relation between \mathbf{y} and \mathbf{x} but with measurement errors, so

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e}. \tag{2.31}$$

This means, for example, that, if our physical model gives us molecular concentrations for \boldsymbol{x} , we are directly measuring those same molecular concentrations, but with error e.

We want to compute the likelihood $P(Y|\Theta_i, I)$ as a function of \boldsymbol{y} , but in terms of what we know, that are the densities of \boldsymbol{x} and \boldsymbol{e} . Here, we assume enough control over the measurement errors to know their probability densities.

Whenever we want to express a probability of a proposition in terms of other propositions, we may want to marginalize their joint probability, in a maneuver to include these propositions just to remove them again:

$$P(Y|\Theta_i, I) = \int \int d\mathbf{x} d\mathbf{e} P(Y, X_t, E|X'_0, \Theta_i, I), \qquad (2.32)$$

We want a probability density on the variable \boldsymbol{y} in terms of the parameters of the model and the measurement error, that's why we invoke the distributions for \boldsymbol{x} and \boldsymbol{e} . With that distribution for \boldsymbol{y} , we'll later input the data value and get the probability of that data value given the model, i.e. the likelihood for that data. Using the product rule and considering that X and E are independent (model independent of measurement errors),

$$P(Y|\Theta_i, I) = \iint dx de P(Y|X_t, E_j, X'_0, \Theta_i, I) P(X_t|X'_0, \Theta_i, I) P(E, \Theta_i, I).$$
(2.33)

Using the fact that $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e}$, we have

$$P(Y|X_t, E, X'_0, \Theta_i, I) = \delta(\boldsymbol{y} - \boldsymbol{x} - \boldsymbol{e}).$$
(2.34)

Putting $\delta(\boldsymbol{y} - \boldsymbol{x} - \boldsymbol{e})$ inside that integral has the effect of singling out the value of \boldsymbol{e} . Call the proposition:

 $E_{y-x} = \{ \text{The measurement errors are between } \boldsymbol{y} - \boldsymbol{x} \text{ and } (\boldsymbol{y} + \boldsymbol{dy}) - (\boldsymbol{x} + \boldsymbol{dx}) \}.$

Then, we end up with

$$P(Y|\Theta_i, I) = \int d\boldsymbol{x} P(X_t | X'_0, \Theta_i, I) P(E_{y-x} | I).$$
(2.35)

If we consider $P(X_t|X'_0, \Theta_i, I) = f_m(\boldsymbol{x}, \boldsymbol{\theta}_i, t)$, the model distribution over \boldsymbol{x} , and $P(E_{y-x}|I) = f_E(\boldsymbol{y} - \boldsymbol{x})$, the error distribution over \boldsymbol{e} ,

$$P(Y|\Theta_i, I) = \int d\boldsymbol{x} f_m(\boldsymbol{x}, \boldsymbol{\theta}_i, t) f_E(\boldsymbol{y} - \boldsymbol{x}).$$
(2.36)

You may recognize this equation as a convolution integral. This relates to the fact that y = x + e, the sum of variables is computed as a convolution at the level of probabilities of propositions asserting those variables.

A simple and usual error model is the following: we assume the measurement error to be distributed as $\mathcal{N}(\boldsymbol{e}|\boldsymbol{0}, \sigma_e^2 \boldsymbol{I})$, a multivariate normal distribution with mean zero and a known, constant, standard deviation over all measured variables and data points (\boldsymbol{I} is the identity matrix of dimension equal to the system's dimension). With that, we have $f_E(\boldsymbol{y} - \boldsymbol{x}) = \mathcal{N}(\boldsymbol{y} - \boldsymbol{x}|\boldsymbol{0}, \sigma_e^2 \boldsymbol{I})$. We now assume the simplest kind of model for the measurement process: a deterministic model. In that case, $f_m(\boldsymbol{x}, \boldsymbol{\theta}_i) = \mu(\boldsymbol{x}, \boldsymbol{\theta}_i)$ is a deterministic function of the data and the parameters. For example, in the case of a linear regression of a one dimensional model, we have $\mu(\boldsymbol{x}, \boldsymbol{\theta}_i) = \theta_1 x + \theta_2$. The result of the convolution will then be trivial

$$P(Y|\Theta_i, I) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{\mu}(\boldsymbol{\theta}_i, t), \sigma_e^2 \boldsymbol{I}).$$
(2.37)

By having the data points (\boldsymbol{y}, t) , we evaluate this likelihood in terms of the parameters $\boldsymbol{\theta}_i$. If we use a different value for the vector of parameters, say $\boldsymbol{\theta}_{i'}$, we obtain another value of the likelihood. The posterior probability density will then be a function of the parameters, given by the prior and likelihood functions.

Let's talk about the data proposition D in order to quantify the likelihood function using a data set obtained from a measurement operation. First, let's suppose that all variables specifying the state of the model are observed in a measurement, so a single measurement gives us a data vector \boldsymbol{w} that is the same dimension as our physical model. This vector represents the observed quantities related to the variables of the physical model. Then we have for the data:

 $D = \{ \text{A measurement observed a set of } d \text{ data points } \{([\boldsymbol{w}_j, \boldsymbol{w}_j + d\boldsymbol{w}], [t_j, t_j + dt])\}$ with j = 0, 1, ..., d - 1, where \boldsymbol{w}_j is a vector for the *j*-th measurement and t_j the measured time at that point. $\}$.

We have $D = \prod_{j=0}^{d-1} D_j$ for the D_j individual measurements. If we consider a string of measurements given a model that's Markovian, we have, using the product rule and the Markovian property

$$P(D|\Theta_i, I) = P(D_0|\Theta_i, I) \prod_{j=1}^{d-1} P(D_j|D_{j-1}, \Theta_i, I).$$
(2.38)
Each term $P(D_j|D_{j-1},\Theta_i,I)$ is, as we calculated, given by

$$P(D_j|D_{j-1},\Theta_i,I) = \int dx P(X_t|X'_0,D_{j-1},\Theta_i,I) P(E_{w_j-x}|I).$$
(2.39)

Here, $t = t_j - t_{j-1}$. How can that conditioned state, \mathbf{x}' , be chosen? The initial time and a possible initial conditioned state are given with the earlier measurement, D_{j-1} . That's the last observation we have of a system that we are modeling as Markovian, so let's use D_{j-1} from now on. In this scenario, we know that at the initial state the system is at \mathbf{w}_{j-1} with an uncertainty given by \mathbf{e} . That's a random initial state for the model,

$$P(X'_0|D_{j-1},\Theta_i,I) = P(E_{w_{j-1}}|I) = f_E(w_{j-1}).$$
(2.40)

We also need to compute the actual initial measurement of the whole data chain, $P(D_0|\Theta_i, I)$. What we do depends on the situation, we have to *model* the initial measurement. When the data is modelled by a normal distribution, it's natural to assume that $P(D_0|\Theta_i, I)$ is a normal distribution. Then, as a standard choice, we'll have the initial mean vector and variance matrix treated as parameters to be inferred, as parts of $\boldsymbol{\theta}_i$.

Then, following our deterministic model example, the likelihood for the data set D is given by

$$P(D|\Theta_i, I) = \mathcal{N}(\boldsymbol{w_0}|\boldsymbol{\theta_0}, \boldsymbol{\theta'_0}) \prod_{j=1}^{d-1} \mathcal{N}(\boldsymbol{w_j}|\boldsymbol{\mu_j}(\boldsymbol{\theta_i}, \Delta t_j, \boldsymbol{w_{j-1}}), \sigma_e^2 \boldsymbol{I}).$$
(2.41)

This will be a function of the data points $(\{w_j, t_j\})$ and the model parameters θ_i . There we have, for this specific case, the whole likelihood function in terms of the model output and the data. The model output, in turn, is a function of the model parameters (what we want to estimate with the posterior).

2.3.3 Incomplete Measurements

Before, We assumed that all system's variables were observed. What happens if we have measurements only of a subset of the system's variables? It gets harder to estimate (even impossible for some kinds of parameters in certain systems). But the difference is that we also have to marginalize over the unobserved variables, to extract the subsection of the model that interacts with the data:

$$P(D_{j}|D_{j-1},\Theta_{i},I) = \iint dx de P(D_{j}|X_{t},X_{0}',E,D_{j-1},\Theta_{i},I)P(E|\Theta_{i},I) \int du P(X_{t}|X_{0}',D_{j-1},\Theta_{i},I).$$
(2.42)

The vector \boldsymbol{u} represents the unobserved variables, and the vectors coming from D have dimension equal to the number of observed quantities. The integral on \boldsymbol{du} must be evaluated over the model's distribution. If it's a deterministic model (or even a possibly time-dependent normal distribution), the integral is the easiest possible, the result is just to ignore the unobserved components. The posterior distribution is the final Bayesian result of the parameter estimation process. Once we have a posterior $P(\Theta_i|D, I)$ for the model parameters, we can summarize it by our decision in order to estimate the parameters in the ways suited to the systems we are analyzing.

3 REACTION NETWORKS

This chapter defines and develops the theory of reaction networks in a way that is most suited to our needs. There is more than one possible focus that we could give to different aspects of reaction networks, and we will emphasize the dynamical properties yielded by the network's structure rather than the structure of reactions themselves.¹² Thus, we carefully describe how transition rates arise from our assumptions and how we can interpret their different terms. We also emphasize the abstract relational nature of the networks as modeling tools, even shifting the meaning of reactions to a more generalized form of interactions resulting from encounters between elements.

We then determine the functional forms of transition rates for both the deterministic and stochastic network kinetics, as well as defining the limit of infinite systems that connect the two descriptions and connect the discrete count-numbers and continuous densities that define the state-space of reaction networks. We do this by first interpreting the stochastic assessment of the networks as arising from modeling uncertainties, under the Bayesian probabilistic paradigm.

We also give some simple examples of well-known systems that are usually defined at the ambiguous level of differential equations and that can be properly specified in terms of local interactions encoded as reactions. We choose systems from different fields of biological modeling.

3.1 Reaction Networks

Reaction networks are an efficient way to model mechanisms of local interactions between components generating a more complex dynamical behavior that shows at the global population level. By using a set of reactions happening at a local level (local in terms of the network's dynamics) and relating to more fundamental mechanisms, we can find and describe the trajectories and properties of the whole system, at a collective level. Reaction networks are then viewed as a set of species composed of identical, individual units that are the elements of each species, and that interact with each other when they encounter each other, through defined reactions. Their interactions, in turn, result in the creation or destruction of elements.¹³

The real nature of these species is not relevant to the framework. Only the relations between their elements. Reaction networks abstract away the differences between elements and group them by how they relate to other elements; the identical elements of a species are defined as the ones that interact in the same ways. This relational construction is essential to the generalizing power of reaction networks; they don't talk about specific physical systems, but abstract entities that can behave in ways that physical systems do.

3.1.1 Definition

There are many ways to introduce reaction networks, and here we define them in a way that is more readable in the language of this work. Consider the reaction network as consisting of three sets: 1) a set of species, that are collections of identical elements interacting in the same ways, 2) a set of complexes, that are linear combinations of species with integer coefficients, divided into reactant and product complexes 3) and a set of reactions, that are maps between complexes of species. In a reaction network, elements follow their own individual trajectories, independently of each other, and when they encounter in groups that are reactant complexes, they can interact in a way that results in a reaction, transforming that group of reactants into a group of products.

1) The set of species is a finite set of N variables representing the system's N different component species, X_i , with i = 1, 2..., N. Each species is composed of many identical elements, measured as a count-number of elements n_i . Given that, we enunciate a key property of a reaction network: The vector of count-numbers, $\mathbf{n} = (n_1, n_2...n_N)^T$, uniquely determines the state of the reaction network. So, its state space is formed by all possible combinations of different count-numbers, and no other property is able to discriminate states, regardless of the possible underlying differences, interactions, and even other elements. In that sense, all elements must be homogeneously mixed on average, without any distances structured among them. Also, underlying mechanisms are modeled as not relevant enough to affect the structure of reactions or the count-numbers of species with effects not covered by reactions. Note that the dynamics in focus with reaction networks are the dynamics of the concentrations of species, that's what defines the network states (so the networks evolve as reactions changing concentrations). As an example, if a reaction network consists of three species, we represent them as X_1, X_2 , and X_3 , and the vector of elements count-numbers as $\mathbf{n} = (n_1, n_2, n_3)^T$.

2) The set of complexes is defined over the set of species as the relevant groups of species. The reaction complexes tell us what combinations of elements are expected to result in reactions (reactant complexes) and what combinations are expected to result from reactions (product complexes). The reactions will transform complexes into other complexes, and that's what constitutes the changes of states. Complexes are collections of species and composed of a certain number of elements, denoted by integer coefficients. Complexes have the general form

$$\sum_{i=1}^{N} c_i X_i. \tag{3.1}$$

In the example of a network of three species, we may have many different sets of complexes, for example this one: $\{X_1 + 2X_2, X_3, 3X_1, \emptyset\}$. Empty complexes as reactants mean that a reaction can occur spontaneously, without encounters (simply creating elements); and empty complexes as products mean that a reaction will result in the suppression of every reactant element (destroying them). How come elements are created from nowhere? In physical systems, this will be a result of an underlying process not explicitly modeled at the level of the network.

3) The reaction set is the set of relations that determine the allowed transitions between states of the reaction network. The reactions are information about how the components interact to dynamically affect their numbers through encounters of reactant complexes. A reaction network will have a countable set of R reactions represented by the numbers r = 1, 2..., R. A reaction r is expressed as dependent on the X_i as

$$\sum_{i=1}^{N} s_{ir} X_i \to \sum_{i=1}^{N} s'_{ir} X_i, \quad r = 1, 2...R.$$
(3.2)

We read a reaction as: at the left of the arrow, we have the reactant complex for this reaction (these elements are used, suppressed from the system), and at the right side we have the product complex (these elements are formed or preserved, appearing as a result of the reaction). Following the example given so far, we could have many possible reactions from the complexes defined, here is an example: $X_1 + 2X_2 \rightarrow 3X_1$, $3X_1 \rightarrow X_3$, $X_3 \rightarrow \emptyset$.

Note that the structure of a reaction as a map between complexes implies by how much the state n can change. The allowed transitions are represented by the coefficients of complexes on a reaction map, denoted by s_{ir} for reactants and s'_{ir} for products, and called stoichiometric coefficients (the nomenclature comes from actual chemical reactions). They mean that, during a reaction r, s_{ir} elements of component X_i are destroyed and s'_{ir} elements are created; or, equivalently, $s'_{ir} - s_{ir}$ elements appear or disappear, depending on the sign. In the reaction network, the reactions happen independent from each other, so when we say that the interactions have a local nature, we mean that the state transitions happen by processes on the level of elements that have no dependence over what's happening away from there and are not directly affected by external processes.

3.1.2 Stoichiometric Coefficients

All stoichiometric coefficients together uniquely define the reactions of a network and their difference $S_{ir} = s'_{ir} - s_{ir}$ uniquely defines the state transitions; to represent the reactions happening, we can think of a stoichiometric operator that acts as $O(n_i) =$ $n_i + (s'_{ir} - s_{ir})$ in all system components, moving from one state to another (it is common to define the system together with a stoichiometric difference matrix, S). Such an operator, in this context, is called a step operator, executing the event of a reaction over any function of n:

$$E_i^{S_{ir}}g(n_1,...,n_i,...,n_N) = g(n_1,...,n_i + S_{ir},...,n_N).$$
(3.3)

Thus, a reaction r happens as the operation $\prod_i E_i^{S_{ir}} \boldsymbol{n}$. See that it allows us to express every state in terms of a given state. All this means, in other words, that "jumps" between states

in the state space are determined by the information contained in the difference between stoichiometric coefficients, that informs how much each reaction will shift components in state space through each transition that happens by the occurrence of a reaction. So, trajectories in state-space are composed of sub-spaces determined by the stoichiometric difference matrix S, given an initial state.

3.1.3 Transition Rates

Reaction networks, as defined above, don't possess any component that identifies their actual dynamical evolution. For that, we need to combine the reaction network with a measure of how often reactions will happen, given a state \boldsymbol{n} . For that, we add a dynamical structure to the reaction networks. We say that reaction networks are associated with a dynamical process of a continuous time passage, and then the system state is a function of time $\boldsymbol{n} = \boldsymbol{n}(t)$. We define the set of R functions, one for each reaction, W_r , and call them transition rates. This means that, when the reaction network is in state \boldsymbol{n} , at a moment of time, W_r will represent the rate at which reaction r will occur, altering the state of the system.

There are two distinct sources of processes affecting W_r . 1) On one side, we have the trajectories of elements in a reaction network converging so as to form reactant complexes in a moment of time, allowing the occurrence of the associated reaction. We have to explicitly model this dependence in order to understand the time-evolution of the network as it is defined. 2) On the other side, we have the fact that not every encounter of reactant elements will result in a reaction. There can be any process not captured by the network's structure that can affect the occurrence of a reaction given that a reactant complex is formed. These two sources of dynamical processes affecting the transition rates are independent, and the first is a process happening in the reaction network while the second is assumed as an underlying source of mechanisms. Underlying here means that these mechanisms can affect the system, but only through reactions, and lying on a different scale (consider for example the sense in which quantum mechanics may be seen as an underlying process giving rise to classical mechanics, from a macroscopic point of view). So we break down these two dependencies in the definition of the transition rates

$$W_r = k_r f_r. aga{3.4}$$

We identify f_r as an event-function, giving the transition rate's dependence on the elements trajectories in the reaction network, the encounter chance of elements forming a reactant complex; and k_r as a reaction rate, being the rate by which reactions locally occur, given an encounter of elements forming a reactant complex.

The reaction rates k_r are usually constant, because otherwise they would represent interactions not being mechanistically explained by the network. But, mathematically, it works fine to have it as a function of n, and it can happen as a result of approximations on reaction networks, such as time-scale related or quasi-steady state approximations, that bundle up a set of interactions into one single composed reaction. They can also be dependent of external parameters, such as a measure of temperature, of element's affinities, element's capacity of seeing each other, etc. Independent of the functional form of the reaction rates k_r , we view them as external rates assigned to reactions. We also expect the event-functions to have the same functional dependency over the reactant complex for any reaction. Therefore, the functional form of f_r will be part of the definition of the reaction network's kinetics. And, given a specific kinetic description of the network, we represent the reaction rates assigned to the reactions, over the mapping arrow:

$$\sum_{i=1}^{N} s_{ir} X_i \xrightarrow{k_r} \sum_{i=1}^{N} s'_{ir} X_i, \quad r = 1, 2...R.$$
(3.5)

3.1.4 Physical Systems

The major restriction of the reaction networks when we think of physical systems, as we mentioned, is that the state of the system must be determined by a single set of numbers that is the quantity of elements of species, their count-numbers. The system must give scalar non-negative measures of concentrations and be homogeneous in all other aspects. Also, the time-scale of the reactions must permit that these conditions keep valid when reactions happen. These traditionally imply spatial homogeneity (as in well-stirred systems) and rapid thermal equilibrium after reactions occur. If space is structured, the state of the system must include information about the spatial coordinates of elements in addition to their numbers; the same goes for temperature or anything more that differs. Fluctuations of quantities, other than the change in the number of elements, like external parameters, as a consequence of reactions or not, usually can't be modelled. So the assumptions of "good behavior" and rapid diffusion must hold.

This kind of restriction has the only purpose of maintaining the state of the system defined by the count-numbers of species alone. If we can represent anything else as a function of n changing the evolution of the network through reaction rates, and justify the connection, then it is not a restriction. The same holds if we can compartmentalize the elements into sub-species that have different values for properties, such as regions representing spatial structure.

Underlying processes that happen, for example, in faster time-scales, or at a smaller spatial scale, could affect the system mainly as changes in reaction rates. We can justify quasi-steady state approximations and time-scale separations using the connection between these processes and the reaction network, then we are able to model different functional forms of reaction rates, dependent on \boldsymbol{n} .

Moreover, we can consider reaction networks as a probabilistic model over physical systems, then we don't suppose well mixing of actual trajectories, but of probabilities of finding the system at a given state. In this way, transition rates are representative of reactions' plausibility, reading as the chances of reactions happening at a given state instead of simply rates. This interpretation is similar to that of a model in statistical mechanics, where systems are designed to have equiprobable microstates instead of actual trajectories.

But, in the end, when modeling real systems, reaction networks are suited to systems composed of "populations", collections of any physical objects that can be considered identical in the scope of the problem. Also systems of physical elements that don't affect each other's trajectories outside of encounters. There can be different kinds of processes affecting the reaction network as external forces (communicating through the count-numbers of species), but the species must be defined with these restrictions.

3.1.5 The Continuous Limit

We have defined the reaction network as a system of interacting species composed of many elements, with a discrete nature, with total counts defining the states of the system through the extensive variable n_i . But the traditional deterministic approach associated with them considers the state variables as continuous intensive concentrations η_i . What is the relation between these two? Concentrations are primarily defined as a quantity per unit of "size" of the system, for example the volume for real biochemical systems. In this case, η_i should be the concentration associated with n_i . We define a general size parameter for the reaction network, Ω , representing a transformation between whole system properties and local system properties that are independent of the system's size (from extensive to intensive properties). In practice, Ω can be "just" a scale resizing, but Ω is theoretically essential to building the bridge between microscopic level dynamics and macroscopic level dynamics. This is the relation between the two levels of state descriptions:

$$\eta_i = \lim_{\Omega, n_i \to \infty} \frac{n_i}{\Omega}.$$
(3.6)

This limit encapsulates a passage from discrete to continuous descriptions. Note that the limit enforces both the system size and n_i to approach infinite to the same degree, otherwise η_i couldn't be defined as it is; n_i being an extensive variable, it grows with the system. So, the deterministic approach is, first of all, building η_i to be a good approximation for n_i/Ω . This enables us to talk about the discrete changes Δn_i in terms of continuous infinitesimal changes $d\eta_i = \Delta n_i/\Omega = \mathcal{O}(1/\Omega)$, because Δn_i is of $\mathcal{O}(1)$ (n_i is a count number). Moreover, $\mathcal{O}(1/\Omega)$ stands as the order of the approximation of the discrete quantity n_i/Ω to the continuous quantity η_i (note that $1/2\Omega$ is the maximum size of the indeterminacy of n_i/Ω by η_i).

3.1.6 Common Systems Modeled as Reaction Networks

Now we give some simple examples of famous biochemical systems. In the examples below and the sections that follow, we pay attention to how reaction networks are 1) a general, abstract, modeling approach encompassing a wide range of physical systems, 2) a unifying framework bridging the gap between theoretical mechanistic modeling (a design level) and mathematical dynamical treatments (a quantitative model), 3) a straightforward generator of deterministic and stochastic analysis and the connection between them, 4) a reliable and easy to understand tool for designing and editing models with many degrees of different complexities, 5) and a readable organization of complex dynamical systems as emergent from simple and clearly defined mechanistic interactions.

These examples are stated in their deterministic kinetic form, as systems of differential equations. In the next section, we'll derive a justification for this connection.

The first example is the Lotka-Volterra system, that describes the dynamics of predators and preys in an ecological environment. The system is commonly written as

$$\frac{d\phi_1}{dt} = \alpha\phi_1 - \beta\phi_1\phi_2$$

$$\frac{d\phi_2}{dt} = k\phi_1\phi_2 - \gamma\phi_2.$$
(3.7)

This can be the result of a reaction network system with three reactions, stated as follows:

$$X_{1} \xrightarrow{\alpha} 2X_{1}$$

$$X_{1} + X_{2} \xrightarrow{\beta} (1 + \delta)X_{2}$$

$$X_{2} \xrightarrow{\gamma} \emptyset, \qquad (3.8)$$

with $k = \beta \delta$. We can easily interpret the reactions as, respectively: the birth of a new prey from a prey, a predator consuming a prey and giving birth to δ new predators, and a predator dying. The relation $k = \beta \delta$ enlightens the connection between β and k that comes from the local interactions. The reaction network has the stoichiometric difference matrix

$$S = \begin{bmatrix} 1 & -1 & 0 \\ 0 & \delta & -1 \end{bmatrix}.$$
(3.9)

From the reaction network, that as we'll see derives the deterministic system above, we can also derive a stochastic version of this system.

Now, it is common to want to add intra-species competition in a Lotka-Volterra model, where the population can't grow indefinitely but is limited by a carrying capacity. It is simple to add that by the use of the reaction network approach. Let's add intra-species competition to the prey population, as the following:

$$X_{1} \xrightarrow{\alpha} 2X_{1}$$

$$2X_{1} \xrightarrow{\epsilon} X_{1}$$

$$X_{1} + X_{2} \xrightarrow{\beta} (1 + \delta)X_{2}$$

$$X_{2} \xrightarrow{\gamma} \emptyset.$$
(3.10)

The second reaction represents intra-species competition, because it says that sometimes the encounter of two preys will result in the death of one of them; they are competing for resources or actually fighting. This additional step, as we'll see in the next section, changes the deterministic system above to

$$\frac{d\phi_1}{dt} = \alpha\phi_1 - \beta\phi_1\phi_2 - \epsilon\phi_1^2$$

$$\frac{d\phi_2}{dt} = k\phi_1\phi_2 - \gamma\phi_2.$$
(3.11)

The carrying capacity is defined as: the concentration of preys that result in the ceasing of net growth due to competition counterbalancing the growth reaction. We have it as $K = \alpha/\epsilon$. Note that a growing population reaching its carrying capacity is mathematically the same as a reversible chemical reaction reaching equilibrium with its reverse reaction. K is here expressed in terms of local parameters.

This approach is in accordance with the usual system of differential equations for this interaction in terms of K. We could add this interaction to the predator population too, but the definition of their carrying capacity in the same way as K would depend on breaking the prey-consumption reaction in two steps, of prey consumption and predator birth (or, if we remove the prey consumption, we could make the two populations symmetrical and end up with a classical competition model). See that the parameter ϵ can have a meaningful interpretation that is often overlooked in usual modeling: an intrinsic rate of intra-species competition. The stoichiometric difference matrix for this new system is

$$S = \begin{bmatrix} 1 & -1 & -1 & 0 \\ 0 & 0 & \delta & -1 \end{bmatrix}.$$
 (3.12)

So note that the reaction network provides a comfortable flow to the modeling approach and also a grounded basis for mechanistic interpretation of the trajectory equations that result from it.

Another example, the SIR compartment model of epidemiology, with S being the susceptible individuals, I being the infected individuals, and R being the recovered or removed individuals. This system can have even actual humans as species. The system is

commonly posed as

$$\frac{ds}{dt} = -\beta si$$

$$\frac{di}{dt} = \beta si - \gamma i$$

$$\frac{dr}{dt} = \gamma i,$$
(3.13)

resulting in $s + i + r = s_0$ if we consider the initial condition $s = s_0$ and i = r = 0. This system is the same as the one derived from the following reaction network with only two reactions

$$S + I \xrightarrow{\beta} 2I$$
$$I \xrightarrow{\gamma} R, \tag{3.14}$$

The reactions are interpreted as: an infected individual infects a susceptible one, and an infected individual gets removed or recovers. We can expand this model to include birth of susceptible individuals and a population death rate that is equal for every compartment:

The last line represents the three death reactions with the same rate. On top of that, we can include an exposed step, allowing for an incubation period with average duration T during which the newly infected individual isn't yet infectious (this means a reaction rate of 1/T). For that, we include the compartment of exposed individuals, E:

Now, the susceptible individual becomes exposed, and the exposed in turn eventually becomes infectious. This variant with the exposed compartment is called the SEIR model.

Now, the deterministic system arising from this is

$$\frac{ds}{dt} = \Delta - \beta si - \mu s$$

$$\frac{de}{tt} = \beta si - \frac{e}{T} - \mu e$$

$$\frac{di}{dt} = \frac{e}{T} - \gamma i - \mu i$$

$$\frac{dr}{dt} = \gamma i - \mu r.$$
(3.17)

These epidemiological models are a good example of the importance of parametric estimation, because the reproduction number R_0 is a valuable metric of how dangerous an infection outbreak actually is, and it is defined in terms of model parameters. For the SEIR model as the above, it is

$$R_0 = \frac{\beta T^{-1}}{(\mu + T^{-1})(\mu + \gamma)}.$$
(3.18)

So, by being able to estimate those parameters using observational data of the model's variables, we are able to estimate the reproduction number. With an estimation procedure over the models from the reaction network, we can do this for either a deterministic or a stochastic epidemic model.

We now through an example of quasi-steady state approximation in the context of molecular dynamics. Consider a substrate molecule S that is consumed to form a product P, with the help of an enzyme E through the reactions

$$S + E \xrightarrow{k_+} C$$

$$C \xrightarrow{k_-} S + E$$

$$C \xrightarrow{k} P + S.$$
(3.19)

The species C is the complex formed by the substrate and the enzyme. The deterministic system associated with this network is

$$\frac{ds}{dt} = k_{-}c - k_{+}se$$

$$\frac{de}{dt} = k_{-}c + kc - k_{+}se$$

$$\frac{dc}{dt} = k_{+}se - k_{-}c - kc$$

$$\frac{dp}{dt} = kc.$$
(3.20)

We observe that the concentration of the substrate-enzyme complex C is approximated constant in the time-scale considered, in a state of quasi-steady state, with the reactions consuming C happening in a faster time-scale and conditioned to C's formation. This can be approximated to a reaction network that doesn't "note" C,

$$S + E \xrightarrow{\kappa_{eff}} P + E,$$
 (3.21)

with E kept with constant concentration e_0 . To calculate k_{eff} , we set $\frac{dc}{dt} = 0$, and note that $e = e_0 - c$. This results in the following equations:

$$\frac{dp}{dt} = k_{eff} e_o s$$

$$\frac{ds}{dt} = -k_{eff} e_0 s,$$
(3.22)

with

$$k_{eff} = \frac{k}{K_M + s},\tag{3.23}$$

where $K_M = \frac{k_-+k}{k_+}$ is the Michaelis-Menten constant. This reduction produces an effective rate that is called the Michaelis-Menten rate of catalyzed product formation. This derivation is compatible with the one from the reduced reaction network. Note that k_{eff} now nonlinearly depends on the concentrations through *s*. The reduced network models underlying mechanisms through their effect over the reaction rate.

3.2 Deterministic System

We follow directly the particular definition of a reaction network from the previous section with the aim to define ways in which we can describe trajectories of the concentrations η_i in time. For this, we assume an infinite system and give a deterministic interpretation to the transition events brought by the reaction sets. We do this by defining the meaning of the transition rates in a particular manner: a reaction having a transition rate W_r will continuously happen through a continuous time passage in a way that the instantaneous amount of reaction events per unit of time is proportional to it. This allows us to define the instantaneous change of concentrations per unit of time in terms of the stoichiometric difference of reactions as

$$\frac{d\eta_i}{dt} = \sum_r S_{ir} W_r(\boldsymbol{\eta}) = \sum_r S_{ir} k_r f_r.$$
(3.24)

This is so regardless of the possible dependence of k_r on the vector of concentrations, η . Implicit in this equation is the assumption of a time scale dt large enough to permit the state of the system to be always defined (in the terms discussed in the last section), but simultaneously a time scale that is small enough to keep the state of the system constant up to $\mathcal{O}(dt)$. Recall that the reaction rate k_r acts as a strength parameter weighting the occurrence of the reaction depending on environmental factors, the affinity factors between reactants, and underlying mechanics affecting the actual occurrence of the reaction, while the event-function $f_r(\eta)$ is dependent only on the reaction structure and the reactant concentrations. The equation above can be read as: the change $d\eta_i$ in concentration of a species during the passage of time dt is equal to the number of reaction events happening during dt, $W_r dt$, times the amount that n_i changes per reaction event, S_{ir} . The number of elements during dt, f_r , times the number of encounters that end up reacting, $k_r dt$.

3.2.1 The Event-function

The problem of directly deriving a deterministic set of equations able to describe the dynamical evolution of the reaction network now reduces to finding the law connecting a given reaction type to its reaction event-function. We'll provide a qualitative reasoning that arrives at the overall reaction occurrence determination that is called law of mass action, which is historically an empirical law that will later have a fully grounded derivation coming as a deterministic limit of a stochastic analysis of the reaction network.

We must expect that $f_r(\boldsymbol{\eta})$ will not depend on species that are not reactant species of r, because their presence is irrelevant to the occurrence of this particular reaction, per construction (but note that k_r might depend on them, if they interfere with environmental or underlying mechanisms also driving the actual occurrence of r). Conversely, $f_r(\boldsymbol{\eta})$ must depend on reactant species' concentrations, because, for example, if a given concentration is zero, the reaction can't possibly occur due to the lack of an "ingredient".

We now argue that $f_r(\eta)$ must be built as separable in terms of functions of different reactant species, and these functions are symmetrical if the species are symmetrical reactants. For example, that a reaction of form $X_1 + X_2 \rightarrow$? will have an event function of the form $f_r(\eta_1, \eta_2) = g(\eta_1)g(\eta_2)$ and $X_1 + 2X_2 \rightarrow$? will have $f_r(\eta_1, \eta_2) = g(\eta_1)h(\eta_2)$. This is because each element of a species is assumed to be independent of the others, so they can't possibly communicate with each other, by construction. Such communication would be viewed as an interaction, thus having its effects considered either as a reaction itself or as a part of the factors bundled inside k_r .

The reasoning above actually leads to a more strict conclusion. We argued for lack of communication between elements of different species, but the whole point also encompasses lack of communication between elements of the same species as well, resulting in a reaction of type $2X_1 \rightarrow$? having an event function of form $f(\eta_1) = g(\eta_1)g(\eta_1) = g^2(\eta_1)$.

Our derivation is now reduced to finding the simplest form of event function, for the reaction of type $X_1 \rightarrow ?$. We argue that it should be the simple form $f_r(\eta_1) = \eta_1$. This comes by once more extending the argument of mutual independence of elements. In this scenario, what would we expect to change if we doubled the concentration η_1 ? Since the elements don't "see" each other, we may visualize this doubled system as two independent superposed systems occupying the same volume, so we would expect to double the number of reactions occurring. This independence in all-terms is conducive to a linear functional form between elements and reaction occurrence, so $f_r(\eta_1) = n_1$. Remember that any multiplicative constant would function as an arbitrary scale already absorbed by the, in principle, unknown reaction rate k_r . Thus, the deterministic derivation is now complete.

Note that we didn't need to invoke any statements about probability, about chances of occurrence, but we needed to invoke the microscopic property of the reaction network, the fact that species are composed of discrete elements interacting in a microscopic, local scale, and that, apart from the reactions themselves, evolve independently of each other. Without this assumption, we could only hope for phenomenological conclusions, since the microscopic details would be inaccessible.

3.2.2 Deterministic Equations

For the general reaction network represented by

$$\sum_{i=1}^{N} s_{ir} X_i \xrightarrow{k_r} \sum_{i=1}^{N} s'_{ir} X_i, \quad r = 1, 2...R,$$

$$(3.25)$$

we arrived at event-functions of the form

$$f_r(\boldsymbol{\eta}) = \prod_i \eta_i^{s_{ir}}.$$
(3.26)

This leads to deterministic differential equations of the form

$$\frac{d\eta_i}{dt} = \sum_r S_{ir} k_r \prod_j \eta_j^{s_{jr}}.$$
(3.27)

The system of equations represented above is what we usually call a mass-action kinetic system, although the term mass-action commonly implies constant reaction rates k_r , that's why we can consider this broader derivation to be a form of generalizing the mass-action system.

The system above can be provided with initial values of concentrations and then integrated in order to output actual trajectories in time for η . But the trajectories themselves are also far from being the only source of information about the system's behavior. Many other tools are also available, such as dynamical stability and state space analyses, for obtaining structural information on systems for which only numerical trajectories are available.

