Information Theoretic Framework for Stochastic Sensitivity and Specificity Analysis in Biochemical Networks

A thesis presented for the degree of

Doctor of Philosophy of Imperial College London

by

Qurat-ul-Ain Azim

Department of Mathematics

Imperial College

180 Queen's Gate, London SW7 2AZ

AUGUST 25, 2016

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Qurat-ul-Ain Azim

Copyright

To my beloved grandfather, *Noor Muhammad (late)*

Abstract

Biochemical reaction networks involve many chemical species and are inherently stochastic and complex in nature. Reliable and organised functioning of such systems in varied environments requires that their behaviour is robust with respect to certain parameters while sensitive to other variations, and that they exhibit specific responses to various stimuli. There is a continuous need for improved models and methodologies to unravel the complex behaviour of the dynamics of such systems. In this thesis, we apply ideas from information theory to develop novel methods to study properties of biochemical networks.

In the first part of the thesis, a framework for the study of parametric sensitivity in stochastic models of biochemical networks using entropies and mutual information is developed. The concept of noise entropy is introduced and its interplay with parametric sensitivity is studied as the system becomes more stochastic. Using the methodology for gene expression models, it is shown that noise can change the sensitivities of the system at various orders of parameter interaction. An approximate and computationally more efficient way of calculating the sensitivities is also developed using unscented transform. Finally, the methodology is applied to a circadian clock model, illustrating the applicability of the approach to more complex systems.

In the second part of the thesis, a novel method for specificity quantification in a receptor-ligand binding system is proposed in terms of mutual information estimates between appropriate stimulus and system response. The maximum specificity of 2×2 affinity matrices in a parametric setup is theoretically studied. Parameter optimisation methodology and specificity upper bounds are presented for maximum specificity estimates of a given affinity matrix. The quantification framework is then applied to experimental data from T-Cell signalling. Finally, generalisation of the scheme for stochastic systems is discussed.

Acknowledgements

My first and foremost thanks are due to my PhD supervisor Dr Vahid Shahrezaei who has been very kind and supportive throughout the duration of my PhD study. It would not have been possible for me to complete this work without his continuous guidance and valuable supervision. He has been very compassionate through all the ups and downs of my PhD journey. His understanding and support for me especially as a new parent has been invaluable. Words cannot do justice to the appreciation and gratitude I hold for him as a supervisor, a mentor and a great human being.

I am grateful to my sponsor, COMSATS Institute of information Technology, Pakistan, for funding my PhD study. Mathematics Department at Imperial College has been very helpful in providing the right kind of environment for research, and I am very thankful for that.

I am also thankful to Professor Greg Pavliotis for his guidance in steering my PhD towards the right direction. I would also like to thank Dr Omer Dushek and Dr Enas Abu Shah from Oxford University for sharing their experimental data with me and for their many helpful discussions and comments on my work.

My sincere thanks are due to my beloved parents, Muhammad Azim and Nasreen Akhtar for their unconditional love, continuous support and constant encouragement. They have been especially helpful during my thesis write-up. I can never thank them enough for supporting me and my son during this crucial period of my PhD and of my life. Their confidence in me is what has brought me to eventually submitting my thesis. Their incomparable love for my son has been immensely helpful in making up for the lack of my attention for him due to my studies, and to keep me going without much worry.

I would also like to thank Daria, Yuko, Ludi and Kasia; my son's preschool carers. Their childcare and supervision for a very important year of my son's life gave me confidence and reassurance, and enabled me to study peacefully.

The love, support and care from my precious friends Sara Tariq, Usman Adeel, Nazia Shareef and Abid Rafique has been pivotal for keeping me sane during the various phases of my PhD. I am grateful to to them for always being there for me, and for helping me a lot especially during my pregnancy and new mother phase. My friend and office mate Din-Houn Lau has been very helpful whenever I have needed him, particularly for taking care of the problems with my computer which I accessed remotely while working from home.

I would also like to thank my uncle Abdul Majid, aunt Nighat Majid, and their kids Shaiza Majid and Israar Majid. The family is like my own family, and their home is like a home away from home. The comfort that I felt having them nearby kept me from being homesick.

If there is one thing that has never changed in any way during my PhD, is the precious love and valuable support I have been receiving from my amazing husband, Muhammad Usman. He has been a great source of motivation and happiness for me, and an awesome father to my son. He has never let me give up, always picked me up at all the difficult times and got me going again.

Last but not least, I am thankful to my precious little son, Muhammad Usayd Bin Usman, for making my life beautiful and worthwhile with his presence. Although my PhD would have been far quicker without him, the challenges I faced with him around have only made me stronger as a person. He has been a source of motivation for me because I hope that some day he can be proud of me and learn from me that I did not give up.

Table of contents

Chapter 1

Motivation and Outline

1.1 Stochasticity in Biochemical Networks

A system is said to be stochastic if its dynamics are partly random. This may be caused by a random force acting at different times or a force acting at random times. In this case, the state of the system cannot be completely determined given the state at a particular time. The best we can do is use the probability that the system is in a certain state and predict the evolution of the probability at later times. This calculation is often very difficult or expensive. Therefore, we usually need to use approximations or computations.

Cellular behaviour that mainly originates from the dynamics of the biochemical reactions is stochastic in nature [107, 113]. Stochasticity arises at the level of intermolecular collisions, which are driven by diffusion. These collisions vary the propensities of the molecules to react, thereby causing the individual reaction events to occur randomly. Many reviews on stochasticity and its modeling in biochemical networks have been presented in literature [41, 107, 131].

Different biochemical networks have different levels of stochasticity. The systems with low copy numbers are generally more stochastic in nature as the random timing of individual reactions is not averaged out. Low copy numbers occur frequently in single cells. For example gene copy number is usually one or two, and transcription factors are often tens in number.

Cells are known to control and exploit stochasticity. Cellular network designs evolve to

reduce fluctuations, which may corrupt the information received by cells and affect required robustness and reliability in cellular behaviour. On the other hand, cellular stochasticity and phenotypic variability can be advantageous in fluctuating environments. Bacteria that have the ability to switch stochastically between states have been known to outperform bacteria that pay the metabolic costs of sensing the environment and then matching the environmental state, thus having an evolutionary advantage [72].

Quantifying stochasticity is not easy and requires measurements at the level of single cells. Usually 'noise' is referred to as the empirical measure of stochasticity. For example, protein noise can be measured by introducing a fluorescent protein downstream of the promoter under consideration and then measuring the coefficient of variation of fluorescence across a population of cells, or over time assuming the system is ergodic. Noise is a measure of the magnitude of a typical fluctuation as a fraction of the mean and is a dimensionless quantity. If N represents a random variable, then noise η can mathematically be defined as

$$\eta = \frac{\sqrt{\langle N^2 \rangle - \langle N \rangle^2}}{\langle N \rangle}$$

1.2 Intrinsic and Extrinsic Stochasticity

Biological noise is commonly divided into two parts, *intrinsic noise* and *extrinsic noise* [115]. The square of total noise in a system can be calculated by taking the sum of squares of intrinsic and extrinsic noise. The part of stochasticity generated by the dynamics of the system under study is called *intrinsic stochasticity*. Variations in the timings of individual reactions give rise to intrinsic noise. Intrinsic noise is enhanced if the number of molecules in the system are low. On the other hand, if the system of interest interacts with other stochastic systems in the environment, the resulting stochasticity is referred to as *extrinsic stochasticity*. If two copies of the system of interest can be created in the same cellular environment as the original system and lodged with different reporters, as pioneered by Elowitz *et al.* [22] to measure noise in gene expression, then both forms of stochasticity can be quantified.

Intrinsic and extrinsic stochasticity are generated by fluctuations in intrinsic and extrin-

sic variables respectively. The intrinsic variables of a system are generally used to describe the copy numbers of the molecular components of the system. On the other hand, extrinsic variables are related to the processes that control parameters of the system. For example in gene expression, the number of mRNAs, number of proteins, and the number of transcribing RNA polymerases are all intrinsic variables. The number of cytosolic RNA polymerases, however, is an extrinsic variable since increase in their number increases the level of gene expression in each copy of the system. Intrinsic noise is a measure of intrinsic stochasticity. It can be quantified by the difference in the values of an intrinsic variable in two copies of the system. Extrinsic noise, on the other hand, is given by the correlation coefficient between intrinsic variables of the two copies of the system. The quantification of intrinsic and extrinsic noise then yields a measure of total noise in the system.

Another source of motivation behind studying biochemical networks from a stochastic perspective is the fact that noise has the ability to change the dynamics of a system. In a oscillatory model like the one studied in [1], noise can destroy a global attractor and at the same point can also stabilise an unstable fixed point. It has also been pointed out that tuning the noise frequencies can diminish or enhance intermittency and mutiscaling. Similarly, in [75], it is shown that for a mixed feedback loop motif, extrinsic noise can make the system switch between steady states when the system is in the bistable region and can also enhance protein production. There have also been studies like [57, 101] that point out that extrinsic noise, in particular, has the potential to diversify as well as offer control over system dynamics.

1.3 Sensitivity Analysis

A biochemical system can undergo different kinds of changes and perturbations [64]. These include environmental changes like temperature, changes in noise [30,31,84,115], changes in molecular concentrations etc. A biochemical model is said to be robust with respect to a particular class of perturbations if it maintains its function under that class of perturbations. Therefore, robustness itself can be used to discriminate between models with behaviours of particular interest [86, 111]. In a robust model, some parameters can be varied over a wide range without affecting the system behaviour significantly. The model is then termed

as being *insensitive* to those parameters.

In complex biological systems, robustness and adaptive evolution go hand in hand with fragility, the antithesis of robustness. In fact, in many cases, the very sources and biological mechanisms that give rise to robustness against a set of perturbations may introduce fragility with respect to others. This gives rise to trade-offs between robustness and fragility of a system [8, 12, 64, 65]. For example, the effectiveness of a drug in producing some desired results and its side effects may be dependent on each other; our efficient metabolisms that evolved for us to survive in conditions of limited food may become a source of obesity and diabetes in a modern lifestyle setup [65]. Thus, the phenomena of robustness and fragility must be studied and analysed together. In particular, the exact parameters that with respect to which robustness is measured should be explicitly specified with each such analysis.

Methods that allow systematic study of dependence of system behaviour on its parameters are called parametric sensitivity analyses. Sensitivity analysis has been a popular area of research [51] for the last few decades because of its significance in parameter estimation, model design and model simplification, and engineering in general [4,98], in addition to its aid in robustness studies. For robustness analysis, sensitivity analysis provides valuable insight into how sensitive the system is to the changes in particular parameters. On the other hand, it helps quantify the dependence of model outputs on model inputs, thus enabling us to determine if the model behaviour corresponds to the predicted behaviour of the experimental system. Experimental and modelling analysis further progresses with sensitivity analysis due to help with decision making in the sense that some insensitive parameters can be fixed as constants and the sensitive ones guide further experiments.

Sensitivity analysis has been applied in many areas of study including economic modelling and decision making [90], biological modelling and analysis [135,137], chemical kinetics [94,103], and environmental modelling study [43]. There are two types of sensitivity analysis: *local sensitivity analysis* and *global sensitivity analysis*. The former methodology addresses the change in the behaviour (gradient) of model outputs with respect to the inputs over a nominal range of parameters. In global sensitivity analysis, on the other hand, these changes are studied over a large range of model parameters, and all the parameters are varied simultaneously to assess the sensitivity. More details on the two methodologies

will be discussed in chapter 3.

As discussed in the previous section, most biological systems exhibit stochasticity. Owing to stochasticity, a deterministic sensitivity analysis setup is usually insufficient for the study of such systems. Although computationally quite expensive as compared to its deterministic version, there has been much emphasis on the need for stochastic sensitivity analysis [18, 33, 54] in recent years.

Robustness and sensitivity trade-offs between different parameters have been mathematically realised as sensitivity summation laws [79]. These summation laws allow us to write the sum of different parameter sensitivities as a constant. This facilitates us to look at the sensitivities and their redistribution among different parameters of the system. It has also been found that complex biological systems are sloppy, i.e, sensitive to only a few parameters and insensitive to a large set of parameters [15, 35, 122]. This sensitivity distribution provides motivation for simpler and more efficient models [122].

1.4 Specificity Problem

Biological systems have a lot of different molecules and species chemically reacting with each other. For organised functioning of a system, there needs to be a setup according to which a certain kind of reactants react with another specific set of reactants. Without this ability of reactants to identify their targets, there can be disorder in the system. For example, a drug needs to be able to be specific in its target cells to be able to generate appropriate responses.

The *specificity* of a ligand for a receptor (or vice versa) in a ligand-recepetor binding system describes how favourable the binding of that ligand is to the receptor as compared to other receptors that may be present in the system. In other words, specificity is the degree of affinity of a ligand for a specific receptor. Real biological systems rarely exhibit complete specificity. This means that there is always cross-talk between different molecules, which is why drugs have side effects. Nonetheless, specificity is essential to organised functioning of biological systems and is termed as one of the pillars of life [70].

Different biochemical reactions take place in the same space like cytosol or a particular organelle like nucleus or mitochondrea. For these reactions to take place in a specific

way, they need to exhibit a certain level of specificity [123]. For example binding reactions or enzymatic reactions have a lot of specificity, which means that the functional partners have much higher affinity than non-functionally related biomolecules. At the same time, certain level of cross-talk between different biological pathways is inevitable, and even useful in certain situations. For example, in yeast signalling, low specificity can be useful as the mating pathway has cross-talk with starvation pathway, which can potentially mediate a physiologically useful interaction between these biological functions. The degree of specificity (or conversely, cross-talk) is dependent on the affinity between the interacting partners. In some particular models, the phenomena and mechanisms contributing towards specificity have been studied in detail [45, 67, 117]. It has also been suggested that specificity is dependent on the topology of a reaction network [82].

Currently, there is no general framework for quantification of specificity, although a statistical mechanics approach has previously been adopted to quantify specificity [55, 128, 134]. These approaches are based on energies of the binding states, and it is argued that the receptor ligand binding with least energy is preferred. On the other hand, a specificity measure based on information transformation in and across different signalling pathways has also been described in literature [2, 68]. However, there is a need for a general framework to quantify specificity. Moreover, quantification of specificity is important for a biologist's point of view. In [56], a two receptor two ligand system is studied. It this work, it has been shown that the because both the ligands bind both the receptors, the overall effect of one of the ligands seems redundant. The method of quantification of specificity can offer a technique to experimental biologists for measuring redundancy in the system by estimating the specificities with and without the second ligand. If, against popular view, the second ligand brings in new information to the system by making it more or less specific, then the usefulness of the second ligand can be established. In chapter 5, we provide a concrete definition of specificity in terms of mutual information and we show how it is useful to obtain insight into some specific biochemical systems.

1.5 Information Theory in Systems Biology

Information theory is the mathematical framework that provides necessary tools and methodologies for analysis of information processing and information transfer. Concepts from information theory have been used in the study of various biochemical networks [127]. For example, in [118, 119], it has been established that in the presence of noise, which affects the reliability of information transmission, a genetic network has an information capacity and that information flow can be optimised in simple and small genetic networks to maximise information transmission. The use of information theory in neuroscience has also been widely known [6] especially in the comparison of information content of experimentally measured neural responses with predictions from the model. The maximum information transfer is also an area of interest in many biological setups [16, 17, 121].

From a systems biology perspective, tools from information theory can be more useful than they first seem to be. Information theoretic analysis can prove to be versatile and offers detailed insight at the level of information processing. Moreover, the ideas of entropy and information can be very intuitive and easy to implement in a wide variety of analytical scenarios. The measure of mutual information, for example, has an added advantage over conventional correlation measures in the sense that it deals with nonlinear relationships between two variables more effectively. Such advantages allow for network inference and reconstruction as well as help in experimental design and decision making [83].

Only a few studies in literature have exploited the advantages of information theory to study sensitivity analysis in biochemical networks. Our main interest lies in the work presented by Ludtke *et al.* [79]. These provide the basics for the information theory based tools developed in the following chapters to study sensitivity analysis. Moreover, we also make use of the concept of mutual information in a novel approach to quantify specificity in chapter 5.

1.6 Thesis Layout

This thesis is planned as follows. Chapter 2 discusses the Mathematical tools, concepts and methodologies used in the analyses in the thesis. This chapter includes a basic gene

expression model and its treatment with some of the mathematical tools introduced. It also contains a primer on the main concepts and ideas in information theory. Chapter 3 is a review of sensitivity analysis methodologies, and includes discussion of various sensitivity analysis methods used in literature. Chapter 4 introduces the extension of information theoretic global sensitivity analysis [79] to a stochastic setup. We present the results of the application of this methodology to different models for different noise levels and observe the effect of stochasticity on parametric sensitivities. We then introduce an efficient and computationally less expensive version of the method and apply it to a circadian clock model. Lastly, chapter 5 introduces an information theoretic method of specificity quantification, and discusses the development of complete methodology. We also present the results of specificity estimates for experimental data related to signalling in T-cells. Chapter 6 is the concluding chapter of the thesis which summarises what has been done in the thesis and points at some directions for related future work.

Chapter 2

Mathematical Tools and Preliminaries

In this thesis, we have made use of several mathematical tools and frameworks that allow for efficient analysis of the problems that we deal with in later chapters. In order to use these methods in the upcoming results chapters, we elaborate on them in this chapter. We start by introducing a simple gene expression model, and as we discuss various mathematical setups, we use this model to elaborate the techniques discussed.

2.1 A Simple Gene Expression Model

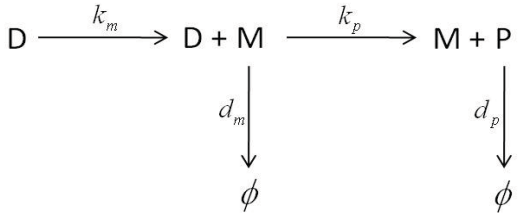

Figure 2.1: The gene expression model with reaction rates

In this section we will consider and briefly state the main analytical results for the simple gene expression model in figure 2.1. This model has been studied extensively in literature [34, 69, 105]. The model includes *transcription* and *translation* along with degradation of mRNAs and proteins. We consider both transcription and translation as first order

processes. According to this model, transcription is initiated and mRNAs are synthesized every $1/k_m$ seconds. Each mRNA is then translated into protein every $1/k_p$ seconds. Both mRNAs and proteins undergo first order degradation at the rates of d_m and d_p per second respectively. We denote respectively the number of mRNAs and proteins present at time t as $M(t)$ and $P(t)$.

2.1.1 Steady-State Characteristics in a Deterministic Setting

The reaction rate equations for the model in figure 2.1 are

$$\frac{dM}{dt} = k_m - d_m M \tag{2.1a}$$

$$\frac{dP}{dt} = k_p M - d_p P \tag{2.1b}$$

These are first order linear differential equations and can be solved easily. We are usually interested in steady-state solutions. It is clear that in the absence of noise, the steady-state values for mRNA and protein numbers are given by the model parameters as

$$M_s = \frac{k_m}{d_m}, \quad P_s = \frac{k_m k_p}{d_m d_p}. \tag{2.2}$$

2.1.2 Steady-State Characteristics in a Stochastic Setting

It can be easily shown that the average behaviour of mRNAs and proteins is similar to their deterministic behaviour. That is, if we denote the mean of mRNA numbers as $\langle M \rangle$ and mean protein numbers as $\langle P \rangle$, then

$$\langle M \rangle = M_s = \frac{k_m}{d_m}, \quad \langle P \rangle = P_s = \frac{k_m k_p}{d_m d_p}. \tag{2.3}$$

We can also obtain expressions for noise in mRNA and protein numbers as

$$\eta_M^2 = \frac{1}{\langle M \rangle} \tag{2.4}$$

$$\eta_P^2 = \frac{1}{\langle P \rangle} + \frac{d_p}{d_m} \frac{1}{\langle M \rangle} \tag{2.5}$$

The system is linear so it can be easily solved for average values. Linear noise approximation as well as master equation can be used to derive the above results. Although the coupled differential equations can be solved directly, the difference in the time-scales of mRNA life-time and protein life-time is usually exploited to simplify calculations. For example in bacteria, the mRNA life-time is of the order of minutes while protein life-time is of the order of hours. This means that mRNA numbers achieve their steady-state long before protein numbers do. Therefore, we often make use of a *quasi steady-state* assumption by putting $dM/dt = 0$ while studying the dynamics of proteins.

2.2 The Master Equation

In biological problems, we often need to study how different states, compositions and reactions of a system evolve with time. This evolution can very well be probabilistic in nature. The master equation is a set of first order differential equations that describe the time evolution of probabilities of a discrete set of states that can be occupied by a system at any given time. Let us suppose that a system can transition randomly between m possible states defined by set $\mathcal{S}$. Let $P_s(t)$ be the probability of the system being in state s. Let the system be in state s at time t and jump to state r in time dt, then the transition rate w_{sr} is defined as such that the probability of this jump happening is $w_{sr}dt$. Here, we focus on the processes that can be described as a discrete state continuous time Markov processes, such as biochemical reactions. We now restrict our discussion to homogeneous Markov processes where the transition rates are independent of time and history of the system. With this in mind, let us look at the probability that the system is in state s at time $t + dt$. This comprises of the probabilities of these different cases: if the system was already in state s at time t ($P_s(t)$); and the system does not jump to another state r in time dt ($\sum_{r\neq s} w_{sr}P_s(t)dt$); or if the system jumps form another state, say r, at time t to the state s in time dt ($\sum_{r\neq s} w_{rs}P_r(t)dt$). Therefore, the probability of the system being in state s at $t + dt$ is given by

$$P_s(t + dt) = P_s(t) + \sum_{r\neq s} w_{rs}P_r(t)dt - \sum_{r\neq s} w_{sr}P_s(t)dt. \tag{2.6}$$

This can be rearranged into the form

$$\frac{P_s(t+dt) - P_s(t)}{dt} = \sum_{r \neq s} w_{rs} P_r(t) - \sum_{r \neq s} w_{sr} P_s(t), \tag{2.7}$$

which, in the limit $dt \to 0$, yields

$$\frac{dP_s}{dt} = \sum_{r \neq s} w_{rs} P_r - \sum_{r \neq s} w_{sr} P_s. \tag{2.8}$$

This equation is called the master equation for state s of the system. For the complete system, the master equations are a set of m linear differential equations in P_is. Technically, the master equation is always solvable for a finite number of states, but in practice, the number of states is usually too large.