3.3 Stochastic System

The stochastic form of reaction networks will be a Markov jump process, with probability distribution governed by the master equation.¹⁴ The central aim of this section is to provide arguments both for this conclusion and for the connection with the deterministic system described in the last section.

We start by noting that reaction networks, as defined in this chapter, are compatible with all system-defining assumptions of a stochastic system governed by the master equation, that we saw in the last chapter. They are systems evolving through a continuous passage of time assumed to have a well-behaved infinitesimal time-step scale, dt; and are systems defined by the total number of discrete state element counts of species, n_i , with states changing through discrete "jumps" between states (the jumps being the reactions occurring and transforming the state vector by a shift of size equal to the stoichiometric difference of species). We only lack a line of reasoning to justify the system having a probabilistic nature, and being Markovian in addition to that.

3.3.1 Bayesian Probabilistic Assessment

The central difference between the stochastic description and the deterministic is that now we are far from the infinite system limit $n, \Omega \to \infty$. This means that we can no longer define the intensive-level concentration η of species. The stochastic kinetic must be set in the level of discrete element count-numbers n_i . A consequence of this is that the definition of W_r as the instantaneous amount of reactions r per time-unit and size-unit in terms of concentrations now breaks down. We need to reinterpret reaction events as actual discrete events happening between the elements of the system, which means we need to actually count the amount of reactions happening inside the entire system. But given that our model is overlooking the actual individual trajectories of each element and their details (per design of the reaction network), we can't know for sure whether a reaction will or will not happen at a given moment in time. That's where a Bayesian probabilistic description enters.

In this view, the elements will follow their unknown trajectories "at will" (actually, through their own underlying mechanical laws, from molecules to actual people), shuffling themselves as they rapidly achieve spatial homogeneity (as the reaction network framework assumes), and sometimes they'll encounter each other, being together at the same place, so as to enable a reaction to occur (think about, for example, molecules bumping each other or preys and predators crossing their paths). This setup doesn't necessarily mean that a reaction will occur, because environmental/underlying factors surrounding the encounter of reactants must be so as to "ignite" the reaction (think about molecular encounters resulting in a chemical reaction or not, or animal encounters resulting in their interaction or not). So, from a Bayesian standpoint, it is important to note that these encounters arise from the actual motions of elements, which can be determined in nature, and what is undetermined is the model's knowledge of whether actual encounters are happening and then resulting in the reactions. And precisely because we are overlooking the mechanistic details surrounding and behind each encounter, we don't know where there will be molecules or animals encountering each other, or whether the molecules are bumping in the right manner, or whether the animals saw each other and engaged in their reaction-interaction.

In order to characterize transition events, or jumps, as a consequence of reactions occurring in the system, we would like to use the following statement:

 $R_{r,n} = \{ \text{Given a state } n, \text{ a reaction } r \text{ occurs during the next interval of time } dt \}.$

But we don't know for sure whether $R_{r,n}$ will be true, so we'll deal with its probability of being true. We know that the system is in a given state n, contained inside a space of size Ω , and the reactions occur locally at a reaction rate k'_r . We rename here the reaction rate because we don't know whether or not this is the same parameter defined for the deterministic description, since we are on a different context, with these parameters meaning different things, in theory. The reaction rate k'_r , being the local rate of occurrence of a given reaction of type r, marks how often an encounter will result in the actual transition taking place. In this scenario, we can't say that $k'_r dt$ will be the amount of reactions that occur somewhere, for two reasons: 1) a given encounter of reactants will either result in 1 reaction occurring or 0 reactions occurring, and 2) we can't simply count what we don't know will happen. So, it only makes sense to think of k'_r as the probability of an encounter of elements of a reactant complex resulting in a transition.

Now, by the reasoning we gave, we can divide the statement $R_{r,n}$ in encounters happening and transitions happening due to encounters, so we can break it into the two following statements:

 $E_{r,n} = \{ \text{Given a state } n, \text{ an encounter of reactants of } r \text{ is occurring} \}.$

 $T_{r,n} = \{ \text{Given a state } n, \text{ encounters result in reactions during the interval } dt \}.$

We say that $R_{r,n} = E_{r,n}, T_{r,n}$. And by the product rule, we have

$$P(R_{r,n}) = P(E_{r,n})P(T_{r,n}|E_{r,n}).$$
(3.28)

In the way k'_r is defined, we can see that $P(T_{r,n}|E_{r,n}) = k'_r dt$, the probability of a transition occurring given an encounter of reactants (note that we are assuming that encounters are preserved at least during the time dt, it is part of the requisites over dt). Finding $P(R_{r,n})$ is now a matter of finding $P(E_{r,n})$, that will give rise to the stochastic form of the event function.

3.3.2 Transition rates and the Markov Property

The most reasonable assumption we can make about where each element is, given that we don't know anything about their individual trajectories, is to suppose a uniform probability over the entire space Ω , so in a way that this probability is simply proportional to the volume considered. But we also need to formalize what we consider to be an encounter. So, we divide the space into minimal cells of size ω_r defined in such a way that elements inside them are considered to be encountering themselves and therefore are candidates to "close the deal" for reaction r. These cells can be viewed as encounter sites, with the size ω_r being a fundamental parameter of the underlying mechanisms allowing the reactions to happen, representing the volume of a characteristic reaction's effective-range. Although we are allowing for different reaction ranges to different types of reactions, it is part of our assumptions to treat each of them as constant parameters, and also the fact that they are well-behaved in the sense that they're small enough in comparison to the system size Ω . The number of such reaction-site cells inside a system of size Ω is Ω/ω_r .

Then, given the sizes ω_r , we conclude that, for each element composing the system, there's a uniform chance ω_r/Ω of finding it inside any small cell of size ω_r and a chance $(\Omega - \omega_r)/\Omega$ of finding it outside a given cell.

Our probabilistic assessment over the cells is diffuse enough to prevent us from incurring in an under-counting of encounters (from, for example, close-by elements that happen to belong to different neighbor cells), because, in a sense, we can visualize every element as having "a portion" ω_r/Ω of it inside every cell (the math can't distinguish this detail, we are not really creating spatial structure here).

We have now the basis to obtain a reasonable probability of having a confluence of reactants inside a cell of size ω_r , thus constituting encounters as we defined them. The encounter happens when we find a group of reactants for a valid reaction in each cell, taken from the entire population of elements given by \boldsymbol{n} .

Take for example the reaction of type $X_1 + 2X_2 \rightarrow ?$. Every time we have an element of type X_1 and two elements of type X_2 inside a cell, we have a valid encounter for reaction r. Consider a particular cell of size ω_r , then if we count how many groups of one X_1 and two X_2 elements that could be there and consider the chances of a given group being there, we have the chances of an encounter be happening there. The number of possible groups is the combination composed of one X_1 and two X_2 from the populations of n_1 elements of type X_1 and n_2 elements of type X_2 . The probability of a particular group of one X_1 and two X_2 being inside a cell is built from the following statement: an element X_1 is inside the cell, and an element X_2 is inside the cell, and an (other) element X_2 is inside the cell. Based on their individual chances and their independence, we conclude that this probability is ω_r^3/Ω^3 (the joint probability of all three elements considered being inside the cell instead of outside). So, for one cell, using event $E_{r,n}$ applied to a single cell $E_{r,n}^{\omega_r}$, in this particular example we would have

$$P(E_{r,n}^{\omega_r}) = \frac{n_1 n_2 (n_2 - 1)}{2} \frac{\omega_r^3}{\Omega^3}.$$
(3.29)

This reads as: an encounter of this type in this cell happens with probability ω_r^3/Ω^3 , but $P(E_{r,n}^{\omega_r})$ considers all $n_1n_2(n_2-1)/2$ possible independent encounters between these elements.

For the general case of a reaction $\sum_{i=1}^{N} s_{ir} X_i \rightarrow ?$, that probability becomes

$$P(E_{r,n}^{\omega_r}) = \prod_i \frac{n_i!}{(n_i - s_{ir})! s_{ir}!} \frac{\omega_r^{s_{ir}}}{\Omega^{s_{ir}}}.$$
(3.30)

And finally, for the total probability of having an encounter for a reaction of type r in the entire system, we sum all independent probabilities of it happening in a single cell, which

$$P(E_{r,n}) = \frac{\Omega}{\omega_r} \prod_i \frac{n_i!}{(n_i - s_{ir})! s_{ir}!} \frac{\omega_r^{s_{ir}}}{\Omega^{s_{ir}}}.$$
(3.31)

Going back to $P(R_{r,n})$, we end up with the following probability of a jump of kind r happening in the system when it is in state n:

$$P(R_{r,n}) = k_r' dt \frac{\Omega}{\omega_r} \prod_i \frac{n_i!}{(n_i - s_{ir})! s_{ir}!} \frac{\omega_r^{s_{ir}}}{\Omega^{s_{ir}}} = \left(\frac{k_r'}{\omega_r^{1 - \sum_i s_{ir}}} \Omega \prod_i \frac{n_i!}{(n_i - s_{ir})! s_{ir}! \Omega^{s_{ir}}}\right) dt. \quad (3.32)$$

Now, a few commentaries before we proceed:

1) Unless $\sum_i s_{ir} \leq 1$, meaning a reaction of type $\emptyset, X_1 \rightarrow ?$, this probability actually depends on the fundamental reaction parameter ω_r . This makes sense, because this parameter defines how common an encounter should be, and for a reaction involving only one element or no elements, there's not really an encounter to happen. Moreover, this parameter is not a part of the structure of the reaction as an interaction of the network, but a part of the underlying mechanics giving rise to the interaction. So this dependence on ω_r should be absorbed by the reaction rate, making us retroactively redefine what we consider to be the reaction rate to

$$k_r = \frac{k'_r}{\omega_r^{1-\sum_i s_{ir}} \prod_i s_{ir}!}.$$
(3.33)

This redefinition comes in order to maintain consistence in the dependencies of the reaction rate and of what we defined as the reaction event-function, despite muddying the clear stochastic interpretations of before. Note that it means that we should expect reaction rates to scale according to the total number of reactants (it makes sense, because reactions requiring more elements should be less probable to happen). Then, we also divided by $\prod_i s_{ir}!$ because, in the deterministic case, we let multiplicative constants to be taken inside the reaction rate (with this, the event-function actually holds only the functional form of the dependence on the architecture of reactions). This redefinition will soon be even more justified when we see that this new reaction rate actually is the deterministic reaction rate, and the equation above already is a connection between deterministic and stochastic formulations.

2) $P(E_{r,n}^{\omega_r})$ depends on the system size Ω , but we are considering reactions that happen on a local scale, independent of the whole system. This comes only from the fact that we are representing these reactions in terms of the total numbers of elements in the system, that are extensive variables. But then, we can't remove the dependence on Ω by just trying to rewrite the expression in terms of intensive variables n_i/Ω . To justify that, note that $P(E_{r,n}^{\omega_r})$ takes in consideration the progressive use of elements from the whole "bag" of n_i elements, making it truly dependent on the system size. See, for example, that reactions using only up to one reactant of each species, called first order reactions, that don't need the successive drawing of elements of any type, can be made independent of Ω . If we could somehow approximate the counting of elements by a counting "with repetition", considering that the "bag" of elements is large enough to not "get smaller" as we use them, then we would also break the necessity of this Ω dependence; that's precisely what happens when we move to the deterministic description, when we are able to make $\Omega \to \infty$.

3) We implicitly assumed a lack of explicit dependence on the time for the jumps. The reactions are built in a matter that doesn't use time as a relevant component in the reactions, but the reaction rates may depend on the absolute value of time (for example, if they are influenced by timed factors). If that is the case, we should adjust the notation accordingly, but nothing would change in the derivation. One thing to note in advance is that an explicit dependence on time would not break the Markovian property of models if the dependence is on the actual time of the jump, not previous times.

4) In our derivation of $P(E_{r,n})$, we could have considered the term $(1 - \omega_r / \Omega)^{(n_i - s_{ir})}$, meaning that elements outside the group should be outside the cell, making it a binomial distribution. But this is not a matter of dividing all elements through a binomial distribution, it is a matter of putting candidate reactants inside the cell, independently of all other elements, and then counting all sets of candidates. The dependence on other elements' situations would break the local and independent nature of this probability.

Now, the probabilities $P(R_{r,n})$ are the probabilities of jumps in a discrete state space from \boldsymbol{n} to all other accessible states (that are made accessible by the set of reactions). This defines transition rates for each reaction as $P(R_{r,n}) = W_{n,n'}^r dt$, with the new state $\boldsymbol{n'} = \boldsymbol{n} + S_r^T$ and S_r^T being the r-th column of the transposed stoichiometric difference matrix, giving the update on \boldsymbol{n} by the jump of type r. These transition rates are given by

$$W_{n,n+S_r}^r = k_r \Omega \prod_i \frac{n_i!}{(n_i - s_{ir})! \Omega^{s_{ir}}}.$$
(3.34)

These stochastic transition rates are also called propensity functions in the biochemistry literature. We see that the transition rates depend only on the state before the jump, always (unless k_r would be made to depend on previous states, something that is possible in principle but would break the Markovian property). That's actually expected, since reactions are defined to be determined by the present number of elements.

This concludes that the system, defined in this way, behaves as a Markovian jump process, and therefore has its probabilistic evolution governed by the master equation derived in the last chapter:

$$\frac{d\Pi(\boldsymbol{n},t)}{dt} = \sum_{r} \left(W_{\boldsymbol{n}-S_{r}^{T},\boldsymbol{n}}^{r} \Pi(\boldsymbol{n}-S_{r}^{T},t) - W_{\boldsymbol{n},\boldsymbol{n}+S_{r}^{T}}^{r} \Pi(\boldsymbol{n},t) \right).$$
(3.35)

Let's pay attention to the changes from the last chapter, from eq. (2.27). Now, the transitions are restricted to the ones allowed by the set of reactions, so the sum now is over

the reactions. The reactions shift the state space in clearly defined quantities, S_r^T . The first term of the sum accounts for the jumps into the present state from the state that has the present state as a possible destination through reaction r, and the second term accounts for the jumps from the present state to the state brought by reaction r. This equation can be rewritten as a function of only the present state if we use the step operator defined before, in equation 3.3 (the operator that shifts states by the stoichiometric differences applied to individual entries). So we have

$$\frac{d\Pi(\boldsymbol{n},t)}{dt} = \sum_{r=1}^{R} \left(\prod_{i=1}^{N} E_i^{-S_{ir}} - 1\right) W_r(\boldsymbol{n}) \Pi(\boldsymbol{n},t),$$
(3.36)

with the transition rates just renamed for convenience, because the reaction label r is enough to denote the transition. Note that each E_i acts only over the *i*-th component of \boldsymbol{n} , shifting it by S_{ir} .

The same remarks made to the system of deterministic differential equations now apply to the master equation. It can be solved, given an initial condition over \boldsymbol{n} , to determine the probability density of states $\Pi(\boldsymbol{n},t)$ for all times; but it can also have its properties analyzed in order to describe the noisy behavior of stochastic reaction networks, with the study of its steady-state solutions, moments, and other analytically useful metrics.

3.3.3 Connection to the Deterministic System

The connection with the deterministic system is made when we equate the stochastic transition rates per unit of size $W_r(\boldsymbol{n})/\Omega$ with the deterministic version of $k_r f_r(\boldsymbol{\eta})$ by performing the limit $\Omega, \mathbf{n} \to \infty$. Recall that the deterministic $k_r f_r(\mathbf{\eta})$ is the number of reactions occurring per unit of time and size, that's why we consider the stochastic transition rates divided by the system size. We must note that, by performing the limit, we are moving from a state of uncertainty to an approximated state of certainty, and that's because the limit washes out our uncertainty along with the approximation. How so? Suppose that the range of uncertainty we associate with n_i is δn_i ; this value, the uncertainty of our belief about n_i , is formalized in probability terms as the standard deviation of the probability associated with n_i . This uncertainty would be translated to η_i as $\delta \eta_i = \delta n_i / \Omega$. Uncertainty being "washed out" is formally put as $\delta \eta_i \to 0$ when $\Omega, n_i \to \infty$. Because the uncertainty about n_i is the sum of independent uncertainties associated with every element of species i, the central limit theorem says that, as the limit goes, the probability associated with the number n_i approximates a Gaussian and $\delta n_i \to \mathcal{O}(\sqrt{n_i})$. Since we must have n_i go with the same order as $\Omega, \, \delta \eta_i \to 0$, as we want, with error of order $\mathcal{O}(1/\sqrt{\Omega})$.

Note that the limit holds a valuable meaning: it is able to transform the probability of a reaction occurring at the microscopic scale into an approximate amount of reactions occurring at the macroscopic scale, without uncertainties associated. Let's illustrate these kinds of transformations with an example: suppose a population of individuals that can be in one of two states A and B, with any individual having probability p of being in state A (meaning that the statement "an individual is in A" is uncertain with probability p). Now, if we take a number Ω of individuals, we may ask, for example: how many of these individuals are in state A at a given moment? We can't know that for sure, because the state of individuals is uncertain. We can argue that, because the total number of individuals in state A is the sum of independent individuals in their states, this value should approach $p\Omega$ as Ω gets larger, but the statement "there are $p\Omega$ individuals in state A among Ω individuals" has a probability associated with it. The point is that, as we make $\Omega \to \infty$, the uncertainty (variance) associated with that last statement goes to zero. If you see that $p\Omega$ is the mean of individuals in state A (and the mean being merely an estimator of this value, an "artificial" expected value for it), you don't need to invoke the central limit theorem to justify the loss of uncertainty, just the law of large numbers (that says that the uncertainty over the mean goes to zero, but doesn't say how it goes). The fact is that we couldn't say that η_i is the mean of n_i/Ω before we perform the limit. What is funny is that, by showing that η_i approximates n_i/Ω with certainty in the limit, we also show that η_i is the mean, at least in the limit. Also, by invoking the central limit theorem we were able to see that the uncertainty over η_i goes as $\mathcal{O}(1/\sqrt{\Omega})$.

So, performing the limit, we reach the connection:

$$\lim_{\Omega, n_i \to \infty} \frac{W_r}{k_r \Omega} = \lim_{\Omega, n_i \to \infty} \prod_i \frac{n_i!}{(n_i - s_{ir})! \Omega^{s_{ir}}} = \lim_{\Omega \to \infty} \prod_i \frac{n_i}{\Omega} \frac{(n_i - 1)}{\Omega} \dots \frac{(n_i - s_{ir} + 1)}{\Omega} = \lim_{\Omega, n_i \to \infty} \prod_i \frac{\eta_i \Omega}{\Omega} \frac{(\eta_i \Omega - 1)}{\Omega} \dots \frac{(\eta_i \Omega - s_{ir} + 1)}{\Omega} = \prod_i \eta_i^{s_{ir}}.$$
(3.37)

The last term is the deterministic event-function $f_r(\boldsymbol{\eta})$. Note how the deterministic derivation is a large size approximation to the stochastic derivation, and all the treatment developed so far in this chapter is applicable to every system modeled as a reaction network.

3.3.4 Simulation of the Limit

To fully appreciate this connection, we consider a simple example where we show a stochastic trajectory, simulated from the stochastic simulation algorithm (SSA), increasingly approaching the deterministic trajectory of the same system as the system size increases. For that, we'll use the simple predator-prey system shown before

$$X_{1} \xrightarrow{\alpha} 2X_{1}$$

$$X_{1} + X_{2} \xrightarrow{\beta} (1 + \delta)X_{2}$$

$$X_{2} \xrightarrow{\gamma} \emptyset.$$
(3.38)

This system exhibits well-known oscillatory trajectories with differing phases (preys go up, then predators go up, then preys go down, then predators go down). The deterministic differential equations is given in equation 3.7, and the full-length master equation is, with implicit dependence on time

$$\frac{d\Pi(n_1, n_2)}{dt} = \alpha \Big((n_1 - 1)\Pi(n_1 - 1, n_2) - n_1\Pi(n_1, n_2) \Big) + \frac{\beta}{\Omega} \Big((n_1 + 1)(n_2 - \delta)\Pi(n_1 + 1, n_2 - \delta) - n_1n_2\Pi(n_1, n_2) \Big) + \gamma \Big((n_2 + 1)\Pi(n_1, n_2 + 1) - n_2\Pi(n_1, n_2) \Big).$$
(3.39)

Each line shows the contribution from one reaction, with the first term being the jump into the state $\mathbf{n} = (n_1, n_2)^T$, and the second term being the jump out of that state.

In figure 3, we show a sample of the stochastic system together with the trajectories of the correspondent deterministic system, for a system size of $\Omega = 1$. We see the contrast between both dynamics with such a low system size. Figure 4 shows the connection between them as the system gets bigger, for the trajectory of preys. We maintain the same initial deterministic concentration and other parameter values through all four scenarios. Note how, at size $\Omega = 100$, both dynamics already are hardly distinguishable.



Figure 3 – Predator-Prey reaction network. Left: a sample of the master equation from the SSA. Right: a numeric solution of the deterministic equation. The system size is $\Omega = 1$, so both scales coincide. Initial values are $n_1 = \eta_1 = 40$ and $n_2 = \eta_2 = 30$. With arbitrary time-scale, we have $\alpha = \gamma = 2$, $\beta = 0.1$, and $\delta = 1$. Source: By the author.

One final aspect to note that is related to this connection is the evolution of the mean number of elements of a species, also given by the master equation. The mean is



Figure 4 – Number of preys compared for the four system sizes $\Omega = 1, 5, 10, 100$. The deterministic solution $(\Omega \eta_1)$ is shown in black and the stochastic samples from the SSA (n_1) are shown in blue. Initial values of concentrations are kept at $\eta_1 = 40$ and $\eta_2 = 30$, and the initial number of individuals are scaled accordingly, $n_1 = 40\Omega$ and $n_2 = 30\Omega$. With arbitrary time-scale, we have $\alpha = \gamma = 2, \beta = 0.1$, and $\delta = 1$. Source: By the author.

defined as

$$\langle n_i \rangle(t) = \sum_{\boldsymbol{n}} n_i \Pi(\boldsymbol{n}, t).$$
 (3.40)

where the sum extends over all possible combination of states n. In the same way, the mean extends to any function of n,

$$\langle g(\boldsymbol{n}) \rangle = \sum_{\boldsymbol{n}} g(\boldsymbol{n}) \Pi(\boldsymbol{n}, t).$$
 (3.41)

To find the equation for the mean, we produce this definition over the master equation. We multiply the whole equation by n_i and sum over all possible states:

$$\frac{d\sum_{\boldsymbol{n}} n_{i}\Pi(\boldsymbol{n})}{dt} = \sum_{r=1}^{R} \left(\prod_{i=1}^{N} E_{i}^{-S_{ir}} \sum_{\boldsymbol{n}} (n_{i} + S_{ir}) W_{r}(\boldsymbol{n}) \Pi(\boldsymbol{n}) - \sum_{\boldsymbol{n}} n_{i} W_{r}(\boldsymbol{n}) \Pi(\boldsymbol{n}) \right),$$
$$\frac{d\langle n_{i} \rangle}{dt} = \sum_{r=1}^{R} \left(\prod_{i=1}^{N} E_{i}^{-S_{ir}} \langle (n_{i} + S_{ir}) W_{r}(\boldsymbol{n}) \rangle - \langle n_{i} W_{r}(\boldsymbol{n}) \rangle \right).$$
(3.42)

note how n_i transforms into $n_i + S_{ir}$ when passing inside the step operation that discounts the shift by S_{ir} . The step operator also doesn't shift anything in means, because they don't depend on the system state \boldsymbol{n} , so it just vanishes. Then the n_i portion of the first term cancels out with the second term, and we end up with

$$\frac{d\langle n_i \rangle}{dt} = \sum_{r=1}^R S_{ir} \langle W_r(\boldsymbol{n}) \rangle.$$
(3.43)

This equation has the same form as the deterministic system from equation 5.3, and with it we can see how $\langle n_i \rangle / \Omega$ is equal to η_i in the limit from the point of view of the master equation. We also see that, if all reactions are of type $\emptyset \rightarrow$? and $X_1 \rightarrow$? (both zero and first order reactions, the ones that give linear transition rates), the equation for the average is exactly the same as the deterministic equation for η_i . We could be tempted to say that, at least in these cases, they are the same even without the limit; even for small systems. But this is wrong! The equality $\eta_i = \langle n_i \rangle / \Omega$ never holds without the limit, it doesn't even make sense. At the level of η_i , we don't have a finite value of Ω (it doesn't exist as a parameter). And if we define η without the limit, we are approximating with error of order $1/\Omega$ (an error we don't even have access to if we don't have access to n_i or Ω). But it is remarkable that the equations of motion have the same form for linear systems, even for small systems, and are equal up to a constant factor. It means that being small doesn't alter the shape of the average dynamics of linear systems, and the same doesn't occur for nonlinear systems.

But, how can we really interpret mean motion? It is the motion we would expect uncertainties to be placed around, an unbiased estimator, while η_i is what we would expect to see in very large systems. But, then, for large systems, η_i also is where we would place uncertainties around (the two kinds of motion converge as the system gets ever larger and uncertainties shrink).

3.3.5 Final Remarks

In conclusion, we present a summarizing discussion about the fundamental differences between the deterministic and stochastic kinetics presented in this chapter:

1) The state variables of the stochastic system, n_i , have three properties that we explored: they are discrete, exact, and uncertain. Meanwhile, the state variables of the deterministic system, η_i , occupy opposite sides in these properties: they are continuous, approximate, and certain. And there is a limit that connects these two worlds. Let's go through these differences once more.

2) Discrete/Continuous: the bridge between discrete and continuous treatments is made possible by the limit. But the limit is a theoretical construct, elements are never infinite in reality. This presupposes that the discrete description is more fundamental than the continuous one. But we are not saying that reality itself has a discrete nature, we are saying that reaction networks have. So, in light of a reaction network model, systems are fundamentally discrete, and whenever we are describing them with continuous variables, we are approximating them. With that said, in the limit, a continuous variable approximates a discrete variable exactly. So when we say that the continuous treatment is an approximation, we say it in relation to the reaction network, assuming finite numbers.

3) Exact/Approximate: So in the limit the deterministic approach is exact. Whenever we say that it is an approximate quantity, we say that in the sense described above. And this is useful to say, because even in theory we may want to consider the limit up to a certain point, and talk about the errors we are allowing ourselves to commit. In this particular case, the approximation error brought by the deterministic description is of $\mathcal{O}(1/\Omega)$, so it is clear that η_i is exact in the actual limit, because $1/\Omega \to 0$.

4) Certain/Uncertain: the whole point of a Bayesian treatment is to quantify our uncertainties about statements, and we do this by associating probabilities to them. So, in this framework, there's a formal process by which uncertainty could become certain, and it involves changing the system in a way that makes our measures of uncertainty go to zero. Regardless of how we actually do this, the most important is to see how it can be done in principle. With that said, note the key roles the law of large numbers and the central limit theorem play in this process. They are fundamentally about the path through which our observations can become certain, and are a connection between processes in individual scales and consequences in population scales. With this, we saw that the uncertainties over η go to zero as $\mathcal{O}(1/\sqrt{\Omega})$. Now, an interesting discussion comes from the fact that, by applying a limiting process over our own assessment of chances, we end up with an equivalent proportion on the system. This can act as a gauge of our uncertainty; think of this example: suppose you toss a coin that you know nothing about, so you state the possible results with probability 1/2. If you toss it "infinitely" and observe that the proportion of heads is 1/4, then that acts as global information demanding you to revise your local assessment of the coin. But then, suppose that you have absolute knowledge of the physical states of the coin, so there are no uncertainties. In this case, does the proportion of heads mean something to you? Yes, it shows you how uncertainties should be assigned by those who don't know the details you know, but know at least this long term behavior. But if you never tossed the coin, it still makes sense to say that the chances of a toss result in heads is 1/2, for you. This example talks about three levels of knowledge. The statement "the next toss will give heads" is evaluated as having 1/2 by those who never saw the coin, or 1/4 by those who saw the long-term proportion it gives, or either 0 or 1 for those who know the details about how the coin will land. No one is wrong, or more wrong, than the others. They are all equally right from their perspectives. So, if we assign chances to a system's local behavior and then we see that its "large frequency" behavior is different from the expected, we are missing important structures happening at the local level that are giving rise to this difference. In this sense, the connection to the deterministic description assures us that our probabilistic assumptions are in the right direction (of course, if the system of interest behaves as our deterministic description).

5) A last remark, about the mean and low values of n_i . Small systems are prone to historical accidents. Take for example the system of preys and predators above. If for some underlying reason we reach $n_1 = 0$, preys will be gone forever, and shortly after that the predators would too. This is never certain to occur, but always possible. The deterministic η could never capture these accidents, while the average $\langle n_i \rangle$ does, in a sense (it takes into account the chances of it occurring and has its value influenced by them). Small size unbalancing the chances of reactions occurring is a fundamental aspect causing divergence in the shapes of their dynamics for nonlinear systems. Suppose, for example, that a reaction needs two elements of species X_1 to occur, but $n_1 = 1$; then that reaction is not allowed to occur. This scenario would be felt by the mean motion, but never by the deterministic motion (that is presupposing an infinite amount of n_1).

4 ESTIMATION AND STOCHASTIC APPROXIMATION ON REACTION NET-WORKS

This chapter develops an introduction to the main methods that we use to analyze the stochastic aspect of reaction networks, and these are actually methods related to Markov jump processes, which are the mathematical translation of the stochastic kinetics of networks.

We introduce the stochastic simulation algorithm for the exact simulation of the master equation, that is based on waiting times between jumps. Then, the system size expansion, that is the only systematic approximation of the master equation, which means it is controlled by a small parameter, the size or volume of the system. This approximation has a widely used first order called linear noise approximation, and it is not only important to the mathematical analysis of noise but also to the theoretical understanding of the connection between probabilistic and deterministic levels through the law of large numbers.

Then we specify the parameter estimation process to statistical models that use the deterministic level of reaction networks, and an introduction to the central Bayesian estimation algorithm of Markov chain Monte Carlo, building each related concept up to the full understanding of the algorithm's rationale.

The chapter ends by introducing the STAN Bayesian statistical language and working out a simple estimation example on a Lotka-Volterra model defined through a reaction network.

4.1 Stochastic Simulation Algorithm

Since solutions to the master equation of Markov jump processes is rarely obtainable, especially for the case of nonlinear propensities, the exact simulation of the stochastic dynamics is extremely useful. The stochastic simulation algorithm (SSA), or Gillespie's algorithm, is a method for exactly sampling trajectories of species governed by the master equation, and it is widely used across many fields, wherever the master equation is present.¹⁵ Equally important is the easiness with which we can understand and use the algorithm. In this work, we make extensive use of the SSA for visualizing the stochastic level of reaction networks and also to generate simulated data for parameter estimation processes.

The SSA is centered around the fact that the waiting time between any jump in the stochastic dynamics follows an exponential distribution weighted by the sum of propensities. The more total propensity the model has (the more reactions in a network), more are the chances for a reaction event to occur at any given time. The derivation of waiting times follows from considering the occurrence of jumps as a counting process, and noting the fact that, in between jumps, all propensities remain constant, because they are only state-dependent (so the SSA is restricted to propensities not explicitly dependent on time). This results in a Poissonian distributed model for jumps, which has exponential waiting times. But note that the Poissonian dynamics is always breaking after a reaction occurs, because then the propensities change; so, after every jump, the waiting time exponential distribution must be updated. Thus, we can write the following:

$$P(A) = \lambda e^{-\lambda\tau},\tag{4.1}$$

for the proposition $A = \{\text{The next reaction occurs after a time } \tau\}$, where λ is the sum of all propensities or transition rates, $\lambda = \sum_{r} W_r(\mathbf{n})$. Then, if we generate a uniform unit random number u_1 , the time for the next reaction will be

$$\tau = -\ln(u_1)/\lambda. \tag{4.2}$$

After that, we determine which reaction will be the result of that jump, considering that each reaction has a probability of being chosen that is proportional to its propensity (its transition rate). So the propositions $B_j = \{$ The next reaction to happen is reaction $j\}$ have probability $P(B_j) = W_j / \sum_r W_r$. By generating another uniform unit random number u_j , we choose the reaction with label equal to the minimum j that satisfies the condition

$$\sum_{r=1}^{r=j} W_r > u_2 \sum_{r=1}^{r=R} W_r.$$
(4.3)

In that way, by starting from any allowed initial state, we can simulate samples up to any given time by updating the model and iterating this procedure for each jump.

This direct stochastic simulation is computationally demanding, due to the need of iterating over every single jump. Many approximations were designed upon the direct and exact SSA, aiming at easier computation. The first and most famous of them is called tau-leaping SSA, which uses a longer and chosen iteration time and approximates the simulated number of jumps occurring in that interval (the model updates with tau-sized leaps composed of many jumps).¹⁶

4.2 System Size Expansion

While the stochastic simulation yields exact samples of the dynamics (in a probabilistic sense), approximations to the master equation are inexact methods to extract structural information from the probability trajectories. We find that the most physically natural, and thus most natural from the modeling perspective, is the system size expansion (SSE).^{13,14} The SSE is a systematic approximation of the master equation. It provides a measure of the approximation error through an expansion parameter dependent on the system size (Ω), and also provides successive orders of approximation, meaning consideration of ever larger systems. This expansion thus acts as a smooth connection between the deterministic and stochastic pictures of the same dynamics represented by the reaction networks, and is the generalization of the limit we performed in the last chapter to achieve this connection.

The expansion is based on what we call the van Kampen's ansatz. It breaks down the uncertainty into a structured noise around the infinite system trajectory (the deterministic limit). The aim is to be able to rely on the limit as the highest degree of approximation and analyze uncertainty as the noise levels around the limit, that would be viewed as the expected trajectory in the absence of noise. The main goal of the ansatz is then to workout how exactly noise scales with the system size. Drawing from the same arguments as in the last chapter: from the central limit theorem, the variance around the mean scales with size as Ω^{-1} . The noise, measured as the deviation around the mean, should then scale as $\Omega^{-1/2}$. With all considerations, we present the ansatz as

$$\frac{n_i}{\Omega} = \eta_i + \Omega^{-1/2} \xi_i, \tag{4.4}$$

where ξ_i are the noise variables, now considered as the focal variables of the model. So, it acts as a transformation from the actual count-numbers n_i to the noise ξ_i , now with the goal to describe the evolution of $\boldsymbol{\xi}$ instead. The ansatz, as expected, implies the order of approximation of the system to the deterministic limit, given its actual size: $\mathcal{O}(\Omega^{-1/2})$. In this context, the limit $\boldsymbol{\eta}$ is viewed as part of the model specification for $\boldsymbol{\eta}$, and can be provided simply as a numerical solution.