Let us now look at the master equation in the context of system composition at a given time. Rather than chasing a state of a system, we now consider the probabilities of the system jumping among different compositions. Suppose that the system is well-mixed. Then the probability of a molecule present in a subvolume ΔV of system volume V is $\Delta V/V$. Let $\mathbf{N}(t) = \{N_1(t), N_2(t), \ldots, N_n(t)\}$ be the composition vector of an n species system at time t, where $N_i(t)$ represents the number of molecules of type i. Let $P(\mathbf{N}, t)$ be the probability distribution of the system having composition $\mathbf{N}(t)$ at time t. Rather than the system jumping from one state to the other, we are now considering the scenario when the system transitions between different compositions via chemical reactions. The transition rates now take the form of reaction propensities and depend upon the nature of reaction. Let $a_r(\mathbf{N})$ be the reaction propensity for a reaction r in the set $\mathcal{R}$ of all reactions, then the master equation assumes the form

$$\frac{dP(\mathbf{N}, t)}{dt} = \sum_{r \in \mathcal{R}} a_r(\mathbf{N} - \nu_r) P(\mathbf{N} - \nu_r, t) - \sum_{r \in \mathcal{R}} a_r(\mathbf{N}) P(\mathbf{N}, t), \tag{2.9}$$

where ν_r is the stoichiometric vector for reaction r representing the change in the number of each species due to that reaction. The above form of the master equation utilising the chemical composition of the system is referred to as the chemical master equation [29]. In this context, it is straightforward to find the reaction propensities. There are three basic

types of chemical reactions that occur. Suppose that κ represents the probability that a given molecule X_i undergoes the reaction, then the reaction propensities are given by:

- $a_r = \kappa N_i$ for reactions of the type $X_i \rightarrow$
- $a_r = \kappa N_i N_j$ for reactions of the type $X_i + X_j \rightarrow$
- $a_r = \kappa N_i(N_i - 1)/2$ for reactions of the type $X_i + X_i \rightarrow$

2.2.1 Chemical Master Equation for Gene Expression Model

We now consider the chemical master equation for simple gene expression model from the previous section. Let M and P denote the number of mRNA and protein molecules, and $P_{m,n}$ simply denote the probability of having m and n molecules of mRNA and proteins respectively at time t. Then equation (2.9) assumes the form

$$\frac{dP_{m,n}}{dt} = [k_m P_{m-1,n} + k_p m P_{m,n-1} + d_m(m+1)P_{m+1,n} + d_p(n+1)P_{m,n+1}] - [k_m P_{m,n} + k_p m P_{m,n} + d_m m P_{m,n} + d_p n P_{m,n}]. \quad (2.10)$$

This can be rearranged to give

$$\frac{dP_{m,n}}{dt} = k_m(P_{m-1,n} - P_{m,n}) + k_p m(P_{m,n-1} - P_{m,n}) + d_m[(m+1)P_{m+1,n} - mP_{m,n}] + d_p[(n+1)P_{m,n+1} - nP_{m,n}]. \quad (2.11)$$

The last equation can then be solved analytically using the moment generating functions [106] to obtain steady state results consistent with (2.3). Also in the limit $d_m \gg d_p$, Shahrezaei and Swain [106] provide analytical expressions for $P_{m,n}$.

However, full analytical solution of the chemical master equations is practically impossible for most realistic systems where the number of possible compositions can be very large or infinite. Therefore, in practice, different mathematical approaches are used to approximate the master equations or individual reactions are rather simulated to obtain the results. Gillespie's stochastic simulation algorithm discussed in the next section is a possible way for producing exact numerical solution.

2.3 Gillespie's Stochastic Simulation Algorithm

2.3.1 The Main Idea

Gillespie's stochastic simulation algorithm [26, 27] enables us to generate an exact realisation of a reaction trajectory whose probabilities follow the master equation. It is easier to understand individual reactions rather than probabilities. The idea behind Gillespie's algorithm is that the next reaction to occur, and the time to next reaction are random variables, the latter being exponentially distributed. In one simulation or realisation of the algorithm, a set of reactions is followed and the corresponding changes in molecular species recorded. Next, we explain how to choose the random reaction time and random reaction.

2.3.2 Selecting Next Reaction and Reaction Time

Random reaction time: At any given time with a particular composition, the reaction propensities serve as reaction probabilities. So the probability of a reaction occurring is the sum of propensities. Let us call this as a_0 given by $a_0 = \sum_r a_r$. The main idea to apply then is to think of the probability of a reaction per unit time as remaining constant until the reaction has occurred. This means that the probability per unit time of no reaction occurring decreases exponentially with time. With this argument, it can be shown that the reaction probability distribution is given by

$$p(t) = a_0 e^{-a_0(t-t_{ref})} \tag{2.12}$$

where t_{ref} is some reference time. This implies that the reaction times to be randomly generated should follow exponential distribution. The idea is to generate a uniform random number r_1 in the interval $[0, 1)$ and then convert it into an exponentially distributed random reaction time using the formula

$$t_{\text{react}} = \frac{1}{a_0} \ln\left(\frac{1}{r_1}\right). \tag{2.13}$$

Random next reaction: Next we need to see which reaction should be fired. This is done by generating another uniform random number r_2 in the interval $[0, 1]$. The proba-

bility of reaction r to fire next is $\frac{a_r}{a_0}$. All the reactions are then labeled and assembled, the assembly remains the same throughout a simulation. Now the number r_2 serves as a pointer to the next reaction that occurs, in a way that r_2 must be less than or equal to $\frac{a_r}{a_0}$ but greater than $\frac{a_{r-1}}{a_0}$. In other words the reaction labeled as N should fire next where N satisfies

$$\sum_{r=1}^{N-1} a_r < r_2 a_0 \leq \sum_{r=1}^{N} a_r. \tag{2.14}$$

2.3.3 The Algorithm

To summarise, the steps of Gillespie's algorithm are given as follows:

1. Initialise time t and store populations and rate constants
2. Calculate reaction propensities for each reaction
3. Generate two uniform random numbers, r_1 and r_2
4. Calculate time to next reaction t_{react} using r_1 in (2.13)
5. Determine next reaction N satisfying (2.14) using r_2
6. Update simulation time to $t + t_{\text{react}}$
7. Update populations according to the reaction fired
8. Repeat from steps 2 to 7 until a stopping criterion is achieved

It is worth noting that if the results from the statistics satisfying chemical master equation are needed then a lot of simulations need to be repeated and the results are then averaged across the simulations. If the stationary distribution is desired then a long simulation is required to obtain a time average. However, both approaches are the same if the stochastic processes being simulated are ergodic. Whatever the approach, the computations via Gillespie algorithm are computationally expensive. Several modifications and adaptations exist in literature including next reaction method [24], tau leaping [28]. Several relatively new adaptations based on partial propensity function also exist [50, 95]. These modifications reduce the computational cost effectively, but compromise on the exactness of solution to the master equation.

2.4 Linear Noise Approximation

In stochastic systems, the master equation with nonlinear transition rates can sometimes be approximated by a Fokker Planck equation, representing linear noise approximation, with linear coefficients that are determined by the stoichiometry and transition rates of the stochastic system. The linear noise approximation [21, 32, 69, 89, 125] is a leading order term in a more general technique termed as system size expansion or Ω-expansion, where Ω represents the system volume. The system size expansion, and hence linear noise approximation, works on the assumption that the system volume is large, and that the variance of the steady state population numbers scales like the system volume.

Assuming that Ω is large and tge fluctualtion scale like the square root of the number of molecules, we write the state X_i representing the number of molecules of species i as

$$X_i = \Omega x_i + \Omega^{1/2}\phi_i, \tag{2.15}$$

where x_i is the steady state deterministic solution to the reaction rate equations and ϕ_i represents a zero-mean random variable that makes the state noisy. If $P(\mathbf{X}, t)$ represents the probability distribution of state $\mathbf{X}$ at time t, then

$$P(\mathbf{X}, t) = P(\Omega\mathbf{x} + \Omega^{1/2}\phi) = \Pi(\phi, t). \tag{2.16}$$

Now let $\mathbf{f}$ be the vector of reaction rates and plugging the result in the master equation yields a complicated expression. In the large Ω limit, we can neglect the terms of $O(\Omega^{1/2})$. If $\mathbf{S}$ denotes the stoichiometric matrix of the system, then collecting the coefficients of Ω^0, we find that

$$\frac{\partial\Pi(\phi, t)}{\partial t} = \sum_{i,k} A_{ik}\frac{\partial(\phi_k\Pi)}{\partial\phi_i} + \frac{1}{2}\sum_{i,k}[\mathbf{B}\mathbf{B}^T]_{ik}\frac{\partial^2\Pi}{\partial\phi_i\partial\phi_k}, \tag{2.17}$$

where

$$A_{ik} = \frac{\partial(\mathbf{S}_i \cdot \mathbf{f})}{\partial x_k}, \tag{2.18}$$

$$\mathbf{B}\mathbf{B}^T = \mathbf{S}\text{diag}(\mathbf{f}(\mathbf{x}))\mathbf{S}^T, \tag{2.19}$$

and $\text{diag}(\mathbf{f}(\mathbf{x}))$ represents a square matrix with $\mathbf{f}(\mathbf{x})$ as diagonal elements and zeros as off-

diagonal elements. The Fokker Planck equation (2.17) now represents the time evolution of Π, and can be solved to obtain an approximation of the solution of master equation. It can be observed that the equation (2.17) is linear, and has Gaussian solution that can be defined by its first two moments. We already know that ϕ is a zero mean random variable. Therefore, we only need the covariance Ξ of ϕ. This is given by the Lyapunov equation [125]

$$\mathbf{A}\Xi + \Xi\mathbf{A}^{\mathbf{T}} + \mathbf{B} = \mathbf{0}, \tag{2.20}$$

which can equivalently be written for the covariance $\mathbf{C}$ of $\mathbf{X}$ as

$$\mathbf{AC} + \mathbf{CA}^{\mathbf{T}} + \Omega\mathbf{B} = \mathbf{0}. \tag{2.21}$$

2.4.1 Linear Noise Approximation for Gene Expression Model

Now for the gene expression model from section 2.1, we have,

$$\mathbf{x} = \begin{bmatrix} M \\ P \end{bmatrix} = \begin{bmatrix} \frac{k_m}{d_m} \\ \frac{k_m k_p}{d_m d_p} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}, \quad \mathbf{f}(\mathbf{x}) = \begin{bmatrix} k_m \\ k_p M \\ d_m M \\ d_p P \end{bmatrix} \tag{2.22}$$

which implies that

$$\mathbf{A} = \begin{bmatrix} -d_m & 0 \\ k_p & -d_p \end{bmatrix}, \quad \mathbf{BB}^{\mathbf{T}} = \begin{bmatrix} k_m + d_m M & 0 \\ 0 & k_p M + d_p P \end{bmatrix}. \tag{2.23}$$

Using these values in (2.20), we find that

$$\mathbf{\Xi} = \begin{bmatrix} \frac{k_m}{d_m} & \frac{k_m k_p}{d_m(d_m+d_p)} \\ \frac{k_m k_p}{d_m(d_m+d_p)} & \frac{k_m k_p}{d_m d_p}\left(1 + \frac{k_p}{d_m+d_p}\right) \end{bmatrix}. \tag{2.24}$$

These results obtained are consistent with known results, for example in [47].

2.5 Main Concepts and Ideas in Information Theory

We first introduce the basic concept of *entropy*. The easiest way to look at it is to think of it as a measure of uncertainty associated with a random variable. It measures, disorder or unpredictability in a random variable. In this manuscript, we will always refer to Shannon's entropy which is a measure of the expected value of the information contained in a message, signal or a random variable. The mathematical definition of entropy now follows.

Definition 2.5.1 *For a discrete random variable X with outcomes x_i, the **Shannon entropy** $H(X)$ is defined as*

$$H(X) = -\sum_x p(x) \log_2 p(x)$$

where $p(x)$ is the probability mass function of outcome x.

The base of the logarithm can be changed. It is set to be 2 here since this entropy is measured in bits. We also use the convention that anything with zero probability does not contribute to entropy since $x \log x \to 0$ as $x \to 0$. It may also be noted that entropy of X can also be interpreted as the expected value of $\log \frac{1}{p(X)}$. Entropy is non-negative and is zero when the outcome is certain.

Definition 2.5.2 *The **joint entropy** $H(X, Y)$ of two discrete random variables (X, Y) with a joint distribution $p(X, Y)$ is given by*

$$H(X, Y) = -\sum_x \sum_y p(x, y) \log_2 p(x, y).$$

Definition 2.5.3 *The **conditional entropy** $H(X \mid Y)$ of two discrete random variables (X, Y) with a joint distribution $p(X, Y)$ is given by*

$$H(X \mid Y) = \sum_y p(y) H(X \mid Y = y).$$

We note that $H(X \mid Y)$ is the average uncertainty in X over all possible discrete values y

that the random variable Y can assume. This quantity may also be represented as

$$H(X \mid Y) = -\sum_x \sum_y p(x, y) \log_2 p(x \mid y).$$

Remark 2.5.4 *It is interesting to note that*

$$H(X, Y) = H(X) + H(Y \mid X),$$
$$H(X) - H(Y \mid X) = H(Y) - H(X \mid Y).$$

Definition 2.5.5 ***Mutual information*** *$I(X, Y)$ is defined as the difference in uncertainty of X with and without knowledge of Y and characterises the influence Y exerts on X. That is*

$$I(X, Y) = H(X) - H(X \mid Y).$$

Likewise, as shown above,

$$I(X, Y) = H(Y) - H(Y \mid X).$$

It can easily be shown that

$$I(X, Y) = -\sum_x \sum_y p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}$$

Remark 2.5.6 *The following identities can easily be deduced from the definitions above:*

$$I(X, X) = H(X)$$
$$I(X, Y) = I(Y, X)$$
$$I(X, Y) = H(X) + H(Y) - H(X, Y)$$

2.6 Bias Correction of Entropy and Information Estimates

Entropy and mutual information can be estimated accurately when we have a lot of samples of the relevant random variables and events. However, with limited computational effort and high dimensional variables, we can practically carry out entropy and information estimation with only a limited subset of the huge number of events involved. The problem becomes more severe in case of stochastic setups. This is because we have to deal with the output distributions for each of the input parameter combinations rather than an output value in the deterministic case. This dimensional explosion results in inability to compute the entropies and information efficiently, and we are often left with the situation with limited sampling. Limited sampling introduces bias in the entropy and information estimates. We cannot get rid of the bias by simply averaging over many information estimates since the magnitude of bias in this case is often of the same order as the information values to be calculated.

Both $H(Y)$ and $H(Y \mid X_i)$ are biased downwards. That is, in case of limited sampling the values calculated are less than the true values for both $H(Y)$ and $H(Y \mid X_i)$ and increase as the number of samples used is increased. The downward bias is due to the fact that limited sampling limits taking into account all the possible uncertainties due to parameter variation. The bias in $H(Y \mid X_i)$ is more than the bias in $H(Y)$ because of further limitation on sampling of values $X_i = x_i$, This results in an upward bias in the estimate of $I(X_i, Y)$.

Bias correction techniques in information estimation have been a hot topic of research recently. Many techniques have been developed to tackle this problem. [91] and [49] present a review of the techniques. Using appropriate techniques, this problem can be tackled efficiently. Below we briefly describe some of the bias correction methods in practice, their advantages, and shortcomings.

2.6.1 Panzeri-Treves Method

Panzeri-Treves method [92] is based on expanding the bias in entropies in powers of $1/N$, where N is the total number of trials. The leading order terms in the analytical expansion

are given by

$$Bias[H(Y)] = \frac{-1}{2N\ln 2}[\bar{Y}-1], \tag{2.25}$$

$$Bias[H(Y \mid X)] = \frac{-1}{2N\ln 2}\sum_x[\bar{Y}_x-1], \tag{2.26}$$

$$Bias[I(X,Y)] = \frac{-1}{2N\ln 2}\left(\sum_x[\bar{Y}_x-1]-[\bar{Y}-1]\right). \tag{2.27}$$

These quantities can be added to the plug-in estimates to obtain the true estimates for entropies and information. In the above formulas, $\bar{Y}$ represents the number of *relevant responses*, and is equal to the number of elements in the response space. But $\bar{Y}_x$ may be different from $\bar{Y}$ as it is the number of relevant responses for the particular stimulus x. The estimate of $\bar{Y}_x$ can be obtained by counting the number of responses observed at least once. This is called the *naive* count.

The bias correction by the Panzeri-Treves method is relatively straight forward and easy to implement. But the biggest disadvantage associated with this method is that this method works only in the *asymptotic sampling regime*. This means that the number of trials is not too small, and in fact large enough so that every possible response is observed many times.

2.6.2 Quadratic Extrapolation Method

Like the Panzeri-Treves method, the quadratic extrapolation method [112] also works only in the asymptotic sampling regime. The method relies on expanding the bias of entropies in a second order expansion in $1/N$. The information is calculated as

$$I_{\text{plug-in}}(X,Y) = I_{\text{true}}(X,Y) + \frac{a}{N} + \frac{b}{N^2}. \tag{2.28}$$

Unlike the Panzeri-Treves method where analytical expressions for the coefficients are present, the coefficients a and b need to be estimated from the data in this method. For this purpose computing the information using fraction of the data of size $N/2$ and $N/4$ and fitting the resulting values to the quadratic expression of (2.28). This provides an estimate for a and b. It takes a full data calculation along with two calculations of size $N/2$ and four

of $N/4$ to average and obtain the values for extrapolation.

2.6.3 Nemenman-Shafee-Bialek Method

The bias correction technique suggested in [88] relies on a Bayesian inference approach. The biggest advantage of the method is that it does not depend on the asymptotic sampling regime and provides considerably accurate results even without a large number of trials. The method is based on the principle that the bias will be minimum when a priori uniform distribution is assumed while estimating a quantity. The entropy estimates are updated step by step after the trials. This involves a number of numerical integrations and function inversions. The method is therefore more demanding to implement than the Panzeri-Treves method or the quadratic extrapolation method.

2.6.4 Shuffled Information Estimator Method

The shuffled information estimation method [85] provides corrected estimates for information only and not for entropies. This method attempts to reduce the bias in the entropy estimates by shuffling the responses and attempting to remove the response correlations. If all the data were independent, the bias will be minimum. Two entropy estimates can be defined in this case: $H_{\text{ind}}(Y \mid X)$, the conditional entropy if all the individual responses for each of the trials were independent of each other for a fixed stimulus, and $H_{\text{sh}}(Y \mid X)$, the conditional entropy obtained when the response correlations are removed by shuffling the data. Both the quantities are estimates for the entropy of the system if there are no correlations and become equal for infinite number of trials. However, in the finite trials case, the bias in $H_{\text{sh}}(Y \mid X)$ much higher than the bias in $H_{\text{ind}}(Y \mid X)$, and of the same order as $H(Y \mid X)$. Using these estimates the shuffled mutual information can be calculated as

$$I_{\text{sh}}(X,Y) = H(Y) - H_{\text{ind}}(Y \mid X) + H_{\text{sh}}(Y \mid X) - H(Y \mid X). \tag{2.29}$$

For large number of trials, $H_{\text{sh}}(Y \mid X) = H_{\text{ind}}(Y \mid X)$, therefore, $I_{\text{sh}}(X,Y) = I(X,Y)$. But in case of a small number of trials, $H_{\text{sh}}(Y \mid X) = H(Y \mid X)$, which means that the quantity $H(Y) - H_{\text{ind}}(Y \mid X)$ dominates in $I_{\text{sh}}(X,Y)$. Hence $I_{\text{sh}}(X,Y)$ is smaller than

$I(X, Y)$ implying a lesser biased estimate of mutual information.

The shuffled estimator method is straightforward and gives efficient results when combined with one of the above-mentioned techniques for entropy bias correction.

2.6.5 Shrinkage Estimator Method

The James-Stein shrinkage estimator approach [39] is based on improving the underlying probability estimates rather than entropy estimates. Once the probability estimates have been corrected the plug-in method should yield correct entropy estimates. The method works by averaging two models, a high dimensional one with large variance and low bias, and another lower dimensional one with small variance and high bias. The probabilities for each response y are calculated by the formula

$$p_y^{shrink} = \lambda t_y + (1 - \lambda) p_y^{ML}, \tag{2.30}$$

where $\lambda \in [0, 1]$ is the shrinkage intensity, t_y is the shrinkage target, and p_y^{ML} is the normal maximum likelihood estimate from frequency counts. It is suggested that the maximum entropy uniform distribution can be considered the target. The expression for λ can be found in [39]. The probabilities are calculated for all response distributions conditioned on each of the stimulus and the entropies can then be obtained.

2.7 The Unscented Transform

2.7.1 Main Idea and Motivation

Unscented transform is a method that enables us to calculate the statistics of an output distribution after the random variable has undergone a nonlinear transformation using only the mean and covariance of the input. The main idea stems from the inadequacies in the approach of extended Kalman filter. In the context of nonlinear estimation via extended Kalman filter, a known nonlinear function is linearised and posterior statistics are calculated while making use of the mean and covariance of a prior distribution. The extended Kalman filter uses only the first order terms in the Taylor series expansion of the nonlinear system.

This introduces significant errors in the estimates of posterior statistics especially when the function is highly nonlinear.

Unlike the extended Kalman filter, the Unscented transform is based on the principle that given only the mean and covariance of the input distribution, it is better to approximate the distribution itself than to approximate the nonlinear function. A set of points, called sigma points, is carefully chosen using the input mean and covariance. This set captures information about the input distribution very well. All the points in the set are then transformed through the nonlinear function and the mean and covariance of the output distribution are then calculated using the transformed sigma points.