With the variable change given by the ansatz, we are able to expand the whole description around $\boldsymbol{\eta}$ in successive orders of $\Omega^{-1/2}$. Given an arbitrary function of \boldsymbol{n}/Ω , $g(\boldsymbol{n}/\Omega)$, we do

$$g(\boldsymbol{n}/\Omega) = g(\boldsymbol{\eta} + \Omega^{-1/2}\boldsymbol{\xi}) = g(\boldsymbol{\eta}) + \Omega^{-1/2} \sum_{j} \xi_{j} \frac{\partial g(\boldsymbol{\eta})}{\partial \eta_{j}} + \mathcal{O}(\Omega^{-1}).$$
(4.5)

Then we define the probability distribution of $\boldsymbol{\xi}$, $\pi(\boldsymbol{\xi})$, in the same way as we defined $\Pi(\boldsymbol{n})$ in the last chapter. The transformation between pictures gives

$$\frac{\partial \Pi}{\partial t} = \frac{\partial \pi}{\partial t} + \sum_{j} \frac{\partial \pi}{\partial \xi_{j}} \frac{\partial \xi_{j}}{\partial t} = \frac{\partial \pi}{\partial t} - \Omega^{1/2} \sum_{j} \frac{\partial \pi}{\partial \xi_{j}} \frac{d\eta_{j}(t)}{dt}.$$
(4.6)

Over $\boldsymbol{\xi}$, the step operator acts as

$$E_i^{S_{ir}}g(\xi_1,\xi_2...,\xi_i...,\xi_N) = g(\xi_1,\xi_2...,\xi_i + \Omega^{-1/2}S_{ir}...,\xi_N),$$
(4.7)

with $S_{ir} = s'_{ir} - s_{ir}$ being the stoichiometric difference. Finally, we have to provide our transition rates in terms of an expansion on n/Ω :

$$W_r(\mathbf{n}) = \Omega \sum_{l=0}^{\infty} \Omega^{-l} W_r^{(l)}(\mathbf{n}/\Omega).$$
(4.8)

For reaction networks, all W_r will be well-behaved and this expansion exists (all the $W_r^{(l)}$ exist). Then, the SSE is executed by applying all transformations above to the master

equation

$$\frac{\partial \Pi(\boldsymbol{n},t)}{\partial t} = \sum_{r=1}^{R} \left(\prod_{i=1}^{N} E_i^{-S_{ir}} - 1 \right) W_r(\boldsymbol{n}) \Pi(\boldsymbol{n},t).$$
(4.9)

After this, we can collect all different orders of approximation by matching the exponents of Ω . The zeroth order, of magnitude $\Omega^{1/2}$, must vanish in order to configure $\Omega^{-1/2}$ as a small parameter and justify the expansion (acting like a consistency check for the ansatz). The condition for that is

$$\frac{d\eta_i(t)}{dt} = \sum_r S_{ir} W_r^{(0)}(\boldsymbol{\eta}).$$
(4.10)

This is precisely the deterministic system derived in the last chapter, with $W_r^{(0)}(\boldsymbol{\eta}) = \prod_j \eta_j^{s_{jr}}$. The next order of approximation, for Ω^0 , represents the first order of the noise $\boldsymbol{\xi}$. It is called linear noise approximation (LNA), and it is linear because the first order of noise is a Gaussian distribution with time-evolving parameters dependent on $\boldsymbol{\eta}$.

4.2.1 Linear Noise Approximation

The LNA is then given by the first order of the expansion on the noise, the following equation on $\pi(\boldsymbol{\xi})$:

$$\frac{\partial \pi(\boldsymbol{\xi})}{\partial t} = \sum_{r} \left(-\sum_{i,j} S_{ir} \xi_j \frac{\partial W_r^{(0)}(\boldsymbol{\eta})}{\partial \eta_j} \frac{\partial}{\partial \xi_i} + \frac{1}{2} \sum_{i,j} S_{ir} S_{jr} W_r^{(0)}(\boldsymbol{\eta}) \frac{\partial}{\partial \xi_i} \frac{\partial}{\partial \xi_j} \right) \pi(\boldsymbol{\xi}).$$
(4.11)

This is a fokker-plank equation on $\boldsymbol{\xi}$, with Gaussian solutions for every instant of time. Thus, for solving this equation, it suffices to find solutions for the first two moments of π , determining the mean and variance of $\boldsymbol{\xi}$. For the equation of the mean, we multiply Eq. (4.11) by $\int_{-\infty}^{+\infty} d\xi_k \xi_k$ to obtain

$$\frac{d\langle\xi_k\rangle}{dt} = \sum_r S_{kr} \frac{\partial W_r^{(0)}(\boldsymbol{\eta})}{\partial \eta_k} \langle\xi_k\rangle.$$
(4.12)

Note that, if we have deterministic initial conditions $(\xi_k(0) = 0), \langle \xi_k \rangle = 0$ for all subsequent times. By proceeding in the same way, and noticing that the evolution is the same for both the second moment and the variance, we have

$$\frac{d\langle\xi_k\xi_l\rangle}{dt} = \sum_r \left(\sum_j \frac{\partial W_r^{(0)}(\boldsymbol{\eta})}{\partial \eta_j} \left[S_{kr}\langle\xi_k\xi_j\rangle + S_{lr}\langle\xi_l\xi_j\rangle\right] + \frac{1}{2}S_{kr}S_{lr}W_r^{(0)}(\boldsymbol{\eta})\right).$$
(4.13)

By solving these equations, we then obtain the time-evolution of the approximate Gaussian noise profiles around the deterministic trajectory η . This solution becomes more precise the larger the system is.

Now, we will build a more operational form of the LNA through matrix notation. We suppose deterministic initial conditions and consider $\langle \xi_k \rangle = 0$. The variance matrix is a NxN matrix defined as $\Sigma_{ij} = \langle \xi_i \xi_j \rangle$. The stoichiometric matrix is NxR matrix defined as $S_{ir} = s'_{ir} - s_{ir}$. Characterizing the input from the deterministic system, we have the Jacobian matrix of the ODE system for η , a NXN matrix defined as $J_{ij} = \sum_r S_{ir} \partial W_r^{(0)} / \partial \eta_j$. Lastly, the diagonal propensity matrix is a RxR matrix defined as $diag(W_r^{(0)})$. Then, we can write Eq. (4.13) in its matrix form:

$$\frac{d\Sigma}{dt} = J\Sigma + \Sigma J^T + D, \qquad (4.14)$$

with $D = S[diag(W_r^{(0)})]S^T$, called diffusion matrix. This equation is a Lyapunov matrix equation over Σ . Solving the deterministic system and the Lyapunov equation then gives the solution to the LNA:

$$\pi(\boldsymbol{\xi}) = \mathcal{N}(\boldsymbol{\xi}|0, \Sigma(\boldsymbol{\eta})). \tag{4.15}$$

As an illustration, consider the Lotka-Volterra system, with N = 2 species and R = 3 reactions:

$$\begin{array}{c} X \xrightarrow{\omega} 2X \\ X + Y \xrightarrow{\gamma} (1 + \delta)Y \\ Y \xrightarrow{\mu} \emptyset. \end{array} \tag{4.16}$$

Denoting $(\eta_X, \eta_Y) = (x, y)$, with the reactions and variables in the same order, the stoichiometric matrix is

$$S = \begin{bmatrix} 1 & -1 & 0 \\ 0 & \delta & -1 \end{bmatrix}.$$

$$(4.17)$$

The Jacobian of the deterministic system is

$$J = \begin{bmatrix} -\gamma y + \omega & -\gamma x\\ \delta \gamma y & \delta \gamma x - \mu \end{bmatrix}.$$
 (4.18)

And the diagonal of the propensities is

$$diag(W_r^{(0)}) = \begin{bmatrix} \omega x & 0 & 0 \\ 0 & \gamma xy & 0 \\ 0 & 0 & \mu y \end{bmatrix}.$$
 (4.19)

This is all it takes to determine the LNA solution.

With the LNA being a Gaussian distribution, we see that, by using it, we lose all structural noise information that can't be captured by a Gaussian. If such information is needed, one must consider the inclusion of higher orders of approximation. But, with the SSE, we see that the magnitude of these effects scales at least as $\mathcal{O}(\Omega^{-1})$.

4.2.2 Relation to Other Approximations

There are other ways to approximate the master equation, but they are not systematic.¹⁷ The Kramers-Moyal expansion is a method that directly applies a Taylor expansion to the equation. The Pawula theorem states that, in order for the approximated

solution to be a proper probability distribution, only the first two expansion terms can be considered. Also, in order to apply the expansion, the count-number variable n is transformed into a continuous variable. So the expansion really becomes just a coarser continuous picture that follows the same propensities as the master equation. The resulting equation for a continuous state variable is a usual Fokker-Planck equation (different from the LNA solution), that has the two expansion terms identified as a drift and a diffusion contributions. This equation, in turn, holds an equivalency to the stochastic differential equation picture represented by an associated Langevin equation. The Fokker-Planck also connects to a path integral formulation. Interestingly, we can instead apply the SSE to this continuous approximation of the master equation and also obtain the LNA, with the drift contribution giving rise to the deterministic trajectory and the diffusion contribution giving rise to the noise. That is because the continuous approximation already is implicit in the SSE too. This expansion is also called a diffusion approximation, and it doesn't regard the system size, so it doesn't connect with the deterministic limit and doesn't measure approximation errors.

Another approximation to the master equation is the method of moment closure.¹⁸ It is a practical, heuristic procedure that truncates the dynamics on n by making its higher order moments equal to zero. Doing so, we are able to close and then solve the EDO system on lesser moments. This is needed because, for the full system, equations like Eq. (4.12) and Eq. (4.13) (but for n instead) generally depend on higher moments, forming an open EDO system. Moment closures can in principle be made more precise by the more rigorous adoption of criteria for truncating moments. Despite the $\boldsymbol{\xi}$ being Gaussian distributed in the LNA, and thus having the first two moments to be nonzero, the LNA doesn't impose restrictions on any moments of n.

4.3 Parameter Estimation on Reaction Networks

Now turning to the problem of parameter estimation, we first note that models of reaction networks have a straightforward statistical implementation, in theory. By somehow obtaining n (for a stochastic model) or η (for a deterministic approximated model), we can provide simple numerical solutions as a model for the data in order to feed the likelihood. The parameters we want to estimate in this case are the reaction rates and also unknown initial states or measurement errors.

The real problem is that we almost never can afford to solve the master equation for reaction networks, not even numerically. Out of the box, a numerical solution for parameter estimation would consist of simulating a large amount of samples for every relevant set of parameter values, which is a prohibitive computational effort. Then, the approximations to the master equation are, in principle, especially useful for this task. More standard stochastic approximation, such as the Langevin equation, unfortunately lead to hard obstacles in terms of matching the evolution of the model with the observed data, which demands complex fixes that are usually called observational bridge constructs. The SSE, on the other hand, can be readily used in the likelihood function, and this is a great feature of the expansion. The reason is that the SSE provides a full model for the evolution of the state distribution between observations. But even the estimation with LNA likelihood can be extremely demanding on computations.

In this work in particular, we will explain and use the process of parameter estimation on reaction networks with the approximated model of the deterministic limit. Although it is a more simplistic approximation that loses all information on the structure of noise around the mean of the model (which is useful for parameter estimation), we can draw on the benefits of it being so easy to implement and also fairly fast. The deterministic model then shows the least the estimation procedure can do in this context, and we will see that it is surprisingly efficient.

Suppose we have a numerical solution for η . Since the model is deterministic, we have the simple form of a likelihood where the model is its mean, and we can choose the most appropriate form of distribution for the observation noise. Since the LNA gives a structured Gaussian distribution of noise, dependent on η , we will choose a Gaussian with constant and diagonal covariance matrix, $\Sigma = diag(\sigma_i^2)$ as an approximation to the form of the likelihood; that can be seen as component observations with independent errors with constant variance. Then, the measurement model we will use is

$$\boldsymbol{\eta} = \boldsymbol{x} + \boldsymbol{e},\tag{4.20}$$

where \boldsymbol{x} is the data and we are implicitly considering only the measured components. The error is then $e \sim \mathcal{N}(0, \Sigma)$. Setting the transition rates and other defining constants as the parameter vector $\boldsymbol{\theta}$, with dimension equal to the number of parameters to estimate, the likelihood becomes

$$P(X|\Theta) = \prod_{k} \mathcal{N}(\boldsymbol{x}_{k}|\boldsymbol{\eta}_{k}(\boldsymbol{\theta}, t_{k}), \boldsymbol{\Sigma}), \qquad (4.21)$$

with k = 1, 2..., K representing the data points, consisting of the pair (t_k, \boldsymbol{x}_k) . In order to initialize the model for $\boldsymbol{\eta}$, we choose the parameters $\boldsymbol{\eta}_0$ as initial states for a time right before the first measurement $t_0 < t_1$, and also estimate the initial states. If we wanted to perform the estimation from the LNA model instead, the main difference would be to sum into e the solution to the LNA.

4.4 Introduction to Markov Chain Monte Carlo

In the parameter estimation process, once the model is ready, we are in theory expected to integrate the kernel of the posterior for every set of values in the multidimensional parameter space; then, in order to extract information from the posterior we have to marginalize and calculate expectations through more integration on the posterior. Our task of estimating parameters transforms into a computational burden of integrating functions on a high dimensional space and which normally feature a slim geometry of probability mass, making integration especially painful. For this reason, direct integration is virtually never a viable option in the Bayesian analysis. One of the main methods to determine probability distributions and widely used in Bayesian parameter inference is the Markov Chain Monte Carlo (MCMC).^{11,19,20} Our goal is to use MCMC to calculate the parameter estimation process on the reaction network models. In this section, we will provide the intuition for this method, from the beginning.

4.4.1 Monte Carlo

A Monte Carlo method is one that in general transforms samples into integrals. This is built upon the law of large numbers, that basically shows us how to view uncertain events as certain events plus an approximation error.

We'll work out the intuitions through one dimensional continuous objects, but they can readily be generalized to more dimensions and discrete spaces. Suppose a data generating process $\{X_i^p\} = \{\text{The variable } x \text{ modeled by the probability density } p(x) \text{ is in } [x_i, x_i + dx]\}$, with probability $P(X_i^p) = p(x_i)dx$. According to the law of large numbers, we can calculate the mean of any function f(x) over a density p(x) by using a set of Nsamples as the approximation

$$\langle f \rangle = \int f(x)p(x)dx \approx \frac{1}{N} \sum_{i=1}^{N} f(x_i),$$
(4.22)

an unbiased estimation with error $\mathcal{O}(N^{-1/2})$. As a particular well-known case, we have $\langle x \rangle \approx \sum_i x_i/N$, called the sample mean. But then, if we view f(x)p(x) as a simple function of a real variable x, this is actually a method for calculating the definite integral of f(x)p(x) over a support set through the sample mean. So the law of large numbers can act as a connection between samples of distributions and deterministic integrals. In particular, for a uniform density over an interval of length L, we have $p(x) = L^{-1}$, and

$$\int_{L} f(x)dx \approx \frac{L}{N} \sum_{i=1}^{N} f(x_i).$$
(4.23)

In this case, we use the uniform samples as a sort of "mining" of function values that in the limit will equally distribute themselves around the function mean. And if we map the area under the curve of f(x) into a rectangle by an area-preserving transformation, that rectangle would have a length of L and a height of $\langle f(x) \rangle$.

By using monte carlo integration, we can focus on just sampling the posterior. It is a much easier task than determining the posterior, marginalizing it, and calculating expectations.
4.4.2 Importance Sampling

The method in Eq. (4.23) presupposes that we draw samples from the distribution p(x), but we may need or want to draw samples from another distribution q(x), for example the standard case of drawing from uniform distributions in algorithms. Then, it would be useful if we could input the sampling from a different distribution q(x) into calculations for p(x). This can be done as the trick

$$\langle f \rangle = \int f(x) \frac{p(x)}{q(x)} q(x) dx \approx \frac{1}{N} \sum_{i=1}^{N} f(x_i^q) \frac{p(x_i^q)}{q(x_i^q)}.$$
 (4.24)

So it is the same as sampling the function f(x)p(x)/q(x) from the q(x) distribution. In this context, we can say that we are giving to each x_i^q an importance, or weight, of p(x)/q(x) in order to calculate the mean of f(x) under p(x).

All this is, in principle, of great value for the parameter estimation process through the posterior distribution. With it, we may sample parameters from the posterior in order to calculate estimators for them, such as the mean, even if we have to sample primarily from another distribution. But then we run into a problem: we can have at most the kernel of the posterior, not the entire density. So, we have the posterior represented by the density $p^*(x) = k(x)/Z$, where k(x) is the kernel and $Z = \int_L k(x) dx$ is the unknown normalization factor of the posterior. But then, since Z is actually an integral, there is now a straightforward way to calculate it:

$$Z = \int_{L} k(x) dx \approx \frac{L}{N} \sum_{i} \frac{k(x_{i}^{q})}{q(x_{i}^{q})}.$$
(4.25)

Thus, by estimating Z itself, we can distribute importance (weight) to values of x in the interval according to an estimated density from the known kernel.

4.4.3 Rejection Sampling

But then, we notice that calculating from narrow distributions by sampling other densities like that may be an inefficient process. If p(x) and q(x) don't match, many samples x_i can have a negligible importance in relation to contributing to the probability mass, especially in high dimensional spaces. That's because the probability mass of kernels is usually concentrated in a narrow subset of the parameter space (called typical set), and it gets more concentrated for higher dimensions. This mismatch is the price we pay in order to sample from a distribution using another distribution. In an estimation task, if we could sample the x_i from the posterior itself, it would be a much more efficient sampling process, optimally efficient in this sense. A way to do this is to reject some x_i^q on the basis of their importance under the kernel. It makes the sample generation less computationally efficient to assure efficiency in the convergence of the integration. This transforms the sampling under q(x) into a proposal of sampling, and a candidate sample is filtered under p(x) (or the kernel). For us, a major advantage of this method is that we don't need to estimate Z, which is a much more inefficient process. We'll see that the acceptance-rejection of x_i^q can be defined with only the kernel.

For each sampled x_i^q , we draw a uniform u in the interval [0, 1], and we accept x_i^q if

$$u < \frac{k(x_i^q)/q(x_i^q)}{max[k(x)/q(x)]},$$
(4.26)

intuitively meaning that, in order to be accepted, x_i^q must fall under the curve of k/q. Thus, we reject the sample if it falls off the curve of the kernel, in a region defined by the constant boundary max[k(x)/q(x)] that ensures to encapsulate the whole curve of k/q. This boundary (and also q) is of course considered only from values inside the support of the kernel. Note that, by using a ratio as the filtering criterion, we don't need information of Z (it is only a scale on the kernel). In order to justify this, consider the propositions:

 $A = \{ A \text{ value was accepted} \},\$

 $X = \{ \text{The sampled value is } x \}.$

Then, P(X|A) = P(A|X)P(X)/P(A) has a density

$$p(x|A) = \frac{(k/(mq))q}{P(A)} = \frac{k(x)}{mP(A)},$$
(4.27)

where we defined m = max[k(x)/q(x)]. P(A), the probability of a proposal being accepted, irrespective of its value, can be calculated by marginalizing P(X, A) over x:

$$P(A) = \int P(X, A)dx = \int P(X|A)P(X)dx = \int \frac{k(x)}{mq(x)}q(x)dx = \frac{Z}{m}.$$
 (4.28)

Then, p(x|A) = k(x)/Z, and the accepted samples are distributed according to the desired density, in our case the posterior $p^*(x)$.

The most widespread picture of a monte carlo integration is done with rejection sampling. Instead of directly calculating the integral Eq. (4.23) from a uniform sampling, the uniform sampling is used as a proposal. Then, the function f(x) itself is used as a kernel for the rejection-acceptance step. The simple integral then equates with the monte carlo estimation of Z. The visualization of this process is one of dots accumulating both inside and outside the curve of f(x); the dots falling inside the function are the accepted ones, and those falling outside are rejected.

4.4.4 Markov Chain

The task of determining a posterior distribution is one of finding its probability mass in the parameter space. We saw that a rejection sampling technique can assure that sampling will efficiently represent the posterior probability mass. But we just shifted the problem to the burden of proposing sample candidates. In high dimension parameter spaces, the probability mass will represent just a slim proportion of the space's volume. This means that a lot of proposals will get rejected if our choice of Q(X) isn't already aligned with the kernel. Thus, we are still left with the pressing goal of electing an efficient proposal distribution, one that listens to the location of the posterior's probability mass.

The idea is to use the posterior's geometry in order to devise a criterion. In general, the probability mass is not scattered over the parameter space, but packed inside a specific typical set. We may guess that the typical set is concentrated around the mode, as is the case of the geometry of a one dimensional Gaussian distribution. But at higher dimensions, it non-trivially spreads away from the mode; because despite the importance of points is decreased, the volume of the typical set increases in regions away from the mode. Thus, the geometry of the high dimension posterior in general resembles a narrow band around the region of large importance. We must devise a sampling method that probes the parameter space for this set and then wanders over it with good mixing.

This suggests that we correlate the sampling process, in an attempt to encode the goal of "getting closer" to the typical set once a sample falls far away from it. More formally, we want, given a sample, to distribute the next sample in a way that actively searches for probability mass. With that, q(x) will shape itself according to the geometry of the posterior, granting a sufficiently high acceptance of proposals. For example, simply proposing samples that are nearby an accepted sample already does wonders in increasing the chances of acceptance, because we can expect that accepted samples are more probably located in good neighborhoods (the posterior mass is not scattered over the parameter space, but concentrated). In other words, if a sample is accepted, there is a higher chance that it is closer to the typical set than rejected samples, because the importance for acceptance is based on the kernel itself.

But if we want to correlate a sample with the previous sample, we want to make the sampling process into a Markovian chain. And since we want to lock it as being distributed as the posterior, it must be in equilibrium. Then the problem is reduced to the coordination of a proposal and an acceptance that result in both the equilibrium state of a Markov chain and the posterior distribution. In theory, no matter where the sampling process starts, it can converge to an equilibrium that mimics the sampling of the posterior. Since we now incur in the drawback of having correlated samples, we must ensure a good sampling mixture in order to use the process for estimations (ensure that the process is really able to capture the whole target distribution, and does not "get stuck" in certain regions).

Another problem to consider is that the acceptance process is not that well defined yet, because the determination of a quantity like m = max(k/q) already is an optimization problem. The idea of probing for the typical set from a current sample can also be used to address this and devise a local acceptance criterion. This process of sampling from a Markov chain in order to calculate expectations from a desired target distribution is what is generally called a Markov chain Monte Carlo sampler (MCMC).

4.4.5 Metropolis-Hastings

The algorithm of Metropolis-Hastings is a MCMC sampler built on a property of reversible Markov chains, an equilibrium constraint called detailed balance. Consider the set of statements about a chain at equilibrium $\{X_i^{(t)}\} = \{\text{The state of the chain } x \text{ is in } [x_i, x_i + dx] \text{ at time } t\}$. Then, in detailed balance,

$$P(X_i^{(t-1)})P(X_j^{(t)}|X_i^{(t-1)}) = P(X_j^{(t-1)})P(X_i^{(t)}|X_j^{(t-1)}),$$
(4.29)

noting that $P(X_i^{(t-1)}) = P(X_i)$, because it is at the equilibrium. This is the same as saying that $P(X_i^{(t-1)}, X_j^{(t)}) = P(X_j^{(t-1)}, X_i^{(t)})$. Under detailed balance, the probability flux of the jump from *i* to *j* is the same as for the jump from *j* to *i*, so there is no net flux in the chain; the transitions are pairwise in equilibrium. When we define a particular chain through its transition probabilities, if we make sure that the chain satisfies detailed balance with the posterior, then if it is a proper posterior, that is the unique equilibrium distribution of the chain. Thus, the requirement is to choose transitions satisfying

$$\frac{P(X_j^{(t)}|X_i^{(t-1)})}{P(X_i^{(t)}|X_j^{(t-1)})} = \frac{k(x_i)}{k(x_j)},$$
(4.30)

where k(x) is the kernel of the posterior. The transition is the product of a proposal and an acceptance given proposal steps, so we must have

$$P(X_j^{(t)}|X_i^{(t-1)}) = q(x_i, x_j)P(A_{ij}),$$
(4.31)

where $q(x_i, x_j)$ is the sampling distribution, now dependent on both x_i and x_j , and $A_{ij} =$ {Given a proposal from x_i to x_j , the jump to x_j is accepted}. This results in

$$\frac{P(A_{ij})}{P(A_{ji})} = \frac{k(x_i)q(x_j, x_i)}{k(x_j)q(x_i, x_j)}.$$
(4.32)

If we chose

$$P(A_{ij}) = min\left(1, \frac{k(x_i)q(x_j, x_i)}{k(x_j)q(x_i, x_j)}\right),$$
(4.33)

then it is a valid distribution for which the condition is always satisfied.

The choice of a sampling proposal distribution $q(x_i, x_j)$ influences the speed of convergence of the chain. The particular Metropolis algorithm chooses it to be symmetrical (and making the acceptance independent of q), $q(x_i, x_j) = q(x_j, x_i)$, often a Gaussian $q(x_i, x_j) = \mathcal{N}(x_j | x_i, \sigma^2)$. In this case, the deviation σ regulates a step-size for proposals, that can't be too large so as to miss the regions of interest and cause a large rejection rate or too small so as to be slow on convergence and mixing. For a multi-dimensional parameter space, there is also a choice involved in the jumps being sequential on each dimension or in form of a batch update (updating all dimensions at once is more efficient). The samples taken before convergence are discarded in the estimation process (the initial samples are called warm up), and parallel exploration with multiple chains is advisable. There are actually many details to address in the practical use of MCMC to carry out the estimation process and also the diagnosis and analysis processes following it. We then rely on a good software that can take care of much of the engineering bits.

4.4.6 STAN

We implement the MCMC method through the STAN statistical programming language.²¹ It is a multi-interface language for custom Bayesian computation through advanced MCMC algorithms, written in C++. STAN targets a log transformation of the posterior distribution (needed for improved computation stability) and can run with two gradient-based MCMC methods for adaptive probing of parameter space. The Hamiltonian MCMC (HMCMC) and its variant No U-turn Sampler (NUTS).^{22,23} The motivation behind HMCMC methods is to interpret the posterior landscape as a potential energy, with a simulated Hamiltonian dynamics imprinting a momentum into the Markov chain, so the jump-size of the proposal adapts according to the gradient of the posterior; that is in contrast with the rigid Gaussian proposal. The NUTS variant implements strategies to improve the covering of the posterior by increasing the awareness of the chain based on the samples already visited; it is able to provide an automatic stop to the posterior's exploration, when the chain "sees" that it's enough.

STAN also contains many diagnosis, analysis, and visualization tools aiming at efficiently automate all tasks that are not related to the modeling aspect of the inference. It is possible to effortlessly check for convergence, mixing, and proportion of effective samples. There are also in-built transformations for constraining parameters and calculating the posterior in the log space.

4.5 Estimation Example: Lotka-Volterra

In order to properly illustrate the parameter estimation process, we consider the Lotka-Volterra model from Eq. (4.16). We generate data using the stochastic simulation algorithm on a medium-sized space of $\Omega = 100$: from a sample trajectory, we extract 30 measurements at random times for both preys and predators.

We then use the deterministic dynamics as a statistical model of the data-generating process, thus the likelihood becomes a Gaussian having the model as the mean and the variance σ^2 as a proxy noise level also to be estimated. For simplicity, we use the generally well-suited exponential priors for all the parameters. The priors reflect the order of magnitude of the parameters, with means equal to one of 0.1, 1, or 10, depending on the parameter.

The statistical task is to estimate the parameters $\boldsymbol{\theta} = (w, \gamma, \delta, \mu, x_0, y_0, \sigma)$, where (x_0, y_0) is the initial state of preys and predators used to initialize the statistical model, and σ is the proxy standard deviation of the measurements. The initial state, for t = 0, was chosen as 200 preys and 100 predators, meaning densities of $(x_0 = 2, y_0 = 1)$. Figure (5) shows the sampled trajectory and measurements together with both the deterministic model and the mean estimated model, along with trace plots of the posterior. The table (2) compares the estimated values with the real parameter values used for data generation.

The estimation process runs with 4 chains. For all chains, the trace plots indicate the expected behavior of the jumps, a "fuzzy caterpillar" shape, of well-mixed exploration of the posterior. Note that, with just the noisy extracted measurements and the statistical model, we are able to, in theory, estimate the particular place of the multi-dimensional parameter space in which the system operates.



Figure 5 – Parameter estimation on the Lotka-Volterra model. Upper left: Stochastic sample with 30 randomly extracted measurements for both preys and predators, compared with the deterministic model. Lower left: Same stochastic sample, compared with the deterministic curve generated with the mean estimated parameters. **Right:** Trace plots for posterior samples of the 4 network parameters and the initial state of the model. The gray region indicates the warm-up iterations.

Source: By the author.

Table 1 – **Estimated parameters.** The estimated posterior yields the mean values and standard deviations shown in the table, compared with the real parameter values. The standard deviation of the measurements σ has no real value because it is a proxy of the actual noise levels coming from the stochastic process.

Parameter	Mean Estimation	Real Value
w	0.57 ± 0.02	0.55
γ	0.18 ± 0.01	0.18
δ	2.04 ± 0.11	2.00
μ	0.91 ± 0.03	0.84
x_0	1.88 ± 0.05	2.00
y_0	1.22 ± 0.05	1.00
σ	0.15 ± 0.02	-

Source: By the author.

5 STOCHASTIC ANALYSIS AND PARAMETER ESTIMATION OF THE GOOD-WIN MODEL FOR GENETIC REGULATION.

This chapter develops a work that was partially presented at the 19th International Conference on Systems Biology, ICSB 2018, hosted at Lyon, France. At the time, the parameter estimation analysis wasn't completed and the overall narrative was slightly different. The work features an application of the stochastic methods developed in previous chapters and taps into a discussion of design principles of modeling that is most relevant for the systems biology field that points to the functional communication between biological systems. In particular, this work analyzes the design principles of biological oscillators through a minimal model generally called the Goodwin oscillator. This chapter is somewhat less mature than the others, since it was the first research and conducted before a more solid understanding of scientific writing and programming.

5.1 Abstract

Gene expression and regulation are intrinsically stochastic processes. Due to low molecular counts and random chemical reactions, intrinsic noise is a major feature to be controlled or exploited by gene network designs. In this work, we develop an analytic modeling of a minimal network of negative feedback through the use of reaction networks, yielding the stochastic approach of the chemical master equation for the probability density of protein levels. The model highlights differences between direct and indirect self-regulation and the appearance of oscillations. We expand the nonlinear equations using Van Kampen's system size expansion and make connections to the deterministic limit. We find a steady-state noise control profile for the indirect feedback strength, and we see an optimum feedback strength value for noise control in the model. We also analyze the model in the case of a Hill-type feedback function, making it a stochastic version of the Goodwin oscillator model. We further study the bifurcation properties of the Goodwin model in terms of the Hill exponent as well as the feasibility of Bayesian parameter estimation in relation to the critical point.

Keywords: Systems Biology, Intrinsic Noise, Circadian Oscillations, Design Principles, Linear Noise Approximation.

5.2 Introduction

5.2.1 Systems Biology

Systems biology is an interdisciplinary field aiming at the modeling and analysis of biological systems. It combines the mathematical and computational descriptions, that are usual to physics and applied mathematics, with the holistic approach of systems thinking. Such an approach enables the quantitative exploration of a system's dynamics and a multiscale treatment of complex biological networks; as, for example, gene expression and regulation networks, protein dynamics, cellular signaling, metabolic pathways, etc.^{24,25} Systems biology became possible with the modern experimental and computational resolutions that enable observation and control of entire molecular systems' dynamics. These observations revealed non-intuitive structures that begged for more quantitative techniques than those available at the time (like features arising from nonlinear dynamics).³

With this view, the behavior and function of a cell population is considered emergent from the behavior of a single cell and the interactions with its environment; and the later is considered emergent from the dynamics of intracellular components networks and the conditions upon them, and so on. The descriptions between scales are alike physical limits, as for instance the thermodynamic limit emerges from statistical mechanics. But systems thinking also suggests multiscale descriptions of function instead of components, so specific functions can be better isolated from neighboring dynamics, like what is explored through the concept of gene networks (so systems are bound by their functional interaction more than their physical interaction). One can systematically follow the functions of a specific gene network up to a phenotypic level without the need to understand the dynamics of an entire genome or the specific marginal interactions making the network possible.

Intracellular processes are modeled as dynamical systems, with parameters and designs made to match experimental observations. The models range from minimal, prototypical models to complex, multi-component simulations and synthetic engineering of biological processes and networks.^{26,27} Minimal models are needed to unveil design principles, network motifs, leading dynamics and properties of more detailed systems, mostly through analytic development.²⁸ These are building blocks to the understanding and prediction of complex systems, connecting structure to behavior. Analytic findings in minimal models can be results of phase space, bifurcation or statistical analysis, and can guide further modeling, experimenting and synthetic engineering of the system.

5.2.2 Gene Expression

Gene expression is the dynamics of synthesis of protein products from one or more genes. Even prokaryotic gene expression and regulation is in detail accomplished by a variety of reactions involving many different components.²⁹ But this complex event can be roughly reduced, functionally, by considering the functions of three central steps:

1) **Transcription:** is when information from the DNA gets transcribed to a messenger RNA by the enzyme RNA polymerase. The enzyme is able to bind to the promoter region of the DNA, then the strands are unzipped to produce the corresponding mRNA with the enzyme reading one of the strands. The new mRNA is released to participate in protein production and, if the cell is an eukaryote, this is preceded by the

additional step of leaving the cell nucleus.

2) **Translation:** is when a protein is synthesized from a mRNA. In the cytoplasm, mRNA binds to the macro-molecule ribossome that reads the mRNA's nucleotides (using freefloating tRNAs) while assembling the aminoacid chain. After being created, the chain folds to the protein conformation and is released.

3) **Regulation:** is the modulation of expression rates done by cellular mechanisms. Genes can regulate the expression of each other by having protein products that initiate processes of regulation. These are the interactions that combine genes in networks and allow their communication, giving complexity to their function. A gene can also regulate itself directly, by having its own products changing the rate of their expression in the process of self-regulation. Regulation is possible in any step of expression, but we focus on transcriptional regulation. The elements that regulate transcription are proteins called transcription factors, and they bind to the DNA, turning the gene on and off. Their dynamics are promoted by parameters of the cell environment and their binding interacts with DNA polymerase and its capacity to engage in transcription, usually obstructing the operator region of the DNA or activating polymerase's binding; doing so, they can be inhibitors or activators of transcription, usually making up for positive and negative feedback. Stochastic regulation is central in determining viral and bacterial behavior and adaptability, and also cellular differentiation and morphogenesis in multicellular organisms.

The regulation of gene expression and, consequentially, the architecture of gene networks, is able to present many features depending on specific environments and functions of genes. Important examples of such features are 1) genetic switches, that control the activation of different cellular environments and account for cellular change of behavior; 2) genetic oscillations, that control cellular rhythms and circadian cycles; 3) signaling and metabolic pathways, that allow orchestrated responses to changes in the environment, like the presence of a substance, and actions upon the environment; 4) operons, that are units of co-transcription of partner genes. Important models include: phage- λ lysis-lysogeny switch, collins toggle switch, bacterial quorum sensing, tryptophan and lactose operons, circadian rhythm generator of drosophila melanogaster and the repressilator.^{13,27,30,31}

Another feature expressed by gene networks is the role of noise.³² Since the process of gene expression involves approximately random chemical reactions among molecules present in low copies subject to random degradation, like mRNA and protein products, and also inside a cell (an unpredictably diverse and ever-changing environment), we can expect that it is crucially affected by both intrinsic and extrinsic noise.^{33,34} And we see that this stochastic nature affects structure and function of gene networks by the attempts to minimize the presence and resist the effects of detrimental noises and to control functional noises.³⁵ Noise can be functional, for example, in processes of cellular fate decisions accounting for robustness of cellular populations, promoting variability in decision triggers. $^{36-40}$

Minimal models of stochastic gene networks include models of single gene expression, as single gene dynamics represent the basic unit of networks. We can expect to understand noise in the network level as emergent from noise at the expression level.⁴¹

5.2.3 Goodwin Oscillator

The oscillatory feature is of great value for genetic processes, and the design principles for genetic oscillators are guides for when to expect and how to create oscillatory networks. In particular, the feature of sustained oscillations by limit cycles, that are robust in relation to external perturbations and portray a stable amplitude; always the result of bifurcations in the parameter space. Design principles for sustained oscillations are at least: 1) a negative feedback loop; 2) a delay between a process and its negative feedback response; 3) and nonlinear dynamics.⁴²

The Goodwin oscillator is a minimal proposal for a genetic oscillator, with just enough complexity to fulfill all necessary design principles.⁴³ In its full three-variable design, it is composed of a transcription reaction that generates mRNA with a nonlinear rate featuring negative feedback; then a translation reaction that generates a protein in the presence of an mRNA; then a reaction where the presence of the protein activates a transcription factor that regulates the transcription reaction rate; and also degradation reactions. It was first proposed in 1960 by Brian Goodwin, with a Hill feedback function modulating the transcription reaction.⁴⁴ The Goodwin oscillator has surprisingly complex behavior, given the simplicity of its equations.