To understand the main idea of unscented transform, let us look at the Taylor series expansion of the nonlinear function. Let us suppose that $\mathbf{x}$ is an $n-$dimensional random variable with mean $\overline{\mathbf{x}}$ and covariance matrix $\mathbf{P}_{\mathbf{x}}$. Let $\mathbf{x}$ be related to another $m-$dimensional variable $\mathbf{y}$ by a nonlinear function $\mathbf{g}$ as

$$\mathbf{y} = \mathbf{g}(\mathbf{x}). \tag{2.31}$$

We want to know what the distribution of $\mathbf{y}$ looks like. We can write (2.31) as

$$\mathbf{y} = \mathbf{g}(\overline{\mathbf{x}} + \delta\mathbf{x}), \tag{2.32}$$

where $\delta\mathbf{x}$ is a zero-mean random variable with the same covariance matrix $\mathbf{P}_{\mathbf{x}}$ as $\mathbf{x}$. The Taylor series expansion of (2.32) can be written as

$$\mathbf{y} = \mathbf{g}(\overline{\mathbf{x}}) + \nabla\mathbf{g}\delta\mathbf{x} + \frac{1}{2}\nabla^2\mathbf{g}\delta\mathbf{x}^2 + \frac{1}{3!}\nabla^3\mathbf{g}\delta\mathbf{x}^3 + \frac{1}{4!}\nabla^4\mathbf{g}\delta\mathbf{x}^4 + \cdots \tag{2.33}$$

where all the derivatives are evaluated at $\mathbf{x} = \overline{\mathbf{x}}$ and ∇ represents gradient of a function. With a bit of algebra and taking expectations [124], we can show that the nth order terms in the series expansions of mean $\overline{\mathbf{y}}$ and covariance matrix $\mathbf{P}_{\mathbf{y}}$ of $\mathbf{y}$ involve products of only nth order derivatives of $\mathbf{g}$ and nth order moments of $\mathbf{x}$. That is

$$\overline{\mathbf{y}} = E[\mathbf{y}] = \mathbf{g}(\overline{\mathbf{x}}) + \frac{1}{2}\nabla^2\mathbf{g}\mathbf{P}_{\mathbf{x}} + \frac{1}{2}\nabla^4\mathbf{g}E[\delta\mathbf{x}^4] + \cdots \tag{2.34}$$

$$\mathbf{P_y} = E[(\mathbf{y}-\overline{\mathbf{y}})(\mathbf{y}-\overline{\mathbf{y}})^T] = \nabla\mathbf{g}\mathbf{P_x}(\nabla\mathbf{g})^T + \frac{1}{2\times 4!}\nabla^2\mathbf{g}\left(E[\delta\mathbf{x}^4] - E[\delta\mathbf{x}^2\mathbf{P_y}] - E[\mathbf{P_y}\delta\mathbf{x}^2] + \mathbf{P_y^2}\right)(\nabla^2\mathbf{g})^T + \frac{1}{3!}\nabla^3\mathbf{g} + E[\delta\mathbf{x}^4](\nabla\mathbf{g})^T + \cdots \quad (2.35)$$

This means that if the second order moments and function derivatives at $\mathbf{x} = \overline{\mathbf{x}}$ are known exactly then the mean and covariance estimates are correct up to second order. The unscented transform enables us to capture the first two moments of $\mathbf{y}$ accurately.

2.7.2 The Unscented Transform Setup

In the scenario presented in the previous section, we want to determine the first two moments of the output distribution of $\mathbf{y}$. To capture the true mean and covariance of the input distribution $\mathbf{x}$, a set of $2N+1$ *sigma points* is deterministically chosen. Although the very first simplex set of points suggested was of $N+1$ points [59], it had issues due to asymmetry and could not capture the higher order effects efficiently. The set of $2N+1$ sigma points, however, does not suffer from symmetry issues [60]. The derivation of those set of points is based on the idea that each sigma point be given a weight so that the output means and covariances are captured accurately upto second order in the form of weighted combinations of function evaluations of the sigma points given that the sigma points accurately describe the input mean and covariance. Let us call the sigma point as $\mathcal{X}_i$ with weight w_i. An example of a set that accurately encodes information about the mean and covariance of input distribution of $\mathbf{x}$ is

$$\begin{aligned}
&\mathcal{X}_\mathbf{0} = \overline{\mathbf{x}} && w_0 = \frac{\kappa}{N+\kappa} && \\
&\mathcal{X}_\mathbf{i} = \overline{\mathbf{x}} + \left(\sqrt{(N+\kappa)\mathbf{P_x}}\right)_i && w_i = \frac{1}{2(N+\kappa)} && i = 1,\ldots,N \\
&\mathcal{X}_\mathbf{i} = \overline{\mathbf{x}} - \left(\sqrt{(N+\kappa)\mathbf{P_x}}\right)_i && w_i = \frac{1}{2(N+\kappa)} && i = N+1,\ldots,2N.
\end{aligned}$$

We note here that the $2N+1$ weights w_i sum up to 1. Also the notation $\left(\sqrt{(N+\kappa)\mathbf{P_x}}\right)_i$

refers to the ith column of the matrix square root of the weighted covariance matrix $(N + \kappa)\mathbf{P_x}$. For the selection of these sigma points, the matrix square root can be calculated using different decomposition methods like Cholesky factorisation scheme. Moreover, since the decomposition i.e. the matrix square root of a positive-semidefinite matrix is not unique, any square root decomposition will suffice. κ is a scaling parameter and we will discuss the effect of κ later in this section.

After the selection of sigma points, the nonlinear function is applied to each of the $2N + 1$ points to obtain $2N + 1$ output points capturing the output distribution

$$\mathcal{Y}_\mathbf{i} = \mathbf{g}(\mathcal{X}_\mathbf{i}), \qquad i = 0, \ldots, 2N. \tag{2.36}$$

The statistics for the output distribution are then approximated as

$$\begin{aligned}
\overline{\mathbf{y}} &\approx \sum_{i=0}^{2N} w_i \mathcal{Y}_\mathbf{i}, \\
\mathbf{P_y} &\approx \sum_{i=0}^{2N} w_i (\mathcal{Y}_\mathbf{i} - \overline{\mathbf{y}})(\mathcal{Y}_\mathbf{i} - \overline{\mathbf{y}})^T, \\
\mathbf{P_{xy}} &\approx \sum_{i=0}^{2N} w_i (\mathcal{X}_\mathbf{i} - \overline{\mathbf{x}})(\mathcal{Y}_\mathbf{i} - \overline{\mathbf{y}})^T,
\end{aligned}$$

where $\mathbf{P_{xy}}$ represents the cross covariance between the input and the output distributions. These estimates for the mean and covariance of $\mathbf{y}$ are accurate upto second order in the Taylor series expansions for (2.34) and (2.35). The accuracy increases to third order if the input distribution is Gaussian.

The construction of sigma points ensures that they are bound by a sphere whose radius depends on κ. The radius can be scaled up or down by making an appropriate choice of κ. This is desirable in some cases because as the dimension of $\mathbf{x}$ increases, the radius of the sphere increases automatically which can then cause non-local effects to become dominant. The issue can then be fixed with scaling down the radius of the sphere, which in effect brings the sigma points closer to the mean.

2.7.3 Scaled Unscented Transform

The problem with the unscented transform setup is that for some choices of κ, the weights of the sigma points can become negative. Since the covariance matrix $\mathbf{P_y}$ is a weighted product of transformed sigma points, negative weights can result in a non-positive semidefinite matrix. This problem was addressed with the introduction of a transformation of sigma points [58] using a scaling parameter α as follows

$$\mathcal{X}'_{\mathbf{i}} = \mathcal{X}_{\mathbf{0}} + \alpha(\mathcal{X}_{\mathbf{i}} - \mathcal{X}_{\mathbf{0}}). \tag{2.37}$$

The scaling parameter α lies in the interval $(0, 1)$ and can be chosen as small as desired. Smaller α ensures that higher order effects are minimised. The unscaled version of unscented transform is a special case of (2.37) when $\alpha = 1$. The choice of α in the interval $(0, 1)$ ensures that the resulting covariance matrix is semi-positive definite while preserving the accuracy of estimates upto second order. The next step after the scaling of sigma points is to select a function $\mathbf{h}$ and another random variable, termed as *auxiliary random variable* in [58], so that

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) = \frac{\mathbf{g}(\overline{\mathbf{x}} + \alpha(\mathbf{x} - \overline{\mathbf{x}})) - \mathbf{g}(\overline{\mathbf{x}})}{\alpha^2} + \mathbf{g}(\overline{\mathbf{x}}). \tag{2.38}$$

The second order accuracy of the statistics of the new auxiliary variable is ensured by the fact that the Taylor series expansions of $\overline{\mathbf{z}}$ and $\mathbf{P_z}$ are the same as those for $\overline{\mathbf{y}}$ and $\mathbf{P_y}$ respectively upto second order. In practice, the auxiliary random variable introduction can be equivalent to scaling the original sigma points as in (2.37) along with the modified weights

$$w'_i = \begin{cases} \frac{w_0}{\alpha^2} + \left(1 - \frac{1}{\alpha^2}\right), & i = 0 \\ \frac{w_i}{\alpha^2}, & i = 1, \ldots, 2N \end{cases} \tag{2.39}$$

An additional parameter is also introduced to direct influence on the weight of the zeroth sigma point. The whole idea of scaled unscented transform can be efficiently combined

and summarised into finding the sigma points according to the following algorithm

$$\begin{aligned}
\mathcal{X}_{\mathbf{0}} &= \overline{\mathbf{x}} \\
\mathcal{X}_{\mathbf{i}} &= \overline{\mathbf{x}} + \left(\sqrt{(N+\lambda)\mathbf{P}_{\mathbf{x}}}\right)_i \qquad && i = 1, \ldots, N \\
\mathcal{X}_{\mathbf{i}} &= \overline{\mathbf{x}} - \left(\sqrt{(N+\lambda)\mathbf{P}_{\mathbf{x}}}\right)_i \qquad && i = N+1, \ldots, 2N.
\end{aligned}$$

with weights

$$\begin{aligned}
w_0^{(m)} &= \frac{\lambda}{N+\lambda} \\
w_0^{(c)} &= \frac{\lambda}{N+\lambda} + (1+\beta-\alpha^2) \\
w_i^{(m)} &= w_i^{(c)} = \frac{1}{2(N+\lambda)}, \qquad && i = 1, \ldots, 2N.
\end{aligned}$$

where

$$\lambda = \alpha^2(N+\kappa) - N$$

and the superscripts (m) and (c) are representative of mean and covariance respectively. β is another parameter that controls the zeroth order weight of the covariance, and encodes any higher order information that one might have for the input distribution. Before the sigma points are sampled using the above scaled method, one needs to have an idea of the values of the parameters, κ, α, and β. The choice should be made keeping in mind that $\kappa \geq 0$ ensures positive semidefiniteness of the covariance matrix, α in the interval $(0, 1)$ is equivalent to choosing appropriate radius of the bounding sphere, and $\beta \geq 0$ provides any extra weight to zeroth order sigma point if need be. Once the setting is complete and sigma points calculated, the original nonlinear function $\mathbf{g}$ is applied to the sigma points in the same way as in (2.36) to obtain their corresponding points in output distribution. The

mean, covariance, and cross covariance of the output can then be obtained as

$$\overline{\mathbf{y}} \approx \sum_{i=0}^{2N} w_i^{(m)} \mathcal{Y}_{\mathbf{i}}, \tag{2.40}$$

$$\mathbf{P}_{\mathbf{y}} \approx \sum_{i=0}^{2N} w_i^{(c)} (\mathcal{Y}_{\mathbf{i}} - \overline{\mathbf{y}})(\mathcal{Y}_{\mathbf{i}} - \overline{\mathbf{y}})^T, \tag{2.41}$$

$$\mathbf{P}_{\mathbf{xy}} \approx \sum_{i=0}^{2N} w_i^{(c)} (\mathcal{X}_{\mathbf{i}} - \overline{\mathbf{x}})(\mathcal{Y}_{\mathbf{i}} - \overline{\mathbf{y}})^T. \tag{2.42}$$

The above estimates for the statistics of the output distribution are accurate upto second order. The accuracy increases to third order if the input distribution is Gaussian.

The unscented transform provides a useful tool to replace Monte Carlo sampling by a deterministic sampling scheme where the number of samples is greatly reduced. We shall show the usefulness of scaled unscented transform in chapters 4 and 5.

2.8 Simulated Annealing

Simulated annealing is a global optimization technique for unconstrained or bound optimization problems [9,63]. The method iterates by generating a new point every time based on a certain probability. The method is based on the physical process of heating a material and cooling it down slowly to minimize the energy of the system. The probability of selecting subsequent points accordingly scales with temperature. The goal of a new point selection is to lower the objective. However, with a certain probability, the points that raise the objective are also selected so that the algorithm does not get trapped in local minima. The probability of accepting worse solution decreases as the iterations progress, thus interpreted as slow cooling of the system. Whenever, a new point is randomly generated, the algorithm probabilistically decides between moving to the new state or staying in the current state. The step is repeated until the probabilities eventually lead the system to the global optimum. The simulated annealing algorithm works as follows

1. Initialise the current state x of the system to x_0

2. Calculate current temperature T

3. Choose a random neighbouring state y

4. Generate a uniform random number r in the interval $(0, 1)$

5. Calculate the state energies $E(x)$ and $E(y)$

6. If $P(E(x), E(y), T) \geq r$, then update the current state to y

7. Repeat steps 2 to 6 until stopping criteria is achieved

The above algorithm is not straightforward since the choice of current temperature, neighbouring state and acceptance probability function P needs to be specified. Unfortunately, there is no general choice that fits all problems and it has to be modified depending on the problem at hand. The guidelines on these choices are not something we would go into detail in this work but can be found in [130].

2.9 Markov Chain Monte Carlo Methods

Markov chain Monte Carlo methods (in short MCMC) are sampling methods from a probability distribution based on constructing a Markov chain whose equilibrium distribution is the target sampling distribution [25, 100]. The sequence of states, the chosen states during the iterative process, possess the Markov property. This means that the system state only depends upon the previous state and not on the states preceding it. MCMC methods have wide application in Bayesian statistics. MCMC methods are also sometimes used as optimization techniques. There is wide collection of sampling, and acceptance rejection techniques while sampling using MCMC methodology, including Gibb's sampling and Metropolis-Hastings. The theory behind MCMC methods is complex in general, and is beyond the scope of this text. In particular, we shall use Metropolis-Hastings algorithm [36,37] in chapter 5 to optimise mutual information in a problem related to specificity in a receptor ligand binding system. In short, the Metropolis-Hastings algorithm uses the acceptance probabilities similar to those in simulated annealing, but the temperature is not reduced.

2.10 Computational Tools and Software

For the analysis presented in chapter 4, the analysis and unscented transform application are implemented in MATLAB. We perform bias correction using pyentropy library in Python [49] with Panzeri Treves method, and Gillespie simulations are performed with Facile and Easystoch [108].

In chapter 5, we perform all the analysis in MATLAB with `fsolve` command for fixed points of the deterministic receptor ligand binding system. A MATLAB toolbox [73] is used for MCMC methods. The stochastic simulations are implemented with Facile and Easystoch [108].

Chapter 3

Sensitivity Analysis of Biochemical Networks: A Review

3.1 Introduction to Sensitivity Analysis

In order to introduce the concept of sensitivity analysis, we first attempt to give a brief account of *robustness*. Robustness is defined as the ability of a biological process to maintain its function against internal and external perturbations in the environment. In most cases, biochemical processes are mathematically described in the form of a model that involves parameters. Robustness is the stability of its behaviour under simultaneous changes in the model parameters. An account of the theory of robustness can be found in [64]. In order to quantify robustness in a model, we must be able to associate the uncertainty or variation in an output to the different sources of variation in the model input. Such an analysis is called sensitivity analysis. Some sensitivity analysis reviews can be found in [94], [18], and [136]. In the following sections, we present some methodologies for performing sensitivity analysis.

In a deterministic setting, a biochemical system with temporal variations can be written as

$$\frac{d\mathbf{X}}{dt} = \mathbf{f}(\mathbf{X}, \mathbf{p}, t), \qquad \mathbf{X}(0) = \mathbf{X}^o, \tag{3.1}$$

where $\mathbf{X}$ is an N-vector of species concentrations and $\mathbf{p}$ is the M-vector of system parameters. Below, we discuss some classical techniques for deterministic sensitivity analysis.

3.1.1 Local Sensitivity Analysis

Local Sensitivity analysis methods calculate partial derivatives of the form $\partial \mathbf{X}/\partial \mathbf{p}$, $\partial^2 \mathbf{X}/\partial \mathbf{p}^2$ etc about the nominal value of the parameter $\mathbf{p}$ in (3.1). These sensitivity measures are essentially gradients around the nominal values of the solution. In large biochemical networks, numerical sensitivity analysis techniques are usually employed. Below are some of the very basic numerical local sensitivity analysis methods from [94].

- The method of *Finite Differences* uses various sets of parameter values, and then the sensitivity coefficient for the j-th parameter is calculated by holding all other parameters fixed and using the formula

$$\frac{\partial \mathbf{X}(t)}{\partial p_j} = \frac{\mathbf{X}(t, p_j + \Delta p_j) - \mathbf{X}(t, p_j)}{\Delta p_j}.$$

  This technique requires M solutions of (3.1).

- Equations (3.1) can be differentiated with respect to parameters p_j to obtain

$$\frac{d}{dt}\frac{\partial \mathbf{X}}{\partial p_j} = \mathbf{J}(t)\frac{\partial \mathbf{X}}{\partial p_j} + \frac{\partial \mathbf{f}(t)}{\partial p_j}, \tag{3.2}$$

  where $\mathbf{J}(t) = (\partial \mathbf{f}/\partial \mathbf{X})_{\mathbf{p}}$. Equations (3.1) and (3.2) can be coupled together and then solved to obtain the required sensitivity measures. This method is called the *Direct Differential Method* [19].

- The *Green's Function Method* [20,46] exploits the fact that equation (3.2) is a linear, inhomogeneous equation with time varying coefficients. The corresponding homogeneous problem can be written as

$$\frac{d}{dt}\mathbf{K}(t, t') = \mathbf{J}(t)\mathbf{K}(t, t'), \tag{3.3}$$

  where $\mathbf{K}(t, t') = 1$ for $t > t'$. $\mathbf{K}$ is an $N \times N$ matrix called the *kernel* or the *Green's function matrix* and its columns are N linearly independent solutions of (3.3). The

full solution of (3.2) can be written as

$$\frac{\partial \mathbf{X}(t)}{\partial p_j} = \mathbf{K}(t,t')\frac{\partial \mathbf{X}(t')}{\partial p_j} + \int_{t'}^{t} \mathbf{K}(t,s)\frac{\partial \mathbf{f}(s)}{\partial p_j} ds \tag{3.4}$$

Usually, an adjoint strategy is employed to solve the above integral efficiently.

- The methods described above have different efficiencies for different problems. There are many other methods [94] as well as modifications of the above methods in practice.

- The metabolic control analysis techniques are also used in systems biology for sensitivity analysis. There has been a lot of work on such techniques. We shall dedicate a complete section to the discussion of these methodologies later in this chapter.

3.1.2 Global Sensitivity Analysis

Global sensitivity analysis techniques involve taking into account the statistical properties of the parameter uncertainty and calculating averaged sensitivities like $\langle \partial \mathbf{X}/\partial \mathbf{p} \rangle$ over the region of parameter uncertainty. Local sensitivity analysis methods are used only when the variations about the reference values of the parameters are small and a low order Taylor expansion of $\mathbf{X}(t)$ is valid. However, if the variations are large, then global sensitivity analysis techniques are to be used. Some of the most common global sensitivity analysis methods [94] are discussed below.

- The spread in parameter values about their true values can be represented in the form of a probability distribution. In this case, the problem of sensitivity analysis reduces to finding the probability distribution of concentrations $\mathbf{X}$ induced by the distribution of parameters. If we assume that $\mathbf{Y}$ and $\mathbf{F}(\mathbf{Y},t)$ are $M+N$ vectors with $Y_i = X_i$, $F_i(\mathbf{Y},t) = f_i(\mathbf{X},\mathbf{p},t)$ for $i = 1,\ldots,N$ and $Y_{i+N} = p_i$, $F_{i+N}(\mathbf{Y},t) = 0$ for $i = 1,\ldots,M$, then the system (3.1) can be written as

$$\frac{d\mathbf{Y}}{dt} = \mathbf{F}(\mathbf{Y},t), \qquad \mathbf{Y}(0) = \mathbf{Y}^o. \tag{3.5}$$

The fact that $F_{i+N}(\mathbf{Y},t) = 0$ for $i = 1,\ldots,M$, implies that $Y_{i+N} = Y^o_{i+N}$ for

$i = 1, \ldots, M$. This means that the initial condition $\mathbf{Y}^o$ are random variables with probability distribution, say, $P_0(\mathbf{Y}^0)$. If $P(\mathbf{Y}, t)$ is the probability distribution of the random variable $\mathbf{Y}$, then the moments of the functions of $\mathbf{Y}$ can be computed as

$$\langle h(\mathbf{y}(t)) \rangle = \int P(\mathbf{Y}, t) h(\mathbf{Y}) d\mathbf{Y}. \tag{3.6}$$

Using the fact that $\int P(\mathbf{Y}, t) d\mathbf{Y} = 1$, it can be shown that $P(\mathbf{Y}, t)$ satisfies the partial differential equation

$$\frac{\partial P(\mathbf{Y}, t)}{\partial t} = -\nabla_{\mathbf{Y}} \cdot (\mathbf{F}(\mathbf{Y}, t) P(\mathbf{Y}, t)), \qquad P(\mathbf{Y}, 0) = P_0(\mathbf{Y}^o). \tag{3.7}$$

The sensitivity information can now be obtained by studying the surface defined by $P(\mathbf{Y}, t)$ in $\mathbf{Y}$ space.

- The *FAST (Fourier Amplitude Sensitivity Test)* method was developed by Schuler *et al.* in [13], [104], and [14]. The method is based on the assumption that each parameter is statistically independent and their values are distributed according to given probability distributions. The sensitivity indices are computed by exploring the M-dimensional parameter space with a search curve given by the parametric equations

$$q_i = G_i \sin(\omega_i s); \qquad i = 1, \ldots, M, \tag{3.8}$$

where parameters are written in the form $p_i = p_i^0 \exp(q_i)$, s is a scalar in $(-\infty, \infty)$, ω_i are a set of incommensurate frequencies that ensure that the curve (3.8) is space filling, and G_i are properly selected transformation functions that depend on the parameter distributions. Thus, the averaging over an ensemble of qs can be reduced to averaging over the variable s. The use of incommensurate frequencies raises a question of computational feasibility. Therefore, integer frequencies are often used which ensures that the concentrations $\mathbf{X}$ are periodic functions of s with period 2π. These integer frequencies are chosen as precisely as possible to mimic a space filling curve. p_i^0 provide a response curve for average concentrations and now that the

concentrations are periodic in s, the following Fourier coefficients can be defined

$$A^j_{k\omega_i}(t) = \frac{1}{\pi}\int_0^{2\pi} X_j(t, \mathbf{q}(s))\cos(k\omega_i s)ds, \qquad k = 0, 1, \ldots, \tag{3.9}$$

$$B^j_{k\omega_i}(t) = \frac{1}{\pi}\int_0^{2\pi} X_j(t, \mathbf{q}(s))\sin(k\omega_i s)ds, \qquad k = 1, 2, \ldots. \tag{3.10}$$

These coefficients provide sensitivity measures. For example, the concentration X_i is insensitive to p_i at the kth harmonic ω_i if $A^j_{k\omega_i}$ and $B^j_{k\omega_i}$ are zero for all i. FAST method is usually very expensive computationally. Several modifications have been proposed to the method. FAST is a variance based approach. Several other variance based approaches are also available in literature. Examples of such methods include Sobol sensitivity analysis method [109] and High dimensional Model representation method [77], [76]. These variance based approaches are computationally very expensive, especially when the input space is high-dimensional.

- Some of the most modern techniques for Global sensitivity analysis of biochemical networks can be found in [96] and [38]. In [96], a global sensitivity analysis techniques is developed based on heat maps and parameter sensitivity spectrum. The method studies the variation of parameters in the whole of solution space rather than focusing on one output variable at a time. A generalised sensitivity summation theorems also presented in the paper. In [38], global as well as local sensitivity analysis techniques are combined to efficiently study the sensitivity of parameters throughout the range of parameter uncertainty.

- There is a global sensitivity analysis approach presented in [79] that uses concepts of entropy and mutual information to determine parametric sensitivities of the system at all orders of interaction. Sensitivity summation laws are also presented in the paper in terms of multivariate mutual informations. This approach is only present for deterministic systems in literature [79]. In chapter 4, we focus on this approach and generalise the methodology for analysis of stochastic systems.

- Several other techniques like Morris screening method [87], weighted average of local sensitivities [5], multiparametric sensitivity analysis [44], and partial rank coef-

ficient analysis [81] have also been applied for global sensitivity analysis in systems biology. An account of the suitability of these methods for different scenarios can be found in [136].

3.2 Control Theoretic Approaches to Sensitivity Analysis

The concept of *trade-off* is very important in the study of robustness. It has been known for quite some time that biological systems have a trade-off between robustness and fragility [74]. If a model becomes stable to perturbations in a set of parameters, it becomes fragile and vulnerable to perturbations in others. There is always a need to optimise the robustness and fragility in models. This is the reason why control theory has been used to study these systems. Optimal control and tuning of parameters has been studied using control theoretic approaches [7, 23, 40, 47, 61, 66, 99, 129]. Nevertheless, there have been questions on applying control theory directly to biological systems. Control theoretic approaches are designed to study the performance of monostable systems and for systems that have been designed to meet certain criteria. Biological systems, on the other hand, are not necessarily monostable. They also evolve with time, which means that their desired state may change from time to time. For a class of biological problems, though, metabolic control analysis (MCA) presents an elegant treatment of the sensitivity analysis of biochemical networks. Some applications can be found in [11, 48, 116]. The basic MCA tools are applied to biochemical networks in [23] to study the sensitivities of the systems at steady-state. These results have been extended to study the behaviour of time-varying systems in [53]. Since the results in [23] can be deduced from those in [53], we present the time-varying results here. To deal with the analysis effectively, we write the equation (3.1) in the form

$$\frac{d\mathbf{X}(t)}{dt} = \mathbf{Q}\mathbf{v}(\mathbf{X}(t), \mathbf{p}, t), \qquad \mathbf{X}(0) = \mathbf{X}^o, \tag{3.11}$$

where $\mathbf{Q}$ is called the *stoichiometric matrix* of the system, $\mathbf{p}$ is the vector of parameters, and $\mathbf{v}$ is the *rate vector*. In order to ease the computation, we split the matrix $\mathbf{Q}$ as

$$\mathbf{Q} = \mathbf{L}\mathbf{Q}_{\mathbf{R}}, \tag{3.12}$$

where $\mathbf{Q_R}$ is obtained by truncating $\mathbf{Q}$ and retaining only its linearly independent rows. In this way, we get rid of the linear dependencies among the state variables that occur due to conservation laws. Moreover, we define

$$\mathbf{q} := [\mathbf{p}, \mathbf{X}^o]^T \tag{3.13}$$

to study the sensitivities with respect to $\mathbf{p}$ and $\mathbf{X^o}$. As usual, there are N species, M parameters and, say, K types of reactions. We now present the following definitions of some quantities of interest in MCA.

Definition 3.2.1 *For a given initial condition $\mathbf{X}(0)$ and a set of parameter values $\mathbf{p}_0$ that together form a vector $\mathbf{q}_0$, the **concentration response coefficients** (or the concentration sensitivity coefficients) are defined as the elements of the $N \times (N + M)$ matrix*

$$\mathbf{R}^{\mathbf{X}}_{\mathbf{q}}(t) = \frac{\partial \mathbf{X}(\mathbf{q}, t)}{\partial \mathbf{q}}|_{\mathbf{q}=\mathbf{q}_0} = \lim_{\mathbf{q}\to\mathbf{q}_0} \frac{\mathbf{X}(\mathbf{q}_0 + \Delta\mathbf{q}, t) - \mathbf{X}(\mathbf{q}_0, t)}{\Delta\mathbf{q}}. \tag{3.14}$$

Definition 3.2.2 *For a given initial condition $\mathbf{X}(0)$ and a set of parameter values $\mathbf{p}_0$ that together form a vector $\mathbf{q}_0$, the **rate response coefficients** (or the rate sensitivity coefficients) are defined as the elements of the $K \times M$ matrix*

$$\mathbf{R}^{\mathbf{v}}_{\mathbf{q}}(t) = \frac{\partial \mathbf{v}(\mathbf{X}(\mathbf{q}, t), \mathbf{p}, t)}{\partial \mathbf{q}}|_{\mathbf{q}=\mathbf{q}_0} = \frac{\partial \mathbf{v}(\mathbf{X}, \mathbf{p}, t)}{\partial \mathbf{X}}\mathbf{R}^{\mathbf{X}}_{\mathbf{q}} + \frac{\partial \mathbf{v}(\mathbf{X}, \mathbf{p}, t)}{\partial \mathbf{q}}, \tag{3.15}$$

where the derivatives are evaluated at $\mathbf{q} = \mathbf{q}_0$ and $\mathbf{X} = \mathbf{X}(\mathbf{q}_0, t)$.

(3.11) can now be differentiated with respect to $\mathbf{q}$ to obtain

$$\frac{d}{dt}\frac{\partial \mathbf{X}(\mathbf{q}, t)}{\partial \mathbf{q}} = \mathbf{L}\mathbf{Q}_\mathbf{R}\left(\frac{\partial \mathbf{v}(t)}{\partial \mathbf{X}}\frac{\partial \mathbf{X}(t)}{\partial \mathbf{q}} + \frac{\partial \mathbf{v}(t)}{\partial \mathbf{q}}\right). \tag{3.16}$$

The equations (3.11) and (3.16) can now be solved as a coupled system to obtain $\mathbf{R}^{\mathbf{X}}_{\mathbf{q}}$. $\mathbf{R}^{\mathbf{v}}_{\mathbf{q}}$ can also be obtained using its definition and the $\mathbf{R}^{\mathbf{X}}_{\mathbf{q}}$ just obtained using (3.15). These computations are usually very difficult to carry out analytically. The numerical solution amounts to the same computational effort as the direct differential method described in the previous section.

3.2.1 MCA Theorems

The MCA treatment of systems is governed by MCA theorems. These theorems describe the so-called trade-offs between sensitivities with respect to various components of the system. The two main theorems (for steady-state) are the *summation theorem* and the *connectivity theorem*. From now on, we shall use $\mathbf{p}$ rather than $\mathbf{q}$ for clarity.

Theorem 3.2.3 ***Summation Theorem*** *If each column of the matrix* $\frac{\partial \mathbf{v}}{\partial \mathbf{p}}(t)$ *lies in the nullspace of* $\mathbf{Q_R}$ *for each* $t \geq 0$*, then*

$$\mathbf{R}^{\mathbf{X}}_{\mathbf{p}}(t) = \mathbf{0}, \quad \textit{and} \quad \mathbf{R}^{\mathbf{v}}_{\mathbf{p}} = \frac{\partial \mathbf{v}}{\partial \mathbf{p}}(t). \tag{3.17}$$

Theorem 3.2.4 ***Connectivity Theorem*** *If* $\frac{\partial \mathbf{v}}{\partial \mathbf{p}}(t) = \frac{\partial \mathbf{v}}{\partial \mathbf{X}}(t)\mathbf{L}$ *for each* $t \geq 0$*, then*

$$\mathbf{R}^{\mathbf{X}}_{\mathbf{p}}(t) = -\mathbf{L}, \quad \textit{and} \quad \mathbf{R}^{\mathbf{v}}_{\mathbf{p}} = \mathbf{0}. \tag{3.18}$$

For ease of understanding, we define the *concentration control coefficients* and the *rate control coefficients* as follows.

$$\mathbf{C}^{\mathbf{X}} = -\mathbf{L}\left(\mathbf{Q_R}\frac{\partial \mathbf{v}}{\partial \mathbf{X}}\mathbf{L}\right)^{-1}\mathbf{Q_R}, \tag{3.19}$$

$$\mathbf{C}^{\mathbf{v}} = -\frac{\partial \mathbf{v}}{\partial \mathbf{X}}\mathbf{L}\left(\mathbf{Q_R}\frac{\partial \mathbf{v}}{\partial \mathbf{X}}\mathbf{L}\right)^{-1}\mathbf{Q_R} + \mathbf{I}. \tag{3.20}$$

From (3.16), it is clear that at steady-state,

$$\mathbf{L}\mathbf{Q_R}\left(\frac{\partial \mathbf{v}}{\partial \mathbf{X}}\frac{\partial \mathbf{X}}{\partial \mathbf{p}} + \frac{\partial \mathbf{v}}{\partial \mathbf{p}}\right) = \mathbf{0},$$

implying that

$$\mathbf{R}^{\mathbf{X}} = -\mathbf{L}\left(\mathbf{Q_R}\frac{\partial \mathbf{v}}{\partial \mathbf{X}}\mathbf{L}\right)^{-1}\mathbf{Q_R}\frac{\partial \mathbf{v}}{\partial \mathbf{p}}.$$

Therefore, we have

$$\mathbf{R}^{\mathbf{X}} = \mathbf{C}^{\mathbf{X}}\frac{\partial \mathbf{v}}{\partial \mathbf{p}}.$$

Similarly, we can also obtain

$$\mathbf{R}^{\mathbf{v}} = \mathbf{C}^{\mathbf{v}} \frac{\partial \mathbf{v}}{\partial \mathbf{p}}.$$

So in terms of the control coefficients, the above summation theorems reduce to the form

$$\sum_{j=1}^{M} C_j^{X_i} = 0 \quad \text{and} \quad \sum_{j=1}^{M} C_j^{v_k} = 1, \tag{3.21}$$

for each $i = 1, \ldots, N$ and $k = 1, \ldots, K$. (3.21) shows that the controls for each reaction are shared among different reaction rates. Similarly, the connectivity theorem can be written as the statement

$$\mathbf{C}^{\mathbf{X}} \frac{\partial \mathbf{v}}{\partial \mathbf{X}} \mathbf{L} = -\mathbf{L} \quad \text{and} \quad \mathbf{C}^{\mathbf{v}} \frac{\partial \mathbf{v}}{\partial \mathbf{X}} \mathbf{L} = \mathbf{0}. \tag{3.22}$$

3.2.2 Frequency Domain Approach to Sensitivity Analysis

A frequency domain approach to metabolic control analysis has been taken in [52] and [78]. This enables us to study the response of a system under arbitrary perturbations in parameters at steady state as opposed to step perturbations. This helps study the dynamic response of the systems where the nominal behaviour is time-varying. We again consider the system in (3.11) and assume that the steady state of interest is $(\mathbf{x}^o, \mathbf{p}^o)$. Linearising the system (3.11) about this steady state, we obtain

$$\dot{\mathbf{x}}(t) = \left[\mathbf{Q}_{\mathbf{R}} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \mathbf{L} \right] \mathbf{x}(t) + \left[\mathbf{Q}_{\mathbf{R}} \frac{\partial \mathbf{v}}{\partial \mathbf{p}} \right] \mathbf{u}(t), \tag{3.23}$$

where $\mathbf{x} = \mathbf{X}(t) - \mathbf{x}^o$ and $\mathbf{u}(t) = \mathbf{p}(t) - \mathbf{p}^o$ so that the linearised system (3.23) has the steady state $(\mathbf{x}, \mathbf{u}) = (\mathbf{0}, \mathbf{0})$. Moreover, linearising $\mathbf{v}(\cdot, \cdot)$ about the nominal value, we obtain

$$\frac{\partial \mathbf{v}}{\partial \mathbf{X}} \mathbf{L}\mathbf{x}(t) + \frac{\partial \mathbf{v}}{\partial \mathbf{p}} \mathbf{u}(t). \tag{3.24}$$

The theory developed in [52] is valid for linear systems at steady states. We define a linear time-invariant input-output system in the form

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \tag{3.25}$$

$$\dot{\mathbf{y}}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t), \tag{3.26}$$

where $\mathbf{u}(t)$ and $\mathbf{y}(t)$ are input and output of the system respectively. For our system, we know from (3.23) that

$$\mathbf{A} = \mathbf{Q_R}\frac{\partial \mathbf{v}}{\partial \mathbf{X}}\mathbf{L}|_{\mathbf{X}=\mathbf{x}^o,\mathbf{p}=\mathbf{p}^o} \quad \text{and} \quad \mathbf{B} = \mathbf{Q_R}\frac{\partial \mathbf{v}}{\partial \mathbf{p}}|_{\mathbf{X}=\mathbf{x}^o,\mathbf{p}=\mathbf{p}^o}. \tag{3.27}$$

Now if we choose our output to be $\mathbf{x}(t)$ then from (3.25) and (3.26), we have

$$\mathbf{C} = \mathbf{I} \quad \text{and} \quad \mathbf{D} = \mathbf{0},$$

but if we are interested in the study of reaction rate vector, then from expression (3.24), we have

$$\mathbf{C} = \frac{\partial \mathbf{v}}{\partial \mathbf{X}}\mathbf{L} \quad \text{and} \quad \mathbf{D} = \frac{\partial \mathbf{v}}{\partial \mathbf{p}}.$$

In [52], it is argued that the frequency response of the input-output system (3.25) and (3.26) is given by

$$\mathbf{H}(i\omega) = \mathbf{C}(i\omega\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}, \tag{3.28}$$

for all real ω. Therefore, we obtain the frequency dependent concentration response coefficients and the rate response coefficients as

$$\mathbf{R}^{\mathbf{X}}(\omega) := \mathbf{H}_{\mathbf{X}}(i\omega) = \left(i\omega\mathbf{I} - \mathbf{Q_R}\frac{\partial \mathbf{v}}{\partial \mathbf{X}}\mathbf{L}\right)^{-1}\mathbf{Q_R}\frac{\partial \mathbf{v}}{\partial \mathbf{p}} \tag{3.29}$$

and

$$\mathbf{R}^{\mathbf{v}}(\omega) := \mathbf{H}_{\mathbf{v}}(i\omega) = \frac{\partial \mathbf{v}}{\partial \mathbf{X}}\left(i\omega\mathbf{I} - \mathbf{Q_R}\frac{\partial \mathbf{v}}{\partial \mathbf{X}}\mathbf{L}\right)^{-1}\mathbf{Q_R}\frac{\partial \mathbf{v}}{\partial \mathbf{p}} + \frac{\partial \mathbf{v}}{\partial \mathbf{p}}. \tag{3.30}$$

The corresponding *control coefficients* can be defined as the response coefficients when $\frac{\partial \mathbf{v}}{\partial \mathbf{p}} = \mathbf{I}$. In general,

$$\mathbf{R}^{\mathbf{X}}(\omega) = \mathbf{C}^{\mathbf{X}}(\omega)\frac{\partial \mathbf{v}}{\partial \mathbf{p}} \quad \text{and} \quad \mathbf{R}^{\mathbf{v}}(\omega) = \mathbf{C}^{\mathbf{v}}(\omega)\frac{\partial \mathbf{v}}{\partial \mathbf{p}}. \tag{3.31}$$

The summation and connectivity theorems assume the following form in the frequency domain.

Theorem 3.2.5 ***Summation Theorem*** *If a vector* $\mathbf{k}$ *lies in the nullspace of* $\mathbf{Q_R}$*, then*

$$\mathbf{C}^{\mathbf{X}}(\omega)\mathbf{k} = \mathbf{0} \quad \textit{and} \quad \mathbf{C}^{\mathbf{v}}(\omega)\mathbf{k} = \mathbf{k}$$

for all $\omega \geq 0$.

Theorem 3.2.6 ***Connectivity Theorem*** *For the control coefficients as described before*

$$\mathbf{C}^{\mathbf{X}}(\omega)\frac{\partial \mathbf{v}}{\partial \mathbf{X}}\mathbf{L} = -\mathbf{I} + i\omega\left(i\omega\mathbf{I} - \mathbf{Q_R}\frac{\partial \mathbf{v}}{\partial \mathbf{X}}\mathbf{L}\right)^{-1}$$

and

$$\mathbf{C}^{\mathbf{v}}(\omega)\frac{\partial \mathbf{v}}{\partial \mathbf{X}}\mathbf{L} = i\omega\frac{\partial \mathbf{v}}{\partial \mathbf{X}}\mathbf{L}\left(i\omega\mathbf{I} - \mathbf{Q_R}\frac{\partial \mathbf{v}}{\partial \mathbf{X}}\mathbf{L}\right)^{-1}$$

for all $\omega \geq 0$.

3.3 Stochastic Sensitivity Analysis

If a biochemical system is stochastic, a deterministic sensitivity analysis is effectively the same as analysing the sensitivities of the mean or other lumped variables of the underlying distributions. In [33], this fact is elaborated. It is argued that deterministic treatment of a system often misses out on the important changes in the underlying distributions that a stochastic treatment might well be able to capture effectively. For example, in case a deterministic analysis tries to address the changes in mean of a distribution as a result of parameter perturbations, the conclusion will be that the system is insensitive to variations in the parameter if the mean remains unchanged. However, there might be potentially very

different distributions exhibiting the same mean. Therefore, a stochastic treatment will be appropriate as this will capture the changes in all the aspects, rather than just the mean, of the underlying distribution. Another very important aspect that backs the use of stochastic sensitivity analysis is that multistable systems have more than one attractors implying that their distributions are bimodal. As before, deterministic sensitivity analysis will only be able to capture shifts in modes of the distribution, it might fail to capture the shift in the weights between the two modalities.

Below, we present three different ideas from literature that deal with the problem of stochastic sensitivity analysis.

3.3.1 Stochastic Sensitivity Analysis Using SSA Realisations

Analogous to the classical sensitivity, the sensitivity of a parameter p_j in a discrete stochastic system can be given as

$$S_j(\mathbf{X}, t) = \frac{\partial f(\mathbf{X}(\mathbf{p}), t)}{\partial p_j}, \tag{3.32}$$

where $f(\mathbf{X}, t)$ is the probability density function of $\mathbf{X}$. It is obvious that this sensitivity measure is closely related to the score function given by

$$\widetilde{S}_j(\mathbf{X}, t) = \frac{\partial \log f(\mathbf{X}(\mathbf{p}), t)}{\partial p_j}, \tag{3.33}$$

which is the gradient of the log-likelihood function. The Fisher Information Matrix (FIM) J given by

$$J = \mathbb{E}[(\nabla_{\mathbf{p}} \log f)(\nabla_{\mathbf{p}} \log f)^T] \tag{3.34}$$

describes the variance of the score function. In [33], the authors present different measures of stochastic sensitivity based on computing the Fisher information matrix. The method in [33] asks for computing Gillespie simulations for the system. These realizations are computed for both perturbed and unperturbed parameters. This results in the computation

of a centered approximation to $\frac{\partial f}{\partial p_j}$ for some j as

$$\frac{\partial f}{\partial p_j} = \frac{f(\mathbf{X}, p_j + \Delta p_j) - f(\mathbf{X}, p_j - \Delta p_j)}{2\Delta p_j}. \tag{3.35}$$

Using these centered differences in (3.34), one can obtain the corresponding Fisher information matrix. According to [33], there are three different measures of sensitivity which may have different values but should have same sensitivity orderings of the parameters for the same initial conditions. The same idea is applied in many other subsequent stochastic sensitivity analysis studies. The above idea uses Gillespie algorithm to produce perturbed and unperturbed probability densities and then calculates FIM. This independent sampling approach has been replaced by more efficient sampling methods like common reaction number [97] in literature to obtain more efficient and less biased results. This technique is also used in more efficient methods for sensitivity analysis based on finite differences. The methodology is further improved in [34] using unbiased techniques for finite difference based sensitivity estimation of stochastic systems. The methodology in [133] takes the finite difference approaches and uses them to calculate second order stochastic sensitivities.

3.3.2 Stochastic Sensitivity Analysis Using Linear Noise Approximation

Another recent method for stochastic sensitivity analysis has been proposed in [69]. The method uses linear noise approximation to the system and spares us of computing realisations of Gillespie's algorithm. In this approach, linear noise approximation helps derive model equations that lead to efficient numerical computation of the Fisher information matrix. The work then focuses on the differences between deterministic and stochastic models for sensitivity analysis and the differences between time series and time point data in the case of a stochastic study. According to [69], the Fisher information matrix can be written in the form

$$J(\mathbf{p})_{k,l} = \frac{\partial \mu^T}{\partial p_k} \Sigma \frac{\partial \mu}{\partial p_l} + \frac{1}{2} trace \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial p_k} \Sigma^{-1} \frac{\partial \Sigma}{\partial p_l} \right). \tag{3.36}$$

This means that in order to calculate the Fisher information matrix, we only need to calculate parameter derivatives of the mean μ and the variance Σ. LNA approximation ensures that we only need to solve systems of ordinary differential equations to obtain the mean and

variance. The mean is given as a solution of an ODE, the associated reaction rate equation. The variance is the product of solutions of ODEs in case we are modelling time series data. In case of time point data, variance can be obtained directly as a solution of an ODE while in the deterministic case, it is a constant.