The system oscillates only in certain regions of the parameter space. A Hopf bifurcation separates the transition between a stable steady-state and limit cycle oscillations. Oscillations occur with high transcription feedback strength, when the model is highly non-linear.⁴⁵

In this work, we investigate the strength of intrinsic noise at steady state for the Goodwin oscillator, examine its properties at the bifurcation point, and compare the feasibility of Bayesian parameter estimation at different distances from the bifurcation, using a deterministic statistical model with stochastic simulated data. We use a general description of the feedback in terms of equilibrium feedback strength, then apply the results for the case of the Hill-type feedback.

5.3 Methods

5.3.1 2D and 3D Models

We consider a genetic system that, at its most basic design, featuring transcription, translation, and regulation, is composed of only the mRNA and protein species. The

protein acts as its own regulator and we will call this model the 2D Goodwin model. This model does not exhibit oscillations; despite the nonlinear negative feedback, it lacks a sufficient delay in the loop from transcription to regulation. Using the reaction network theory to represent the reactions of the model, we have

$$\emptyset \xrightarrow{F(n_2/\Omega)} X_1, \quad X_1 \xrightarrow{\gamma_1} \emptyset$$
$$X_1 \xrightarrow{k_2} X_1 + X_2, \quad X_2 \xrightarrow{\gamma_2} \emptyset.$$
(5.1)

We represent the mRNA as X_1 and the protein as X_2 . The first reaction is a transcription, and the functional shape of the regulation is modelled with the activation function of protein density $F(n_2/\Omega)$, where n_2 is the protein count-number. In order to configure it as a negative feedback, we impose that 1) $F(n_2/\Omega) > 0$ for every n_2 , as it is a reaction rate, and 2) $\frac{dF(\phi_2)}{d\phi_2} < 0$ for every ϕ_2 , so that $F(n_2/\Omega)$ is a monotonically decreasing function of n_2 , characterizing the negative, repressing, feedback loop (where ϕ_2 is the deterministic limit density of proteins). The second reaction is the degradation of mRNA with a rate of γ_1 . The third reaction is the translation, with rate k_2 , and the last is the degradation of proteins, with rate γ_2 .

The oscillatory behavior is made possible with the inclusion of a third species that acts as a transcription factor for indirect regulation of the protein production. The Goodwin oscillator, that we will call 3D Goodwin model, is represented by the reaction network

$$\emptyset \xrightarrow{F(n_3/\Omega)} X_1, \quad X_1 \xrightarrow{\gamma_1} \emptyset$$
$$X_1 \xrightarrow{k_2} X_1 + X_2, \quad X_2 \xrightarrow{\gamma_2} \emptyset$$
$$X_2 \xrightarrow{k_3} X_2 + X_3, \quad X_3 \xrightarrow{\gamma_2} \emptyset.$$
(5.2)

Now, the transcription regulation is no longer a function of the protein X_2 , but of the transcription factor X_3 : $F(n_3/\Omega)$. X_3 is activated by the presence of the protein, with a rate of k_3 , and is assumed to degrade with the same rate as the protein, γ_2 . The indirect regulation may, with sufficient non-linearity, provide the needed delay to sustain oscillations.

5.3.2 Deterministic Description

The reaction network models yield a deterministic description of a system of ordinary differential equations, given by

$$\frac{d\phi_i}{dt} = \sum_r S_{ir} \kappa_r \prod_j \phi_j^{s_{jr}},\tag{5.3}$$

where ϕ_i are the density of species, s_{jr} is the stoichiometric coefficient of reactant j in reaction r, κ_r is the reaction rate of reaction r, and S_{ir} is the stoichiometric difference of

species *i* in reaction *r*. Using this system of equations, we perform a dynamic stability analysis of the models. For that, we linearize the system around the fixed points and determine the behavior through the Lyapunov exponents, that are the eigenvalues of the Jacobian of the linearized system. A supercritical Hopf bifurcation can happen when the system exhibits complex conjugate Lyapunov exponents, with a phase transition occurring with the passage from $Re(\lambda) < 0$ to $Re(\lambda) > 0$, i.e. from a stable fixed point to an unstable fixed point with the appearance of a stable limit cycle around it.⁴⁶ This is the bifurcation known to occur for the 3D Goodwin model, having the critical point for a parameter set that yields $Re(\lambda) = 0$ for the pair of complex conjugate exponents.

5.3.3 Stochastic Description

There is also the more fundamental stochastic description, with time evolution of the densities' probabilities given by a master equation of Markov jump processes,⁴⁷ with transition rates

$$W_r = \kappa_r \Omega \prod_i \frac{n_i!}{(n_i - s_{ir})! \Omega^{s_{ir}}},\tag{5.4}$$

where Ω is the system size, the intracellular volume in this case. The connection to the deterministic system is achieved by the limit $\mathbf{n}, \Omega \to \infty$, with $\boldsymbol{\phi} = \mathbf{n}/\Omega$. In order to analytically treat the stochastic system, we use the linear noise approximation (LNA), the first order of the system size expansion for the noise around the deterministic trajectory, accomplished with the transformation $\mathbf{n} = \Omega \boldsymbol{\phi} + \Omega^{1/2} \xi$.^{14,48} The parameter ξ is the noise around the deterministic trajectory, under the LNA it follows time-dependent Gaussian distributions, and we can calculate the level of noise as a solution to a Lyapunov equation for the variance matrix of ξ in terms of the deterministic solution.

5.3.4 Noise Analysis

The measure of noise magnitude relative to the density size is the index of dispersion of each species, $D(n_i) = \frac{\langle \Delta n_i^2 \rangle}{\langle n_i \rangle}$.⁴⁹ We can evaluate noise strength in relation to the unstructured Poisson distribution, that always has D = 1. Thus, sub-Poissonian noises have $\langle \Delta n_i^2 \rangle < \langle n_i \rangle$ and super-Poissonian noises have $\langle \Delta n_i^2 \rangle > \langle n_i \rangle$. We will evaluate the steady-state index of dispersion for the Goodwin model in terms of ξ :

$$D_s(n_i) = \frac{\langle \Delta n_i^2 \rangle^*}{\langle n_i \rangle^*} = \Omega^{-1/2} \frac{\langle \xi_i^2 \rangle^*}{\phi_i^*}$$
(5.5)

. In the results, we will plot the index of dispersion as a function of the feedback strength at steady-state.⁵⁰ The measure of feedback strength is the derivative of the steady-state feedback function $F(\phi^*/\Omega)$, the parameter

$$w = -\frac{1}{\gamma_2} \frac{dF(\phi_2^*)}{d\phi_2} > 0.$$
(5.6)

Thus, the repression of regulation increases with w. Valuable metrics defined in terms of w include the strength that minimizes the noise levels of the system, which is of evolutionary interest for processes that require dynamical control and precision, and the critical value of w, meaning the least feedback strength that results in oscillations.

5.3.5 Hill Feedback

At first, the feedback function $F(n/\Omega)$ can be any function of n/Ω that is positive and monotonically decreasing. It characterizes a negative feedback loop. But the standard activation function of feedback in the Goodwin model is the Hill function⁵¹

$$F(\phi) = \frac{a}{1 + b\phi^m}.$$
(5.7)

The parameter m is called Hill exponent and is a measure of nonlinearity and feedback intensity, a higher m means high nonlinearity and low feedback intensity.

In the following, we show a possible model of time-scale reduction that results in a Hill function for the context of genetic regulation.⁵² Consider a DNA operator site O that can be either free for transcription or bounded by a repressor R. It takes m molecules of the repressor to inactivate the transcription expressed by that operator. The reactions of binding/unbinding form an equilibrium pair:

$$O + mR \xrightarrow{k_+} OR_m \xrightarrow{k_-} O + mR.$$
(5.8)

The mass-action equations for these reactions is given by

$$\frac{d[O]}{dt} = -k_{+}[O][R]^{m} + k_{-}[OR_{m}]$$

$$\frac{d[R]}{dt} = -mk_{+}[O][R]^{m} + mk_{-}[OR_{m}]$$

$$\frac{d[OR_{m}]}{dt} = +k_{+}[O][R]^{m} - k_{-}[OR_{m}].$$
(5.9)

At the steady-state, we have a detailed-balance relation,

$$k_{-}[OR_{m}] = k_{+}[O][R]^{m}.$$
(5.10)

With this equation, we build the following ratio in terms of [R]

$$f = \frac{[O]}{[O] + [OR_m]} = \frac{1}{1 + K^{-1}[R]^m},$$
(5.11)

where $K = k_{-}/k_{+}$ is the dissociation constant. The rate f represents the fraction of operators that is free for transcription at each time, but also represents the fraction of time each operator is free (that we know with uncertainty and model probabilistically). The transcription occurs only if the operator is free from the repressor molecules, so we make it's rate proportional to this fraction,

$$F([R]) = k_1 f([R]) = \frac{k_1}{1 + K^{-1}[R]^m}.$$
(5.12)

Each transcription event occurs in a slower time-scale and the process of regulation is taken to be in equilibrium always; this is the assumption that justifies the reduction and then allows the non-mass-action reaction rate. The regulation rate of the Goodwin model then takes the following Hill shape:

$$F(n/\Omega) = \frac{k_1}{1 + K^{-1}(\frac{n}{\Omega})^m}.$$
(5.13)

The hill exponent m is also called the degree of cooperation, representing the number of repressor molecules needed to form a *complex* bounded to the operator.

5.3.6 Parameter Estimation

We also perform a Bayesian parameter estimation on the 3D Goodwin oscillator with a Hill regulation function, focusing on the Hill exponent as a measure of nonlinearity and of different noise levels. Using m as the critical parameter of choice, we analyze the noise levels at different distances from the critical point modeled as the estimated standard deviation of the densities. Our statistical model considers the deterministic system with constant Gaussian noise as the data generating model while we use the stochastic system as the real model behind a simulated set of data.

As a simulation of the real data generating process, we extract, at random time intervals, observations from a sample of the stochastic description of the Hill feedback 3D Goodwin model. For that, we use the stochastic simulation algorithm (SSA) of the master equation. The data sets are made with 40 observations each, from 3 different intensities of feedback nonlinearity: subcritical regime far from the critical point $(m - m_c \ll 0)$, subcritical regime near the critical point $(m - m_c \text{ close to } 0^+)$, and supercritical regime near the critical point $(m - m_c \text{ close to } 0^-)$. The three scenarios represent different magnitudes of noise and investigate the feasibility of estimation near the bifurcation.

The estimation task is performed through the STAN language for Bayesian computation.²¹ It uses a Hamiltonian Markov Chain Monte Carlo algorithm to estimate a posterior distribution for a vector of model parameters $\boldsymbol{\theta}$.²² With exponential priors matching the expected scales of parameters, we model the likelihood of the data as a Gaussian with mean equal to the deterministic description of the reaction network model. The Gaussian has a constant independent variances that we estimate along with the reaction rates of the Hill 3D Goodwin model. Thus, $\boldsymbol{\theta} = (b, d, \epsilon, \gamma_1, m, k_1, K, \sigma_1, \sigma_2, \sigma_3)$ and we have the likelihood of the data, given the model, as

$$P(\boldsymbol{x}|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\phi}(\boldsymbol{\theta}), \boldsymbol{\sigma}), \qquad (5.14)$$

where $\boldsymbol{x} = (x_1, x_2, x_3)$ are the observed data points. The posterior is then estimated with the kernel $P(\boldsymbol{\theta}|\boldsymbol{x}) \approx P(\boldsymbol{x}|\boldsymbol{\theta})P(\boldsymbol{\theta})$.

5.4 Results

5.4.1 2D Model

We first reparameterize the model with dimensionless parameters, defining $b = k_2/\gamma_1$ and $\epsilon = \gamma_2/\gamma_1$ and changing the temporal scale as $t \to \gamma_1 t$. The parameter ϵ is the mRNAto-protein mean lifetime ratio. Usually, mRNA degradation is much faster than that of proteins. That can make $\epsilon \ll 1$ in real genetic systems. The parameter b measures the mean number of proteins produced by each mRNA by translations and is called the protein burst size. From the reaction network, we then have the following deterministic system:

$$\frac{d\phi_1}{dt} = -\phi_1 + \frac{\epsilon F(\phi_2)}{\gamma_2}$$

$$\frac{d\phi_2}{dt} = b\phi_1 - \epsilon\phi_2.$$
(5.15)

The fixed point is of the form

$$\phi_1^* = \frac{\epsilon}{b}\phi_2^* = \frac{\epsilon}{\gamma_2}F(\phi_2^*) \tag{5.16}$$

The Jacobian matrix for this system in the fixed point is then

$$\begin{bmatrix} -1 & -\epsilon w \\ b & -\epsilon \end{bmatrix},\tag{5.17}$$

where we use the previously defined steady-state strength of feedback.

The eigenvalues of (5.17) are

$$\lambda_{\pm} = -\frac{1}{2}(1+\epsilon) \pm \frac{1}{2}\sqrt{(1+\epsilon)^2 - 4\epsilon(1+wb)}.$$
(5.18)

We see that they are not real numbers for $(1 + \epsilon)^2 < 4\epsilon(1 + wb)$, it means that the phase portrait might be a spiral; but the real part is always negative, so the steady-state is always a stable fixed point.

The reaction network also yields a stochastic dynamics that, as the LNA approximation, results in a Fokker-Planck-type equation with time-dependent Gaussian solutions on the joint density of the model:

$$\frac{\partial \Pi(\xi_1,\xi_2,t)}{\partial t} = \left(-\frac{dF(\phi_2)}{d\phi_2} \frac{\partial(\xi_2\Pi)}{\partial\xi_1} + \frac{F(\phi_2)}{2} \frac{\partial^2\Pi}{\partial\xi_1^2} \right) + \left(-k_2 \frac{\partial(\xi_1\Pi)}{\partial\xi_2} + \frac{k_2\phi_2}{2} \frac{\partial^2\Pi}{\partial\xi_2^2} \right) \\ + \left(\gamma_1 \frac{\partial(\xi_1\Pi)}{\partial\xi_1} + \frac{\gamma_1\phi_1}{2} \frac{\partial^2\Pi}{\partial\xi_1^2} \right) + \left(\gamma_2 \frac{\partial(\xi_2\Pi)}{\partial\xi_2} + \frac{\gamma_2\phi_2}{2} \frac{\partial^2\Pi}{\partial\xi_2^2} \right).$$
(5.19)

Assuming a determined initial state, the first moment of the noise variable ξ is always zero, and the second moments evolve as a Lyapunov system of equations. In steady-state,

we can write the system as

$$b\langle\xi_1^2\rangle^* + w\epsilon b\langle\xi_1\xi_2\rangle^* - \epsilon\phi_2^* = 0$$

$$b\langle\xi_1\xi_2\rangle^* - \epsilon\langle\xi_2^2\rangle^* + \epsilon\phi_2^* = 0$$

$$(1+\epsilon)\langle\xi_1\xi_2\rangle^* + w\epsilon\langle\xi_2^2\rangle^* - b\langle\xi_1^2\rangle^* = 0.$$
(5.20)

Solving for the index of dispersion of n_2 , we get:⁵⁰

$$D_s(n_2) = \frac{\langle \Delta n_2^2 \rangle^*}{\langle n_2 \rangle^*} = \Omega^{-1/2} \frac{\langle \xi_2^2 \rangle^*}{\phi_2^*} = \Omega^{-1/2} \left(1 + \left(\frac{b}{1+\epsilon}\right) \frac{(1-w)}{(1+bw)} \right)$$
(5.21)

This result shows two noise contributions: the unregulated Poissonian transcription noise represented by the factor 1 and the translation noise represented by the factor $b/(1 + \epsilon)$.⁵³ We see that the negative regulation acts to decrease the noise, so negative feedback may function as a noise control feature.

Figure (6) shows Eq. (5.21) for $\Omega = 1$. A curve for $\epsilon = 0.01$ and $\epsilon = 1$ are represented. We see that both curves cross the Poissonian threshold for w = 1, as expected. For the super-Poissonian case, the noise control is greater for higher values of ϵ . For the sub-Poissonian case, a substantially greater noise control can achieved for a lower value of ϵ . We can visualize that D_s is a monotonically decreasing function of w, which can effectively decrease fluctuations to a minimum value of

$$D_s(w \to \infty) = \frac{\epsilon}{1+\epsilon}.$$
(5.22)

The infinite strength fluctuations equate to ϵ for $\epsilon \ll 1$, showing a baseline persistent level mediated only by the life-time ratios.

5.4.2 3D Model

For the deterministic description of the 3D model, with dimensionless parameters, we introduce $d = k_3/\gamma_1$, measuring the creation of protein repressors:

$$\frac{d\phi_1}{dt} = \frac{\epsilon F(\phi_3)}{\gamma_2} - \phi_1$$

$$\frac{d\phi_2}{dt} = b\phi_1 - \epsilon\phi_2$$

$$\frac{d\phi_3}{dt} = d\phi_2 - \epsilon\phi_3.$$
(5.23)

For the fixed point, we have

$$\phi_2^* = \frac{b}{\gamma_2} F(\phi_3^*), \quad \phi_1^* = \frac{\epsilon}{b} \phi_2^*, \quad \phi_3^* = \frac{d}{\epsilon} \phi_2^*.$$
 (5.24)

Now, we note that

$$w = -\frac{1}{\gamma_2} \frac{dF(\phi_3^*)}{d\phi_3}$$
(5.25)



Figure 6 – Noise levels by equilibrium feedback strength. Above: The steady-state index of dispersion for protein products D_s as a function of the steady-state feedback strength w, for the direct regulation 2D model. b = 20, d = 10and the blue and red lines have $\epsilon = 1$ and $\epsilon = 0.01$ respectively. Below: D_s as a function of the re-scaled steady-state feedback strength bdw, for the indirect regulation 3D model. For the blue case, $bdw_{min} \approx 2.98$; for the red case, $bdw_{min} \approx 0.45$. Dashed lines are incorrect solutions after the bifurcation, where LNA assumptions break. Source: By the author.

and write the Jacobian of the system for the fixed point:

$$\begin{bmatrix} -1 & 0 & -\epsilon w \\ b & -\epsilon & 0 \\ 0 & d & -\epsilon \end{bmatrix}.$$
 (5.26)

The eigenvalues are solutions to the following equation

$$\lambda^3 + (1+2\epsilon)\lambda^2 + (\epsilon^2 + 2\epsilon)\lambda + \epsilon^2 + db\epsilon w = 0.$$
(5.27)

For the particular case of $\epsilon = 1$, it's easily solvable, and the eigenvalues are

$$\lambda_R = -1 - (dbw)^{1/3}, \quad \lambda_{\pm} = -1 + \frac{1}{2} (dbw)^{1/3} \pm \frac{i}{2} \sqrt{3} (dbw)^{1/3}.$$
 (5.28)

There is a real eigenvalue that is always negative, and there is a pair of conjugate complex eigenvalues. The real part of these eigenvalues is zero for the critical value of the feedback strength: $w_c = 8/db$. This is a critical value for a supercritical Hopf bifurcation. With $w > w_c$, the system exhibits limit cycle oscillations.

This motivates us to look for a general form of w_c for any ϵ , and we do this by searching in equation (5.27) for imaginary solutions, i.e., solutions of the type $\lambda = ai$. This procedure gives:

$$w_c = \frac{2 + 4\epsilon + 2\epsilon^2}{db}.$$
(5.29)

With $a = Im(\lambda) = \sqrt{\epsilon^2 + 2\epsilon}$; the value $2\pi/a$ is an approximation of the limit-cycle's period near the bifurcation point. We have, thus, found a critical value for the feedback strength at the fixed-point.

Numerically, we will analyze the rescaled parameter bdw. The critical rescaled strength, bdw_c , has a minimum value of 2 and increases with the small parameter ϵ , being usually close to 2.

The stochastic dynamics of the 3D model, with the LNA approximation, reduces to the equation

$$\frac{\partial\Pi(\xi_1,\xi_2,t)}{\partial t} = \left(-\frac{dF(\phi_3)}{d\phi_3}\frac{\partial(\xi_3\Pi)}{\partial\xi_1} + \frac{F(\phi_3)}{2}\frac{\partial^2\Pi}{\partial\xi_1^2}\right) + \left(-k_2\frac{\partial(\xi_1\Pi)}{\partial\xi_2} + \frac{k_2\phi_2}{2}\frac{\partial^2\Pi}{\partial\xi_2^2}\right) \\
+ \left(-k_3\frac{\partial(\xi_2\Pi)}{\partial\xi_3} + \frac{k_3\phi_2}{2}\frac{\partial^2\Pi}{\partial\xi_3^2}\right) + \left(\gamma_1\frac{\partial(\xi_1\Pi)}{\partial\xi_1} + \frac{\gamma_1\phi_1}{2}\frac{\partial^2\Pi}{\partial\xi_1^2}\right) \\
+ \left(\gamma_2\frac{\partial(\xi_2\Pi)}{\partial\xi_2} + \frac{\gamma_2\phi_1}{2}\frac{\partial^2\Pi}{\partial\xi_2^2}\right) + \left(\gamma_2\frac{\partial(\xi_3\Pi)}{\partial\xi_3} + \frac{\gamma_2\phi_3}{2}\frac{\partial^2\Pi}{\partial\xi_3^2}\right).$$
(5.30)

This equation, together with the deterministic steady-state, results in the following six dimensional linear system for the steady-state second moments:

$$w\epsilon \langle \xi_{3}^{2} \rangle^{*} - d \langle \xi_{1}\xi_{2} \rangle^{*} + (1+\epsilon) \langle \xi_{1}\xi_{3} \rangle^{*} = 0$$

$$w\epsilon \langle \xi_{2}\xi_{3} \rangle^{*} - b \langle \xi_{1}^{2} \rangle^{*} + (1+\epsilon) \langle \xi_{1}\xi_{2} \rangle^{*} = 0$$

$$-b \langle \xi_{1}\xi_{3} \rangle^{*} - d \langle \xi_{2}^{2} \rangle^{*} + 2\epsilon \langle \xi_{2}\xi_{3} \rangle^{*} = 0$$

$$b \langle \xi_{1}\xi_{2} \rangle^{*} - \epsilon \langle \xi_{2}^{2} \rangle^{*} + \epsilon \phi_{2}^{*} = 0$$

$$d \langle \xi_{2}\xi_{3} \rangle^{*} - \epsilon \langle \xi_{3}^{2} \rangle^{*} + d\phi_{2}^{*} = 0$$

$$-w\epsilon b \langle \xi_{1}\xi_{3} \rangle^{*} - b \langle \xi_{1}^{2} \rangle^{*} + \epsilon \phi_{2}^{*} = 0$$
(5.31)

Solving the system for n_2 's index of dispersion, we get, for $\Omega = 1$:

$$D_s(n_2) = \frac{\langle \xi_2^2 \rangle^*}{\phi_2^*} = \frac{2\epsilon(1+\epsilon)(1+\epsilon+b) + dbw \left[(1+2\epsilon)(wb+\epsilon) + 1 + \epsilon + b \right]}{2\epsilon(1+\epsilon)^2 + dbw \left[(1+\epsilon)(1+2\epsilon) + 1 - dbw \right]}.$$
 (5.32)

Figure (6) shows equation (5.32), D_s in terms of the rescaled steady-state strength (bd)w, for $\epsilon = 1$ and $\epsilon = 0.01$. There is a value of the feedback strength that makes the protein dispersion diverge to infinity. This value obeys the equation

$$d^{2}b^{2}w^{2} - db(1 + (1 + \epsilon)(1 + 2\epsilon))w - 2\epsilon(1 + \epsilon)^{2} = 0.$$
(5.33)

We see that the critical noise strength we previously calculated, w_c , is precisely the solution of the equation above, meaning that, in the critical point, the noise becomes divergent. We could expect this result as a fluctuation precursor of the bifurcation, resulting from increasing near-critical time-correlations of different instants' fluctuations in steady-state.⁵⁴ This analytic finding motivates the study of bifurcation noise in such systems; we ask whether a feature like this could be of any functional use by the network, like, for example, the use of noise-induced oscillations for systems near the bifurcation point, still in the subcritical dynamics. We note the importance of fluctuation precursors of bifurcations in the prediction of critical behavior in real systems. It's also an interesting result that we could reach a critical point calculation by using the framework for the system's noise.

Beyond the critical point, in the supercritical dynamics, the stochastic calculations lose their meaning, because the assumption of one stable fixed point is violated after the Hopf bifurcation. The most interesting aspect of figure (6) lies in the subcritical region though; we note that there is an optimal control value of $w = w_{min}$, one that minimizes the fluctuations, unlike what we have seen in the 2D model featuring direct regulation. Thus, indirect regulation by transcription factors introduce the feature of noise suppression through regulation. We see that, for lower values of ϵ , a lower feedback strength is needed to achieve optimal control, and the control is more efficient. We can derive an analytic expression for w_{min} , but the expression is loaded and provides no further insights. Changes in b and d have shown less sensitivity than changes in ϵ .

Now, assuming a Hill-type feedback reaction rate, we write

$$F(\phi_3^*) = \frac{k_1}{1 + K^{-1}(\phi_3^*)^m}$$
(5.34)

For simplicity, we relabel $\phi_3^* = \phi$, so the derivative is

$$F'(\phi) = \frac{-k_1 K^{-1} m \phi^{m-1}}{(1 + K^{-1} \phi^m)^2}$$
(5.35)

and then

$$w = -\frac{F'(\phi)}{\gamma_2} = \frac{k_1 K^{-1} m \phi^{m-1}}{\gamma_2 (1 + K^{-1} \phi^m)^2}.$$
(5.36)

We know that, in the fixed point,

$$\phi_1^* = \frac{\epsilon^2}{bd} \phi = \frac{\epsilon k_1}{\gamma_2 (1 + K^{-1} \phi^m)}.$$
(5.37)

Reorganizing these expressions, we have w(m) explicitly stated in terms of $m, \epsilon, k_1, K, b, d$ and γ_2 as

$$w = \frac{m\epsilon}{bd} - \frac{m\epsilon^2 \gamma_2 \phi}{b^2 d^2 k_1} = w_0 - \alpha \phi \tag{5.38}$$

with $w_0 > 0$, $\alpha > 0$ and ϕ implicitly calculated from

$$\frac{\epsilon \gamma_2}{bdk_1} \phi = \frac{1}{1 + K^{-1} \phi^m}.$$
(5.39)

We see that w_0 is the maximum strength value, achieved if the steady-state concentration of repressor proteins is zero. This makes sense because that is when the feedback rate is the most sensitive to a change in the repressor's density. If $w_c > w_0$, the phase space has no bifurcations regardless of ϕ . So we set a minimum value of m that allows bifurcations:

$$m_{min} = \frac{2 + 4\epsilon + 2\epsilon^2}{\epsilon}.$$
(5.40)

For small values of ϵ , we have $m_{min} \approx 2\epsilon^{-1}$, which is a ridiculous number. We also see that the minimum value of m_{min} is 8, achieved for $\epsilon = 1$. Also, for usual parameter values of genetic systems, w is a very slowly decreasing function of ϕ , so it's usually near its maximum value for these systems.

The maximum re-scaled strength is $bdw_0 = m\epsilon$. For low values of m, the Hill feedback gives a maximum w below the optimal value for noise control. The numerical cases shown in figure (6) give m = 45 as the optimal Hill exponent for the case $\epsilon = 0.01$ and m = 3 for the case $\epsilon = 1$. Thus, the system can more easily achieve the optimal noise-control dynamics for similar degradation rates of the maker-molecule and the product-molecule, unlike the ones of mRNA and protein products.

5.4.3 Parameter Estimation

We explore the parameter estimation process on the Hill 3D Goodwin model for three scenarios in order to compare a low noise condition with a high noise condition and also a subcritical regime with a supercritical regime. For that, we fix parameter values apart from the Hill exponent m, as shown in the real value column of table (2). A medium value of $\epsilon = 0.3$ is chosen to ensure a modest lower bound for the possibility of bifurcation into oscillations. We then calculate, based on our analysis, the critical value of m and the value that minimizes the protein noise level at equilibrium. The critical value is $m_c \approx 11.6$ and the minimum noise is achieved with $m_n \approx 0.42$. With these results, we choose: 1) the low noise subcritical regime, far from the bifurcation, has $m_1 = 2$; 2) the high noise subcritical regime, near the bifurcation, has $m_2 = 10$; 3) the supercritical regime near the bifurcation has $m_3 = 13$. For each of the three cases, we take the 40 measurements of mRNA, proteins, and transcription factors from samples of the simulated master equation of the stochastic process. We choose a mesoscopic dynamics with $\Omega = 100$, aiming at a sufficient stability in the behavior of the stochastic samples.

Figures (7), (8), and (9) show the estimation setup for the three scenarios. For the more stable, low-noise case of m = 2, the estimated parameters are fairly accurate and reflect the reality of the data generating dynamics. The high-noise, but still subcritical, case of m = 10 can also be estimated with the deterministic model, but not completely so. The less sensitive parameters from the Hill function, m and K^{-1} , can't be so precisely defined. The supercritical case, for m = 13, features oscillations. Then, the deterministic model has a hard time adjusting to the oscillations in face of the unpredictable noise levels, so the estimation process becomes unreliable.

The three cases yield practically the same equilibrium densities for all three species, but they mainly differ on the oscillatory pattern before the equilibrium, that is more accentuated for higher values of m. These oscillations become sustained once beyond the bifurcation, thus breaking the stable fixed point. Table (3) shows analytically calculated values of equilibrium density and index of dispersion for proteins, with the expected deviation around the mean for the subcritical regimes. The estimated standard deviations underestimate the noise levels calculated with the LNA, even for the case of m = 2. But the estimation is unreliable for noise levels, because it is made with only 1 sampled trajectory. In table (2), however, we see that the magnitude of deviations behaves as expected between the cases, with the lowest estimated value being for the case m = 2 and the highest estimated value for the case of m = 13, the supercritical regime; especially σ_3 , associated with the more abundant species X_3 (transcription factors).

According to the LNA, the protein standard deviation at equilibrium near the bifurcation (m = 10) are ≈ 4 times larger than far from the bifurcation (m = 2) (≈ 16 times the noise levels). According to the estimation of the proxy deviation of the deterministic model, they are ≈ 2.2 times larger.



Figure 7 – Parameter estimation for m = 2, low noise far from bifurcation. Estimation is fairly accurate in this case, with low estimated noise levels. The dynamics is mesoscopic with $\Omega = 100$. Upper Left: The stochastic process with the 40 measurement points, mRNA (X_1) is red, protein (X_2) is green, and transcription factor (X_3) is blue. The deterministic trajectories for the real parameter values are shown in black. Lower Left: The stochastic trajectories are compared with the process for the mean estimated parameter values. Right: Trace-plots of parameters for 4 chains, with warmup shown in gray, featuring $\boldsymbol{\theta} = (b, d, \epsilon, \gamma_1, m, k_1 \epsilon^{-1}, K^{-1})$ and the proxy standard deviations of the statistical model for mRNA and protein. Source: By the author.



Figure 8 – Parameter estimation for m = 10, subcritical regime near the bifurcation. Estimation is less accurate than the first case, especially m and K. Estimated noise levels are considerably higher. The dynamics is mesoscopic with $\Omega = 100$. Upper Left: The stochastic process with the 40 measurement points, mRNA (X_1) is red, protein (X_2) is green, and transcription factor (X_3) is blue. The deterministic trajectories for the real parameter values are shown in black. Lower Left: The stochastic trajectories are compared with the process for the mean estimated parameter values. Right: Trace-plots of parameters for 4 chains, with warmup shown in gray, featuring $\boldsymbol{\theta} = (b, d, \epsilon, \gamma_1, m, k_1 \epsilon^{-1}, K^{-1})$ and the proxy standard deviations of the statistical model for mRNA and protein.

Source: By the author.

Table 2 – **Estimated parameters.** The estimated posterior yields the mean values and standard deviations shown in the table, compared with the real parameter values. The standard deviation of the measurements σ has no real value because it is a proxy of the actual noise levels coming from the stochastic process.

Parameter	(m=2)	(m=10)	(m=13)	Real Value
b	0.84 ± 0.10	0.96 ± 0.19	0.72 ± 0.08	0.95
d	0.81 ± 0.11	0.90 ± 0.19	0.71 ± 0.08	0.82
ϵ	0.29 ± 0.04	0.33 ± 0.07	0.24 ± 0.03	0.30
γ_2	0.57 ± 0.06	0.52 ± 0.09	0.60 ± 0.05	0.55
m	2.12 ± 0.25	5.39 ± 0.97	17.73 ± 2.77	2/10/13
$k_1 \epsilon^{-1}$	0.69 ± 0.10	0.59 ± 0.13	0.62 ± 0.06	0.70
K^{-1}	9.01 ± 0.95	6.15 ± 1.18	2.48 ± 2.24	10.5
σ_1	0.04 ± 0.01	0.14 ± 0.02	0.14 ± 0.02	—
σ_2	0.06 ± 0.01	0.13 ± 0.02	0.16 ± 0.02	—
σ_3	0.12 ± 0.02	0.15 ± 0.02	0.28 ± 0.03	_

Source: By the author.

5.5 Discussion

In this analysis we have advanced the exploration of the 3D Goodwin model, designed under the theory of reaction networks, in the same lines as the more easily solvable 2D model. We determined the equilibrium critical point for a general feedback functional form, in terms of the steady-state feedback strength parameter, and then a



Figure 9 – Parameter estimation for m = 13, low noise far from bifurcation. The estimation is not accurate, especially m and K^{-1} , and the noise levels are the highest and are confounded with the oscillations. The dynamics is mesoscopic with $\Omega = 100$. Upper Left: The stochastic process with the 40 measurement points, mRNA (X_1) is red, protein (X_2) is green, and transcription factor (X_3) is blue. The deterministic trajectories for the real parameter values are shown in black. Lower Left: The stochastic trajectories are compared with the process for the mean estimated parameter values. Right: Trace-plots of parameters for 4 chains, with warmup shown in gray, featuring $\boldsymbol{\theta} = (b, d, \epsilon, \gamma_1, m, k_1 \epsilon^{-1}, K^{-1})$ and the proxy standard deviations of the statistical model for mRNA and protein.

Source: By the author.

Table 3 – **Calculated protein levels and feedback.** We see increasing values of feedback strength at equilibrium, while the protein density remains practically the same. Noise levels are unavailable for the supercritical case, where the LNA assumptions break. As expected, the noise increases as the model approaches the bifurcation. The standard deviation of ξ_2 is calculated as $\sqrt{\phi_2^* D_s}$.