3.3.3 Summation Theorems for Stochastic Sensitivity Analysis

MCA-like theorems have been developed in [62, 101]. The works present an extension of deterministic metabolic control analysis to the stochastic regime. Summation theorems for control coefficients related to sensitivities of species concentrations and reaction fluxes have been described. In [101], it has been argued that the summation and connectivity theorems are valid in the presence of extrinsic noise but in this case, and the MCA coefficients accommodate the effect of noise. It is also suggested that because the summation and connectivity theorems hold true in deterministic as well as stochastic cases, the control of the system is shared among all the reaction rates. This enables us to exploit noise dependency to exert control on the system, which means that noise can be regarded as a mechanism that allows us to steer the system in a desired direction. Thus noise can become an effective source of regulatory mechanism for the system. In addition to this, in [62], the summation theorem for reaction fluxes is shown to depend on measurement time window ε. If the control parameter for the species X_i relative to the parameter p_j is defined as

$$C_{p_j}^{X_i} = \frac{d \log X_i}{d \log p_j},$$

then the summation theorem for the mean $\langle X_i \rangle$ of species X_i is given by

$$\sum_{j=1}^{M} C_{p_j}^{\langle X_i \rangle} = 0.$$

A similar theorem has been derived for the concentration coefficient of covariation V_{ij}^X and is given as

$$\sum_{j=1}^{M} C_{p_j}^{V_{ij}^X} = 0.$$

Furthermore, if the reaction flux J for ith reaction is defined as

$$J_i = \left(\frac{1}{\varepsilon}\right) \text{Number of reaction events that occur during } \varepsilon,$$

then the summation theorems for reaction fluxes are given as

$$\sum_{j=1}^{M} C_{p_j}^{\langle J_i \rangle} = 1, \tag{3.37}$$

$$\sum_{j=1}^{M} C_{p_j}^{V_{ik}^J} = \frac{\partial \log V_{ik}^J}{\partial \log \varepsilon}. \tag{3.38}$$

3.4 Concluding Remarks

In this chapter, we presented a brief review of some sensitivity analysis methodologies in practice. There are local and global sensitivity analysis methods. The suitability of these methods is dependent on the problem at hand. In addition to this, some sensitivity analysis approaches take into account the effects of stochasticity of the system. In the next chapter, we shall introduce a novel stochastic sensitivity analysis methodology that possesses global properties, and makes use of concepts from information theory.

Chapter 4

An Efficient Information Theoretic Framework for Global Stochastic Sensitivity Analysis

Sensitivity analysis has been an active field of research due to its applications in robustness analysis, model design and simplification, and decision making [4, 94, 98, 103]. We also know that biochemical networks are stochastic in nature [107, 113]. In this chapter, we deal with the problem of stochastic sensitivity analysis of biochemical networks using tools and ideas from Information theory. Chapter 3 presented a review of the tools and methodologies in practice to perform parametric sensitivity analysis. Since we know that most biochemical systems are stochastic in nature, we would like to perform sensitivity analysis in a stochastic setup. The methodology devised in the next few sections is based on the technique used in [79]. The use of information theory allows for efficient generalisation from the deterministic setting in this case.

4.1 The Main Idea

The sensitivity analysis technique in [79] is based on defining a sensitivity measure for each of the model parameters as the mutual information between the parameter and the output variable. We know that mutual information between an input and output variable

is basically the difference between uncertainties in the output variable with and without the knowledge of input variable. This results in quantifying the effect exerted by the input parameter on the output variable. Mathematically, we define sensitivity as:

$$s_i = \frac{I(X_i, Y)}{H(Y)}, \tag{4.1}$$

where s_i is the sensitivity measure corresponding to the parameter X_i of the network, $H(Y)$ represents the entropy of output variable Y, and $I(X_i, Y)$ is the mutual information between X_i and Y with respect to changes in all model parameters. The method of computing this sensitivity is straightforward entropy and information calculation using probability distributions. All the parameters of the model are varied simultaneously about their nominal values and the output is calculated for each of the combinations of these parameters. Since only deterministic settings are dealt with in [79], this output computation is done by solving the coupled reaction rate equations corresponding to the system. As we are dealing with a stochastic setting, we obtain the outputs by using Gillespie's stochastic simulation algorithm. We then obtain the probability distribution for the variable Y which yields $p(y)$, the probability of Y having the value y. We also compute $p(x_i, y)$, the joint probability distribution of X_i and Y assumes the value (x_i, y). This can be done using the joint probability histogram obtained from the values of X_i and Y. These values can simply be plugged into the following formulas to obtain the desired quantities

$$H(Y) = -\sum_{y} p(y) \log_2 p(y), \tag{4.2}$$

$$I(X_i, Y) = \sum_{x_i} \sum_{y} p(x_i, y) \log_2 \frac{p(x_i, y)}{p(x_i)p(y)}. \tag{4.3}$$

Alternatively, $I(X_i, Y)$ can also be computed by the formula

$$I(X_i, Y) = H(Y) - H(Y \mid X_i), \tag{4.4}$$

where

$$H(Y \mid X_i) = \sum_{x_i} p(x_i) H(Y \mid X_i = x_i).$$

4.2 The Concept of Noise Entropy and Summation of Sensitivities

4.2.1 Discretisation Entropy

In order to estimate entropies and mutual informations in a discrete manner given in (4.2) and (4.4), we need to discretise both the input and output variables. In a deterministic setting, the knowledge of input completely determines the output. Therefore, $H(Y \mid X_i) = 0$ in that case. But when we discretise a system, there is only a finite number of bins that we can discretise the input into. The result is some uncertainty in the system. This leads to the idea of discretisation entropy. Let us consider the deterministic setting again. Suppose our inputs are x and y that are both in the interval $[0, 9)$. We also suppose that our output is of the form z = $x^2 + y$. We bin the inputs as $[0, 3) \cup [3, 6) \cup [6, 9)$ and the output as $[0, 45) \cup [45, 90)$. In an ideal world with a continuous setting, we will have the output estimating formula at hand. In this case, the conditional entropy $H(Y \mid X_i)$ will not be zero. But in the discrete case we might only know the output bin number for a certain combination of bin numbers for inputs. There is a certain level of uncertainty associated here. For example for $x = 6, y = 6$, both in bin 3, the output is 42 which is in bin 1 for the output. But at the same time, for the bin 3 inputs, $x = 8, y = 8$, the output 72 lies in output bin 2. Therefore, the knowledge of input bins does not completely identify the output bin number. The uncertainty arising here is termed as discretisation entropy, and we denote it as H_Δ.

4.2.2 Intrinsic Noise Entropy

In a stochastic setting the idea of discretisation entropy generalises to that of noise entropy. There is no longer only the discretisation scheme that introduces uncertainty in the output given input, there is another source for the entropy - the intrinsic noise, or simply the stochasticity of the system. In a stochastic setting, an estimate of $H(Y \mid X_i)$ is that of the noise entropy, denoted as $H_\mathcal{N}$, of the system. The noise entropy has two contributing factors, the discretisation entropy H_Δ and the intrinsic noise entropy H_{int}.

Just as in the deterministic case, the input bin numbers cannot fully identify the output bin numbers in stochastic framework. The noise might just push the outputs across the

boundaries of the binning scheme, hence making the output bin number uncertain. The contributing factors of H_{Δ} and H_{int} lead us to think that the noise entropy in a stochastic setting is greater than the corresponding entropy in a deterministic case. In fact, one is tempted to write that $H_{\mathcal{N}} = H_{\Delta} + H_{int}$ could be the case. This means that the stochastic noise entropy has two components, due to the contributions from discretisation scheme and intrinsic noise in the system. Although the idea makes sense intuitively, this is not generally true.

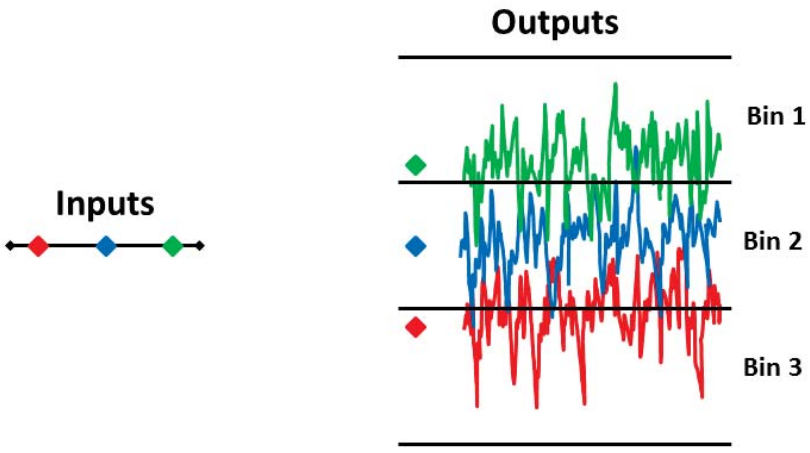

Figure 4.1: Examples of deterministic versus stochastic outputs for inputs in a particular bin. In deterministic case, the output entropy conditioned on the particular input bin is 1.58 with equal number of outputs in each of he three output bins. The outputs in stochastic case are time series with some of the output values crossing the boundaries of the bins. The number of outputs in each bin are therefore different resulting in smaller conditional output entropy as compared to the deterministic case. This shows that stochastic noise entropy can be less than deterministic discretisation entropy

Let us consider a completely deterministic system. Only the discretisation scheme applied for entropy estimations introduces uncertainty. Now, let us add stochasticity into the system. The intrinsic noise will make the response more haphazard and the response might jump across the boundaries of the binning scheme. It seems possible that this will make the response more uncertain and the entropy will be greater. But, at the same time, the discretisation scheme and stochasticity might work in opposite or competing ways. This will be

the case when the stochastic noise entropy will turn out to be less than the discretisation entropy. To illustrate this concept properly, let us look at figure 4.1, an example of how noise can work counter intuitively. The outputs are binned into three bins. Let us assume the case where three inputs are sampled from the same bin to be able to relate to noise entropy. This setting effectively works as conditioning on the input coming from a particular bin. In the deterministic setting, the green input maps to bin1, the blue input to bin 2, and the red one to bin 3. The noise entropy in this deterministic case is $\log_2(3) = 1.585$. This entropy is based on the equal probability of the output lying in each bin being $\frac{1}{3}$. Now let us add noise to the system. Each output, that was previously a point in output bins now corresponds to time point data. The figure shows that the data corresponding to input can jump across the bin boundaries. Now the probability of an output lying in each bin is not equal anymore. The figure shows that there are far more time point data in bin 2 as compared to bins 1 and 3. This results in a higher probability of output lying in bin 2. The resulting stochastic noise entropy is less than 1.585, thus proving the fact that stochastic noise entropy can be less than deterministic noise entropy.

4.2.3 Higher Order Sensitivities

In order to understand the higher order sensitivities, we need to define multivariate mutual information. Until now we have looked at bivariate mutual information. Reminding ourselves of the definition of mutual information between two variables, we define $I(X, Y)$ as

$$I(X, Y) = H(Y) - H(Y \mid X)$$

which can also be written as

$$I(X, Y) = H(X) + H(Y) - H(X, Y), \tag{4.5}$$

where $p(x, y)$ is the joint probability distribution of X and Y and

$$H(X, Y) = -\sum_x \sum_y p(x, y) \log_2 p(x, y).$$

Extending this to three variables X, Y, Z, we write the mutual information as

$$I(X,Y,Z) = H(X) + H(Y) + H(Z) - H(X,Y) - H(Y,Z) - H(Z,X) + H(X,Y,Z).$$

In analogy to this, we define multivariate mutual information among parameters $X_1, X_2, \ldots, X_n$ with joint probability distribution $p(x_1, x_2, \ldots, x_n)$, denoted by $I_n(X_1; X_2; \ldots; X_n)$ as

$$I_n(X_1; X_2; \ldots; X_n) = \sum_{x_1,\ldots,x_n} p(x_1, x_2, \ldots, x_n) \log_2 \frac{p(x_1, x_2, \ldots, x_n)}{p(x_1)p(x_2)\cdots p(x_n)} \tag{4.6}$$

which can also be expressed as [110]:

$$I_n(X_1; X_2; \ldots; X_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{X \subset \{X_1, X_2, \ldots, X_n\}, |X|=k} H(X) \tag{4.7}$$

where

$$H(X_1, X_2, \ldots, X_n) = -\sum_{i=1}^{n} \sum_{x_i} p(x_1, x_2, \ldots, x_n) \log_2 p(x_1, x_2, \ldots, x_n).$$

For this section and the following, where higher order interactions with multivariate informations are implied, we shall denote the separation of variables by a semicolon rather than a comma. That is, a comma will separate different components of a variable argument, while a semicolon will be used to separate individual arguments. We shall also use a subscript for I to denote the number of variables whose mutual information is implied. With this notation, figure 4.2 provides an illustration that helps understand the concept of multivariate mutual information.

After defining the concept of multivariate mutual information, we look at the quantity $I_2(X_1, X_2, \ldots, X_n; Y)$ in our context of sensitivity analysis. This is bivariate mutual information, as the index shows, with $n-$dimensional input parameter space. The quantity encompasses up to nth order multivariable interactions, due to different input variables interacting with each other, as well as in their combined influence on the output Y. In addition to individual parameter contributions, the quantity also includes the joint effect of, for example, the pair (X_1, X_2) on the output Y. Even if the inputs themselves are uncorrelated,

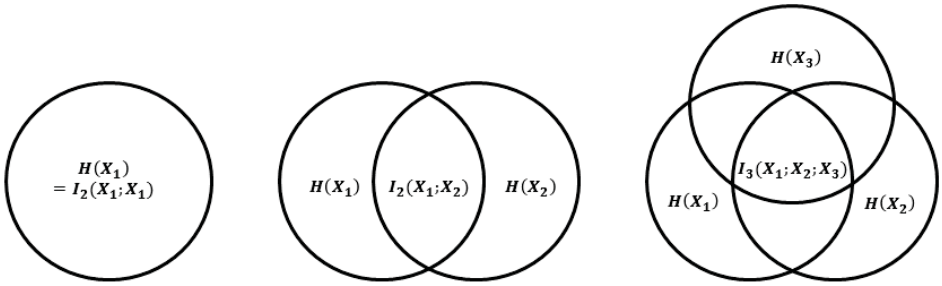

Figure 4.2: Illustration of multivariate mutual information and individual entropies. Mutual information between two variables X_1 and X_2 is the information shared by $H(X_1)$ and $H(X_2)$. In the same way, the three variable mutual information $I_3(X_1; X_2; X_3)$ is the amount of information shared among $H(X_1)$, $H(X_2)$, and $H(X_3)$

their joint influence on the output can result in an influence on the information estimates.

We now have a close look at different orders of interaction among inputs. Let us split the term $I_2(X_1, X_2; Y)$ into different orders of interaction (proof to follow later):

$$I(X_1, X_2; Y) = (\text{influence of } X_1 \text{ on } Y + \text{influence of } X_2 \text{ on } Y) + (\text{influence of } (X_1, X_2) \text{ on } Y)\,.$$

Similarly,

$$\begin{aligned} I(X_1, X_2, X_3; Y) = {} & (\text{influence of } X_1 \text{ on } Y + \text{influence of } X_2 \text{ on } Y + \text{influence of } X_3 \text{ on } Y) \\ & + (\text{influence of } (X_1, X_2) \text{ on } Y + \text{influence of } (X_2, X_3) \text{ on } Y + \text{influence of } (X_3, X_1) \text{ on } Y) \\ & + (\text{influence of } (X_1, X_2, X_3) \text{ on } Y)\,. \end{aligned} \quad (4.8)$$

These influences are termed *higher order sensitivities*. Since parameter pairs, and triplets and, $n-$tuples can influence an output, there will be correlations among these parameter combinations when conditioned on an output value. That is, if the output is $y \in Y$, this output will induce correlations between input pairs and $n-$tuples. These output induced influences are then averaged for each possible output value, to yield the influence of input pair on the output. Therefore, in their effect, these influences from equation (4.8) are

actually multivariate conditional mutual informations for $n \geqslant 2$, defined as

$$I_n(X_1; X_2; \ldots; X_n \mid Y) = \sum_y p(y) \sum_{x_1,\ldots,x_n} p(x_1, x_2, \ldots, x_n \mid y) \log_2 \frac{p(x_1, x_2, \ldots, x_n \mid y)}{p(x_1 \mid y)p(x_2 \mid y) \cdots p(x_n \mid y)} \quad (4.9)$$

Higher order sensitivities and parameter interactions are important for analysis of a model, and can even guide model development and model simplification. This is because complex systems are known to be sloppy, i.e., sensitive to only some parameters, while insensitive to a large number of other parameters [15, 35, 122]. The function of sloppy or insensitive parameters, or even parameter combinations, can then be replaced by constants or a smaller combination of new, less sloppy parameters. Moreover, in some cases, first order sensitivity estimates may not provide any insight into the analysis. For example, in case of boolean biocatalytic XOR gate [93], the first order sensitivities are zero, and all the information of the system is carried by second order parameter interactions between the inputs. This is because for an XOR gate, the output is true when only one of the inputs is true, and false otherwise. This means that none of the inputs, only by themselves, can give an idea of the output. Therefore, output entropy conditioned on one of the input variables is the same as the output entropy itself, resulting in zero first order sensitivities. In this case, all the information is contained in the second order sensitivities.

We shall now show how the term $I_2(X_1, \ldots, X_n; Y)$ can be decomposed into several terms representing different orders of interaction among the input parameters while influencing an output. We first derive the relations for second order sensitivity estimates. Third and higher sensitivities will follow from there. If we condition (4.5) on a third parameter, we obtain

$$I_2(X; Y \mid Z) = H(X \mid Z) + H(Y \mid Z) - H(X, Y \mid Z). \quad (4.10)$$

It is also a well known fact that

$$H(X, Y) = H(X) + H(Y \mid X). \quad (4.11)$$

Now, let us rewrite (4.5) with two dimensional input (X_1, X_2) :

$$I_2(X_1, X_2; Y) = H(X_1, X_2) + H(Y) - H(X_1, X_2, Y). \tag{4.12}$$

Using (4.11) in (4.12) implies that

$$I_2(X_1, X_2; Y) = H(X_1) + H(X_2 \mid X_1) + H(Y) - H(X_1, X_2, Y).$$

Writing the second term and the combination of third and fourth terms on the right with the help of equations (4.4) and (4.11) respectively, we get

$$I_2(X_1, X_2; Y) = H(X_1) + H(X_2) - I_2(X_1; X_2) - H(X_1, X_2 \mid Y).$$

We then split the first two terms on the right using (4.4) to obtain

$$I_2(X_1, X_2; Y) = I_2(X_1; Y) + H(X_1 \mid Y) + I_2(X_2; Y) + H(X_2 \mid Y) - I_2(X_1; X_2) - H(X_1, X_2 \mid Y).$$

Now combining the second, fourth and sixth terms on the right and using (4.10) on the combination yields

$$I_2(X_1, X_2; Y) = I_2(X_1; Y) + I_2(X_2; Y) + I_2(X_1; X_2 \mid Y) - I_2(X_1; X_2). \tag{4.13}$$

The equation (4.13) the general form of all the interactions that $I_2(X_1, X_2; Y)$ involves. The first two terms on the right hand side provide the first order sensitivities, the fourth term is the input correlation, while the third term is purely second order sensitivity. The simple form of (4.13) allows a huge advantage over variance based approaches to sensitivity analysis because it can easily accommodate for correlated inputs, which is not straightforward to do in the variance based methods. However, in what follows next, we shall restrict ourselves to cases where inputs are uncorrelated, and can therefore write (4.13) as

$$I_2(X_1, X_2; Y) = I_2(X_1; Y) + I_2(X_2; Y) + I_2(X_1; X_2 \mid Y). \tag{4.14}$$

It is important to note though that we cannot put the last term in (4.13) to zero when

the equation itself is conditioned on another variable. The whole concept of higher order interaction lies in the fact that even though parameters themselves may be uncorrelated, they can still interact in their effect on the output. One can expect in this case that the last term in (4.13) when conditioned on another variable, say Z. In that case (4.13) becomes

$$I_2(X_1, X_2; Y \mid Z) = I_2(X_1; Y \mid Z) + I_2(X_2; Y \mid Z) + I_2(X_1; X_2 \mid Y, Z) - I_2(X_1; X_2 \mid Z). \tag{4.15}$$

The formula (4.14) provides a basis for deriving expressions for higher order sensitivities. As an illustration, we next derive the third order sensitivity terms in what follows. Let us consider the term $I_2(X_1, X_2, X_3; Y)$ as before, which includes upto third order sensitivities. Then

$$I_2(X_1, X_2, X_3; Y) = I_2(X_1; Y) + I_2(X_2, X_3; Y) + I_2(X_1; X_2, X_3 \mid Y)$$

Now using (4.14) and (4.15) on the second and third terms on the right, we obtain

$$\begin{aligned} I_2(X_1, X_2, X_3; Y) = I_2(X_1; Y) &+ [I_2(X_2; Y) + I_2(X_3; Y) + I_2(X_2; X_3 \mid Y)] \\ &+ [I_2(X_1; X_2 \mid Y) + I_2(X_1; X_3 \mid Y) + I_2(X_2; X_3 \mid X_1, Y) \\ &- I_2(X_2; X_3 \mid X_1)]. \end{aligned}$$

Rearranging the terms according to the order of sensitivities, we obtain

$$\begin{aligned} I_2(X_1, X_2, X_3; Y) = &[I_2(X_1; Y) + I_2(X_2; Y) + I_2(X_3; Y)] \\ &+ [I_2(X_1; X_2 \mid Y) + I_2(X_1; X_3 \mid Y) + +I_2(X_2; X_3 \mid Y)] \\ &+ I_2(X_2; X_3 \mid X_1, Y) - I_2(X_2; X_3 \mid X_1). \end{aligned}$$

Due to the fact that when all the first and second order sensitivities are subtracted out from $I_2(X_1, X_2, X_3; Y)$, what remains is purely third order sensitivity denoted by $I_3(X_i; X_j; X_k \mid Y)$, we can deduce from the last equation that

$$I_3(X_i; X_j; X_k \mid Y) = I_2(X_j; X_k \mid X_i, Y) - I_2(X_j; X_k \mid X_i). \tag{4.16}$$

Since multivariate mutual information is symmetric, the third order sensitivity is also symmetric with respect to its arguments. Following the above procedure and repeated application of (4.14) and (4.15), we can also derive the expressions for higher order sensitivities.

4.2.4 Summation of Sensitivities

In the discussion from previous section, we have shown that for an input output system with n inputs $X_1, \ldots, X_n$ and an output Y, the mutual information can be split into different orders of sensitivities of the output. That is, we can write

$$I_2(X_1, \ldots, X_n; Y) = \sum_{i_1} I_2(X_{i_1}; Y) + \sum_{i_1 < i_2} I_2(X_{i_1}; X_{i_2} \mid Y) + \cdots + \sum_{X \subset \{X_1, \ldots, X_n\}, |X| = n} I_n(X \mid Y) \tag{4.17}$$

Now we go back to the basic definition of mutual information from (4.4). Treating the $n-$tuple $(X_1, X_2, \ldots, X_n)$ as an input, we can write that

$$I_2(X_1, \ldots, X_n; Y) = H(Y) - H(Y \mid X_1, \ldots, X_n). \tag{4.18}$$

The second term on the right, for a discretised version of the system, becomes noise entropy $H_{\mathcal{N}}$ as discussed earlier. From (4.17) and (4.18), what follows is sometimes referred to as sensitivity summation theorem which states that

$$H(Y) = H_{\mathcal{N}} + \sum_{i_1} I_2(X_{i_1}; Y) + \sum_{i_1 < i_2} I_2(X_{i_1}; X_{i_2} \mid Y) + \cdots + \sum_{i_1, \ldots, i_n} I_n(X_{i_1}; \ldots; X_{i_n} \mid Y). \tag{4.19}$$

This becomes easier to understand when we normalise all the sensitivities and the noise entropy by output entropy. The sensitivity summation law then takes the form

$$H^*_{\mathcal{N}} + \sum_{i_1} I^*_2(X_{i_1}; Y) + \sum_{i_1 < i_2} I^*_2(X_{i_1}; X_{i_2} \mid Y) + \cdots + \sum_{i_1, \ldots, i_n} I^*_n(X_{i_1}; \ldots; X_{i_n} \mid Y) = 1, \tag{4.20}$$

where asterisks represent normalised quantities. In what follows, we shall drop the asterisks for convenience and explicitly mention if the estimates are not normalised. In the examples section later in this chapter, this summation law proves to be very helpful in order

to compare different cases with varying levels of stochasticity. For ease of comparison and consistency with (4.1),we shall be normalising all sensitivities and $H_{\mathcal{N}}$ with $H(Y)$.

4.2.5 Total Sensitivity Indices

For the purpose of sensitivity analysis, it is usually more useful to have a sensitivity measure for every parameter, rather than sensitivities at all orders of interaction. Just like in variance based approaches, we need to define some quantity for each parameter that combines the effect of that parameter at all orders. An obvious choice is essentially the sum of all sensitivities that involve that parameter. This is the amount of information that the particular parameter adds to the system. In mathematical terms, the total sensitivity index for parameter X_i, denoted here by s_i^{total}, can be defined as

$$s_i^{\text{total}} = I_2(X_1, X_2, \ldots, X_n; Y) - I_2(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n; Y), \tag{4.21}$$

Because we can split each of the above terms on the right into varying orders of interactions, the resulting difference is only the terms with sensitivities involving the parameter X_i. We shall also be normalising the total sensitivity indices for each parameter with $H(Y)$. It is worth pointing out here, that although the sums of all normalised sensitivities is 1, the sum of all normalised total sensitivity indices is not equal to 1 since in that case some sensitivity terms are repeated.

4.2.6 High Dimensional Parameter Space

With high dimensional parameter space, the terms $H(Y \mid X_1, \ldots, X_n)$ and $I_2(X_1, \ldots, X_n; Y)$ are not easily computable. We need more and more parameter samples to estimate this quantity accurately. The same is true for higher order sensitivity estimates. As the dimensionality of the arguments in the information measure increases, we get more and more limited by the computational expense of sampling. This makes it difficult to compute all the quantities in (4.19). First and second order sensitivities can be reasonably estimated with a reasonable number of samples. But the main problem that arises in this case is that the present setup is only sufficient for sensitivity analysis, and without the knowledge of

all orders of sensitivities, the noise entropy from (4.19) will remain unknown. We would also like to have an idea of how much of the output uncertainty is due to stochasticity in the system.

The solution to this problem, as suggested in [79], is to perform a Monte Carlo estimate of the noise entropy, which, in their case is the discretisation entropy. We extend the same idea for stochastic systems. The idea is that if we look closely at the definition of noise entropy $H_{\mathcal{N}}$, we find that it is the averaged output entropy conditioned on all the input parameters. With our discretised system where the parameters are conditioned to be in certain bins rather than having a particular value, this means that the average is taken over all possible parameter bin combinations. For system with n parameters, all discretised into b_i number of bins, this implies that the average has to be taken over b_i^n possible occurrences. This number is above 50000 for as little as 4 parameters, each discretised into 15 bins. To overcome this problem, only a few bin combinations are randomly selected. For each of these bin combinations, input samples are produced in such a way that all those samples correspond to the same bin combination. The outputs for each of the samples for a bin combination are calculated and the entropy of the outputs estimated. For each particular bin combination, the output entropy is estimated using the sample outputs in order to quantify the effect of discretisation and noise on the output. This output entropy conditioned on the particular bin combination is averaged over all the bin combinations to give an estimate of noise entropy. It is shown in [79] that very few of these bin combinations suffice for a reasonable estimate of noise entropy. For our case of a stochastic system, the only change we make is that the output distribution for each sample is considered rather than an output value. This makes sure that the noise entropy includes the effect of discretisation as well as stochasticity in the system.

4.3 The Algorithm

4.3.1 Estimation of Sensitivities

Suppose that our input variable X is $M_{in}-$dimensional and the dimension of output variable Y is M_{out}. The sensitivity analysis methodology is based on the following main steps:

1. **Generating samples for input parameters:** This is done by generating N samples for each of the input parameters. Depending on the shape of the input distribution we want to look at, we can do one of the following two:

   - Generate uniformly distributed random numbers in the interval $[(1-\theta)K_i, (1+\theta)K_i]$ for each input variable X_i, where K is the vector of nominal values of X. Here θ and K are user input, and θ specifies the range of parameters. The choice of θ determines the global characteristics of the analysis and is made depending upon the range of parameter space to be covered.
   - Generate independent lognormally distributed random numbers with means K_i and σ_i^2 for each of the parameters X_i. Here μ and σ^2 are user input. σ^2 incorporates the effect of θ by setting the variance scaled by θ

2. **Performing stochastic simulations:** Using the Gillespie algorithm, a stochastic simulation is performed for each of the N parameter samples. We then obtain a probability distribution in each case and for each of the output variables using M independent steady state output values corresponding to each of the N parameter sets.

3. **Binning the inputs and outputs:** All the input parameters are binned into b_i number of bins and all outputs into b_o number of bins. The interval over which binning for a parameter X_i is done is set to be $[\max(0, \bar{X}_i - 3\sigma_{X_i}), \bar{X}_i + 3\sigma_{X_i}]$, where $\bar{X}_i$ and σ_{X_i} are respectively the mean and standard deviation of the N sampled values for X_i. The outputs are binned similarly, except in this case, the mean and standard deviation are taken over NM normalised output values. The output normalisation is done in order to compare results with different stochasticity levels. The outputs are normalised either with corresponding deterministic values or with averaged stochastic outputs.

4. **Bias correction:** The binned inputs and outputs are then used to obtain bias corrected estimates for mutual information $I_2(X_1, \ldots, X_{M_{in}}; Y_j)$, noise entropies $H_{\mathcal{N},j}$ and output entropies $H(Y_j)$ for each $j = 1, \ldots, M_{out}$. The first order sensitivities can be estimated directly using mutual information, the second order entropies can be estimated using the formula (4.14) and for a sufficiently large N and small M_{in}, the

subsequent higher order sensitivities can be estimated using the formula

$$I_k(X_{i_1};\ldots;X_{i_k} \mid Y) = I_2(X_{i_1},\ldots,X_{i_k};Y_j) - \sum_{X\subset\{X_{i_1},\ldots,X_{i_k}\},|X|=1} I_2(X;Y_j) - \sum_{X\subset\{X_{i_1},\ldots,X_{i_k}\},|X|=2} I_2(X \mid Y_j) - \ldots - \sum_{X\subset\{X_{i_1},\ldots,X_{i_k}\},|X|=k-1} I_{k-1}(X \mid Y_j) \tag{4.22}$$

4.3.2 Estimation of Noise Entropy

In case of large M_{in}, the quantities $H_{\mathcal{N}}$ and $I_2(X_1,\ldots,X_{M_{in}};Y_j)$ are difficult to estimate. Only the first few orders of sensitivities are calculated in this case and the following procedure is followed to approximate noise entropy $H_{\mathcal{N}}$.

1. **Generating input data for noise entropy estimation:** We generate N_1 number of parameter samples as before and then bin them. The next step is to sample another N_2 uniformly distributed parameter samples from each of the N_1 bin combinations. This is done to assess the effect that the discretisation scheme might have on the entropy of the system.

2. **Performing stochastic simulations:** Using the Gillespie algorithm, a stochastic simulation is performed for each of the N_1N_2 samples. We then obtain a probability distribution in each case and for each of the output variables using M independent steady state output values corresponding to each of the N_1N_2 samples.

3. **Estimation of noise entropy:** We bin the outputs using a significantly higher value of M in this case. Bias corrected estimates for a conditional output entropy from N_2M binned outputs is estimated for each of th N_1 unique bin combination samples. The entropies are then averaged over N_1 bin combinations to yield $H_{\mathcal{N}}$

4.4 Analysis of Gene Expression Model

In this section, we perform the sensitivity analysis and discuss our findings for the four parameter gene expression model from chapter2 in a stochastic setting. We use the method-

ology outlined in section 4.3 to investigate the influence of different parameters on the outputs. Our output in this case is proteins in number of molecules. We perform 10000 simulations for our analysis and vary each parameter, sampling from a uniform distribution, between 0.5 and 1.5 times the nominal value. All the inputs are discretised into 15 bins and the output is discretised into 20 bins.

The stochastic analysis is different from deterministic one in the sense that we have to deal with output distributions rather than an output value for each input parameter combination. While in deterministic case, the output entropy or information is taken over the distribution of fixed output values coming from inputs, we take the entropies over the superimposed distribution of individual distributions coming from each sample input. In a stochastic setting this can become a tedious and computationally expensive task for even simple models like the one under consideration. Even if we take as little as 50 outputs to account for the distribution coming from one input sample, the overall task becomes 50 times more complicated as compared to the deterministic one. But there is one advantage, the estimates for entropies and mutual informations start approaching the actual unbiased values quicker, i.e. with relatively less number of samples in the stochastic case.

4.4.1 Reproducibility of Results

For the sensitivity analysis to follow, we use 10^4 input samples i.e. 10^4 Gillespie simulations to perform the analysis, and use 50 stochastic repetitions i.e., 50 independent timepoints from the Gillespie simulation to generate the output distribution per sample from the stationary state. But first, we need to justify that these numbers are sufficient for the analysis. Let us first consider the reproducibility of results using 10^4 input samples. In figure 4.3, the blue bars have the errorbars on the mean of output entropy, mutual information, and noise entropy estimates from 9 individual analyses. The errors are small with standard deviations of 0.008, 0.0068, and 0.012 respectively, and therefore hard to see on the graph. On the other hand, the deep red bars show the same analysis with 9×10^4 input samples. The results match very well between the original sample and a sample that is 9 times bigger than that. The mutual information estimate is always biased upward while entropy estimate is always biased downward. So as we increase the number of samples,

the mutual information should go down and the entropies should go up. As our input space is multidimensional, it takes a large number of input samples to get rid of the bias. The results in figure 4.3 are bias corrected and show that the results for the two sets of 10^4 and 9×10^4 inputs samples are reasonably close for us to stick to only 10^4 samples.

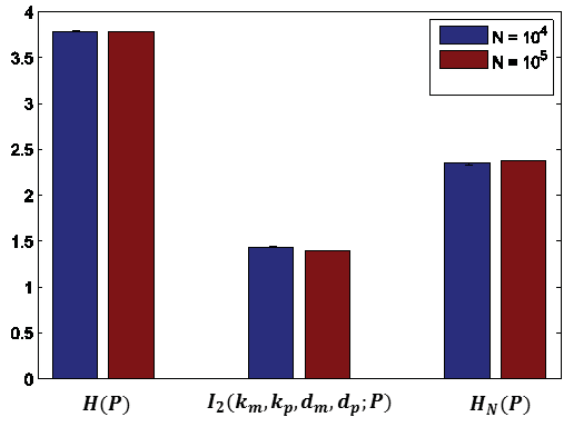

Figure 4.3: Output entropy $H(P)$, mutual information $I_2(k_m, k_p, d_m, d_p; P)$, and noise entropy $H_{\mathcal{N}}$ for different sizes of input sample space. The blue bar shows the mean of 9 independent analyses with $N = 10^4$. The errorbars show the standard deviation over the runs. $M = 49$. Nominal parameter values are $k_m = 0.1$, $k_p = 0.1$, $d_m = 0.01$ and $d_p = 0.001$

Another parameter to set for our experiments is the stochastic repetitions, M. For our experiments, very few output bins are non-empty. As we bin our outputs in 20 bins, only a few stochastic repetitions should in principle be enough to crowd them. Figure 4.4 shows the results for $M = 10$, $M = 50$, and $M = 90$. The results are strikingly similar, and even $M = 10$ gives good results. We set M to be 50 for our experiments to err on the side of caution.

4.4.2 Mutual Information, Entropy and Sensitivity Estimation

Now that we have justified the use of $N = 10^4$ and $M = 50$, we proceed with estimating mutual informations, sensitivities at different orders and noise entropies. We do the analysis in a stochastic setting as well as in a deterministic setting for different noise levels. We vary the noise levels by changing the number of mRNA and protein molecules. We fix the ratio

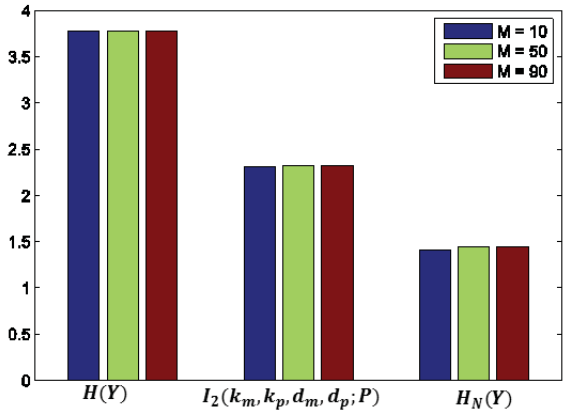

Figure 4.4: Output entropy $H(P)$, mutual information $I_2(k_m, k_p, d_m, d_p; P)$, and noise entropy $H_{\mathcal{N}}$ for different number of stochastic repetitions with $N = 10^4$. Nominal parameter values are $k_m = 0.1$, $k_p = 0.1$, $d_m = 0.01$ and $d_p = 0.001$

($\frac{k_p}{d_p}$ in the deterministic setting) of proteins to mRNAs to be 20 and change the parameter k_m in order to get different levels of protein and mRNA molecules. Also, for the deterministic case we set $N = 10^5$ as the results exhibit a variation in case of lesser number of samples. Further to make the comparison meaningful, we make sure the binning scheme that we use for the deterministic case is the same as that applied during the binning of stochastic inputs and outputs.

Figure 4.5 shows the estimates of output (protein) entropy with different levels of noise. We notice that the output entropy does not change significantly as the number of molecules or stochasticity varies across a range. There is only a slight increase in output entropy for large noise. We emphasise here that this is purely the entropy of the output without any conditioning on the inputs.

The next set of results shown in figure 4.6 highlight the differences between deterministic and stochastic analyses. The blue bars represent the mutual information $I_2(k_m, k_p, d_m, d_p; P)$ and noise entropy $H_{\mathcal{N}}$ for deterministic cases. As expected, the levels of both these quantities remain unaffected for the deterministic case. The reason lies in the behaviour of noise entropy as the mutual information is just output entropy minus the noise entropy. This is because the noise entropy in this case is only the discretisation entropy and since there is

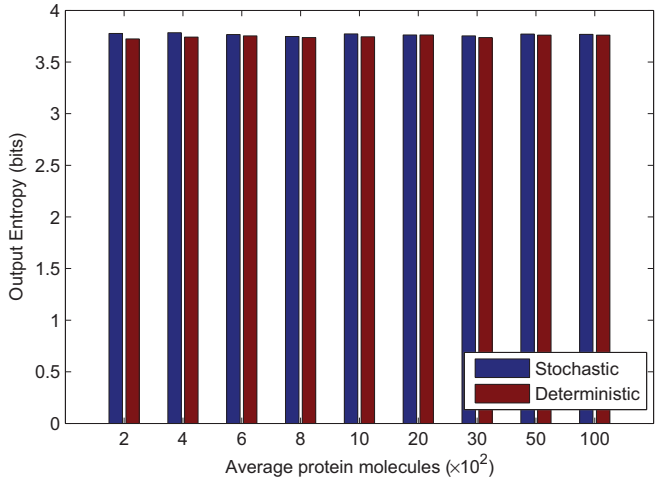

Figure 4.5: Output entropies $H(P)$ for deterministic and stochastic systems at different noise levels. For deterministic analysis, $N = 10^5$ and for stochastic analysis $N = 10^4$ and $M = 50$. Nominal parameter values are $k_p = 0.1$, $d_m = 0.01$ and $d_p = 0.001$

no intrinsic noise, the dicretisation entropy does not change with the number of molecules present in the system. However, for stochastic estimation of noise entropy, the intrinsic noise of the system comes into play in addition to the uncertainty due to binning. That is why as the number of protein molecules increase, or as the stochasticity goes down, the stochastic noise entropy goes down. At some point it even gets less than the corresponding deterministic noise entropy. This is the case $k_m = 0.3$ where the discretisation and noise start counteracting each others influence on the output. There can be many interesting phenomena that can be happening at this critical point, but this might need further mathematical analyses to verify. Although we have established in section 4.2.2 that stochastic noise entropy can be less than its deterministic counterpart, we are unsure when the system from figure 4.6 approaches the deterministic results. The verification of the system converging to its deterministic version requires that another noise entropy estimate with very high number of mRNAs and proteins be carried out. Our understanding is that at high levels of noise the outputs haphazardly cross binning boundaries to ultimately provide similarly populated output bins. This yields high levels of noise entropy. At lower enough noise levels, however, as in the case similar to that of figure 4.1, the stochastic noise entropy goes down. And for even lower noise levels, we believe that the stochastic time series

will almost be a straight line to give the same stochastic noise entropy as the discretisation entropy. The results in figure 4.6 may seem simple they can have a variety of implications. Reminding ourselves of equation 4.17, the most interesting finding is that stochasticity has the ability to decrease the sensitivity of the system. We shall elaborate this point in the discussion of the results that follow.

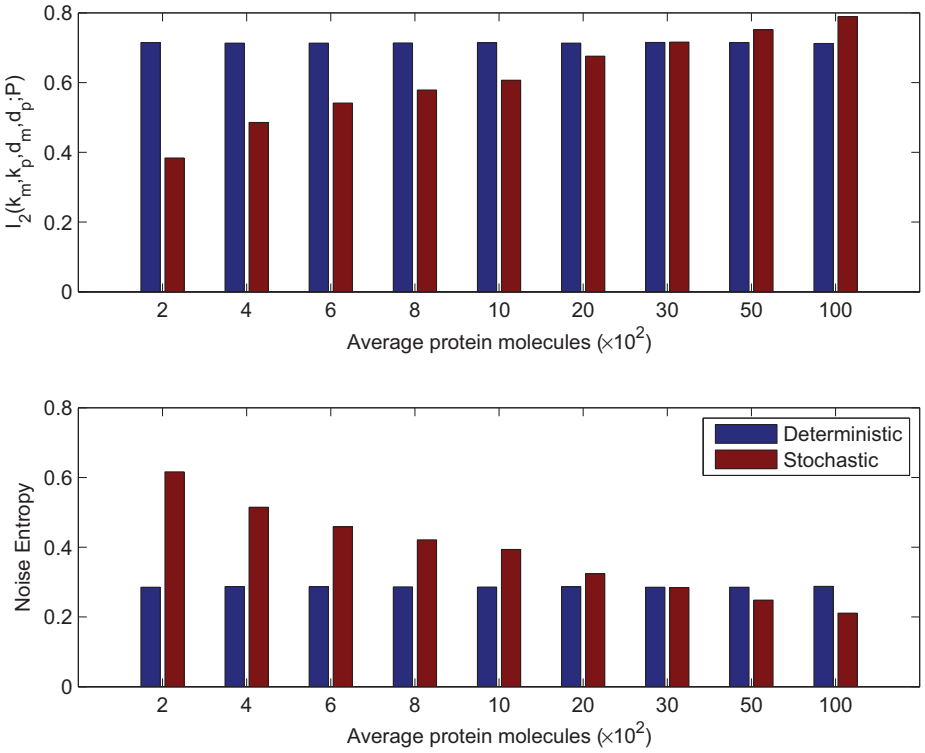

Figure 4.6: Normalised mutual information $I_2(k_m, k_p, d_m, d_p; P)$ and noise entropy $H_{\mathcal{N}}$ for deterministic and stochastic systems at different noise levels. The stochastic noise entropy is much higher than deterministic noise entropy at high noise levels, and it keeps going down as the noise goes down. For deterministic analysis, $N = 10^5$ and for stochastic analysis $N = 10^4$ and $M = 50$. Nominal parameter values are $k_p = 0.1$, $d_m = 0.01$ and $d_p = 0.001$

Figure 4.7 shows the effect of changing noise levels on the first order sensitivities of the system. Both the deterministic and stochastic analyses show that the system is almost equally sensitive to all the four parameters as far as the first order sensitivities are concerned. However, it is worth noting that the stochastic sensitivity level increases gradually,

albeit equally for all parameters, as the noise in the system decreases. At one point during the process of decreasing noise, this sensitivity even increases beyond the deterministic level of first order sensitivities.