Parameter	m=2	m=10	m=13
w	0.75	3.75	4.87
ϕ_2^*	0.35	0.36	0.36
D_s	0.21	3.32	-
$\sqrt{\langle \xi_2^2 angle}$	0.27	1.1	-

Source: By the author.

minimal value of the Hill exponent that permits the transition to oscillatory behavior for the case of a Hill feedback; the Hill exponent being a measure of nonlinearity in the feedback and also an important parameter relating to the underlying dynamical process happening at a faster time-scale. We then determined the noise behavior of the model by using the LNA expansion on the stochastic description of the model and criticized the model in the context of genetic regulation. We also performed a Bayesian parameter estimation on the 3D model using the deterministic description, exploring its feasibility at different distances from the bifurcation, considering the noise levels and the appearance of oscillations. For the deterministic description of the model, we found an analytical expression for the critical feedback strength at equilibrium, w_c . With the LNA, we found the equilibrium subcritical noise levels through the index of dispersion D_s , in terms of w, and we saw that noise diverges at the critical point. The expression for D_s shows that there is a minimum noise level that in theory can be achieved for regulation setups aiming at suppressing noise, which can be a desirable feature for intracellular processes; an optimal noise-suppressing feedback is not present for the 2D model.

In the context of genetic regulation, the mRNA lifetime can be several orders of magnitude smaller than the proteins lifetime. We found that it can be an issue for oscillatory designs of the Goodwin model, since the minimum Hill exponent permitting oscillations, previously determined to be $m_{min} = 9$ for equal lifetimes,⁵⁵ was found to be $m_m in \approx 2\epsilon^{-1} >> 1$ in this context. In order for Goodwin oscillations to be applicable to gene regulation, we must have similar lifetimes of mRNA and proteins or we should be able to further justify the rise of large Hill exponents, currently seen as the number of binding entities of some sort (be it a phosphorylation, or operator binding, or else).⁵⁶

The divergence of noise at the bifurcation is an interesting way to relate two different analytical approaches. Despite using the deterministic system at the core of its calculations, the LNA is able to provide its own clean expression for critical values as the point of noise divergence. We understand this divergence as due to the divergent sensitivity of the system to perturbations on the critical point, so the noise reverberates accordingly.⁵⁴ The LNA approach is unable to analyze the noise levels of the supercritical oscillatory regime, because it breaks the unique stable steady-state assumption.

The Bayesian parameter estimation process is a promising method for determining reaction network models, and we saw how it can accurately estimate parameters far from the bifurcation. For systems near the bifurcation or in the oscillatory scenario, we acknowledge the need for a better statistical model instead of the simple deterministic description with constant independent Gaussian noise. The LNA itself is a good candidate to more accurately describe the noise structures for these more complex cases.⁵⁷ A LNA statistical model can be numerically inputted even for oscillations, with the benefit of adapting the noise levels according to the dynamical evolution of the system, following the Lyapunov system of equations for the correlation matrix of the model. This better statistical assessment can also be valuable to estimating processes for smaller, more stochastically unstable systems in microscopic conditions ($\Omega \approx 1$), as is often the case for intracellular processes.

6 A UNIFYING MECHANISTIC FRAMEWORK OF EVOLUTIONARY DYNAM-ICS USING REACTION NETWORKS

This work encompasses all the generalization of reaction network theory that is pertinent to the context of eco-evolutionary dynamics and then uses the resulting framework to reinterpret, unify, and extend the standard modeling theories of the field. Many comprehensive examples are worked in order to illustrate the consistency and robustness of the framework. We consider this to be the main result of this thesis and it features both the methodology and its application to theory development and specific biological problems. For example, it discusses the interpretation of the Price equation for trait evolution, a generalization and connection between evolutionary game theory and adaptive dynamics, and also provides a critique to modeling approaches for the ecological problem of the evolution of parental investment.

6.1 Abstract

Evolutionary game theory is a successful modeling framework for evolutionary dynamics. It is able to simulate reproduction that is shaped by an ecological environment where individual success depends quantitatively on the behaviors of others, and populations can evolve in time through the replicator equation. But game theory's proportion-based results provide an incomplete representation of population dynamics which is fundamentally dependent on densities. In this work, I use reaction network theory, an established framework for modeling populations in general, as a generalization of evolutionary game theory that defines interactions at the stochastic level of individuals, generating a density-based dynamics. Using the hawk-dove game as an example, I highlight the consequent limitations of evolutionary game theory and propose network-based solutions. In particular, games can easily lead to divergent densities, and the solution to this problem is built upon the properties of the eco-evolutionary process. Then, as a unification of frameworks, I explore the integration between evolutionary game theory, replicator dynamics, the Price equation, and adaptive dynamics, and work out examples to better understand the framework. In special, I present a null model of parental care evolution as an example of trait evolution.

Keywords: Eco-evolution, Replicator Dynamics, Adaptive Dynamics, Price Equation, Density-Dependent Selection, Markov Jump Process.

6.2 Introduction

In a general sense, evolutionary processes are emergent dynamics at the level of populations that happen as a consequence of how interactions at an individual level shape the characteristics of evolving units. Many forces are simultaneously at play mostly in nonlinear and non-intuitive ways, and the quantitative treatment of those forces often is necessary to the description of the many possible outcomes and conditions from which they arise. The understanding of evolutionary processes through the lens of mathematical modeling goes a long way in helping us to formalize concepts and sort out these complex scenarios. However, mathematical models of evolution traditionally rely on a more abstract grounding, without direct connection to the individual level interactions determining the birth and death of evolving units.² To the models, concepts as reproduction, mutation, and selection are important only by how they define characteristics of individuals and populations and how they determine and are determined by births and deaths. In order to understand how the actual environment of evolving populations helps in determining their evolutionary fate, mathematical models must account for how evolving units are created and destroyed differentially. That is depending on their adaptive capacities and the environments made relevant through their individual interactions. The modeling task in this case is to try and capture the most relevant interactions determining birth and death.

Adapting from the discussion in Dieckmann et. al., 2006.⁵⁸ these are some requirements that a grounded framework of evolutionary dynamics should intend to satisfy: (1) Mechanistic justifications. All forces considered in shaping evolution at the population level must arise from explicitly defined processes happening at the individual level.⁴ (2) Explicit eco-evolution. The ecological environment directly determines adaptation through births and deaths, and models of evolution at least imply what they consider as the relevant factors of this environment; these should be made explicit.⁵⁹ (3) Density-dependence. More fundamental than proportion (or frequency) dependence, the outcome of evolution depends on the actual size of populations, and even when infinite population limits are considered, evolution is sensible to their densities.⁶⁰ (4) Stochastic foundation. When assuming average individual interactions instead of tracking every individual trajectory, all deterministic models must arise from a limit of a stochastic process, and this process should be available for the cases where the deterministic assumptions break.⁶¹ In addition to these requirements, I consider what is not necessarily a requirement, but a desirable goal: design-level modeling. To improve communication between research fields and enhance model conception, it is desirable to encapsulate the mathematical details behind a design-level interface, where one can think and talk about the processes without the need to understand or worry about what makes the model work in the details.

Evolutionary game theory (EGT) is a traditionally successful modeling framework for eco-evolutionary dynamics, and its methods are evolving and have a great potential to compose a grounded framework of evolutionary dynamics.^{62,63} Models in EGT are built in two steps, first define a match-maker game delivering payoffs to individuals as results of their interactions (ecological scale), then use payoffs in a replicator dynamics as modulators of reproductive rates (evolutionary scale). This process enables us to investigate the relative success and stability of strategies in face of a defined strategy pool. However, standard EGT is rooted in some detrimental limitations. (1) It does not have a clear unique connection with an individual-level stochastic foundation. (2) The replicator dynamics is based on proportions and often assumes the equilibrium of population densities, which are the actual quantities shaped by births and deaths, thus being mechanistically relevant to the process and also intrinsically tied to it. (3) The rigid match-making of games is not dynamical in nature, but based on a specific sorting of all individuals at certain points in time, usually pairwise sorting.

Another example is how the Price equation is such a source of misunderstanding in the way that trait evolution is supposed to play out, because it does not make it clear how we can interpret it under the light of the mentioned requirements.⁶⁴ In addition, it is important to understand how the many different techniques can relate to each other inside an unifying picture. Page and Nowak (2002)⁶⁵ explore the mathematical correspondences between deterministic equations of evolutionary dynamics, promoting unification at this level

Reaction networks originated in the modeling of chemical reactions, but the framework itself is a powerful abstract reasoning system suited for all kinds of collective dynamics.^{12,66} They spread to be widely used across many fields of biology, such as molecular biology³⁴ and epidemiology.⁶⁷. Reaction networks define mechanisms of local interactions, based on encounters, that collectively generate a population-level dynamics, in a relational manner. By treating network dynamics as a set of interaction rules, the theory functions as a smooth connection between verbal design-level reasoning with quantitative nonlinear outputs. It is much like EGT, but reaction networks are natively dynamical, and cannot only correspond to EGT, but also effortlessly address its shortcomings. It has been used in this context before; Veloz et al. $(2014)^{10}$ offered a reaction network model for payoff matrices, by modeling network units as decisions instead of individuals. Taylor and Nowak $(2006)^{68}$ also used for EGT ideas found in reaction networks. The use of the mass-action law as a tool is also standard in population dynamics.

In this work, I introduce reaction networks adapted to the context of evolutionary dynamics, then show how both modeling steps of EGT can be understood as particular cases of two reaction network models, one at an ecological scale, and another at an evolutionary scale. I illustrate how we simply cannot sustain the assumption of densities being independent of replicator dynamics, by analyzing the density instability of the Hawk-Dove game; then I proceed to show how we can correct the problem with reaction networks and also gain insights into the underlying evolutionary processes.

6.3 Methods

6.3.1 Reaction Networks for Evolutionary Dynamics

The central idea is to represent a macroscopic population system through a set of relational rules of interaction described at the microscopic level of individuals. The fact that interactions have a relational nature holds a major correspondence with evolutionary theory: the theory leaves open the specification of what an individual is, what matters is how it relates, not its identity. For example, it does not matter what a replicator is, as long as it replicates. Thus, as with evolutionary processes, any unit, or coherent collection of subunits, can be subjected to the same theory, as long as its external relations are preserved. What reaction networks do is to describe all relevant interactions available to these (abstract) individuals with the aim to transform those interactions into populationlevel dynamics. For that, we require that populations be defined as collections of *identical* individuals, in the sense that they interact in the same ways. So, if an event results in an individual changing its relational nature, it must become part of another population, in the same way as chemical components can transform themselves with reactions or individuals transform from susceptible to infected in epidemiology models. As a result, the state of the system is not about individual trajectories, it is about the sizes of populations, and the interactions are considered through probabilistic arguments.

For the purposes of this work, I will define reaction networks in a simple intuitive manner, although there are more mathematically sound definitions.⁸ I also maintain a terminology that is more suited to evolutionary theory. Consider a set of N populations with individuals represented as X_i , with i = 1, 2..., N. They are each composed of n_i identical individuals that are distributed inside a space with size Ω (n_i 's are natural numbers). The individuals interact with surrounding peers, changing their numbers according to a set of R interactions (or reactions) defined as

$$\sum_{i=1}^{N} s_{ir} X_i \to \sum_{i=1}^{N} s'_{ir} X_i, \quad r = 1, 2...R.$$
(6.1)

This means that s_{ir} elements of species X_i are needed for the occurrence of interaction r, that in turn results in s'_{ir} elements of population X_i emerging from the interaction. The interaction r then changes the amount of X_i by $(s'_{ir} - s_{ir})$. How interactions physically happen is not determined by the network, but is a part of the model specification and its assumptions. As an example, a simple interaction representing replication would be $X_1 \rightarrow 2X_1$. Another example, an interaction representing a competition between two individuals resulting in the death of one of them would be $X_1 + X_2 \rightarrow X_1$.

The state of the network is determined solely by individual count-numbers, n_i , combined into a state vector $\mathbf{n} = (n_1, n_2, ..., n_N)$. We can view the state description by \mathbf{n} as the information about how many individuals of each type are present in the

system at a given moment, and it is sufficient to uniquely determine the system. Each interaction r, when it happens, provokes a state transition where each n_i is transformed into $n_i + (s'_{ir} - s_{ir})$. State transitions are the pace at which the system can evolve in time as a result of interactions. As an example, consider the network formed by two populations, X_1 and X_2 . At a given moment, the network state is $\mathbf{n} = (n_1, n_2)$. If a competition interaction $X_1 + X_2 \to X_1$ happens in the system, it undergoes a transition to the new state $\mathbf{n'} = (n_1, n_2 - 1)$, meaning that a X_2 individual was destroyed. If, then, a replication interaction $X_1 \to 2X_1$ happens, the system undergoes another state transition to $\mathbf{n''} = (n_1 + 1, n_2 - 1)$, because a X_1 individual was created.

6.3.1.1 Stochastic Kinetics

With the network defined as above, we have no information about how often interactions will occur, only the static description of the allowed transitions of the system. For that, we add to the reaction network a *kinetic description* by specifying the rate at which interactions will take place. Assuming a kinetic of continuous time passage, we can introduce transition rates for each interaction, W_r . Since we do not know individual's trajectories, this transition rate means that we should expect a probability of $W_r dt$ for interaction r to occur during an infinitesimal interval of time dt. Knowledge about the detailed wanderings of individuals is not required if we can guarantee a transition rate W_r .

Apart from interactions, individuals will have trajectories independent of each other. To understand the derivation of transition rates, it is useful to divide them into two contributions. (A) The chance that required individuals will encounter themselves within an interaction range σ_r , and (B) the chance that, given the encounter A, those individuals actually engage in the interaction. It is also default to assume that interactions are given with σ_r that is constant in time. As an example, consider the following picture of interaction $X_1 + X_2 \rightarrow X_1$: X_1 is a predator consuming a prey X_2 . Then, A represents the chances of the predator finding itself close enough to the prey (where enough is given by the interaction range), and B represents the chances of the predator actually finding and capturing the nearby prey.

The reasoning above results in a form of stochastic mass-action law.¹³ For a general interaction r composing a network as defined, the transition rate is given by

$$W_r(\boldsymbol{n},\Omega)dt = (k_r dt)\Omega \prod_i \frac{n_i!}{(n_i - s_{ir})!\Omega^{s_{ir}}}.$$
(6.2)

I will call the function $k_r(n)$ an interaction rate. The term $k_r dt$ encapsulates the contribution B above, but is not technically equivalent to it, because k_r absorbs the interaction range parameter from A:

$$k_r = \frac{k_r^0}{\sigma_r^{1-\sum_i s_{ir}} \prod_i s_{ir}!},$$
(6.3)

where k_r^0 is the contribution from B. The other terms describe the rest of contribution A, and they depend only on the properties of the interaction itself. Interaction rates k_r are constant rates by default or their form as a function of the state \boldsymbol{n} is part of the system specification. We represent it over the arrow of an interaction. So, taking again the example of the competition interaction, now with the k_r represented, we have $X_1 + X_2 \xrightarrow{k_r} X_1$, with transition rate given by $W_r = k_r n_1 n_2 / \Omega$.

Since W_r depends only on the current state \mathbf{n} , reaction networks behave as a Markov jump process with stochastic transition rates given by $W_r(\mathbf{n})$, considering that transition rates between states not linked by reactions are equal to zero. Markov jump processes are stochastic systems with discrete state space and continuous time that obey the Markov property saying that transitions depend only on the current state. They evolve according to a master equation for the probability density of states, given by

$$\frac{\partial \Pi(\boldsymbol{n},t)}{\partial t} = \sum_{r} \left(W_r(\boldsymbol{n}-\boldsymbol{s}_r)\Pi(\boldsymbol{n}-\boldsymbol{s}_r,t) - W_r(\boldsymbol{n})\Pi(\boldsymbol{n},t) \right), \tag{6.4}$$

where the vector $\mathbf{s}_{\mathbf{r}}$ is such that $(\mathbf{s}_{\mathbf{r}})_i = (s'_{ir} - s_{ir})$. This equation yields a complete and exact stochastic evolution of the reaction network as a Markov jump process.⁶⁹ Together with an initial condition, it can (1) be solved analytically for a narrow range of mostly linear systems, or it can (2) be simulated with a stochastic simulation algorithm (also called Gillespie's algorithm), or it can (3) be approximated by the Kramers-Moyal expansion as a diffusion approximation resulting in a Fokker-Planck equation (or the associated Langevin stochastic differential equation), (4) be approximated by moment-closure techniques, or (5) be approximated by a systematic system size expansion over Ω having as the first stochastic order what is called a linear noise approximation.¹³ The later expansion also gives a connection to the deterministic description, as its zeroth order is a deterministic system of differential equations. To establish a connection to the deterministic limit, for a finite time domain, the system-size expansion assumes we can write the expansion

$$W_r(\boldsymbol{n},\Omega) = \Omega \sum_{i=0}^{\infty} \frac{W_r^{(i)}(\boldsymbol{n}/\Omega)}{\Omega^i}.$$
(6.5)

This forms a class of transition rates that are well-defined in terms of densities, and is always valid for mass-action rates. For non-mass-action rates, when k_r is no longer constant, it is sufficient to define $k_r(\boldsymbol{n}, \Omega) = k_r(\boldsymbol{n}/\Omega)$. For the following, we assume these requirements.

6.3.1.2 Deterministic Kinetics

We can achieve a deterministic description from the stochastic kinetics by performing a limit of large system in a way that approximates the count-numbers n_i by a continuous density of individuals. Following the system-size expansion, we write the state of the network in terms of a continuous density $n_i = \Omega \eta_i + \sqrt{\Omega}$. Then, for the limit of infinite populations and infinite size, we have

$$\eta_i = \lim_{\Omega, n_i \to \infty} \frac{n_i}{\Omega}.$$
(6.6)

The approximation is of $\mathcal{O}(\Omega^{-1})$ and is exact in the limit.^{13,70,71} By performing this limit over the stochastic kinetic for n_i given by the master equation, using the system size expansion, we achieve deterministic kinetics for η_i given by a set of differential equations for the time-evolution of the densities vector, $\boldsymbol{\eta}$:

$$\frac{d\eta_i}{dt} = \sum_r (s'_{ir} - s_{ir}) W^d_r(\boldsymbol{\eta}), \tag{6.7}$$

where $W_r^d(\boldsymbol{\eta})$ are the deterministic transition rates for densities, given by

$$W_r^d(\boldsymbol{\eta}) = \lim_{\Omega, n_i \to \infty} \frac{W_r(\boldsymbol{n})}{\Omega} = k_r \prod_i \eta_i^{s_{ir}}.$$
(6.8)

Since the deterministic kinetics are a result of a limit over the system size Ω , this size is no longer a parameter of the model. As \boldsymbol{n} and Ω increase in the stochastic kinetics, the noise around the deterministic trajectories gets smaller as the continuous approximation gets more exact.

I now work an example to show the transformation from stochastic to deterministic descriptions. Consider a system of one population X. Individuals reproduce as exually with a constant rate $k_1 = \omega$, they also die with a constant rate $k_2 = \mu$, and they fight for space and resources, resulting in death, with a constant rate $k_3 = \gamma$. The reaction network that represents this system is

$$\begin{array}{c} X \xrightarrow{\omega} 2X \\ X \xrightarrow{\mu} \emptyset \end{array}$$
$$2X \xrightarrow{\gamma} X. \tag{6.9}$$

This is the design-level of reaction networks, the model is now fixed. The mathematical description that comes next is already determined by the system specification. The master equation giving the evolution of the probability density of states n is

$$\frac{\partial \Pi(n,t)}{\partial t} = \omega(n-1)\Pi(n-1,t) + \mu(n+1)\Pi(n+1,t) + \frac{\gamma}{\Omega}(n+1)n\Pi(n+1,t) - \left(\alpha n + \mu n + \frac{\gamma n(n-1)}{\Omega}\right)\Pi(n,t).$$
(6.10)

The transition rates are $W_1 = \omega n$, $W_2 = \mu n$, and $W_3 = \gamma n(n-1)/\Omega$. The differential equation over the density η obtained as deterministic kinetics for this network is then

$$\frac{d\eta}{dt} = \omega\eta - \mu\eta - \gamma\eta^2. \tag{6.11}$$

This equation is equivalent to the logistic model $\dot{\eta} = \eta r(1 - \eta/K)$, and we derive the growth rate $r = (\omega - \mu)$ and carrying capacity $K = (\omega - \mu)/\gamma$ as functions of the fundamental rates of birth, death, and competition. We see that the system we are modeling of replication, death, and competition is described in the deterministic limit of infinite system by a logistic model.

6.3.1.3 Modeling transition rates

The two sources of contribution to transition rates are important for their design. The shape of the encounter rates is determined by the network structure, by how the interactions themselves are designed. What determines them is the availability of individuals needed for the interactions, the reactant complexes in chemical terminology. The other contribution is encoded into the interaction rates, and it has a major role in our interpretation of them. In the chemical domain, they are usually constant contributions, especially because one needs to provide physical justifications for more complex shapes, such as time-scale separations and quasi steady state approximations.²⁷ But for population dynamics, interaction rates are designed as probabilistic functions relating to the real interactions as a sort of activation function. They regulate the intensity by which possible encounters between individuals result in the outcome of their interactions. This is the same modeling role as payoff functions have in influencing birth and death rates in EGT, and k_r can also be subjected to a game model that determines its functional shape.

Take as example a mating interaction of the form $X_1 + X_2 \xrightarrow{k_r} b(X_1 + X_2)$. Suppose X_1 are female individuals and X_2 are males, and they interact to produce (b-1) offspring, with b > 1. The interaction rate for this is $W_r = k_r n_1 n_2$, it is independent of the terms in the outcome, the products, that enter only as part of coefficients in the equations describing the evolution of states. The $n_1 n_2$ factor is an encounter rate contribution, and is present by the structure of the interaction. Now, if we know from the system that this particular mating interaction is somehow affected by the ratio of males to females, being more or less likely to happen as a function of this ratio, we must have $k_r = k_r(n_2/n_1)$. It might be inversely proportional to it, then $k_r = \lambda n_1/n_2$. The value of this proportionality constant λ is particular to the system, and in principle can be further modeled as a function of the environment and the species or even estimated as a part of a fit to data.¹³

One particular family of interactions that is especially important to evolutionary dynamics is composed of interactions of type $X \xrightarrow{k_r}$?, involving a single individual. Simple births and deaths are interactions of this type, that are not exactly interactions; are more properly *self-interactions*. These do not need an encounter to happen, so they are always susceptible to occur. The reaction rates are $W_r = k_r n$, and k_r has a more definite interpretation of a frequency, with $1/k_r$ being an average time of occurrence. Then, for example, if an individual X has an average life-time of T before spontaneously dying, it is subject to the death interaction $X \xrightarrow{1/T} \emptyset$. We interpret this rate as a constant propensity to die that results on average life-times of T units of time; since the Markov jump process produces Poisson distributed independent interactions, this life-time will have a standard deviation of T, and the population in the deterministic limit decays to 1/e of its initial size after T. In the same way, if an individual X reproduces itself in cycles happening at a frequency of ω per unit of time, it is subject to the birth interaction $X \xrightarrow{\omega} 2X$. By default, ω is also a constant propensity, but as a function of the state, $\omega = \omega(\mathbf{n})$, it encodes the birth propensities as dependent of the surrounding population dynamics. By that, we design interactions directly influencing what we may call the fitness or reproductive success of individuals.

6.4 Results

6.4.1 Reaction networks for evolutionary game theory

I first define two more features that are compatible with reaction networks in the way we defined. (1) Interactions can produce quantities that are not part of the system state as individuals. (2) In a hierarchical manner, we can design models for interaction rates, shaping rates by the results of a game at the ecological level; this game can itself be a reaction network, one that is simulating rates for a network at the evolutionary level.

I generalize equation (6.1) to

$$\sum_{i=1}^{N} s_{ir} X_i \xrightarrow{k_r} \sum_{i=1}^{N} s'_{ir} X_i + c_r \Lambda, \qquad (6.12)$$

where Λ can be any kind of quantity with an amount λ and c_r is how much of it is produced by interaction r. Since Λ is never required for interactions, λ does not strictly need to be part of \boldsymbol{n} , λ can be a real number, and transition rates do not depend on λ . A produced quantity is passively accumulated as the system evolves in time. Its stochastic description generalizes equation (6.4) to

$$\frac{\partial \Pi(\boldsymbol{n},t,\lambda)}{\partial t} = \sum_{r} \left(W_{r}(\boldsymbol{n}-\boldsymbol{s}_{r})\Pi(\boldsymbol{n}-\boldsymbol{s}_{r},t,\lambda-c_{r}) - W_{r}(\boldsymbol{n})\Pi(\boldsymbol{n},t,\lambda) \right).$$
(6.13)

Thus, in the deterministic limit we defined, the production of Λ is given by

$$\frac{d\lambda}{dt} = \sum_{r} c_r k_r \prod_i \eta_i^{s_{ir}}.$$
(6.14)

In the same way, more than one quantity can be produced by a network, and λ can be generalized to a vector $\boldsymbol{\lambda}$.

I then define a game network, using the example of a simple birth interaction $X \xrightarrow{\omega} 2X$. This interaction is a part of a population model, and we wish to build a model for the interaction rate ω , in order to define it as a function of the state \boldsymbol{n} in accordance

with our design intentions to represent a system of interest. Whenever we define a function $\omega(\mathbf{n})$, we are implicitly designing a model for it; the simplest case being a constant, that is a modeling choice nonetheless. The process is to model $\omega(\mathbf{n})$ with another reaction network, using probabilistic argumentation to connect the game network to $\omega(\mathbf{n})$. The aim is to define mean-field interactions shaping the propensities of the interaction $X \xrightarrow{\omega} 2X$ to happen as a function of the surrounding environment represented by the state \mathbf{n} . This is what EGT implicitly accomplishes, with payoffs acting as state-dependent propensities to breed and die. The quantities Λ produced by interactions are central to this, as they can act as a measure of interaction intensity. Suppose Λ is produced in the game network for ω , then $\omega \propto d\lambda/dt$ is a model for the interactions that drive the occurrence of the birth $X \xrightarrow{\omega} 2X$, interpreted as increasing the probability of its occurrence. In game theory terminology, Λ is payoff for the birth interaction.

The stochastic kinetics of reaction networks describe the network state at the level of individual counts, represented here by \boldsymbol{n} . The infinite size limit of deterministic trajectories equivalently represents population densities, $\boldsymbol{\eta}$. But the characteristic of replicator dynamics is to represent birth and death population dynamics using relative proportions, defined as $p_i = n_i / \sum_i n_i$, with deterministic counterpart $\rho_i = \eta_i / \sum_i \eta_i$.⁷² This level of representation loses information on the total sizes of populations $\sum_i n_i$ (or $\sum_i \eta_i$), and only a subset of systems is by any means reducible to relative proportions. Fixing densities at equilibrium does not work as an assumption, since we cannot control density-level dynamics beforehand. Therefore, replicator dynamics is an inherently incomplete representation of populations.

6.4.1.1 Evolutionary-level replicator dynamics

The replicator dynamics models the evolutionary level taking the game as input, and can be viewed as a network. To derive it as such, consider a network of N populations named X_i , i = 1, 2, ...N, with count-numbers vector $\mathbf{n} = (n_1, n_2...n_N)$. The X_i individuals self-reproduce with rate $\omega_i(\mathbf{n})$ and die with rate $\mu_i(\mathbf{n})$. At birth, each X_i mutates into X_j with probability $q_{ij}(\mathbf{n})$ (a fraction q_{ij} of offspring from X_i mutates into X_j); then the probability of not mutating is $q_{ii}(\mathbf{n}) = 1 - \sum_j q_{ij}$. These result in the network

$$X_i \xrightarrow{q_{ij}\omega_i} X_i + X_j$$

$$X_i \xrightarrow{\mu_i} \emptyset. \tag{6.15}$$

In the first interaction, X_i produces X_j , and for every population X_i there are N types of these birth interactions; one for each type of individual produced. In the second, X_i transforms into nothing, it dies. This system yields a stochastic description resulting in the master equation (6.4), but we focus here on the deterministic limit. The limit (6.6)
results in the following system for the densities η_i :

$$\frac{d\eta_i}{dt} = -\eta_i \mu_i(\boldsymbol{\eta}) + \sum_j \eta_j \omega_j(\boldsymbol{\eta}) q_{ji}(\boldsymbol{\eta}).$$
(6.16)

By expressing this evolution in terms of proportions, we have

$$\frac{d\rho_i}{dt} = -\rho_i \mu_i(\boldsymbol{\eta}) + \sum_j \rho_j \omega_j(\boldsymbol{\eta}) q_{ji}(\boldsymbol{\eta}) - \rho_i \sum_j \rho_j(\omega_j(\boldsymbol{\eta}) - \mu_j(\boldsymbol{\eta})).$$
(6.17)

If the system depends on densities only through proportions (a somewhat strict restriction), we have $\omega_i(\boldsymbol{\eta}) = \omega_i(\boldsymbol{\rho})$, and the same for the other rates; then, the system is closed on the proportions and we have a standard replicator-mutator equation. If, as a more relaxing assumption, all rates are of the form $\omega_i(\boldsymbol{\eta}) = g(\sum_i \eta_i)\omega_i(\boldsymbol{\rho})$ for any function $g(\sum_i \eta_i)$ of the population size that is the same for all rates, then the phase portrait of the system can still be obtained at the level of proportions, but not its time evolution, that depends on g.

If we discard the possibility of mutations (all $q_{ij} = 0$ for different populations and $q_{ii} = 1$), the equation (6.16) for densities becomes

$$\frac{d\eta_i}{dt} = \eta_i(\omega_i - \mu_i) = \eta_i F_i \tag{6.18}$$

if we define a growth function $F_i = (\omega_i - \mu_i)$. And from that we arrive at the particular replicator equation, that has the general form:

$$\frac{d\rho_i}{dt} = \rho_i \left((\omega_i - \overline{\omega}) - (\mu_i - \overline{\mu}) \right) = \rho_i (F_i - \overline{F})., \tag{6.19}$$

with $\overline{\omega} = \sum_i \rho_i \omega_i$ and equivalently for $\overline{\mu}$. This is a replicator equation with F_i interpreted as an abstracted measure of fitness of populations. Equation (6.19) does not impose any particular restrictions on the functions F_i , and holds for any network of this kind, possibly resulting in different forms of F_i . Equations (6.18) and (6.19) highlight that densities are driven by F_i while proportions are driven by how F_i deviates from its mean among the populations.

Not only birth and death, but a baseline competition is also fundamental at the evolutionary scale, because it encodes the indirect dispute for limited resources; a dispute that fills the important role of containing growth by being the primary consequence of resource scarcity. These forces are absent from the standard replicator dynamics and, with competition, we can instead define a logistic form of replicator equation. Consider the following network, as a general form of (6.15):

$$\begin{array}{c} X_i \xrightarrow{q_{ij}\omega_i} X_i + X_j \\ \\ X_i \xrightarrow{\mu_i} \emptyset. \end{array}$$

$$X_i + X_j \xrightarrow{\gamma_{ij}} X_i. \tag{6.20}$$

Where γ_{ij} are competition rates. Defining $f_i = \omega_i - \mu_i$ as an intrinsic growth rate, this network yields growth rates of the form $F_i = f_i - \sum_j \gamma_{ji} \eta_j$. A strong sign of the limitations of the proportions-only description, the network results in an explicitly density-dependent replicator equation for the proportions:

$$\frac{d\rho_i}{dt} = \rho_i(F_i - \overline{F}) = \rho_i(f_i - \sum_j \gamma_{ji}\eta_j - \sum_j \rho_j f_j + \sum_j \eta_j \sum_k \gamma_{jk}\rho_k).$$
(6.21)

The densities follow $\dot{\eta}_i = \eta_i F_i$ with the new F_i . With competition, the replicator dynamics yields a more complex trajectory. But if we assume a constant baseline rate of competition between all individuals, $\gamma_{ij} = \gamma$, the competition term of F_i becomes the same for every population, a selective pressure of $\gamma \sum_j \eta_j$, not affecting the dynamics of proportions. For such case, densities still depend on γ , but the proportions follow the dynamics $\dot{\rho}_i = \rho_i (f_i - \overline{f})$, the same as before. However, the growth of densities stops whenever $f_i = \gamma \sum_j \eta_j$. This happens when f_i is the same for all populations, and it coincides with the equilibrium of proportions, given by the standard replicator dynamics. This is a means of modeling forces that successfully limit population size and allow evolution to occur as we expect at the level of densities, but without changing the replicator dynamics.

6.4.1.2 Ecological-level games

The actual game of EGT is a deterministic method for modeling interaction rates as a mean field of interactions that shape the functional forms of birth and death rates in the replicator dynamics. It is a model of propensities for birth and death in which individuals interact without changing their numbers. They can also be defined as reaction networks. For this, interactions output payoff quantities instead of creating and destroying individuals.

Suppose then the evolutionary-level reaction network of N populations X_i that includes birth and death interactions $X_i \xrightarrow{w_i} 2X_i$ and $X_i \xrightarrow{\mu_i} \emptyset$ with every $w_i(\mathbf{n})$ and $\mu_i(\mathbf{n})$ as state dependent functions, and possibly competition with a constant rate γ . I define a game for the rates ω_i and μ_i as the following: no individuals are born or die, and the result of interactions is the production of quantities that function as propensities to reproduce or die, payoffs; then, the interaction rates we are modeling will be proportional to the average payoffs per individual, gathered during the time span of the game, T. The interactions considered in the game are ecological situations that continuously happen during the lives of individuals in that environment. Naturally, each different behavioral strategy adopted by individuals defines a different relational population. The general standard form of game interactions is

$$\sum_{i=1}^{N} s_{ir} X_i \xrightarrow{k_r} \sum_{i=1}^{N} s_{ir} X_i + \sum_{i=1}^{N} \left(c_{ir}^G G_i + c_{ir}^L L_i \right).$$

$$(6.22)$$

The result of the interactions is the production of the payoff quantities G_i and L_i for each population. These represent gain payoff values of g_i as propensity of birth rates and loss

payoff values l_i as propensity of death rates. The payoffs are modeled as the gains and losses per participating individual, for each population, as results of each interaction. Their time evolution is given by equation (6.14). The constants c_{ir}^G and c_{ir}^L are the amounts of payoff produced by each interaction. Thus, the birth and death rates are given as

$$\omega_i(\boldsymbol{\eta}, T) = \frac{1}{\eta_i T} \int_0^T \frac{dg_i}{dt} dt$$
$$\mu_i(\boldsymbol{\eta}, T) = \frac{1}{\eta_i T} \int_0^T \frac{dl_i}{dt} dt.$$
(6.23)

When the game models a non-changing mean environment, payoffs are not functions of the time t, and the interaction rates reduce to

$$\omega_i = \frac{1}{\eta_i} \sum_r k_r \prod_i \eta_i^{s_{ir}} c_{ir}^G,$$

$$\mu_i = \frac{1}{\eta_i} \sum_r k_r \prod_i \eta_i^{s_{ir}} c_{ir}^L.$$
 (6.24)

Consider then the game of two populations with pair-wise interactions outputting a general payoff matrix

$$\begin{array}{cccc}
X_1 & X_2 \\
X_1 & \left(\begin{array}{ccc} (a^G - a^L) & (b^G - b^L) \\
X_2 & \left(c^G - c^L \right) & (d^G - d^L) \end{array} \right).
\end{array}$$
(6.25)

This matrix means that whenever there is an interaction of type $X_1 + X_1$, each X_1 involved gets $(a^G - a^L)$, and the same goes for $X_2 + X_2$ interactions resulting in a payoff $(d^G - d^L)$. For $X_1 + X_2$ interactions, the X_1 individual gets $(b^G - b^L)$ while X_2 gets $(c^G - c^L)$. The superscripts G and L mean gained and lost payoff, that contribute respectively to birth and death rates.