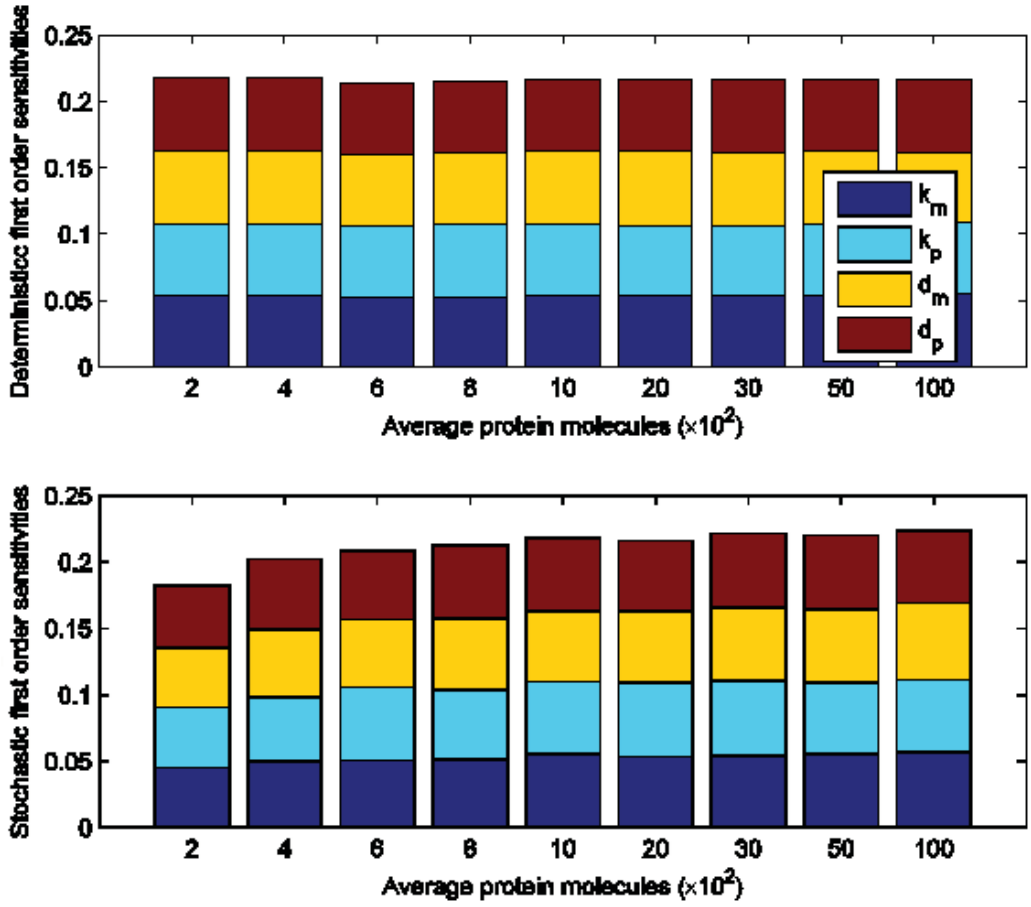

Figure 4.7: Normalised first order sensitivities for deterministic and stochastic systems at different noise levels.A change in the sensitivities is observed in the stochastic case as the level of stochasticity changes. For deterministic analysis, $N = 10^5$ and for stochastic analysis $N = 10^4$ and $M = 50$. Nominal parameter values are $k_p = 0.1$, $d_m = 0.01$ and $d_p = 0.001$

The effect of increase in sensitivity with decrease in stochasticity is more pronounced in the case of second order sensitivities. This is shown in figure 4.8, although we see that no two parameter pair dominates others while influencing the output. We observe similar trends in third order sensitivities and fourth order sensitivity of parameters as shown in

figures 4.9 and 4.10. There is a slight decrease in fourth order stochastic sensitivity for the least noise case but the actual value of the sensitivity index is too small when compared with other quantities to imply anything significant. However, it is worth noting that the parameter interactions at the first order do not cover the system functionality in its entirety, and that there are significant parameter interactions at higher orders.

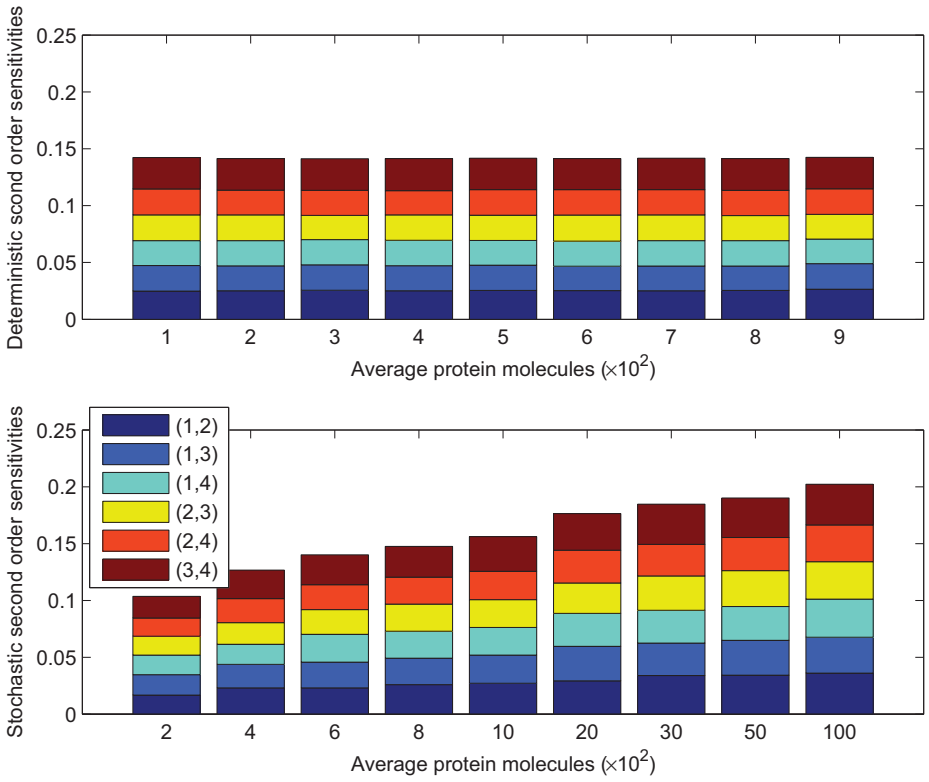

Figure 4.8: Normalised second order sensitivities for deterministic and stochastic systems at different noise levels. For deterministic analysis, $N = 10^5$ and for stochastic analysis $N = 10^4$ and $M = 50$. Nominal parameter values are $k_p = 0.1$, $d_m = 0.01$ and $d_p = 0.001$. The indices 1,2,3, and 4 on the legend represent the parameters k_m, k_p, d_m, and d_p respectively

Figure 4.11 shows the comparison of the contributions of parameters at different orders of interaction. In the deterministic setting, the first order and third order sensitivities are slightly higher than second and fourth order sensitivities. But the trend remains consistent as expected throughout the varying number of protein molecules. In the stochastic setting, on the other hand, the first order sensitivities dominate at high levels of noise while at

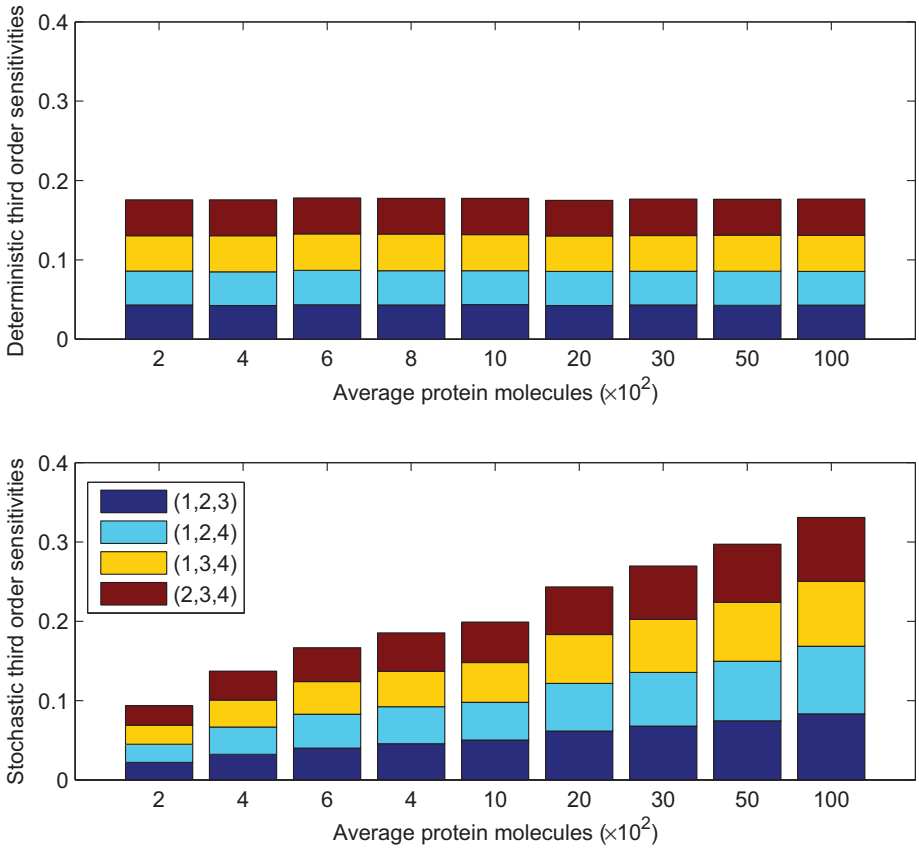

Figure 4.9: Normalised third order sensitivities for deterministic and stochastic systems at different noise levels. Like the first and second order sensitivities, third order parameter sensitivities also increase with decreasing noise. For deterministic analysis, $N = 10^5$ and for stochastic analysis $N = 10^4$ and $M = 50$. Nominal parameter values are $k_p = 0.1$, $d_m = 0.01$ and $d_p = 0.001$. The indices 1,2,3, and 4 on the legend represent the parameters k_m, k_p, d_m, and d_p respectively

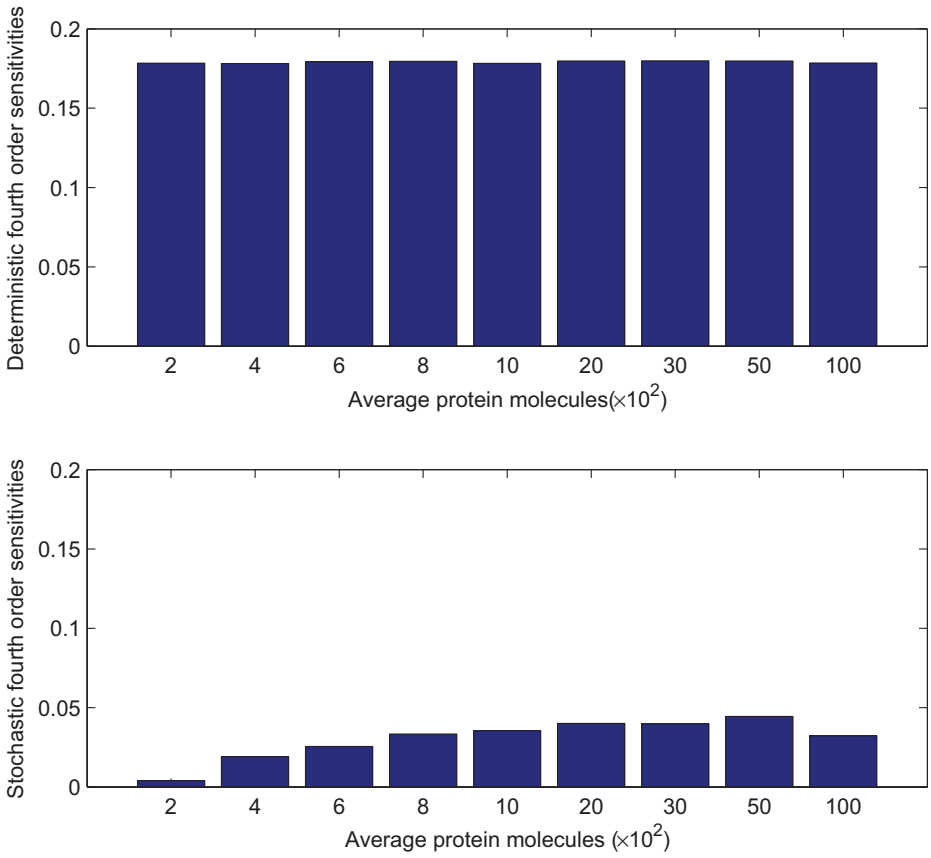

Figure 4.10: Normalised fourth order sensitivity for deterministic and stochastic systems at different noise levels. Deterministic fourth order sensitivities do not change while we observe a small change in the level of fourth order sensitivities as noise levels change. For deterministic analysis, $N = 10^5$ and for stochastic analysis $N = 10^4$ and $M = 50$. Nominal parameter values are $k_p = 0.1$, $d_m = 0.01$ and $d_p = 0.001$

lower stochasticity levels, the third order contributions come into a dominant position. Also the first and second order interactions become equally important. As compared to the deterministic case, the fourth order interaction remains the least throughout the different levels of stochasticity.

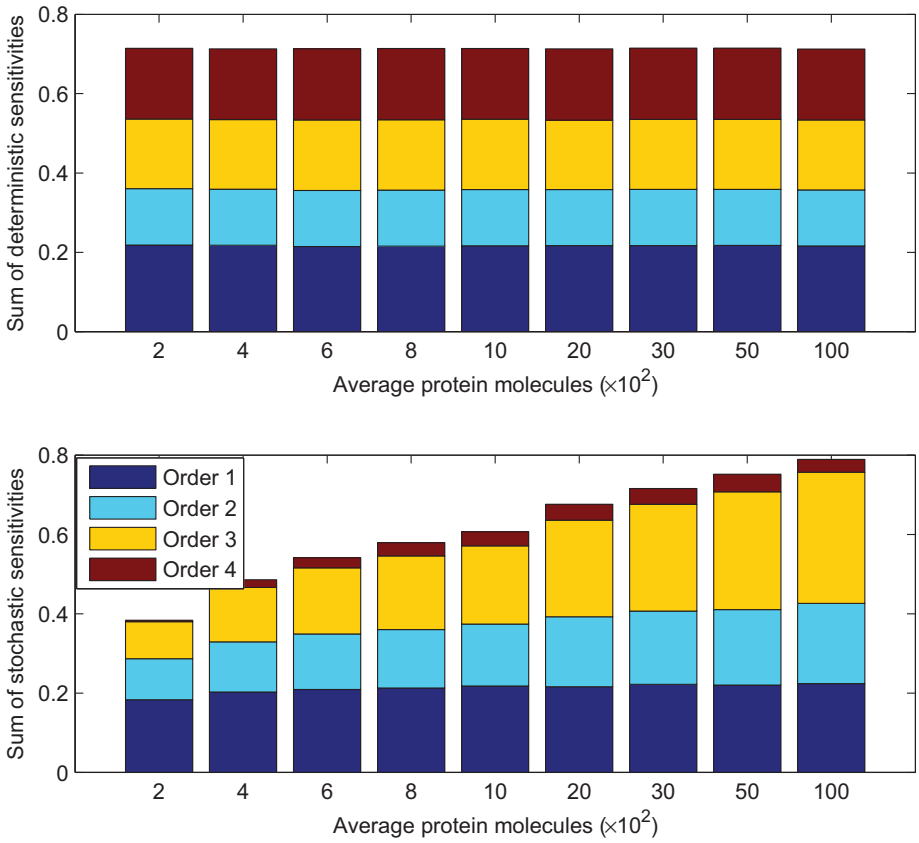

Figure 4.11: Normalised sum of all orders of sensitivities for deterministic and stochastic systems at different noise levels. For all levels of average protein molecules, the deterministic results remain the same. In the stochastic case, parameter sensitivities increase as noise levels go down. For deterministic analysis, $N = 10^5$ and for stochastic analysis $N = 10^4$ and $M = 50$. Nominal parameter values are $k_p = 0.1$, $d_m = 0.01$ and $d_p = 0.001$

Next we look at the total sensitivity indices of the four parameters. This is essentially the total effect a parameter can exert, independently as well as in combinations with other parameters, on the output. Figure 4.12 shows that the output is equally sensitive to all the

input parameters, and the sensitivity for all parameters decreases with increasing levels of noise.

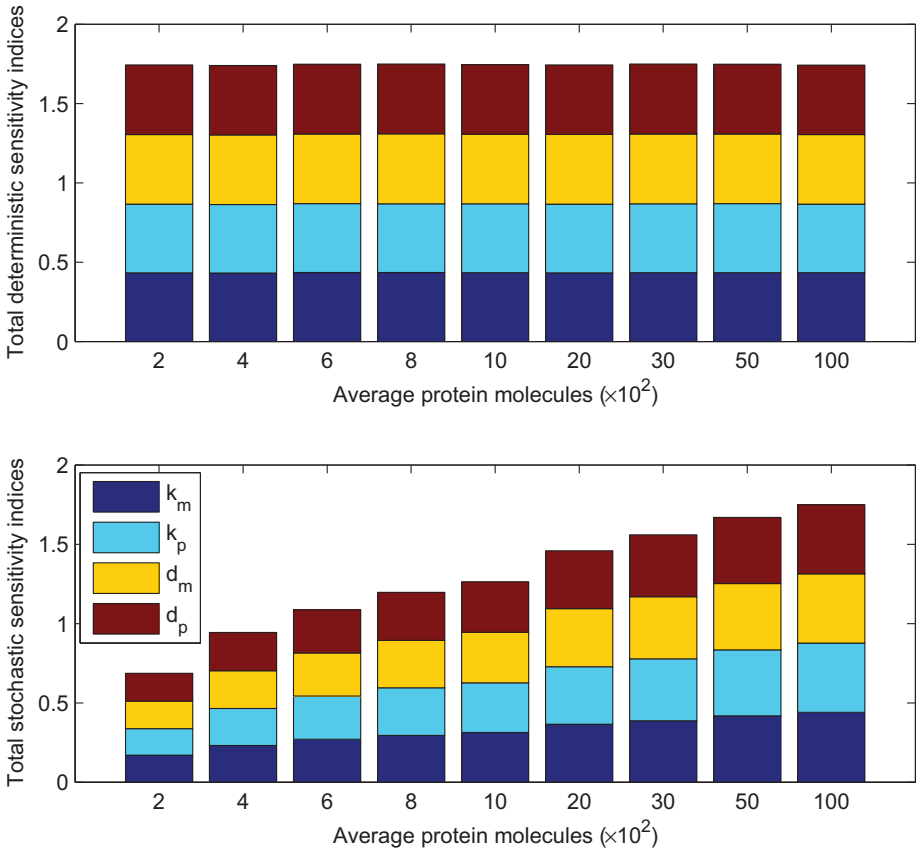

Figure 4.12: Changes in normalised total sensitivity indices for deterministic and stochastic systems at different noise levels. For deterministic analysis, $N = 10^5$ and for stochastic analysis $N = 10^4$ and $M = 50$. Nominal parameter values are $k_p = 0.1$, $d_m = 0.01$ and $d_p = 0.001$

In figure 4.13, we summarise all the results from the deterministic and stochastic sensitivity analyses as presented in the summation law 4.19 of sensitivities. The difference between stochastic and deterministic points of view are highlighted and the dominance of intrinsic noise effects over parametric sensitivities is shown.

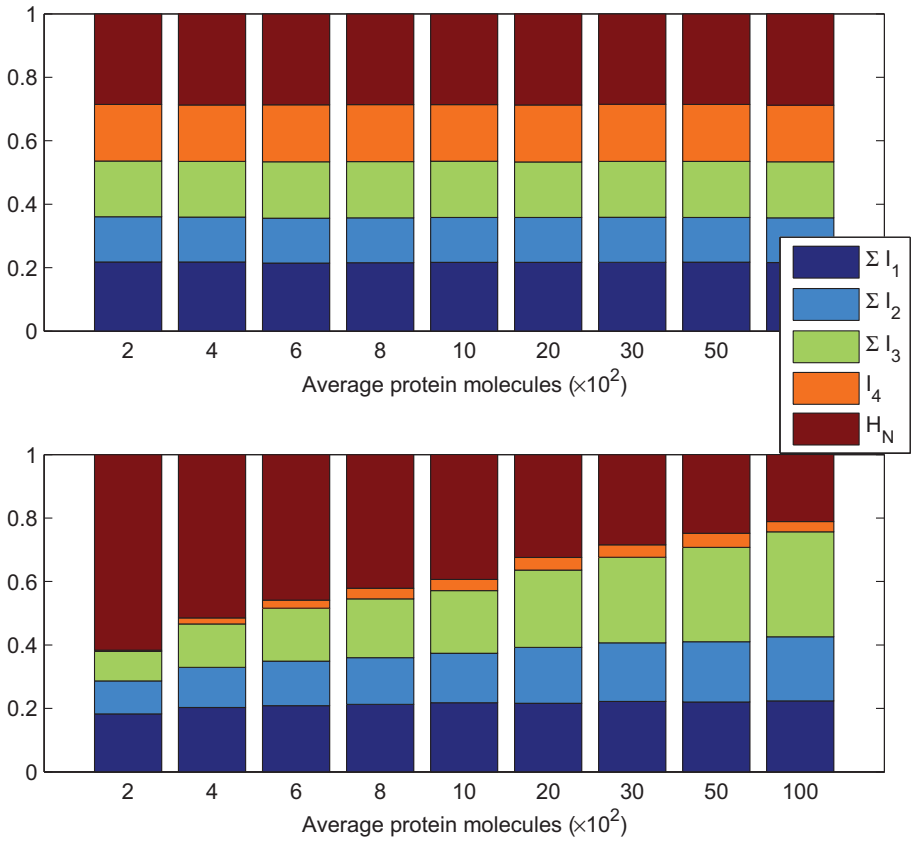

Figure 4.13: Noise can change sensitivities of parameters. This is demonstrated in sensitivity summation law for deterministic and stochastic systems at different noise levels. For deterministic analysis, $N = 10^5$ and for stochastic analysis $N = 10^4$ and $M = 50$. Nominal parameter values are $k_p = 0.1$, $d_m = 0.01$ and $d_p = 0.001$. The subscripts in the legend denote different orders of sensitivity

4.4.3 Uniform vs Lognormal Input Samples

We have looked at the sensitivity analysis in detail in the previous section. But the sensitivity analysis performed so far was based on uniform input samples. We want to see what happens when the input parameters are not uniformly distributed. The difference is mainly based on how local we want the sensitivity analysis to be. Further, in case of, say lognormally distributed input samples, more samples are drawn around the mean or nominal parameter values as compared to uniformly distributed input. But infact, in the current scenario of information theoretic sensitivity analysis, the shape of the parameter distributions is not as important as the entropy of those distributions. Two different distributions with similar entropies will yield similar results for the sensitivity analysis. We observed the same phenomenon when we tried to compare the previously studied case of uniformly distributed parameters with the case where the parameter distributions are lognormal. We chose the case where the nominal values yield deterministic nominal values of outputs as $mRNA = 10$ and $P = 1000$. We also let the variance of the underlying normal distribution be 10 percent of the nominal values.