The network for this payoff matrix is

$$2X_1 \xrightarrow{k} 2X_1 + a^G G_1 + a^L L_1$$

$$X_1 + X_2 \xrightarrow{k} X_1 + X_2 + b^G G_1 + b^L L_1 + c^G G_2 + c^L L_2$$

$$2X_2 \xrightarrow{k} 2X_2 d^G G_2 + d^L L_2.$$
(6.26)

For this game, all interaction rates k are equal and there are only interactions composed of two individuals; however, these are not necessary constraints for reaction networks. Then,

the payoffs are time-independent, so we are able to obtain instantaneous birth and death rate estimates

$$\omega_1 = k(\eta_1 + \eta_2)(a^G \rho_1 + b^G \rho_2)$$

$$\mu_1 = k(\eta_1 + \eta_2)(a^L \rho_1 + b^L \rho_2),$$
(6.27)

and analogously for X_2 . Neglecting mutations in the evolutionary scale, from equation (6.19), this yields a replicator dynamics that has the same phase portrait as yielded by EGT. There is one difference, though: the time-evolution of both populations is proportional to the factor $(\eta_1 + \eta_2)$, the total density. This results in a dynamical evolution that is affected by densities, not just proportions.

Consider a population that, under the rule of a standard replicator birth-death network at the evolutionary scale, evolves according to game-shaped rates of equation (6.27). This population has an artificially unstable density growth. Whenever $\omega_i > \mu_i$, densities tend to grow indefinitely, and whenever $\omega_i < \mu_i$, densities tend to decrease indefinitely. For the case of a Hawk-Dove game, $\omega_i > \mu_i$ always for both Hawks and Doves, so densities will rapidly diverge.

If we instead consider the logistic evolutionary scale, given by a network with a constant baseline competition with rate γ , it is still not sufficient to regulate population growth in this model. We define the average intrinsic growth function for proportions as

$$\overline{\phi(\boldsymbol{\rho})} = k \Big(\rho_1 \Big((a^G - a^L) \rho_1 + (b^G - b^L) \rho_2 \Big) + \rho_2 \Big((c^G - c^L) \rho_1 + (d^G - d^L) \rho_2 \Big) \Big), \quad (6.28)$$

with $\phi_i = f_i/(\eta_1 + \eta_2)$. Then, the total density $(\eta_1 + \eta_2)$ is at a finite equilibrium when the following condition is met:

$$\overline{\phi(\boldsymbol{\rho})^{eq}} = \gamma. \tag{6.29}$$

This is an artificially narrow condition, indicating an inconsistency in the model formulation. The factor $\overline{\phi(\boldsymbol{\rho})^{eq}}$ is determined for a given game independently of densities or γ . If $\overline{\phi(\boldsymbol{\rho})^{eq}} < \gamma$, the equilibrium total density is zero, and if $\overline{\phi(\boldsymbol{\rho})^{eq}} > \gamma$, it still diverges.

This problem points to the issue that the evolutionary birth and death rates, ω_i and μ_i , must also include the influence from background births and deaths ω_0 and μ_0 that are independent of the standard encounters modeled by the payoff matrix. In theory, the game payoffs are supposed to represent propensities deviating from this background that encompasses the influences from other activities and baseline fertility/lifespan of individuals.⁶³ This means for us to add the following set of interactions to the standard game in (6.26):

$$X_i \xrightarrow{\omega_0} X_i + G_i \tag{6.30}$$
$$X_i \xrightarrow{\mu_0} X_i + L_i,$$

assuming, for simplicity, the background to be constant and the same for all populations (and this is mathematically equivalent to making $\omega_i \rightarrow \omega_0 + \omega_i$ at the evolutionary network). Then, this inclusion, in the same way as the competition interaction, does not alter the replicator dynamics. Instead of equation (6.29) being needed, the equilibrium total density is now well-defined and given by

$$\left(\eta_1 + \eta_2\right)_{eq} = \frac{(\omega_0 - \mu_0)}{(\gamma - \overline{\phi(\boldsymbol{\rho})^{eq}})},\tag{6.31}$$

and it is a simple measure of total population size of an ecological system of this kind at equilibrium. But it demands that $\overline{\phi(\boldsymbol{\rho})^{eq}} < \gamma$ at the equilibrium of proportions, otherwise populations would be on average too successful to be restrained by an intensity γ of competition.

6.4.2 Stochastic Models

Consider the logistic network

$$\begin{array}{c} X_i \xrightarrow{\omega_i} 2X_i \\ X_i \xrightarrow{\mu_i} \emptyset \end{array}$$

$$X_i + X_j \xrightarrow{\gamma} X_i, \tag{6.32}$$

with i = 1, 2, ...N and $\omega_i(\boldsymbol{n})$, $\mu_i(\boldsymbol{n})$, and $\gamma(\boldsymbol{n})$ being state dependent. This is the logistic evolutionary-scale network. Its stochastic evolution, by equation (6.4), is

$$\frac{\partial \Pi(\boldsymbol{n},t)}{\partial t} = \sum_{i=1}^{N} \left((n_i - 1)\omega_i^{-}\Pi^{-} + (n_i + 1)\mu_i^{+}\Pi^{+} + (n_i + 1)\gamma^{+} \sum_{j=1}^{N} (n_j + 1 - \delta_{ij})\Pi^{+} - n_i(\omega_i + \mu_i + \gamma \sum_{j=1}^{N} (n_j - \delta_{ij}))\Pi \right),$$
(6.33)

where the superscripts (+) and (-) represent a function calculated at state \boldsymbol{n} with the exception of n_i being respectively $n_i + 1$ and $n_i - 1$. The deterministic game at the ecological scale functions as a mean-field model for interaction rates for the stochastic evolution too. We can compute the game-simulated interaction rates as $\omega_i(\boldsymbol{\eta}) = \omega_i(\boldsymbol{n}/\Omega)$ with approximation error of $\mathcal{O}(\Omega^{-1})$. This process potentially yields a highly complex stochastic evolution, with nonlinear transition rates.

6.4.3 Stabilizing the Hawk-Dove game

The Hawk-Dove game (HD) is an important model of animal conflict, representing an ecological setting where individuals, when encountering, can choose to engage a dispute or avoid it. Hawks portray an aggressive behavior of engaging while Doves are more cautious and choose to avoid disputes.⁷³ The simplest form of the Hawk-Dove game is expressed by the payoff matrix

$$\begin{array}{ccc}
H & D \\
H & \left(\frac{(v-c)}{2} & v \\
D & \left(\begin{array}{c} 0 & \frac{v}{2} \end{array} \right) \\
\end{array} \right)$$
(6.34)

with c > v. An interpretation is that when individuals do engage in a dispute, the winner gains a payoff v in terms of reproductive advantage (increased birth rate) and the loser gains a negative payoff c in terms of injuries (increased death rate). It has the corresponding reaction network:

$$2H \xrightarrow{k} 2H + \frac{v}{2}G_H + \frac{c}{2}L_H$$
$$H + D \xrightarrow{k} H + D + vG_H$$
$$2D \xrightarrow{k} 2D + \frac{v}{2}G_D.$$
(6.35)

(6.35)

The game assumes equal interaction rates
$$k$$
 for all kinds of encounters. If we want to consider background birth and death rates, we deviate from the standard game by also including the interactions

$$\begin{array}{ll} H \xrightarrow{\omega_0} H + G_H & H \xrightarrow{\mu_0} H + L_H \\ D \xrightarrow{\omega_0} D + G_D & D \xrightarrow{\mu_0} D + L_D. \end{array}$$
 (6.36)

This game is used as an ecological-level simulation of birth and death rates. The logistic evolutionary network, of which the usual replicator dynamics is a particular case, is represented for hawks and doves as

$$\begin{array}{cccc} H \xrightarrow{\omega_{H}} 2H & D \xrightarrow{\omega_{D}} 2D & H \xrightarrow{\mu_{H}} \emptyset \\ D \xrightarrow{\mu_{D}} \emptyset & 2H \xrightarrow{\gamma} H & 2D \xrightarrow{\gamma} D \\ H + D \xrightarrow{\gamma} H & H + D \xrightarrow{\gamma} D, \end{array}$$
(6.37)

with ω_H , μ_H , ω_D , and μ_D modeled by the game (6.35-6.36). This network yields, for example, a replicator equation for hawks and doves:

$$\frac{d\rho_H}{dt} = \rho_H \rho_D (f_H - f_D)$$
$$\frac{d\rho_D}{dt} = \rho_H \rho_D (f_D - f_H), \qquad (6.38)$$

with

$$f_H = \omega_H - \mu_H = k(\eta_H + \eta_D) \left(\rho_H \frac{(v-c)}{2} + \rho_D v\right)$$

$$f_D = \omega_D - \mu_D = k(\eta_H + \eta_D) \Big(\rho_D \frac{v}{2} \Big).$$
 (6.39)

Having the same phase portrait as the simple HD game, it also results in the equilibrium values of $\rho_H^{eq} = v/c$ and $\rho_D^{eq} = (c - v)/c$. Apart from the speed of the time-evolution of proportions, that depends on $(\eta_H + \eta_D)$, all differences we implemented are felt only at the level of densities, preserving the same structure of the HD game. For this game, we have from equation (6.28):

$$\overline{\phi(\boldsymbol{\rho})^{eq}} = \frac{kv(c-v)}{2c}.$$
(6.40)

If $\overline{\phi(\boldsymbol{\rho})^{eq}} > \gamma$, the theory predicts that the densities diverge. If $\overline{\phi(\boldsymbol{\rho})^{eq}} < \gamma$, the densities stabilize, according to equation (6.31), with size

$$(\eta_H + \eta_D)_{eq} = \frac{2ck(\omega_0 - \mu_0)}{2c\gamma - v(c - v)}.$$
(6.41)

Then, the standard HD model, obtained by reducing the network with $(\omega_0 - \mu_0) = \gamma = 0$, results in the usual replicator dynamics for the proportions ρ_H and ρ_D , but with densities rapidly diverging to infinity (Fig. 10-A). If we move to the logistic network, with the baseline competition rate $\gamma > 0$, the equilibrium proportions are the same, but now the densities can go to zero if $\overline{\phi(\rho)^{eq}} < \gamma$ (Fig. 10-B) or still diverge if $\overline{\phi(\rho)^{eq}} > \gamma$ (Fig. 10-C). Only by also adding the simulation of payoffs for independent births and deaths, with $(\omega_0 - \mu_0) > 0$, we are able to stabilize the densities when respecting the condition $\overline{\phi(\rho)^{eq}} < \gamma$ (Fig. 10-D).

In addition, we can consider the stochastic version of the logistic evolutionary network with birth and death rates modelled by the same HD game. The stochastic evolution is given by equation (6.4). The network (6.37) defines 8 possible state jumps. As an example, the birth transition for hawks happens with rate

$$W_{H\to 2H} = \omega_H(\boldsymbol{n}/\Omega)n_H = \left(\omega_0 + \frac{kv(n_H + 2n_D)}{2\Omega}\right)n_H.$$
(6.42)

The unpredictability of the stochastic network decreases as the size of the system increases, and that process is fundamentally dependent on the sizes of populations, n. For small systems, the deterministic dynamics is unable to represent the behaviors of the fundamental level, however for large systems they become similar dynamics (Fig. 11). The system-size expansion representing this convergence pictures the stochastic dynamics as composed of size-dependent noise around the deterministic trajectories, that gets washed out in the limit by the law of large numbers.

This analysis highlights the existence of three model layers of the populations dynamics, each one abstracted from the layer before. (1) The fundamental, stochastic layer of individual-counts represented by \boldsymbol{n} , (2) the deterministic density layer represented by $\boldsymbol{\eta}$, and (3) the deterministic proportions layer represented by $\boldsymbol{\rho}$ and the replicator dynamics.



Figure 10 – Densities and proportions in the Hawk-Dove game. Left: Densities; Right: Proportions. Initial state for hawks and doves is (0.4, 0.6), with densities matching proportions. Equilibrium proportions are the same, but their time-evolution depends on the total density. We have $\overline{\phi(\rho)^{eq}} = 0.1$. (A) Default replicator dynamics. Densities quickly diverge, with unbounded growth, while proportions approach the expected equilibrium. (B) Dynamics with a constant background competition of rate $\gamma = 0.3 > \overline{\phi(\rho)^{eq}}$. Now, population size decreases to zero while proportions remain approaching the expected equilibrium values. (C) With a lower competition rate $\gamma = 0.05 < \overline{\phi(\rho)^{eq}}$, densities also diverge. (D) Dynamics with both competition of $\gamma = 0.3$ and background birth and death rates for hawks and doves, $(\omega_0 - \mu_0) = 0.6$. The model reflects a more realistic evolutionary setting and the densities are finally stable. Other parameter values: k = 1, v = 1, c = 1.25, and arbitrary units of time. Source: By the author.

6.4.4 Price Equation and Adaptive Dynamics

Now we turn to the task of unifying the developed ideas with adaptive dynamics. I begin with an analysis of the Price equation. Given any population dynamics, we may be interested in the evolution of traits belonging to individuals. The population dynamics



Figure 11 – Stochastic evolution and the system size. Stochastic samples for 4 different sizes, $\Omega = (10, 100, 500, 2000)$, compared to the deterministic limit. All cases depict the same density dynamics of the stable hawk-dove game, with the model specified as in fig. (10-D). Black lines depict the deterministic limit, red lines are stochastic densities of hawks and blue lines are stochastic densities of doves.

Source: By the author.

affect the distribution of the trait among populations and, in turn, the trait can be a factor affecting the outcome of the dynamics. In order to assign values of a trait to populations and understand its evolution, we need to provide a *trait model* reflecting how the interaction rates depend on trait values and how trait values can change inside populations as time passes.

Suppose we have a quantitative trait z assuming the value z_i for population X_i on a network with N populations with proportions following the replicator dynamics. This model has a mean trait value defined as $\overline{z} = \sum_i \eta_i z_i / \sum_i \eta_i = \sum_i \rho_i z_i$. If we provide a model for any possible changes dz_i/dt , the evolution of the mean \overline{z} is completely determined by the evolution of η . The dynamical representation at the level of the trait values is expressed by the Price equation, obtained by deriving the definition of \overline{z} in relation to time:

$$\frac{d\overline{z}}{dt} = cov(z, (\omega - \mu)) + \frac{\overline{dz}}{dt} + \overline{\omega\delta z}, \qquad (6.43)$$

with $cov(z, (\omega - \mu)) = \sum_i \rho_i(z_i - \overline{z}) \left((\omega_i - \overline{\omega}) - (\mu_i - \overline{\mu}) \right)$, $\frac{d\overline{z}}{dt} = \sum_i \rho_i dz_i/dt$, and $\overline{\omega \delta z} = \sum_i \rho_i \omega_i \delta z_i$ for $\delta z_i = \sum_j (z_j - z_i) q_{ij}$. This equation is a mathematical equality. Covariance and mean in this context are referring to simple functions of the distribution of z over the network, they are summaries of the distribution of z on the populations weighted by their

proportions ρ_i .⁷⁴ In order to derive this equation, we have to suppose that the trait z has different values z_i distributed over different individuals, and the equation itself relates this distribution of z to the sources of its changes, ultimately driven by interactions of birth and death. Changes in \overline{z} are separated in three contributions. (1) The covariance term: populations with higher birth rates or lower death rates among the network will increase in proportion, and the trait value that they bear will become more present, driving \overline{z} towards them. (2) The mean derivative term: if, over time, each z_i can change its value according to any trait model, the corresponding change in \overline{z} is the mean of the changes in each z_i . (3) The mutation term: because mutations occur between populations, some offspring of X_i might be a X_j , bearing the trait value z_j instead of z_i ; the change in \overline{z} is modulated by the difference between z_j and z_i for every mutation.

If we provide a model of the interaction rates, for example as $\omega_i = \omega_i(\eta, z)$, $\mu_i = \mu_i(\eta, z)$, and $q_{ij} = q_{ij}(\eta, z)$ in the case of birth-death-mutation interactions, the values z_i will affect the population dynamics of η that in turn will affect the distribution of the z_i , driving changes in z. The Price equation captures this interplay and is able to provide the resulting evolution of z. However, the variance in z comes from the fact that different populations possess different trait values z_i and will tend to decrease when a population is able to dominate in proportion. In the case where the system is dominated by one population, the only contribution that is still able to drive changes in z is the second term, of the mean derivative of z_i , that is the only term accounting for the appearance of new values of the trait; the other terms depend on the permanence of different populations in order to be effective.

But the mean derivative term, $\overline{\frac{dz}{dt}}$, is rather strange in this context, because it implies that individuals can change trait values throughout their lives. In the way we defined the model, each population is representative of the value z_i , so it is more natural for individuals to change trait values by mutating from a X_i into a X_j . Thus, the mean derivative term vanishes and \overline{z} changes only due to births and deaths. Additionally, if we further assume random mutations, so that mutations are equally likely to increase or decrease trait values, $\delta z_i = 0$ and the mutation term also vanishes. Thus, usually, the Price equation reduces to $\dot{\overline{z}} = cov(z, F)$, and is independent of the q_{ij} .

If we assume that mutations almost never occur, with $q_{ij} \ll 1$ for any $i \neq j$, we consider that equilibrium states of η are attained for all $F_i = 0$ while mutations do not occur. Then, density equilibrium states always have cov(z, F) = 0, whether this equilibrium consists of coexistence between populations or of a single dominating population. Whenever a mutation occurs in this state, with a new population X_m appearing with trait value z_m , there are two possible outcomes: (1) $F_m < 0$ and the mutant population fails to grow, or (2) $F_m \geq 0$ and the mutant population has the opportunity to grow. For the second case, the growth of the new mutants will depend on the behavior of all F_i as functions of η , as η_m grows and other populations possibly shrink. If the mutants successfully grow in the population, a new equilibrium for \overline{z} is reached, and this is a general sketch of how the trait evolution plays out.

To further explore the evolution of the trait z, we consider all populations to be equal apart from the difference in the trait z. This results in $\omega_i(\boldsymbol{\eta}, \boldsymbol{z}) = \omega(\boldsymbol{\eta}, \boldsymbol{z})$ for all i and the same for all other interaction rates. Then, the network starts with a single resident population at equilibrium, X_r with trait value z_r . When a mutation occurs, this equilibrium is disturbed and new mutants X_m appear with trait value z_m . We suppose that all mutations occur in small steps $|z_m - z_r| = \delta$; and $(z_m - z_r) > 0$ means that the mutation increases the trait value and $(z_m - z_r) < 0$ means it decreases. We have for example the possible populations for positive trait values: X_1 with trait $z_1 = 0$, X_2 with trait $z_2 = \delta$, X_3 with trait $z_3 = 2\delta$, and so on. Thus, for the model of the two populations X_r and X_m once the mutants arise, we have the function $F(\eta_r, \eta_m, z_r, z_m)$ determining the outcome of both the population dynamics and adaptive dynamics.

With the rise of the mutant, having a proportion ρ_m , the initial trait average is $\overline{z} = z_r + \rho_m \delta$. If the mutants fail to grow, \overline{z} goes back to z_r . If the mutants grow, \overline{z} is able to evolve up to z_m . Expanding F around the resident trait value z_r and neglecting $\mathcal{O}(\delta^2)$ terms, the Price equation becomes

$$\frac{d\overline{z}}{dt} = \rho_m (1 - \rho_m) \delta^2 \frac{\partial F}{\partial z_m} \Big|_{z_m = z_r}.$$
(6.44)

Using $\dot{\eta} = \eta F$, we are able to determine the time-evolution of the invasion dynamics behind this process, with the condition for mutants growth being $\frac{\partial F}{\partial z_m}\Big|_{z_m=z_r} > 0$ for $z_m > z_r$ or $\frac{\partial F}{\partial z_m}\Big|_{z_m=z_r} < 0$ for $z_m < z_r$. If the mutants dominate over the residents $(\rho_m \to 1)$, η_m stops again at a new equilibrium with $\overline{z} = z_m$, until new mutants appear. When the resident population cannot be invaded by either increasing or decreasing z, z_r is an equilibrium value for the trait evolution, represented by $\frac{\partial F}{\partial z_m}\Big|_{z_m=z_r} = 0$. The stochastic counterpart for the adaptive dynamics as described here arises from considering $\overline{z} = \sum_i z_i n_i / \sum_i n_i$ for the same networks. For multi-trait evolution, the condition for evolutionary equilibrium in this case is expressed in terms of the gradient $\nabla_m F$ and the evolutionary trajectory is given by successive increments of $\delta^{(j)}$ for any trait $z^{(j)}$ that is changed by a mutation.

Adaptive dynamics usually rely on two general assumptions: (1) mutations are rare so as to always occur only when populations are at a dynamical equilibrium, so dynamical and adaptive time-scales are separated; and (2) population interactions happen in a way that no coexistence of different trait values can hold, so at dynamical equilibrium there is only one dominant trait value.⁷⁵ These assumptions work to provide a purely adaptive evolutionary model, without interference from the detailed competition dynamics between populations with different traits, between residents and new mutants. With this simplification, we can model trait evolution as a Markov jump process over the trait values. The jump probabilities depend on the mutation rates, q_{ij} , and on the probability of fixation of the new mutants, which is non-zero only for $F_m > 0$. The resulting deterministic limit is called canonical equation of adaptive dynamics.⁵⁸ It can be summarized as:

$$\frac{dz_r}{dt} = C\eta_r \frac{\partial F_m}{\partial z_m}\Big|_{z_m = z_r}.$$
(6.45)

The positive parameter C encapsulates the mutation rate and trait variance, it can depend on z_r but is usually assumed to be constant in models. This equation highlights the fact that the evolutionary trajectory of the trait z generally depends only on the selection gradient. It is similar to the Price equation as we defined, although it describes a fundamentally different system. In the canonical equation, there is no dynamical evolution of proportions, so any variance over the trait z cannot come from its distribution among the different coexisting populations. In fact, it comes from the mutation rates, now expressed as a probability distribution over the values of z, as the probability of mutating from the resident trait value to adjacent values. The dependence over F does not come from the dynamical evolution $\dot{\eta}$ in this scenario, but from the probability of fixation once the mutant appears in the population. Thus, the canonical equation models evolution on the trait space only, assuming instantaneous equilibrium over the dynamical state space, while trait evolution from the Price equation is about changes in trait due to changes in the dynamical population evolution. However, both models capture, in agreement, the direction of trait evolution as given by $\frac{\partial F}{\partial z}$, the selection gradient.

6.4.4.1 Games as Generations

In order to provide an F that is suited to trait evolution through equation (6.45), we now make use of our generalization to define games as generations. An issue with games as simulations of interaction rates is that the payoffs must be further specified in order to refer to actual populations in biology. Otherwise, games remain as theoretical explorations of possible scenarios, designed to capture the general features of frequency dependent evolution. So this is a proposal of an additional design for games, in which we model birth rates from a simulated generation of individuals. The evolutionary time scale is one generation per unit of time, and the payoff gathered is the number of offspring per individual in one generation.

Consider an evolutionary network composed of birth, death, and competition interactions:

$$\begin{array}{ccc} X_i \xrightarrow{\omega_i} 2X_i \\ X_i \xrightarrow{1} \emptyset \\ X_i + X_j \xrightarrow{\gamma} X_i, \end{array} \tag{6.46}$$

assuming a constant competition rate γ . The death rate is 1 because each individual lives once per generation, so the average lifetime, that is the inverse of the death interaction rate,

is the same as the time scale. In this setting, the birth rates ω_i will be equal to the number of offspring each individual produces per generation. Inside a generation, individuals will never increase in number, because birth interactions increase only birth payoff. But they die at their natural death rates. The total payoff at the end of the generation, when all individuals are dead, will be the amount of offspring produced and equal to the birth rate. It is the same as the birth rate because at each time step, that is the expected time that individuals stay alive, each individual produces on average the payoff's amount of offspring, and that is how many birth reactions each individual undergoes on average per unit of time, the birth interaction rate itself. The initial state of the simulated generation is the present state of the evolutionary network, and the birth rate can be density dependent.

I illustrate this modeling strategy with an example of simple sexual reproduction. The network consists of the populations of females X_1 and males X_2 . We then simulate generations in order to calculate the birth rates. We use the following simulation network at the ecological level, that is the game:

$$X_1 + X_2 \xrightarrow{m} X_1 + X_2 + \beta_1 W_1 + \beta_2 W_2$$
$$X_1 \xrightarrow{\mu_1} \emptyset$$
$$X_2 \xrightarrow{\mu_2} \emptyset. \tag{6.47}$$

 W_1 and W_2 are payoff variables representing the number of offspring at a given time, w_1 and w_2 . The coefficient β_i considers the average number of offspring from each mating event that reaches maturity and belongs to the *i*-th sex. The mating interaction rate m(n)modulates the shape of the mating transition rate (mating function); I am considering the baseline mating dynamics of constant m, but other functional forms are responsible for other kinds of mating functions, and they need a proper mechanistic justification for their use. To maintain clarity, we use the deterministic limit for this example:

$$\frac{d\eta_1}{dt} = -\mu_1 \eta_1$$

$$\frac{d\eta_2}{dt} = -\mu_2 \eta_2$$

$$\frac{dw_1}{dt} = m\beta_1 \eta_1 \eta_2$$

$$\frac{dw_2}{dt} = m\beta_2 \eta_1 \eta_2,$$
(6.48)

and the birth rates are given by

$$\omega_i = \frac{1}{\eta_i} \int_0^\infty \frac{dw_i}{dt} dt.$$
(6.49)

All parameters can be functions of the state, but here we consider them as constants. For an initial state (η_1, η_2) , this system has an analytic solution. The output of the simulation is the ω_i , given by

$$\omega_1 = \frac{m}{(\mu_1 + \mu_2)} \beta_1 \eta_2,$$

$$\omega_2 = \frac{m}{(\mu_1 + \mu_2)} \beta_2 \eta_1. \tag{6.50}$$

The ratio ω_2/ω_1 is a measure of relative fitness between males and females. Fitness per individual, for each sex, is proportional to the density of the other sex, and they reach equality when $\beta_1 n_2 = \beta_2 n_2$ in the evolutionary level. In particular, for equal offspring share, the equilibrium is for equal number of males and females. In the case of females, the time evolution is given by

$$\frac{d\eta_1}{dt} = \eta_1 \left(\frac{m}{(\mu_1 + \mu_2)} \beta_1 \eta_2 - 1 - \gamma(\eta_1 + \eta_2) \right), \tag{6.51}$$

and the replicator equation for proportions is

$$\frac{d\rho_1}{dt} = (\eta_1 + \eta_2)\rho_1\rho_2 \frac{m(\beta_1 p_2 - \beta_2 p_1)}{(\mu_1 + \mu_2)},\tag{6.52}$$

with $\rho_2 = 1 - \rho_1$. The equilibrium state is then

$$\eta_i = \frac{\beta_i}{\frac{m\beta_1\beta_2}{\mu_1 + \mu_2} - \gamma(\beta_1 + \beta_2)}$$
(6.53)

As expected, the equilibrium state for proportions, from the replicator equation, is independent of γ :

$$\rho_i = \beta_i / (\beta_1 + \beta_2). \tag{6.54}$$

The mating transition rate is $W_m = m\eta_1\eta_2$. For females, the mating function, that is the mating transition rate per individual female, is proportional to $m\eta_2$ (and for males it is proportional to $m\eta_1$). This is the mating function that naturally arises from male-female encounters for otherwise independent individuals. A commonly used mating function is the harmonic mean⁷⁶, that models a decrease in mating that is the result of increased density (implying increased mating selection, for example), making the chances of mating events inversely proportional to the population size, so we have $m \propto 1/(\eta_1 + \eta_2)$. Every justifiable functional form of the mating rate m gives a different mating transition rate and thus an alternative mating scenario.

6.4.5 Examples

Finally, we turn to four different examples with the aim to further establish the use and potentials of this perspective.

6.4.5.1 Generalized Lotka Volterra

A generalized Lotka-Volterra $\operatorname{system}^{72}$ can be modeled at the design level with a reaction network as

$$X_i \xrightarrow{\omega_i} 2X_i, \tag{6.55}$$

$$X_i \xrightarrow{\mu_i} \emptyset. \tag{6.56}$$

$$X_i + X_j \xrightarrow{\kappa_{ij}} X_i \tag{6.57}$$

$$X_i + X_j \xrightarrow{\gamma_{ij}} 2X_i \tag{6.58}$$

$$X_i + X_j \xrightarrow{m_{ij}} 2X_i + 2X_j \tag{6.59}$$

$$X_i + X_j \xrightarrow{c_{ij}} X_i + 2X_j. \tag{6.60}$$

In these models, all interaction rates are constant. Interactions (6.55) are birth, (6.56) are death, (6.57) are competition, (6.58) are predation or parasitism, (6.59) are mutualism, and (6.60) are commensalism. This network yields deterministic description for densities and replicator dynamics (6.19) governed by the growth function $F_i = (\omega_i - \mu_i) + \sum_j \eta_j (\gamma_{ij} - \gamma_{ji} + m_{ij} + m_{ji} + c_{ij} - \kappa_{ji})$. Some generalizations of this model consist simply of making any of the interaction rates depend on the network's state. Arditi-Ginzburg models, for example, consider that prev consumption rates and birth rates are density dependent.^{77,78} These would model population size effects, such as density of prevs and predators, over their rates of birth and predation, and would need an additional justification for being that way.

6.4.5.2 Stochastic Rock-Paper-Scissors Game

This is just another example of the modeling of a standard game and the connection between stochastic and deterministic equations. Consider the reaction network version of the simple rock, paper, an scissors game given by the payoff matrix

The corresponding reaction network simulation for this game is

$$R + P \xrightarrow{k} R + P + \Lambda_R^L + \Lambda_P^G$$

$$R + S \xrightarrow{k} R + S + \Lambda_R^G + \Lambda_S^L$$

$$P + S \xrightarrow{k} P + S + \Lambda_P^L + \Lambda_S^G.$$
(6.62)

We interpret this game as: when R and P cross paths, there is a tendency, represented by k, for P to reproduce and, with the same intensity, also for R to die. And equivalently for the pairs S - R and P - S.

For this example, I consider a deterministic game simulation giving the rates of a stochastic birth and death evolution. In other words, this is a model of stochastic populations evolving in accordance with transition rates obtained as average payoffs from a deterministic game of mean-field propensities. The stochastic evolution is given by equation (6.4), with birth and death transition rates

$$W_{R \to 2R} = n_R f_R(\boldsymbol{n}) = n_R \frac{k n_S}{\Omega}$$
(6.63)

and

$$W_{R \to \emptyset} = n_R \mu_R(\boldsymbol{n}) = n_R \frac{k n_P}{\Omega}$$
(6.64)

for rocks, and equivalently for papers and scissors. These transition rates provide a stochastic phase portrait equivalent to replicator dynamics for the relative proportions: $p_R = n_R/(n_R + n_P + n_S)$, for rocks, and equivalently for papers and scissors. If we make $\Omega \to \infty$ while keeping η the same, we recover the deterministic replicator evolution. Figure (12) shows the evolution of rock proportion p_R and the simplex representation of a sample of the stochastic process for two different system sizes, comparing with the deterministic limit. The deterministic phase portrait is independent of the population size, but the stochastic process is inevitably dependent. This game conveniently avoids the instability of densities, because of its cyclic nature.



Figure 12 – Rock-Paper-Scissors Game. Stochastic samples for small system ($\Omega = 5$) and large system ($\Omega = 100$), compared with the deterministic limit. Left: rock proportion, p_R , over time. Right: simplex representation showing joint oscillations, with scissors on the left bottom, rock on the right bottom, and paper on top. Other parameter values: k = 1, initial condition $(n_R, n_P, n_S) =$ $(30\Omega, 8\Omega, 8\Omega)$, and arbitrary unit of time. Stochastic samples were obtained by the stochastic simulation algorithm. Source: By the author.

6.4.5.3 Predator-Prey with Hiding Preys

This is a sketch of a more complex ecological situation. The purpose is just to show the modeling pipeline, not in the rigorous analysis of results. Consider the example of the following toy ecological system: there is a population of a species that is prey to the population of a predator species. But preys have the capacity to flee and hide once they are attacked by a predator. Also, when a prey is consumed, the predator rests from the hunt in order to produce offspring. Preys only grow when they are feeding from a substrate plant species, but in order to eat they expose themselves to predators.

I model this system with a reaction network composed of eleven interactions. Exposed and hiding preys are respectively X_o and X_h . Hunting and breeding predators are respectively Y_o and Y_p . The plant substrate is S. A fraction q of predator attacks is evaded by preys, so it is a measure of prey's adaptation to predators. Preys stay hidden for an average time T_h , after which they expose themselves again in order to eat and reproduce. After consuming a prey, predators do not hunt for an average time T_p , after which they come back to hunting with a new offspring predator. All preys have an intrinsic death rate of μ_x , and μ_y for predators. The full network is:

$$\begin{aligned} X_o + Y_o &\xrightarrow{(1-q)\alpha} Y_p \\ X_o + Y_o &\xrightarrow{q\alpha} X_h + Y_o \\ & Y_p &\xrightarrow{1/T_p} 2Y_o \\ & X_h &\xrightarrow{1/T_h} X_o \\ & X_h, X_o &\xrightarrow{\mu_x} \emptyset \\ & Y_p, Y_o &\xrightarrow{\mu_y} \emptyset \\ & S + X_o &\xrightarrow{\omega} 2X_o \\ & S &\xrightarrow{g} 2S \\ & \emptyset &\xrightarrow{g_0} S. \end{aligned}$$

The first two interactions are predator's attack, happening at a rate α . Exposed preys reproduce when they eat S. Plants grow by reproducing at a rate g, but also grow by random external seeds at a constant rate g_0 . The model is fully specified at the design level. For our analysis, we'll explore the deterministic system of this network. Representing

(6.65)

the densities as lower-case letters, we have

$$\frac{dx_o}{dt} = -\alpha x_o y_o + \frac{x_h}{T_h} - \mu_x x_o + wsx_o$$

$$\frac{dx_h}{dt} = q\alpha x_o y_o - \frac{x_h}{T_h} - \mu_x x_h$$

$$\frac{dy_o}{dt} = -(1-q)\alpha x_o y_o + 2\frac{y_p}{T_p} - \mu_y y_o$$

$$\frac{dy_p}{dt} = (1-q)\alpha x_o y_o - \frac{y_p}{T_p} - \mu_y y_p$$

$$\frac{ds}{dt} = g_0 + gs - wsx_o.$$
(6.66)

The total density of preys is given by $x = x_o + x_h$, and $y = y_o + y_p$ for predators. With this model, I show a simple numerical analysis of the prey adaptation to the predators, measured through the parameter q, while maintaining other parameters fixed. Figure (13) shows four scenarios with increasing values of prey adaptation, for a given initial state. When preys are mostly unable to flee (13-A, q = 0.1), the model yields oscillations. With some capacity to flee (13-B and C, q = 0.4 and q = 0.7), the model no longer exhibits oscillations and predators still dominate in proportion. For a high enough capacity to flee (13-D, q = 0.9), predators become extinct. I used an initial state with densities summing to 1, but note that this number changes depending on the dynamics.

6.4.5.4 Model of Parental Care Evolution

As an example of trait evolution, a reframe a model for parental care evolution, from Kokko and Jennions (2008),⁷⁹ later updated by Fromhage and Jennions (2016).⁸⁰ This model aims at drawing phase portraits for the expected duration of care, τ and $\tilde{\tau}$, respectively for females and males. It considers that males and females change between states of time-in and time-out, for seeking mates and providing parental care after mating. By counting all potential offspring, the simulation estimates the birth rates of males and females as a function of the trait. Then we can use these in the evolutionary network (6.46), composed of birth, death, and background competition for resources. I will call females X and males \tilde{X} . The simplest generation, simulating birth rates, is

$$X_{I} + \tilde{X}_{I} \xrightarrow{m} X_{O} + \tilde{X}_{O} + \frac{\beta}{2} (W_{1} + W_{2})$$

$$X_{O} \xrightarrow{1/\tau} X_{I}$$

$$\tilde{X}_{O} \xrightarrow{1/\tilde{\tau}} \tilde{X}_{I}$$

$$X_{I}, X_{O}, \tilde{X}_{I}, \tilde{X}_{O} \xrightarrow{\mu} \emptyset.$$
(6.67)

 X_I represent females in time-in and X_O represent females in time-out, the same for males. When mating occurs, females go time-out for an expected time of τ , and males for $\tilde{\tau}$. After caring for offspring, they go back to time-in. We assume half of the offspring to grow as



Figure 13 – **Predator-prey with hiding preys**. Time evolution of predators, preys, and plants, for the initial state (0.1, 0.3, 0.6), with varying prey capacity to hide, q. (A) q = 0.1, predators dominate and the model eventually oscillates. (B) q = 0.4, no more oscillations. (C) q = 0.7, preys are about to dominate in number. (D) q = 0.9, predators are extinct as a result of prey adaptation. Other parameter values: $T_p = 5$, $T_h = 1$, $\alpha = 0.1/r$, $\omega = 0.2/r$, $\mu_x = 0.1$, $\mu_y = 0.05$, g = 0.1, $g_0 = 0.5r$, and r = 0.1 is a parameter measuring the order of interaction range. Source: By the author.

male and half as female. All death rates are equal to μ . The coefficient β is a function of the trait,

$$\beta = b e^{-\alpha/(C(\tau,\tilde{\tau}))}.$$
(6.68)

The parameter b is the brood size, and the exponential term models the probability of the offspring to reach maturation, as a function of the caring times. $C(\tau, \tilde{\tau})$ is a care function, the amount of care given to offspring as a result of the time spent caring; here I consider a simple, linear care function $C = \tau + \tilde{\tau}$. α is an abstract parameter indicating the intensity of the offspring's need of care. The functional form of β is a probability of survival for each offspring and represents diminishing returns for the amount of care. This model also considers a parameter for the intensity of sexual selection, k, by leaving a portion of males out of the mating dynamics; Only \tilde{n}/k males are allowed to breed, the rest being denied by forces of sexual selection. This results in a bias over the operational sex ratio, defined as the ratio between males and females that are engaged in the mating dynamics.

Since males and females are born at the same rate for every generation, the population size is equally split in half for each population at the evolutionary level. I evaluate the evolutionary trajectories of average care duration, $(\tau, \tilde{\tau})$, using the assumptions of adaptive dynamics. Thus, the total population density size, $\eta_X + \eta_Y = N/2 + N/2 = N$, is assumed to be in equilibrium composed by the resident populations when each mutation arises. The equilibrium is given by the dynamics at the level of the evolutionary network, with a constant competition rate γ :

$$N^{eq} = \frac{\frac{1}{2}(w_X + w_Y) - 1}{\gamma}.$$
(6.69)

This equilibrium density is the measure of size for the resident population whenever mutants arise. Equation (6.69) is solved implicitly, because the simulated birth rates w_X and w_Y depend on the equilibrium density. Once we solve it, we are able to evolve the care duration traits $(\tau, \tilde{\tau})$ through the phase portrait, by using equation (6.45). for that, I consider the interactions of new mutants X' and Y' with trait values τ' and $\tilde{\tau}'$ appearing among the resident population. Female mutants interact as

$$X'_{I} + \tilde{X}_{I} \xrightarrow{m} X'_{O} + \tilde{X}_{O} + \frac{\beta}{2} (W'_{1} + W_{2})$$
$$X'_{O} \xrightarrow{1/\tau'} X'_{I}$$
$$X'_{I}, X'_{O} \xrightarrow{\mu} \emptyset.$$
(6.70)

Male mutants interact analogously. Thus, we are able to compute the relative success of mutants over residents through $\frac{\partial F_m}{\partial z_m}\Big|_{z_m=z_r}$ using the birth rates as given by equation (6.49) for the deterministic system of the simulated generation. Then, we can numerically integrate the birth rates for two scenarios, varying the parameter k, the intensity of sexual selection for males, considering the mutants to occupy 1% of the total population of their sex. Figure (14) shows the dynamics without and with the presence of sexual selection. (14-A), without sexual selection (k = 1), is a static fitness landscape from the resident point of view, showing care duration mapped onto the reproductive rate $(F_r, which is the$ same for both sexes) with maximum indicated in red. (14-B) is a phase portrait of the evolutionary trajectories for the same case. The results are divergent, and equilibrium states are not states of maximum reproductive success of the resident population, but states of maximum joint success of male and female mutant populations. The evolutionary result depends on the initial proportions of care between males and females. This symmetric system is unstable to any differences in the sex populations. (14-C) and (14-D) show the dynamics with sexual selection (k = 1.2) for the same initial states; it results in female-only care, despite the little change in the point of maximum reproductive success.

Kokko and Jennions, in their first formulation of the model, did not verify the Fisher condition on mating rates, so males and females had different total mating rates, and therefore different total fitness even for equal offspring share. We see that the reaction network model, however, trivializes the Fisher condition, because mating rates are bounded to be equal, due to the joint mating interaction. Fromhage and Jennions, in their updated



Figure 14 – Trait evolution for the care duration of females and males. Left: Resident reproductive success per care duration. Right: Evolution of average care duration for a set of initial conditions. (A) and (B) show the case without sexual selection (k = 1). The maximum value of F_r is shown in red at approx. (0.33, 0.33). Equilibrium states of the evolutionary trajectories depend on the initial state. (C) and (D) show the case with sexual selection on males (k = 1.2); Maximum F_r at (0.38, 0.31). Other parameter values: $M = 1, \mu = 0.01, \alpha = 0.5$, competition rate $\gamma = 1$ with step $d\tau = 0.01$ and arbitrary units of time. The constant from equation (6.45) is $C = 10^{-6}$. Source: By the author.

model, respected the Fisher condition, and imposed dependence on the sex ratio of time-in sub-populations (the operational sex ratio $\tilde{\eta}_I/\eta_I$), which leads to a different formulation of mating rates. Their formulation considers $m = m'/\sqrt{\tilde{\eta}_I\eta_I}$, but without further justification besides imposing dependence on the operational sex ratio. Also, these existing models do not consider the possible effects of varying population density.

In this model, there is a trade-off between staying in time-in or time-out, where individuals invest in generating more offspring or assuring the survival of a given offspring. The equilibrium is when the increase in average offspring from leaving care is the same as the increase from staying with the current offspring. The basic model above can be further expanded to include and analyze many other interactions, such as parental uncertainty, synergy between male and female parental care, different death rates, etc.

6.5 Discussion

In this work I developed a methodology for the use of reaction networks as a tool for a more fundamental version of evolutionary game theory and its connection to trait evolution through adaptive dynamics. The network description generalizes the replicator dynamics to a density-level description, and I argued that the evolution of densities is important for population dynamics and is tied to the evolution of proportions. Evolutionary game theory deals with an ecological scale game-simulation of reproduction rates and an evolutionary scale represented by the replicator dynamics; both modeling steps are better understood in terms of networks. Reaction networks provide a design-level model specification, and is well-suited for communication between fields and between modelers and experimental scientists.⁸¹ Interactions portray individual relational structures giving rise to global dynamics at the level of populations, for both a fundamental stochastic dynamics and the deterministic limit of infinite system.

The methods were applied to show that the standard games derived from simple payoff matrices result in unstable population densities, and that birth-death simulations driven solely by pairwise interactions between strategies are dynamically artificial. Baseline birth-death-competition must be accounted as the building block of an evolutionary network, and it was shown that this architecture is sufficient to stabilize densities arising from standard games. I applied these developments to the hawk-dove game to show that it features divergent densities and to show how to stabilize them. I also used the HD game to show the connection between stochastic and deterministic network dynamics, mediated by the size of the system.

A good picture of the network-based EGT is as follows: there is an evolutionary model, with a larger timescale, where individuals from a population can reproduce or die, and are subjected to an intrinsic competition stemming from limited resources indirectly shared by them as they coexist in the same environment. This model is encoded by networks such as (6.20) and (6.32). The interaction rates present in this network are influenced by the ecological scenario of the populations, where during their lives they engage in all sorts of different specific events depending on their individual strategies and choices. A game network is used to model how this ecological dynamics, happening in a smaller timescale and not directly leading to individual reproduction and death, help in shaping birth and death rates. The game works as a mean-field approximation to the actual ecological setting as a function of given population sizes. Usually, in EGT, these two scales are represented by a payoff matrix and a replicator equation. However, the game network can encode a large variety of interaction structures, and is by no means limited by pairwise interactions or by all interactions happening with the same rate. And the evolutionary network primarily deals with the evolution of densities instead of just proportions, resulting in a type of density-based logistic form of evolutionary dynamics, generalizing the replicator dynamics.

Because of how proportions are defined, they are naturally modeled with timeevolution driven by the relative fitness of a population compared to the average fitness among populations $(F_i - \overline{F})$; for a proportion ρ_i to increase, it naturally forces all other proportions to decrease respecting the constraint $\sum_i \rho_i = 1$. Densities evolve according to the fitness alone (F_i) , which is itself affected by the interactions with other individuals; because the individual fitness is what really shapes the growth of a population. Thus, the dynamics of proportions, when considered alone, is implicitly providing an artificial source of selection, while also always neglecting the mechanistic impact of densities in shaping the intensity with which interactions take place. Moreover, many replicator dynamical models will yield infinite or zero densities, and this is not something observable at the level of proportions.

The reaction network approach to EGT has the potential to connect to other types of models in population dynamics. Simulating for the same logistic evolutionary network, an ecological game can also be for example a Lotka-Volterra network, or a network for an epidemiology compartment model. This sort of universality is in line with the correspondence in equations explored by Page and Nowak, 2002.⁶⁵

The approach of evolutionary dynamics is also not restricted to static strategies, and it can be extended to the evolution of continuous traits as an integration with adaptive dynamics.⁷⁵. In that case, by extension of the fundamental theorem of natural selection, traits evolve in the direction of the adaptive landscape, that can be informed by a game-theoretical model. We developed this approach, stemming from a discussion of the Price equation and the canonical equation of adaptive dynamics. I presented a solid example of parental care evolution to illustrate the evolution of the time of care using a game-informed numerical model of adaptive dynamics.

It is possible to perform Bayesian parameter estimation over reaction network rate parameters. The framework aims to connect models to real observation data of network elements.^{13,57,82} The methods can be adapted to this context of evolutionary dynamics and can provide estimates of interaction rates through observations of population sizes. This has the potential to be an additional line of integration with empirical grounds and a robust test to modeling predictions.

We developed here an approach that is centered around the individual. Traits and mutations are considered as given. These mechanisms are further expanded by considering their genetic foundation. I ask whether genetic models can also be integrated with the framework we develop, with more rich dynamics driving the behavior of traits and mutations. We can say the same about group selection and inclusive fitness, that have its own issues of compatibility.^{83,84} Genetics and group selection comprise both ends towards which we can aim to expand the reach of population dynamics in the context of evolution. My desire

for grand models and integration between all relevant dynamics at play follows the same broad lines of systems biology. In this perspective, models not only are useful to describe reality, but also on their own, as means of synthetically producing realities. The robust study of evolving simulated environments should always be a background goal with models of evolution, aiming at producing ever more complete modeling frameworks.¹

In conclusion, reaction networks help in grounding and integrating evolutionary game theory to stochastic foundations, density-dependent logistic population dynamics, a wide set of general interactions not limited by standard pairings, a design-level modeling approach, and potentially more, like observational data. The simplicity of reaction networks is remarkable. I used networks to analyze and solve the problem of diverging densities of evolutionary game theory, applying the methods to the hawk-dove game. For that, I calculated the equilibrium of stable densities of a logistic evolutionary network in terms of baseline birth-death-competition rates and the average game payoffs.

7 INDIVIDUAL SPECIALIZATION AND GENERALIZATION IN PREDATOR-PREY DYNAMICS

This chapter presents a the model for a work done in collaboration with my collaborator, the ecologist Dr. Rafael Rios Moura. We invested in applying the ecoevolutionary framework with reaction networks to elaborate the first mechanistically robust model of individual specialization and generalization of predators. This first development made extensive use of the design-level proposal of modeling with reaction networks in order to define the relevant parameters and interactions. We intend to further expand on this application with a model for trait-evolution alike adaptive dynamics, using nested networks and game-like modeling.

The work is concerned with the question of what are the mechanistic factors driving the heterogeneity in predation strategy in regards with the amount of variation in prey consumption. When can we expect to see coexistence between individuals consuming only one type of prey and individuals consuming many prey kinds? What are the conditions for different strategies to arise? The model is designed using the predator-prey dynamics as a building block. The Python code supporting the results is available at <<u>https://nbviewer.jupyter.org/github/LR-GUI/IndividualSpecialization/blob/main/Code.ipynb></u>.

7.1 Introduction

Darwin struggled with his contemporaries to break the scientific tradition of seeing variation as deformations of ideal immutable forms.⁸⁵ By introducing the theory of evolution, the variation of heritable components underlying organismal form became the basis of the evolutionary process, inaugurating a new scientific tradition widely accepted today.⁸⁶ Recently, ecologists are struggling with their contemporaries to show that individuals' niche differ in ways that change our comprehension of eco-evolutionary processes and species coexistence.^{87,88} Niche theory is still based on the assumption that individuals are ecologically equivalent within a population or even a species.⁸⁹ This ongoing debate has become central to the current discussion between the extended evolutionary synthesis and the modern synthesis.⁹⁰

The ecological niche is currently defined as a hypervolume in an n-dimensional space composed of the conditions and resources required for a population to persist indefinitely in the environment.⁹¹ Recently, the concept of ecological niche was adapted to accommodate the individual niche, which can be defined as the ecological interactions of an organism with all components of its environment.^{88,92} Many empirical evidence support the statement that individual niche vary within populations.^{93,94} Hence, populations and species are actually groups of heterogeneous rather than homogeneous individuals

presenting variation in how they compete, select habitats, reproduce, and forage; among others.⁸⁸ This knowledge is not neglected by ecologists, but they historically attributed individual variation in resource use to differences between sexes, ageing, or ontogenetic stages.⁹⁵ However, individuals can differ ecologically regardless of these conditions.⁹⁴ Thus, individual specialization emerged as a concept of intraspecific niche variation that occurs when individuals only use a subset of the resources exploited by the population regardless of sex, age class or discrete morph.^{88,93,94}

A current challenge for the theory of individual specialization is to predict how and when individuals are expected to differ.⁹⁶ Based on empirical evidence, individuals use an average of 47% ($\pm 19.7\%$) of the resource types exploited by the population, and individual niche is, on average, 66% ($\pm 20.9\%$) as broad as the population niche.⁹⁴ Therefore, not all individuals exhibit identical niche widths. Interestingly, the strength of individual specialization can also change in space and time^{97–101} and vary up to ninefold across populations of the same species.⁸⁷ Considering predator-prey interactions, some predators in a population may specialize on the capture of one taxon or a few taxa from those available as prey, while other conspecifics may exhibit a generalist diet, capturing the same prey types consumed by the whole population.¹⁰² These differences can emerge from genetic variation¹⁰³ or through flexibility in learning many different foraging behaviors¹⁰⁴, but it is still unclear when a population is expected to exhibit strong or weak levels of individual specialization.

There is empirical evidence supporting that the strength of individual specialization can be driven by intraspecific competition.⁹⁴ Among-individual niche variation is expected to increase under intense intraspecific competition when conspecifics diverge from the use of a common optimal resource and begin to consume alternative prey.^{94,105} Therefore, individual predators should be more specialized in the capture of distinct prey items. An alternative scenario frequently ignored is that not all individuals need to specialize in a different prey type to coexist. A population may consist of a mixture of generalist and specialist individuals.¹⁰² When individuals share identical rank preferences, but differ in how much they are likely to add novel prey types to their diet, patterns of nestedness in forager diet are expected to occur under high population density of predators.¹⁰⁶ This model predicts how predator behavioral flexibility should affect individual niche variation. However, it did not consider how prey traits, such as its nutritional value and reproductive rates, may influence the coexistence of specialist and generalist individuals.

The theory of individual specialization is still largely based on verbal arguments with mixed support from empirical evidence.^{87,88,93,94} Other theories based on verbal assumptions produced logically inconsistent predictions and, consequently, are not supported by empirical evidence, mainly due to an absence of a formal mathematical model.⁷⁹ Therefore, in this study, we advanced the theory of individual specialization by formally modeling its main assumptions using predator-prey dynamics.¹⁰⁷ This modeling approach can be applied to individuals as well as species. However, we modeled few foraging strategies and prey types that are likely to be observed in populations rather than communities. In addition, species may share phylogenetic relationships that may strongly influence how they interact,^{108,109} except for communities with few competing predators of the same genus.¹¹⁰ In a population, all individuals are conspecifics and, consequently, share a similar evolutionary history. Thus, the model is better suited to describe individual specialization and generalization than community patterns.

In this study, we investigated when generalists and specialists are expected to co-occur in a population if individuals exhibit different efficiencies in prey capture and consume prey with distinct nutritional values and reproductive rates. Thus, we can clearly predict when and in which direction the strength of individual specialization is expected to change. We also tested if the central claim that populations are generally composed by heterogeneous individuals in resource use, rather than just generalists, still holds.

7.2 Methods

7.2.1 Lotka-Volterra Model

Our model is based on a Lotka-Volterra dynamics of prey and predators. In the simple Lotka-Volterra model, the system consists of a prey species X and a predator species Y, both participating in events that remove or include new individuals in the population. Prey reproduce at a rate ω , resulting in the inclusion of a new prey from an existing prey: $X \xrightarrow{\omega} 2X$. Predators die at a rate μ , resulting in the removal of the predator: $Y \xrightarrow{\mu} \emptyset$. Predators and prey interact at a rate α , with the predator consuming the prey for reproduction, resulting in the removal of the prey and the inclusion of δ predators: $X + Y \xrightarrow{\alpha} (\delta + 1)Y$. This structure of interactions results in a deterministic continuous limit composed of a system of differential equations over the densities of prey and predators. It provides the dynamical evolution of the populations in terms of the interaction rates:

$$\frac{dx}{dt} = x(\omega - \alpha y)$$
$$\frac{dy}{dt} = y(\delta \alpha x - \mu).$$
(7.1)

This is the usual Lotka-Volterra system of differential equations, and it results in an oscillatory pattern of prey and predator populations over time. The oscillations are sustained around unstable fixed points given by

$$(y^*, x^*) = \left(\frac{\omega}{\alpha}, \frac{\mu}{\delta\alpha}\right). \tag{7.2}$$

Thus, the typical density of predators at equilibrium is larger for high prey birth rate and smaller for high predation rate; a balance between prey influx and prey consumption. For prey, it is larger for high predator mortality rate and smaller for high predator birth rate; a balance between predator removal and predator influx. Their success is dynamically tied to each other.

7.2.2 Specialization Model

We are interested in a model that can explore the patterns of interaction between predators with different degrees of specialization in foraging strategies. Generalist predators are expected to prey upon a diverse range of species, while specialist predators constrain themselves to a limited niche. We assume that a generalist strategy prioritizes the ease in finding consumable prey types, relying less on availability of each type of prey, and a specialist strategy prioritizes the capacity to efficiently hunt its selected prey; thus, evolving better hunting performance.

In order to gain insights and analytically understand the determining factors behind specialist and generalist predator strategies, and their coexistence, we model prey with different nutritional values and reproductive rates. Therefore, we consider a system with two kinds of prey, X_A and X_B , in which X_A is *n* times more nutritive than X_B , with n > 1, so it provides *n* times more resources for predators. The more nutritive prey X_A also has a reproduction rate that is *p* times that of X_B ; when p < 1, X_A reproduces at a slower rate and, when p > 1, X_A reproduces at a faster rate. Each prey is consumed by a specialist predator; the predator Y_A that consumes X_A and the predator Y_B that consumes X_B , at the same rates (same efficiencies). Then, there is a generalist predator, Y, that consumes both prey at a rate k times smaller than the specialist predators, with k > 1. Apart from the predation events, predators are subject to a spontaneous death and prey may spontaneously reproduce, at rates respectively of μ and ω . This model is represented by the following set of events:

$$X_{A} \xrightarrow{p\omega} 2X_{A}, \quad Y_{A} + X_{A} \xrightarrow{k\alpha} (1 + n\delta)Y_{A}, \quad Y_{A} \xrightarrow{\mu} \emptyset,$$

$$X_{B} \xrightarrow{\omega} 2X_{B}, \quad Y_{B} + X_{B} \xrightarrow{k\alpha} (1 + \delta)Y_{B}, \quad Y_{B} \xrightarrow{\mu} \emptyset,$$

$$Y + X_{A} \xrightarrow{\alpha} (1 + n\delta)Y, \quad Y + X_{B} \xrightarrow{\alpha} (1 + \delta)Y, \quad Y \xrightarrow{\mu} \emptyset.$$
(7.3)

This model yields, as a deterministic limit on densities, the following system of differential equations for species over time:

$$\frac{dy_A}{dt} = y_A(k\alpha n\delta x_A - \mu),$$

$$\frac{dy_B}{dt} = y_B(k\alpha \delta x_B - \mu),$$

$$\frac{dy}{dt} = y(\alpha n\delta x_A + \alpha \delta x_B - \mu),$$

$$\frac{dx_A}{dt} = x_A(p\omega - \alpha y - k\alpha y_A),$$

$$\frac{dx_B}{dt} = x_B(\omega - \alpha y - k\alpha y_B).$$
(7.4)

Variable	Description
Y_A	Specialist predator that consumes X_A
Y_B	Specialist predator that consumes X_B
Y	Generalist predator that consumes both prey types
X_A	More nutritious prey
X_B	Less nutritious prey
α	Predator-prey base competition rate
ω	Prey base birth rate
μ	Predators death rate
δ	Predators base brood size
k	Specialists hunting efficiency factor $(k > 1)$
n	A/B Prey nutritiveness ratio $(n > 1)$
p	A/B Prey reproduction rate ratio

Table 4 – Model Variables.

Source: By the author.

This is a system of two separated Lotka-Volterra dynamics, $Y_A - X_A$ and $Y_B - X_B$, that are coupled by the presence of the generalist Y. Table (4) shows descriptions of all model variables. In our analysis, we found the fixed points of this system and explored the behavior of the state space through stability analysis and the possibility of invasions over stable configurations. For that, we calculated the Jacobian an its eigenvalues and associated eigenvectors for each fixed point, exploring the different possible initial states and parameter values. For the analytic calculations, we used the symbolic computations python package "Sympy". Detailed implementation and results are found as a supplementary material.

7.2.3 N-prey extension

In order to better evaluate emerging thresholds depending on the number of available prey, it is of interest to generalize our model to an arbitrary number of prey, maintaining the structure of predation. The generalization consists of adding new prey X_N with their respective specialist predator Y_N , and making the generalist Y also able to consume the new prey. For the purposes of this work, it is not necessary to differentiate their nutritive values or reproduction rates. For each new prey, the following interactions are added to the system:

$$X_N \xrightarrow{p_N \omega} 2X_N, \quad X_N + Y_N \xrightarrow{k\alpha} (1+\delta)Y_N,$$

$$X_N + Y \xrightarrow{\alpha} (1+\delta)Y, \quad Y_N \xrightarrow{\mu} \emptyset,$$
(7.5)

with their corresponding terms added to the time-evolution of densities.

7.3 Results

The time-evolution of the model exhibits oscillatory dynamics in the same manner as the standard Lotka-Volterra model. Analogously to (7.2), densities in the specialization model oscillate around unstable fixed points. But, instead of a single fixed point, there are 7 nontrivial unstable steady-states, shown in table (5), each representing a mode of the compound predator-prey dynamics. The way parameters relate to define these possible modes gives us clear insights into what mechanisms are the most relevant in determining predators strategies.

During the analysis of the different modes represented by each fixed point, we noted a fundamental threshold present in the system. The intensity of increased efficiency of specialists against the generalist, represented by parameter k, is determinant for the success of generalists. If k > 2, the specialists being at least twice as efficient as the generalist, then the generalist never succeeds in the population. For k < 2, where the specialists are not so efficient, the generalist can exploit the presence of specialists and succeed. Coexistence between all predator types is only possible for the threshold case of k = 2. The threshold value of 2 relates to the number of available prey types, as we will see.

The points in table (5) refer to the following scenarios:

1 and 2: Simple Lotka-Volterra involving the generalist predator Y and one of the prey types; it can only happen for k < 2. In this case, whenever the generalist is together with both prey types, the less fertile one dies out, and the dynamics reduce to 1 if p > 1 or 2 if p < 1. Thus, if there are no specialists present, the generalist strategy cannot sustain itself as a generalist, since one of the prey always dies out. Furthermore, because of the lesser efficiency (k > 1), the generalist would succumb if the specialist would invade 1 or 2. The generalist cannot sustain both prey populations because the greater benefit and increase in density that it gets from the presence of an additional source of fitness is sufficient to create excessive predation, driving the less fertile prey to extinction. This prey cannot keep up with the increased rates of predation.

3 and 4: These are simple Lotka-Volterra dynamics, involving one of the specialists with its prey. Featuring only one of the prey, the specialists can never be directly invaded by Y, even for k < 2. But, if k < 2, the absent prey can independently appear and enable the generalist as a successful strategy, resulting in mode 6 (from 4) or 7 (from 3). That is because, with an additional available prey, generalists have a chance to fare better than specialists. The access to more prey features a trade-off with low predation efficiency.

5: Both specialists with both prey types, but in an uncoupled manner, meaning two independent Lotka-Volterra dynamics existing in parallel. For k < 2, the system can be exploited by generalists seeking to attack both prey (and causing dependency between the two dynamics). In that case, the predator consuming the less fertile prey dies out, irrespective of how nutritious the prey is or how much fertile it is. Only the relative fertility of prey types determine which specialist dies. Thus, the generalist would invade and cause the extinction of a specialist, resulting in mode 6 (if p < 1) or 7 (if p > 1). This means that preying upon sources with faster reproduction (represented by p) is a determining factor for the success of a specialist strategy in face of generalists, while preying upon highly nutritious sources (represented by n) is not. However, if k > 2, above the threshold of specialists efficiency, the generalist strategy fails to thrive and this mode is stable (Figure 1D).

6 and 7: The more interesting dynamics featuring both the generalist and one of the specialists with the two prey is only possible for k < 2. Since the specialist that consumes the less fertile prey dies out in the presence of the generalist, 6 happens if p < 1 (Figure 1A, Y_A dies out) and 7 happens if p > 1 (Figure 1B, Y_B dies out). While k < 2 remains valid, this mode is stable. These modes are the ending result of successful generalists appearing in populations of specialists in order to take advantage of multiple prey availability. We see that the presence of specialists, and thus the competition between predators, hinders the capacity of the generalist to extinguish a prey; which is counter-intuitive, since there are more types of predators competing for the same prey.

When k is exactly 2, all populations can coexist around a fixed point that is degenerate on the densities of predators:

$$(Y_A, Y_B, Y, X_A, X_B) = \left(\frac{p\omega - \alpha, y^*}{2\alpha}, \frac{\omega - \alpha y^*}{2\alpha}, y^*, \frac{\mu}{2\alpha\delta n}, \frac{\mu}{2\alpha\delta}\right).$$
(7.6)

For this coexistence, the density y^* can be any value for which $y^* < \min(\omega/\alpha, p\omega/\alpha)$, depending on the initial state. k = 2 is a very narrow condition, but, if $k \approx 2$, we can expect a long transient state of coexistence, meaning that any strategy would die out only after a long time of coexistence (Figure 1C).

The determinant factor deciding which specialist strategy is best was found to be the difference in the reproductive rate of prey. An increase in the absolute intensity of prey reproduction is not relevant because it also drives the success of predators, while the difference in reproduction determines which prey falls behind first, prejudicing itself and its specialist. Surprisingly, the nutritive value of prey, affecting the reproductive success of predators, is important only to modulate prey's equilibrium density sizes.

7.3.1 N-prey model

The specific value of k = 2 for the threshold of the success of the generalist predator, at first, can seem somewhat arbitrary. To investigate its origin, we added more prey to the system, according to (7.5). We found that this threshold is actually k = N, where N is the number of prey and specialists available. Thus, the generalist is successful over specialists

Fixed Point $(\mathbf{Y}_{\mathbf{A}}, \mathbf{Y}_{\mathbf{B}}, \mathbf{Y}, \mathbf{X}_{\mathbf{A}}, \mathbf{X}_{\mathbf{B}})$	Description
$1. \left(0, 0, \frac{p\omega}{\alpha}, \frac{\mu}{\alpha\delta n}, 0\right)$	$Y - X_A$ Lotka-Volterra dynamics
2. $\left(0, 0, \frac{\omega}{\alpha}, 0, \frac{\mu}{\alpha\delta n}\right)$	$Y - X_B$ Lotka-Volterra dynamics
3. $\left(\frac{p\omega}{\alpha k}, 0, 0, \frac{\mu}{\alpha \delta k n}, 0\right)$	$Y_A - X_A$ Lotka-Volterra dynamics
4. $\left(0, \frac{\omega}{\alpha k}, 0, 0, \frac{\mu}{\alpha \delta k}\right)$	$Y_B - X_B$ Lotka-Volterra dynamics
5. $\left(\frac{p\omega}{\alpha k}, \frac{\omega}{\alpha k}, 0, \frac{\mu}{\alpha \delta k n}, \frac{\mu}{\alpha \delta k}\right)$	$Y_A - X_A$ and $Y_B - X_B$ uncoupled dynamics
6. $\left(0, \frac{(1-p)\omega}{\alpha k}, \frac{p\omega}{\alpha}, \frac{\mu(k-1)}{\alpha\delta kn}, \frac{\mu}{\alpha\delta k}\right)$	No Y_A , possible for $p < 1$
7. $\left(\frac{(p-1)\omega}{\alpha k}, 0, \frac{\omega}{\alpha}, \frac{\mu}{\alpha\delta kn}, \frac{\mu(k-1)}{\alpha\delta k}\right)$	No Y_B , possible for $p > 1$

Table 5 – Fixed Points.

Source: By the author.

for k < N. Then, if N - 1 < k < N, the specialist of the least fertile prey dies out. For N - 2 < k < N - 1, the specialist of the second least fertile prey also dies out, and so does the least fertile prey. Then, the pattern repeats up to 1 < k < 2, where there are only the generalist, the two most fertile prey, and only the specialist of the most fertile prey. This result shows that, without its specialist, a prey depends on the efficiency of other specialists to be high enough to survive in face of the generalist. That is because k is able to gauge the amount of competition a generalist faces so as to not grow too much and over-consume its prey.

For example, for N = 3 prey and the corresponding specialist predators, the generalist, consuming all three prey types, is successful only for k < 3, and it dies whenever k > 3 (Figure 2D). Therefore, it is harder for the specialists to win over the generalist if there are 3 prey instead of 2, as we could expect. In that case, the specialist consuming the least fertile prey dies out for 2 < k < 3 (Figure 2B). The second least fertile prey and the specialist in consuming it also die out if k < 2 (Figure 2A). Hence, the less efficient the specialists are, the less they can collectively sustain themselves in face of a successful generalist. Coexistence of all specialists and the generalist only happens for exactly k = 3 (Figure 2C).

7.4 Discussion

In our model, the emerging populations were composed of heterogeneous individuals, both generalists and specialists, in terms of foraging behavior, as supported by empirical evidence.^{93,94} Surprisingly, prey nutritional value was not an important factor determining



Figure 15 – Different modes with varying k and p. Predators and prey types dying out are highlighted in black. (A) (k=1.5,p=0.5). With less efficient specialists, the generalist is successful. The least fertile prey is X_A , so its specialist Y_A dies out. (B) (k=1.5,p=1.5). Now the least fertile prey is X_B , then the specialist Y_B dies out. (C) (k=2.1,p=1.5). The efficiency of specialists is slightly above the threshold, the generalist slowly dies out, after a long transient state. (D) (k=2.5,p=1.5). The specialists are securely efficient, then the generalist cannot succeed and dies out rapidly. Other parameter values: $\omega = 0.2$, $\alpha = 0.5$, $\delta = 1$, $\mu = 0.3$, n = 2.5. The equilibrium states do not depend on initial densities. Source: By the author.

the predator-prey dynamics. In the model with two prey types, when specialists were less efficient than the generalist, the one who consumed the highly nutritious and less fertile prey was extinguished. Only the specialist who consumed the highly fertile and less nutritious prey coexisted with the generalist and the other prey. The addition of new prey types as well as specialists on capturing them also led to the extinction of the forager consuming only the less fertile prey. Therefore, prey fertility was an important factor determining the coexistence of specialists and generalists in a population. However, prey fertility is empirically understudied and overlooked compared to their nutritional value^{111–113} The generalist was only extinguished from the population when specialists were N times more efficient than it, in which N is the number of prey. Therefore, the coexistence of all specialists in consuming only one type among all available resources and the generalist is expected to be rare in nature, because it occurs in very restricted conditions (when k = N), as well as the absolute individual specialization or generalization.



Figure 16 – **Different modes with** 3 **prey types**. Predators and prey types dying out are highlighted in black. (A) (k=1.5). For k < 2, both the least fertile prey and its specialist die out (X_C and Y_C), together with the specialist of the second least fertile prey (Y_B). (B) (k=2.5). For 2 < k < 3, only the specialist of the least fertile prey dies out (Y_C). (C) (k=3). For k = 3, it is the threshold case where all predators and prey types coexist. (D) (k=3.5). For k > 3, specialists are efficient enough and the generalist no longer can succeed. Other parameter values: $\omega = 0.2$, $\alpha = 0.5$, $\delta = 1$, $\mu = 0.3$, n = 2.5, p = 1.5, q = 0.5. Source: By the author.

In predator-prey dynamics, if specialists are not present, and there is no heterogeneity among predators, a generalist behavior cannot sustain itself, since it is conducive to extinguish the less fertile prey types. The balance between a generalist and some specialists is necessary to ecologically stabilize the population composed of prey and predators. In addition, the co-occurrence of a generalist with all possible specialists is not stable, because to specialize in preying upon less fertile prey is a bad strategy in face of generalists. Therefore, most populations in nature are expected to be composed of a mixture of some specialists and generalists. This prediction is supported by the average intermediate individual specialization level observed in empirical assessment in populations (reviewed by Araújo et al. 2011).⁹⁴

When new prey types become available for a predator that consumes only one type of resource (*i.e.*, ecological opportunity), it can be a good strategy to exploit the additional prey, even if the predator is less efficient in capturing it. Therefore, in an evolutionary

time, individuals may diverge from the consumption of the prey with the highest rank preference in the population and specialize in exploiting alternative prey types.^{105, 106} This ecological opportunity may occur when competing species are excluded from the community⁸⁷ or when there is variability in other factors, such as patch size, microhabitat diversity, resource diversity, and environmental stability.^{93,94} Based on the current theory of individual specialization, it is expected an increase in individual foraging variation, allowing niche divergence.^{94, 105, 106} However, we found that the addition of new prey types may increase the challenge for specialists overcome the generalist, because they will need to be much more efficient than the generalist. In addition, the generalist is more prone to add a new prey type to its diet than specialists, which will reduce its risk of extinction. Therefore, ecological opportunity is expected to reduce individual specialization instead of increasing it.

The central trait of prey contributing to the success of specialist predators is the reproduction rate. Other traits, such as the nutritious quality of the prey, contribute only to the density of prey themselves, and do not affect the patterns of predation. Furthermore, it is not the absolute intensity of prey reproduction that determines the success of specialist predators, but the relative difference in reproduction rates of available prey. From the point of view of the prey, the dynamics imposes that a high reproduction rate will increase their value as prey; thus, making them more prone to become targets of predators. However, an increased fertility also prevents prey from being extinguished in the presence of generalist predators. In this scenario, the prey must race among themselves to not be the least fertile ones, because they would become candidates for extinction due to the unrestricted growth of generalists who are capable of consuming a wide variety of prey. This is a new insight for advance the theory of individual specialization and guide empiricists to investigate how the relative fertility of prey influence individual variation in diet.

To be successful as a strategy, a generalist must satisfy a lower bound in the efficiency of catching new prey. And as with prey fertility, this lower bound in hunting efficiency is given in terms of the difference of efficiency between different predator strategies, and not the absolute efficiency in hunting prey. The capacity and opportunity to consume more types of prey is beneficial to generalists, and an increased variety of prey types reduces the lower bound for a successful generalist hunting efficiency. For the point of view of specialists, the more pressing goal is to be able to consume the more fertile prey types. In general, our model predicted that populations should present generalists as well as specialists. However, it is important to highlight that other factors may constrain the occurrence of broader generalists that consume the same set of prey types captured by the whole population,⁹³ due to neurological, physiological or morphological constraints.^{97,114–116} Hence, a next step to advance the predictions of our model is to assess the outcomes of predator-prey dynamics when intermediate generalists in prey consumption compete against more restricted specialists in the absence of the broader generalists. Sketching a more theoretical reason for the threshold k = N, we note that, at equilibrium, every prey becomes an equally advantageous source of food for each living individual predator, irrespective of how nutritive it is or how fertile it is. That happens because all the differences are made even when we take into consideration the relative densities of predators and prey oscillating at equilibrium and, therefore, the intensity of competition that they face. Thus, if each specialist has access to 1 source of food, and the generalist has access to N sources, the specialists must be able to exploit that single source at least with N times the efficiency of the generalist to be up to the challenge.

In conclusion, our model of predator-prey dynamics presents a solid starting point for the quantitative assessment of the mechanistic factors behind specialization and generalization of prey consumption. The focus on relative efficiency of predators and relative fertility of prey types can improve the understanding of more complex food web dynamics. Further explorations of this model should explore the generalization to the dynamics featuring intermediate levels of generalists, under more complex predator-prey dynamics containing more types of individuals, and also the relaxation of the constraint imposing an equal efficiency among different specialists. The investigation of trait-evolution in this context, especially of predator efficiency and prey reproduction rate, is also an interesting addition.
8 CONCLUSION

This thesis presented the main narrative weaved in the formative years of my PhD. First of all, it is the development of a robust and comprehensive framework of modeling for population models in biology, with a greater focus on applications to the context of eco-evolutionary dynamics, but also specifically suited to biochemical models. The framework was built over a Bayesian probabilistic paradigm and through the use and expansion of the theory of reaction networks as a design-level modeling method that also yields dynamical differential equations for both stochastic and deterministic scales.

Our main result was the theoretical development and application of the basis of the framework to build a foundation to population models of evolution. By extending the reach of reaction networks, we were able to generalize evolutionary game theory, putting density-dependence under the spotlight, and integrating the evolutionary scale with the ecological scale. With a grounded foundation, we could also incorporate the evolution of traits through adaptive dynamical modeling at the evolutionary scale. As an example of the framework's potential, we exposed and solved the problem of diverging densities arising from the standard hawk-dove game.

We also applied the stochastic and statistical branch of the framework to the analysis of noise, bifurcation, and parameter estimation of the Goodwin model for biochemical oscillations. We solved for the critical feedback strength by both analyzing the bifurcation of the deterministic equations and the divergence of noise in the LNA solution to the stochastic system. We saw how the negative feedback can serve as a feature of noise control in self-regulating genes. By applying a Hill-type feedback, we solved for the minimum Hill exponent, representing a threshold non-linearity, that allows the occurrence of oscillatory behavior, and concluded that genetic systems require additional structure to explain the presence of oscillations, not just what is captured by a Hill feedback. We then explored the problem of Bayesian parameter estimation on the model by using the deterministic system as a data generating process contrasted with simulated data from stochastic samples of the system.

Finally, we applied the eco-evolutionary branch of the framework to the mechanistic assessment of the ecological problem of the evolution of predator-prey behavior, specifically the specialist-generalist dynamics promoting heterogeneity among predation strategies. We found that, contrary to intuition, it is not the prey nutritional value that determines the success of specialists, but relative prey reproductive rates, i.e., specialists should opt to consume prey with the highest reproductive rate. That is especially important in the presence of generalist predators, because they pose the threat of out-competing specialists. But generalists can only survive if specialists are not too good in capturing their chosen prey, leaving room for generalists to take their share.

Overall, the framework functions at least as a solid base over which to build models, ranging from design, to mathematical description, and to statistical assessment of data. All by a fully integrated and fairly automated methodology. Future developments include a refinement of the parameter estimation process, especially to include stochasticbased likelihood, for example using the LNA solution, and also the continuous use of the framework to address problems in theoretical ecology, evolutionary dynamics, and biochemistry.

REFERENCES

1 KRAKAUER, D. C. *et al.* The challenges and scope of theoretical biology. **Journal of Theoretical Biology**, v. 276, n. 1, p. 269–276, May 2011. DOI: 10.1016/j.jtbi.2011.01.051.

2 DOEBELI, M.; ISPOLATOV, Y.; SIMON, B. Towards a mechanistic foundation of evolutionary theory. **eLife**, v. 6, Feb. 2017. DOI: 10.7554/eLife.23804.

3 LAZEBNIK, Y. Can a biologist fix a radio?—Or, what I learned while studying apoptosis. **Cancer Cell**, v. 2, n. 3, p 179-82, Sept. 2002. DOI: 10.1016/S1535-6108(02)00133-2.

4 GERITZ, S. A. H.; KISDI, E. Mathematical ecology: why mechanistic models? **Journal of Mathematical Biology**, v. 65, n. 6-7, p. 1411–1415, Dec. 2012. DOI: 10.1007/s00285-011-0496-3.

5 GREGORY, P. C. P. C. Bayesian logical data analysis for the physical sciences: a comparative approach with Mathematica support. Cambridge: Cambridge University Press, 2005. 468 p.

6 GARDINER, C. Handbook of stochastic methods: for physics, chemistry, and the natural sciences. 3rd ed. Switzerland: Springer, 2004.

7 JAYNES, E. T. **Probability theory:** the logic of science. Cambridge: Cambridge University Press, 2003.

8 FEINBERG, M. Foundations of chemical reaction network theory. Cham: Springer International Publishing, 2019. v. 202. DOI: 10.1007/978-3-030-03858-8.

9 BAEZ, J. C.; POLLARD, B. S. A compositional framework for reaction networks. Reviews in Mathematical Physics, v. 29, n. 09, p. 1750028, Oct. 2017. DOI: 10.1142/S0129055X17500283.

10 VELOZ, T.; RAZETO-BARRY, P.; DITTRICH, P.; FAJARDO, A. Reaction networks and evolutionary game theory. **Journal of Mathematical Biology**, v. 68, n. 1-2, p. 181–206, Jan. 2014. DOI: 10.1007/s00285-012-0626-6.

11 GELMAN, A. et al. Bayesian data analysis. 3rd ed. Boca Raton: CRC, 2013.

12 VELOZ, T.; RAZETO-BARRY, P. Reaction networks as a language for systemic modeling: on the study of structural changes. **Systems**, v. 5, n. 2, p. 30, Mar. 2017. DOI: 10.3390/systems5020030.

13 SCHNOERR, D.; SANGUINETTI, G.; GRIMA, R. Approximation and inference methods for stochastic biochemical kinetics - a tutorial review . **Journal**

of Physics A: mathematical and theoretical, v. 50, n. 9, p. 093001, 2017. DOI: 10.1088/1751-8121/aab38b.

14 KAMPEN N.G. Van. Stochastic processes in physics and chemistry. 3rd ed. Amsterdam: Elsevier, 2007.

15 GILLESPIE, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational **Physics**, v. 22, n. 4, Dec. 1976. DOI: 10.1016/0021-9991(76)90041-3.

16 GILLESPIE, D. T. Stochastic simulation of chemical kinetics. Annual Review of Physical Chemistry, v. 58, n. 1, p. 35, 2007. DOI: 10.1146/annurev.physchem.58.032806.104637.

17 RISKEN, H. **The fokker-planck equation**. Berlin, Heidelberg: Springer, 1984. (Springer series in synergetics, v. 18). ISBN 978-3-540-61530-9.

18 SCHNOERR, D.; SANGUINETTI, G.; GRIMA, R. Comparison of different moment-closure approximations for stochastic chemical kinetics. Journal of Chemical Physics, v. 143, n. 18, p. 11, 2015. DOI: 10.1088/1751-8121/aa54d9.

19 BROOKS, S. *et al.* Handbook of Markov Chain Monte Carlo. Boca Raton: Chapman and Hall/CRC. 2011. ISBN 9781420079418.

20 GAMERMAN, D; LOPES, H. F. Markov Chain Monte Carlo. 2nd ed. Boca Raton: Chapman and Hall/CRC. 2006. ISBN 9781584885870.

21 CARPENTER, B. *et al.* Stan: a probabilistic programming language. Journal of Statistical Software, v. 76, n. 1, 2017. DOI: 10.18637/jss.v076.i01.

22 BETANCOURT, M. A conceptual introduction to Hamiltonian Monte Carlo. Jan. 2017. Available from <u>http://arxiv.org/abs/1701.02434</u>. Accessible at: 25 June. 2021.

23 HOFFMAN, M. D.; GELMAN, A. The No-U-Turn Sampler: adaptively setting path lengths in hamiltonian monte carlo. Journal of Machine Learning Research, v. 15, n. 47, p. 1593–1623, 2014.

24 KLIPP, E. *et al.* Systems biology in practice. Weinheim: Wiley, 2005. DOI: 10.1002/3527603603.

25 BRESSLOFF, P. C. Stochastic processes in cell biology. 2nd ed. Switzerland: Springer International Publishing, 2021. v. 41. DOI: 10.1007/978-3-030-72515-0. 26 MACKEY, M. *et al.* The utility of simple mathematical models in understanding gene regulatory dynamics. **In Silico Biology**, v. 12, n. 1,2, July 2015. DOI: 10.3233/ISB-140463.

27 INGALLS, B. Mathematical modeling in systems biology: an introduction. Cambridge: MIT Press, 2013, 424 p. ISBN 9780262018883.

28 ALON, U. An introduction to systems biology. Boca Raton: Chapman and Hall/CRC, 2006. DOI: 10.1201/9781420011432.

29 ALBERTS, B. *et al.* Essential cell biology. 5th ed. New York: Garland Science, 2015. ISBN 978-0-393-69109-2.

30 HASELTINE, E. L.; ARNOLD, F. H. Synthetic gene circuits: design with directed evolution. Annual Review of Biophysics and Biomolecular Structure, v. 36, n. 1, June 2007. DOI: 10.1146/annurev.biophys.36.040306.132600.

31 ELOWITZ, M. B.; LEIBLER, S. A synthetic oscillatory network of transcriptional regulators. **Nature**, v. 403, n. 6767, Jan. 2000. DOI: 10.1038/35002125.

32 ELDAR, A.; ELOWITZ, M. B. Functional roles for noise in genetic circuits. **Nature**, v. 467, n. 7312, p. 167-173, Sept. 2010. ISSN 0028-0836.

33 MUNSKY, B.; NEUERT, G.; OUDENAARDEN, A. van. Using gene expression noise to understand gene regulation. **Science**, v. 336, n. 6078, Apr. 2012. DOI: 10.1126/science.1216379.

34 THATTAI, M.; OUDENAARDEN, A. van. Intrinsic noise in gene regulatory networks. **Proceedings of the National Academy of Sciences**, v. 98, n. 15, p. 8614–8619, July 2001. DOI: 10.1073/pnas.151588598.

35 SANCHEZ, A.; CHOUBEY, S.; KONDEV, J. Regulation of noise in gene expression. **Annual Review of Biophysics**, v. 42, n. 1, May 2013. DOI: 10.1146/annurev-biophys-083012-130401.

36 RASER, J. M. Noise in gene expression: origins, consequences, and control. Science, v. 309, n. 5743, Sept. 2005. DOI: 10.1126/science.1105891.

37 RAJ, A.; OUDENAARDEN, A. van. Nature, nurture, or chance: stochastic gene expression and its consequences. **Cell**, v. 135, n. 2, Oct. 2008. DOI: 10.1016/j.cell.2008.09.050..

38 MAHESHRI, N.; O'SHEA, E. K. Living with noisy genes: how cells function reliably with inherent variability in gene expression. **Annual Review of**

Biophysics and Biomolecular Structure, v. 36, n. 1, June 2007. DOI: 10.1146/annurev.biophys.36.040306.132705.

39 AUSTIN, D. W. *et al.* Gene network shaping of inherent noise spectra. **Nature**, v. 439, n. 7076, Feb. 2006. DOI: 10.1038/nature04194.

40 KÆRN, M.; ELSTON, T. C.; BLAKE, W. J.; COLLINS, J. J. Stochasticity in gene expression: from theories to phenotypes. **Nature Reviews Genetics**, v. 6, n. 6, June 2005. DOI: 10.1038/nrg1615.

41 MCADAMS, H. H.; ARKIN, A. It's a noisy business! genetic regulation at the nanomolar scale. **Trends in Genetics**, v. 15, n. 2, Feb. 1999. DOI: 10.1016/S0168-9525(98)01659-X.

42 NOVÁK, B.; TYSON, J. J. Design principles of biochemical oscillators. **Nature Reviews Molecular Cell Biology,** v. 9, n. 12, p 981-991, 2008. DOI: 10.1038/nrm2530.

43 GONZE, D.; RUOFF, P. The Goodwin oscillator and its legacy. Acta Biotheoretica, Mar. 2020. DOI: 10.1007/s10441-020-09379-8.

44 GOODWIN, B. C. Oscillatory behavior in enzymatic control processes. Advances in Enzyme Regulation, v. 3, p. 425-428, 1965. DOI: 10.1016/0065-2571(65)90067-1.

45 WOLLER, A.; GONZE, D.; ERNEUX, T. The Goodwin model revisited: Hopf bifurcation, limit-cycle, and periodic entrainment. **Physical Biology**, v. 11, n. 4, Aug. 2014. DOI: 10.1088/1478-3975/11/4/045002.

46 STROGATZ, S. H. Nonlinear dynamics and chaos : with applications to physics, biology, chemistry, and engineering. 2nd ed. Boca Raton: CRC Press, 2015. ISBN 9780813349107.

47 GILLESPIE, D. T. A rigorous derivation of the chemical master equation. **Physica A:** statistical mechanics and its applications, v. 188, n. 1-3, Sept. 1992. DOI: 10.1016/0378-4371(92)90283-V.

48 TAO, Y.; JIA, Y.; DEWEY, T. G. Stochastic fluctuations in gene expression far from equilibrium: Ω expansion and linear noise approximation. Journal of Chemical Physics, v. 122, n. 12, Mar. 2005. DOI: 10.1063/1.1870874.

49 OZBUDAK, E. M.; THATTAI, M.; KURTSER, I.; GROSSMAN, A. D.; OUDENAARDEN, A. van. Regulation of noise in the expression of a single gene. **Nature Genetics**, v. 31, n. 1, May 2002. DOI: 10.1146/annurev-biophys-083012-130401.

50 BRESSLOFF, P. C. Stochastic switching in biology: from genotype to phenotype. **Journal of Physics A:** mathematical and theoretical, v. 50, n. 13, Mar. 2017. DOI: 10.1088/1751-8121/aa5db4.

51 SANTILLÁN, M. On the use of the hill functions in mathematical models of gene regulatory networks. Mathematical Modelling of Natural Phenomena, v. 3, n. 2, p. 85–97, Jan. 2008. DOI: 10.1051/mmnp:2008056.

52 SMITH, S.; GRIMA, R. Model reduction for stochastic reaction systems. *In:* HOLCMAN, D. (ed.). Stochastic processes, multiscale modeling, and numerical methods for computational cellular biology. Cham: Springer International Publishing, 2017. DOI: 10.1007/978-3-319-62627-7_7.

53 PAULSSON, J. Models of stochastic gene expression. **Physics of Life Reviews**, v. 2, n. 2, June 2005. DOI: 10.1016/j.plrev.2005.03.003.

54 FRANZ, S. Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools. Journal of Physics A: mathematical and general, v. 37, n. 40, Oct. 2004. DOI: 10.1088/0305-4470/37/40/B03.

55 WOLLER, A.; GONZE, D.; ERNEUX, T. Strong feedback limit of the Goodwin circadian oscillator. **Physical Review E -** statistical, nonlinear, and soft matter physics, v. 87, n. 3, Mar. 2013. DOI: 10.1103/PhysRevE.87.032722.

56 KIM, J. K. Protein sequestration versus Hill-type repression in circadian clock models. **IET Systems Biology**, v. 10, n. 4, Aug. 2016. DOI: 10.1049/iet-syb.2015.0090.

57 KOMOROWSKI, M.; FINKENSTÄDT, B.; HARPER, C. V.; RAND, D. A. Bayesian inference of biochemical kinetic parameters using the linear noise approximation. **BMC Bioinformatics**, v. 10, n. 1, p. 343, Dec. 2009. DOI: 10.1186/1471-2105-10-343.

58 DIECKMANN, U.; LAW, R. The dynamical theory of coevolution: a derivation from stochastic ecological processes. **Journal of Mathematical Biology**, v. 34, n. 5-6, p. 579–612, May 1996. DOI: 10.1007/BF02409751.

59 PELLETIER, F.; GARANT, D.; HENDRY, A. Eco-evolutionary dynamics. Philosophical Transactions of the Royal Society B: biological sciences, v. 364, n. 1523, p. 1483–1489, June 2009. DOI: 10.1098/rstb.2009.0027.

60 BERTRAM, J.; MASEL, J. Density-dependent selection and the limits of relative fitness. **Theoretical Population Biology**, v. 129, p. 81–92, Oct. 2019. DOI: 10.1016/j.tpb.2018.11.006.

61 CHAMPAGNAT, N.; FERRIÈRE, R.; MÉLÉARD, S. Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. **Theoretical Population Biology**, v. 69, n. 3, p. 297–321, May 2006. DOI: 10.1016/j.tpb.2005.06.009.

62 HUANG, W.; HAUERT, C.; TRAULSEN, A. Stochastic game dynamics under demographic fluctuations. **Proceedings of the National Academy of Sciences**, v. 112, n. 29, p. 9064–9069, July 2015. DOI: 10.1073/pnas.1418745112.

63 ARGASINSKI, K.; BROOM, M. Interaction rates, vital rates, background fitness and replicator dynamics: how to embed evolutionary game structure into realistic population dynamics. **Theory in Biosciences**, v. 137, n. 1, p. 33–50, Apr. 2018. DOI: 10.1007/s12064-017-0257-y.

64 VEELEN, M. van. The problem with the Price equation. **Philosophical Transactions of the Royal Society B:** biological sciences, v. 375, n. 1797, p. 20190355, Apr. 2020. DOI: 10.1098/rstb.2019.0355.

65 PAGE, K. M.; NOWAK, M. A. Unifying evolutionary dynamics. Journal of Theoretical Biology, v. 219, n. 1, p. 93–98, Nov. 2002. DOI: 10.1006/jtbi.2002.3112.

66 ANGELI, D. A tutorial on chemical reaction network dynamics. **European Journal of Control**, v. 15, n. 3-4, p. 398–406, Jan. 2009. DOI: 10.3166/ejc.15.398-406.

67 FEARNHEAD, P.; GIAGOS, V.; SHERLOCK, C. Inference for reaction networks using the linear noise approximation. **Biometrics**, v. 70, n. 2, p. 457–466, June 2014. DOI: 10.1111/biom.12152.

68 TAYLOR, C.; NOWAK, M. A. Evolutionary game dynamics with non-uniform interaction rates. **Theoretical Population Biology**, v. 69, n. 3, p. 243–252, May 2006. DOI: 10.1016/j.tpb.2005.06.009.

69 KAMPEN, N. V. The master equation. *In:* KAMPEN, N. G. van **Stochastic processes in physics and chemistry**. Amsterdam: Elsevier, 2007. DOI: 10.1016/B978-044452965-7/50008-8.

70 KAMPEN, N. V. The expansion of the master equation. *In:* KAMPEN, N. G. van **Stochastic processes in physics and chemistry**. Amsterdam: Elsevier, 2007. DOI: 10.1016/B978-044452965-7/50013-1.

71 KURTZ, T. G. The Relationship between stochastic and deterministic models for chemical reactions. **The Journal of Chemical Physics**, v. 57, n. 7, p. 2976–2978, Oct. 1972. DOI: 10.1063/1.1678692.

72 HOFBAUER, J.; SIGMUND, K. Evolutionary game dynamics. Bulletin of the American Mathematical Society, v. 40, n. 04, July 2003. DOI: 10.1090/S0273-0979-03-00988-1.

73 SMITH, J. M.; PRICE, G. R. The logic of animal conflict. **Nature**, v. 246, n. 5427, Nov. 1973. DOI: 10.1038/246015a0.

74 VEELEN, M. van. On the use of the Price equation. Journal of Theoretical Biology, v. 237, n. 4, p. 412–426, Dec. 2005. DOI: 10.1016/j.jtbi.2005.04.026.

75 BRÄNNSTRÖM, A.; JOHANSSON, J.; FESTENBERG, N. von. The hitchhiker's guide to adaptive dynamics. **Games**, v. 4, n. 3, p. 304–328, June 2013. DOI: 10.3390/g4030304.

76 BESSA-GOMES, C.; LEGENDRE, S.; CLOBERT, J. Discrete two-sex models of population dynamics: on modelling the mating function. Acta Oecologica, v. 36, n. 5, p. 439–445, Sept. 2010. DOI: 10.1016/j.actao.2010.02.010.

77 ARDITI, R.; GINZBURG, L. R. Coupling in predator-prey dynamics: ratiodependence. **Journal of Theoretical Biology**, v. 139, n. 3, Aug. 1989. DOI: 10.1016/S0022-5193(89)80211-5.

78 ABRAMS, P. A.; GINZBURG, L. R. The nature of predation: prey dependent, ratio dependent or neither? **Trends in Ecology & Evolution**, v. 15, n. 8, p. 337–341, Aug. 2000. DOI: 10.1016/S0169-5347(00)01908-X.

79 KOKKO, H.; JENNIONS, M. D. Parental investment, sexual selection and sex ratios. **Journal of Evolutionary Biology**, v. 21, n. 4, p. 919–948, July 2008. DOI: 10.1111/j.1420-9101.2008.01540.x.

80 FROMHAGE, L.; JENNIONS, M. D. Coevolution of parental investment and sexually selected traits drives sex-role divergence. **Nature Communications**, v. 7, n. 1, p. 12517, Nov. 2016. DOI: 10.1098/rspb.1992.0141.

81 ROSSBERG, A. G. *et al.* Let's train more theoretical ecologists – here is why. **Trends in Ecology & Evolution**, v. 34, n. 9, p. 759–762, Sept. 2019. DOI: 10.1016/j.tree.2019.06.004.

82 GOLIGHTLY, A.; WILKINSON, D. J. bayesian sequential inference for stochastic kinetic biochemical network models. Journal of Computational Biology, v. 13, n. 3, p. 838–851, Apr. 2006. DOI: 10.1089/cmb.2006.13.838.

83 VEELEN, M. van; GARCÍA, J.; SABELIS, M. W.; EGAS, M. Group selection and inclusive fitness are not equivalent; the Price equation vs. models and statistics. **Journal of Theoretical Biology**, v. 299, p. 64–80, Apr. 2012. DOI: 10.1016/j.jtbi.2011.07.025. 84 BIRCH, J. The inclusive fitness controversy: finding a way forward. **Royal** Society Open Science, v. 4, n. 7, p. 170335, 7 2017. DOI: 10.1098/rsos.170335.

85 DARWIN, C. The origin of species by means of natural selection: or, the preservation of favored races in the struggle for life. New York: AL Burt, 2009.

86 FUTUYMA, D. J. **Evolution** . 3rd ed. Sunderland, Massachusetts: Sinauer Associates Inc., 2013. ISBN 9781605351155.

87 COSTA-PEREIRA, R.; RUDOLF, V. H. W.; SOUZA, F. L.; ARAÚJO, M. S. Drivers of individual niche variation in coexisting species. **Journal of Animal Ecology**, v. 87, n. 5, Sept. 2018. DOI: 10.1111/1365-2656.12879.

88 INGRAM, T.; COSTA-PEREIRA, R.; ARAÚJO, M. S. The dimensionality of individual niche variation. **Ecology**, v. 99, n. 3, Mar. 2018. DOI: 10.1002/ecy.2129.

89 CHASE, J. M.; LEIBOLD, M. A. Ecological niches: linking classical and contemporary approaches. **Biodiversity and Conservation**, v. 13, n. 9, Aug. 2004. DOI: 10.1023/B:BIOC.0000029366.24837.fc.

90 TRAPPES, R. *et al.* **How individualized niches arise**: mechanisms of niche construction, niche choice, and niche conformance. 2021. Available from <u>https://ecoevorxiv.org/wahcy/</u>. Accessible at: 25 June 2021.

91 HUTCHINSON, G. E. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology, v. 22, Jan. 1957. DOI: 10.1101/SQB.1957.022.01.039.

92 TAKOLA, E.; SCHIELZETH, H. Hutchinson's ecological niche for individuals. 2021. Available from: <u>https://ecoevorxiv.org/zrcbg/</u>. Accessible at: 25 June 2021.

93 BOLNICK, D. *et al.* The ecology of individuals: incidence and implications of individual specialization. **American Naturalist**, v. 161, n. 1, Jan. 2003. DOI: 10.1086/343878.

94 ARAÚJO, M. S.; BOLNICK, D. I.; LAYMAN, C. A. The ecological causes of individual specialisation. **Ecology Letters**, v. 14, n. 9, Sept. 2011. DOI: 10.1111/j.1461-0248.2011.01662.x.

95 SCHOENER, T. Resource partitioning. *In:* KIKKAWA, J.; ANDERSON, D. J. (eds.). **Community Ecology** – pattern and process. Oxford: Blackwell Scientific Publications, 1987.

96 DALL, S. R. X.; BELL, A. M.; BOLNICK, D. I.; RATNIEKS, F. L. W. An evolutionary ecology of individual differences. **Ecology Letters**, v. 15, n. 10, Oct. 2012. DOI: 10.1111/j.1461-0248.2012.01846.x.

97 ARAÚJO, M. S.; GONZAGA, M. O. Individual specialization in the hunting wasp Trypoxylon (Trypargilum) albonigrum (Hymenoptera, Crabronidae).
Behavioral Ecology and Sociobiology, v. 61, n. 12, Sept. 2007. DOI: 10.1007/s00265-007-0425-z.

98 ARAÚJO, M. S.; COSTA-PEREIRA, R. Latitudinal gradients in intraspecific ecological diversity. **Biology Letters**, v. 9, n. 6, Dec. 2013. DOI: 10.1098/rsbl.2013.0778.

99 EVANGELISTA, C.; BOICHE, A.; LECERF, A.; CUCHEROUSSET, J.
Ecological opportunities and intraspecific competition alter trophic niche specialization in an opportunistic stream predator. Journal of Animal Ecology, v. 83, n. 5, Sept. 2014. DOI: 10.1111/1365-2656.12208.

100 NEWSOME, S. D. *et al.* The interaction of intraspecific competition and habitat on individual diet specialization: a near range-wide examination of sea otters. **Oecologia**, v. 178, n. 1, May 2015. DOI: 10.1007/s00442-015-3223-8.

101 COSTA-PEREIRA, R.; TAVARES, L. E. R.; CAMARGO, P. B. de; ARAÚJO, M. S. Seasonal population and individual niche dynamics in a tetra fish in the Pantanal wetlands. **Biotropica**, v. 49, n. 4, July 2017. DOI: 10.1111/btp.12434.

102 POWELL, E. C.; TAYLOR, L. A. Specialists and generalists coexist within a population of spider-hunting mud dauber wasps. **Behavioral Ecology**, v. 28, n. 3, May 2017. DOI: 10.1093/beheco/arx050.

103 MARCHETTI, K. Host-race formation in the common cuckoo. Science, v. 282, n. 5388, Oct. 1998. DOI: 10.1126/science.282.5388.471.

104 OTTERBECK, A.; LINDÉN, A.; ROUALET, E. Advantage of specialism: reproductive output is related to prey choice in a small raptor. **Oecologia**, v. 179, n. 1, Sept. 2015. DOI: 10.1007/s00442-015-3320-8.

105 SVANBÄCK, R.; BOLNICK, D. I. Intraspecific competition drives increased resource use diversity within a natural population. **Proceedings of the Royal Society B:** biological sciences, v. 274, n. 1611, Mar. 2007. DOI: 10.1098/rspb.2006.0198.

106 SVANBÄCK R., B. D. Intraspecific competition affects the strength of individual specialization: an optimal diet theory method. **Evolutionary Ecology Research**, v. 7, n. 7, p. 993–1012, 2005.

107 ABRAMS, P. A. The evolution of predator-prey interactions: theory and evidence. **Annual Review of Ecology and Systematics**, v. 31, n. 1, Nov. 2000. DOI: 10.1098/rspb.1992.0141.

108 EKLÖF, A.; STOUFFER, D. B. The phylogenetic component of food web structure and intervality. **Theoretical Ecology**, v. 9, n. 1, Feb. 2016. DOI: 10.1007/s12080-015-0273-9.

109 BROUSSEAU, P.; GRAVEL, D.; HANDA, I. T. Trait matching and phylogeny as predictors of predator-prey interactions involving ground beetles. **Functional Ecology**, v. 32, n. 1, Jan. 2018. DOI: 10.1111/1365-2435.12943.

110 MOURA, R. R.; PITILIN, R. B.; BUSCHINI, M. L. Specialisation in prey capture drives coexistence among sympatric spider-hunting wasps. **Ecological Entomology**, v. 44, n. 5, Oct. 2019. DOI: 10.1111/een.12743.

111 ESTES, J. A.; RIEDMAN, M. L.; STAEDLER, M. M.; TINKER, M. T.; LYON, B. E. Individual variation in prey selection by sea otters: patterns, causes and implications. **Journal of Animal Ecology**, v. 72, n. 1, Jan. 2003. DOI: 10.1046/j.1365-2656.2003.00690.x.

112 TINKER, M. T.; BENTALL, G.; ESTES, J. A. Food limitation leads to behavioral diversification and dietary specialization in sea otters. **Proceedings of the National Academy of Sciences**, v. 105, n. 2, Jan. 2008. DOI: 10.1073/pnas.0709263105.

113 HAN, C. S.; JÄGER, H. Y.; DINGEMANSE, N. J. Individuality in nutritional preferences: a multi-level approach in field crickets. **Scientific Reports**, v. 6, n. 1, July 2016. DOI: 10.1038/srep29071.

114 BENCE, J. R. Feeding rate and attack specialization: the roles of predator experience and energetic tradeoffs. Environmental Biology of Fishes, v. 16, n.
1-3, June 1986. DOI: 10.1007/BF00005164.

115 HOWARD, J. J. Temporal pattern of resource use and variation in diets of individual grasshoppers (Orthoptera: Acrididae). Journal of Insect Behavior, v. 6, n. 4, July 1993. DOI: 10.1007/BF01049524.

116 FERMON, Y.; CIBERT, C. Ecomorphological individual variation in a population of Haplochromis nyererei from the Tanzanian part of Lake Victoria. **Journal of Fish Biology**, v. 53, n. 1, July 1998. DOI: 10.1111/j.1095-8649.1998.tb00110.x.

Appendix

APPENDIX A – STAN MODELS

The models in STAN are composed of blocks that allow us to define functions, specify the data to be inputted, define parameters and generated quantities, and also the statistical model itself. In our applications, the functions we define are the deterministic systems of differential equations yielded by reaction networks, and the model is simply composed by the prior definitions and the data likelihood calculated over the integrated deterministic model. In our case, the distribution of the likelihood is a normal distribution, and we use exponential priors. The STAN code was processed using the R language with the RSTAN interface.

A.1 Example Model of Lotka-Volterra

}

```
functions {
  real[] LV(real t, real[] y, real[] theta, real[] x_r, int[] x_i) {
    real dydt[2];
    real X = y[1];
    real Y = y[2];
    real w = theta[1];
    real g = theta[2];
    real d = theta[3];
    real u = theta[4];
    dydt[1] = X*(w-g*Y);
    dydt[2] = Y*(d*g*X-u);
    return dydt;
  }
}
data {
  int<lower=1> N;
  real ts[N];
  real y[N, 2];
  real t0;
```

```
transformed data {
  real x_r[0];
  int x_i[0];
}
parameters {
  real<lower=0> y0[2];
  vector<lower=0>[2] sigma;
  real<lower=0> theta[4];
}
model {
  real mu[N,2];
  mu = integrate_ode_rk45(LV, y0, t0, ts, theta, x_r, x_i);
  y0 ~ exponential(0.1);
  sigma ~ exponential(2);
  theta[1] ~ exponential(2);
  theta[2] ~ exponential(10);
  theta[3] ~ exponential(1);
  theta[4] ~ exponential(2);
  for (t in 1:N)
    y[t] ~ normal(mu[t], sigma);
}
```

A.2 3D Goodwin Model

```
functions {
  real[] G3D(real t, real[] y, real[] theta, real[] x_r, int[] x_i) {
    real dydt[3];
  real X = y[1];
  real Y = y[2];
  real Z = y[3];
  real b = theta[1];
  real d = theta[2];
```

```
real e = theta[3];
    real g = theta[4];
    real m = theta[5];
    real k = theta[6];
    real K = \text{theta}[7];
    dydt[1] = k / (1 + K * Z ^ m) - g * X;
    dydt[2] = b * g * X - e * g * Y;
    dydt[3] = d * g * Y - e * g * Z;
    return dydt;
  }
}
data {
  int<lower=1> N;
  real ts[N];
  real y[N, 3];
  real t0;
}
transformed data {
  real x_r[0];
  int x_i[0];
}
parameters {
  vector<lower=0>[3] sigma;
  real<lower=0> theta[7];
}
transformed parameters {
  real<lower=0> y0[3];
  y0[1] = 1;
  y0[2] = 0;
  y0[3] = 0;
}
model {
```

```
real mu[N,3];
mu = integrate_ode_rk45(G3D, y0, t0, ts, theta, x_r, x_i);
sigma ~ exponential(2);
theta[1] ~ exponential(1);
theta[2] ~ exponential(1);
theta[3] ~ exponential(1);
theta[4] ~ exponential(1);
theta[5] ~ exponential(1);
theta[6] ~ exponential(1);
theta[7] ~ exponential(1);
for (t in 1:N)
   y[t] ~ normal(mu[t], sigma);
}
```