Figure 4.14 shows the similarity between the results from both the analyses. The values of sensitivities for the lognormal input are slightly higher than those for the uniform input. This is because the parameter entropy has slightly higher values for the lognormal case. The remaining trend is very similar to the previous case. However, we get negative fourth order sensitivity value for the lognormal input case, which is why in figure 4.14, the sum of sensitivities and noise entropy seem to be more than 1. In reality, the sum is 1 with a minus sign for the fourth order sensitivity. The value is very small although we are not sure why this could be the case. We cannot be sure but studies like [132] point out the fact that Shannon entropies and mutual informations can sometimes accommodate redundant terms. Although we have not tested their suggestions, they provide a potential solution to the problem of negative mutual informations.

4.4.4 Gene Expression Model with Negative Feedback

In this section, we apply the methodology to a gene expression model with negative feedback [114]. The model is shown in figure 4.15. In this model of gene expression, the DNA,

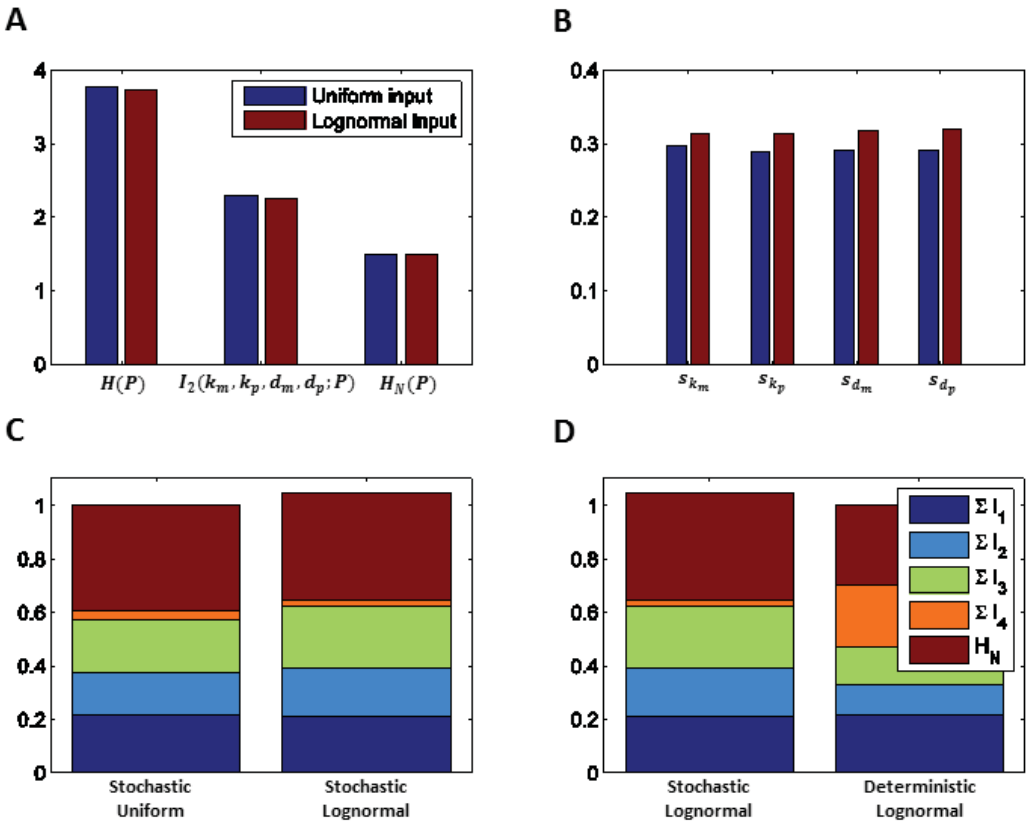

Figure 4.14: Comparison between sensitivity analyses with uniformly versus lognormally distributed input parameters. Deterministic analysis is also performed for reference. For deterministic analysis, $N = 10^5$ and for stochastic analysis $N = 10^4$ and $M = 50$. Nominal parameter values are $k_m = 0.1$, $k_p = 0.1$, $d_m = 0.01$ and $d_p = 0.001$. (A) Stochastic analysis of output entropy, mutual information and noise entropy. (B) stochastic total sensitivity indices. (C) Summation theorem demonstration for stochastic setting with uniformly and lognormally distributed input samples. (D) Summation theorem demonstration for stochastic and deterministic setting with lognormally distributed input samples

denoted as D is suppressed by the protein P. In other words, the protein binds to the DNA turning it to an off state, D_{off} which cannot take part in transcription anymore. The model therefore involves a reversible reaction of proteins binding to DNA, and two additional reaction rates v_1 and v_2. Working with this model, we will assume that the average number of DNA molecules present is 1, implying that $D_{off} = 1 - D$. We shall also restrict our analysis to a region of parameter space in which the overall system dynamics do not change and no bifurcations take place. The deterministic ODE system is as follows:

$$\begin{aligned} \frac{dD}{dt} &= v_2(1-D) - v_1 PD \\ \frac{dM}{dt} &= k_m D - d_m M \\ \frac{dP}{dt} &= k_p M - d_p P + v_2(1-D) - v_1 PD \end{aligned} \tag{4.23}$$

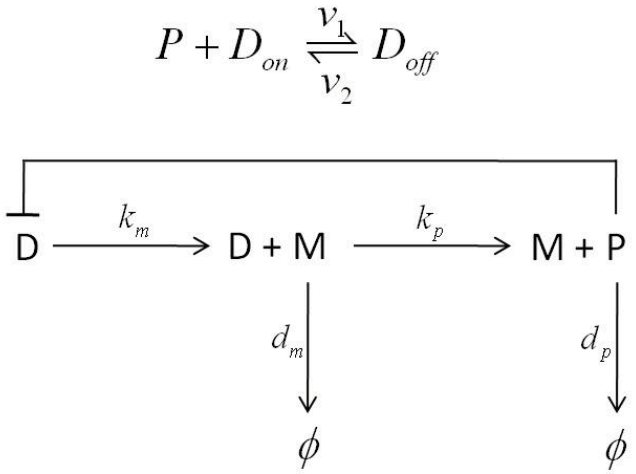

Figure 4.15: Gene expression model with negative feedback. The model involves transcription, translation, and a negative feedback loop.

Figure 4.16 shows results of application of sensitivity analysis methodology to the six parameter gene expression model from figure 4.15. We choose protein levels as our output variable. The analysis is performed twice for two different levels of stochasticity. As for the

four parameter model, the variation in the levels of stochasticity is achieved by changing the nominal value of the transcription rate k_m, where lower transcription rate means lower number of steady state protein and mRNA levels which further implies higher level of noise. The parameters used to obtain the results are shown in the caption of the figure. The analysis with negative feedback model further illustrates our claim that stochasticity has the ability to increase or decrease parameter sensitivity.

Panel A of the figure 4.16 shows that even though the output entropy is same, the total mutual information and noise entropy differ significantly for the two levels of noise used. The high noise/stochasticity levels correspond to high levels of noise entropy, thereby pushing the total mutual information to a lower level as compared to the case when the noise is lower. The total mutual information, as discussed before, encompasses the different orders of sensitivity, so a lower mutual information implies that the sensitivity of the system has been suppressed by noise at some or all orders. It should be noted here that for the six parameter case with limited computational ability, the total mutual information is obtained as the difference between output entropy and noise entropy. Individual sensitivity estimates have been made only upto second order.

Panel B of the figure 4.16 shows the first order sensitivities of all parameters in the negative feedback model. It is clear that for our choice of nominal parameters, the system is more sensitive to k_m, k_p, d_m, and d_p as compared to v_1 and v_2. This is expected since for our purposes, our choices of nominal values for v_1 and v_2 are not entirely free. The choices are made dependent on other parameters so that the values of D_{on} and D_{off} are both maintained at a steady state rate of $\frac{1}{2}$. Further, it can be noted that at individual parameter levels (first order) the protein levels are more sensitive to all the six parameters in case of low stochasticity. The most interesting observation here is that at high noise, the parameters v_1 and v_2 become more sensitive. This is as expected since they control the strength of the negative feedback which is relevant at high noise. Similar observation is made for total sensitivity indices of parameters, as shown in panel C. Panel D is an illustration of the sensitivity summation theorem. The comparison between different levels of sensitivities are also shown.

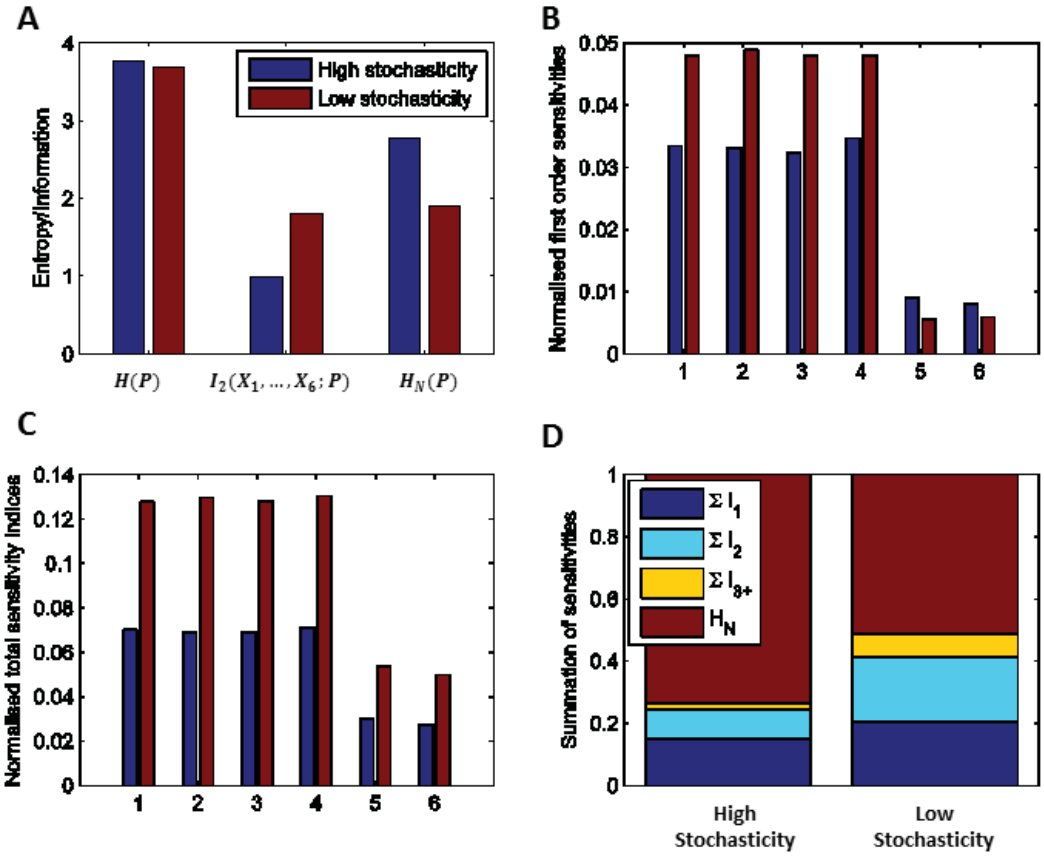

Figure 4.16: Sensitivity analysis of the six parameter gene expression model from figure 4.15 for different levels of stochasticity. For the analysis, we use $N = 15000$ and $M = 50$. Nominal parameter values are $k_m = 0.1$ for high stochasticity and $k_m = 10$ for low stochasticity, $k_p = 0.2$, $d_m = 0.01$, $d_p = 0.01$, $v_1 = 0.1$ and $v_2 = \frac{k_m k_p}{2 d_m d_p} v_1$. (A) Stochastic analysis of output entropy, mutual information and noise entropy. (B) Normalised first sensitivity indices. (C) Normalised total sensitivity indices. (D) Summation theorem demonstration for low and high stochasticity with lognormally distributed input samples

4.5 Possible Modifications of the Methodology

The methodology developed and presented in the previous sections is in the form that yields the most accurate results. However, this analysis is computationally very expensive. Performing tens and hundreds of stochastic simulations can prove very difficult. And although we use bias correction, most of the bias correction techniques require that the number of samples be in asymptotic sampling regime. We have only performed the analysis for a six parameter system. However, the complexity can grow very rapidly as the number of input parameters increases. The first thing that is affected is the accuracy of the estimates of noise entropy and mutual information. These two can be very inaccurate if the analysis is performed with a number of samples that is not enough. Then the accuracy of higher orders of sensitivities becomes doubtful. The methodology under discussion then needs to be modified with different tools.

The first option one could try is to replace the Gillespie simulation with something like linear noise approximation that is less time consuming and less taxing on the processor. We have tried using this modification, and it yields results in very good agreement with the ones presented in the previous section. Although faster than the stochastic simulations, the idea fails for systems with high levels of noise and nonlinearity in the model, and especially for oscillatory systems.

We have also used Unscented transform successfully to improve the computational effort required for the method under discussion. We present more details on this later in this chapter. In short, the application of Unscented transform enables us to capture the essence of the output distribution without having to run Gillespie simulations for all the samples. The Gillespie simulations are only performed for a carefully chosen set of input parameter values, and these simulations are only $2M_{in} + 1$ in number. The results obtained are quite accurate, and there is a huge computational relief in switching from tens of thousands of Gillespie simulations to only a few.

4.6 Application of Unscented Transform for Sensitivity Analysis

As discussed in the previous section, one of the most effective ways to improve the computational efficiency of our sensitivity analysis methodology is to use the unscented transform to capture output distribution effectively. The section 2.7 details the application and computational details of the methodology. For our purposes, in this section we describe how we incorporate the methodology effectively at the step of performing stochastic simulations.

Our methodology discussed earlier in this chapter adopts a Monte Carlo based approach to analyse different samples of parameter inputs through Gillespie's simulations. The simulations are expensive and take up most of the time for sensitivity analysis. Instead of adopting a Monte Carlo based approach, we use unscented transform, a weighted set of sigma points that capture the input space effectively and only pass those selected points through stochastic simulations. We then approximate the output distribution using the outputs corresponding to the sigma points. From the output distribution that the unscented transform yields, we generate samples for sensitivity estimates. Below we describe the procedure in detail.

- We choose sigma points using (2.37) for scaled unscented transform. We assume that the input parameters are lognormally distributed. But we know that the unscented transform has best accuracy when the input distribution is Gaussian. To still be able to get best results through unscented transform, we choose our Gaussian prior to be $\Xi = \ln(X)$ for parameters X. That is, we actually find the sigma points, denoted by Φ_k, of the underlying Gaussian distribution Ξ.

- The next step is to choose an appropriate nonlinear function or the form of the output we desire from the results of stochastic simulations. This choice varies with problem at hand. Our main goal is to be able to reproduce samples of the output distribution. In general, nonlinear function will do the following to each of the sigma points:

  (i) Take exponential of each coordinate of the sigma point (because sigma should follow Gaussian distribution for best accuracy, and we are assuming that our inputs are lognormal)

  (ii) Apply Gillespie simulation to the resulting points

(iii) From the output time series for each input, calculate means and variances for the logarithms of each output variable, covariances between logarithms of each output variable

If the outputs of the original problem of sensitivity analysis are different species, like protein numbers in the gene expression model, it is best to assume a lognormal output distribution from the unscented transform. Another important factor that affects our choice of the function is the fact that we need to sample various stochastic outputs corresponding to each input. Therefore, we need to incorporate the covariance among the stochastic outputs into the nonlinear function. We shall describe the exact choice of this nonlinear function with every example we apply unscented transform to.

- With the outputs corresponding to all the sigma points, the output distribution Γ with mean μ_Γ, covariance P_Γ, and cross covariance $P_{\Xi,\Gamma}$ is reconstructed using the scheme (2.40). We note that the dimension of Γ is $M_{out}(M_{out}+3)/2$, with M_{out} elements for means of (logarithms of) output variables and $M_{out}(M_{out}+1)/2$ elements for covariance of (logarithms of) output variables.

- We now need to be able to perform sensitivity analysis as with the full scale methodology. For this purpose, we construct a vector of means, say Z that is a combination of means from input distribution and the means of the outputs from the previous step. That is $Z = [\mu_\Xi, \mu_\Gamma]$. Similarly, another concatenated covariance matrix is constructed using the input covariance, input-output cross covariance and output covariance in the form $P_Z = \begin{bmatrix} P_\Xi & P_{\Xi,\Gamma} \\ P_{\Xi,\Gamma}^T & P_\Gamma \end{bmatrix}$. N Gaussian samples from these means and covariances are then generated.

- The samples are split into input and output parts. We then take exponential of the input samples to get the final N lognormal input samples.

- Since the outputs parts of the samples also include the covariance among output samples, another M Gaussian outputs are sampled using this information for each of the N original samples. Exponential function is then applied to each of them yielding NM positive lognormal output samples.

- The usual sensitivity analysis apparatus is now complete. We then proceed as usual to bin the inputs and outputs and obtain bias corrected estimates for sensitivity analysis.

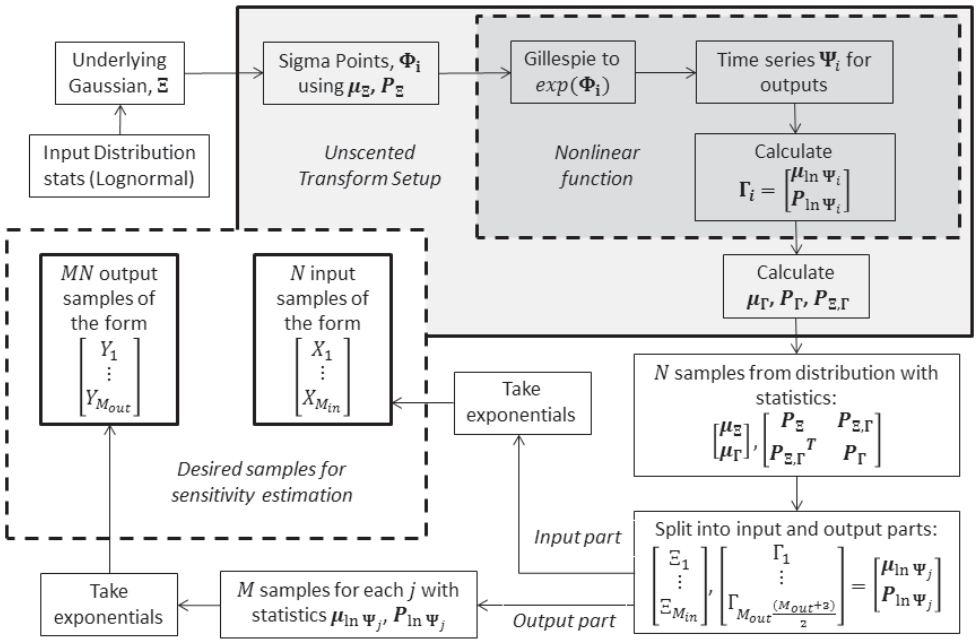

Figure 4.17: Schematic diagram for incorporating unscented transform in the sensitivity analysis methodology; $i = 1, \ldots, 2M_{in} + 1$, and $j = 1, \ldots, N$. Starting, just as in the method with Monte Carlo sampling, with a lognormal input distribution the unscented transform provides the first two moments of the output distribution without the need for MN Gillespie simulations. Samples from the input and output distributions are generated to eventually yield N input and MN output samples for sensitivity analysis

Figure 4.17 presents a detailed illustration of the above methodology. It is worth noting that the above method is only useful in the situations when only sensitivity analysis and output entropies are required. The above modification is not meant for estimating noise entropy. Estimating noise entropy is different from output entropy in the sense that it requires averaging over thousands of entropy estimates, each for a particular input bin combination. So even though unscented transform can be used to replace the averaging step, this does not

have a significant computational advantage as hundreds or thousand of samples still need to be generated and for each bin combination. All these require a stochastic simulation run. The other way around using unscented transform is to apply unscented transform twice, once taking care of the averaging step, and next when computing entropies is required. So for a system with m input parameters, this arrangement will require $(2m + 1)^2$ stochastic simulations. The repeated approximation here will affect the accuracy of the estimates, as we are approximating twice. In addition to this, we already know that unscented transform yields most accurate results when the prior or input distribution is Gaussian. But if we go ahead with using unscented transform twice as mentioned, the second application requires us to sample from particular bin combinations, thus restricting sampling from uniform distribution with fixed boundaries dictated by the binning scheme. This additionally affects the accuracy of the results obtained. Due to these shortcomings of the application of unscented transform in our particular scenario, we restrict ourself to only performing sensitivity analysis with this approximation.

4.6.1 Sensitivity Analysis of the Gene Expression Models with Unscented Transform

Figures 4.18, 4.19, and 4.20 show the comparison between results of sensitivity analysis when unscented transform and Monte Carlo approaches are used on the four parameter gene expression model without feedback. For these analyses, the difference between computational expenses is huge. The steps of random sampling of 10^4 parameter combinations and stochastic simulations for all these samples in the usual Monte Carlo based approach are replaced by calculating only 9 sigma points in parameter space, performing stochastic simulations for them to get the output distribution and then sampling from that distribution in the unscented transform based approximation. The output binning and bias corrected estimates of information and entropy are made in a similar way in both methods. Thus the stochastic simulations are reduced a thousand times with unscented transform for this particular example. This computational advantage is huge considering the results from both approaches match very well.

From the perspective of applying unscented transform to obtain the results in this case,

the first step is the choice of nonlinear function, the estimates of which are to be made for all the sigma points. We have assumed that the parameters are lognormally distributed therefore the sigma points are calculated in the associated normal distribution and the nonlinear function then applied to the exponential function of the sigma points. Similar approach has been used in [120]. We then choose the output of the nonlinear function to generate a vector of the form $\left[\overline{\ln m} \quad \overline{\ln P} \quad \text{Var}_{\ln m} \quad \text{Var}_{\ln P} \quad \text{Cov}_{\ln m,\ln P}\right]$. The reason for this choice is that when we generate Gaussian samples from the output distribution, the exponential of those samples will be lognormally distributed, which is a more correct choice for the distributions of mRNAs and protein. Even if we were to choose the vector $\left[\overline{m} \quad \overline{P} \quad \text{Var}_m \quad \text{Var}_P \quad \text{Cov}_{m,P}\right]$ and generate lognormal samples for outputs, we would have to go through the route of associated normal distribution for sampling purposes. Another reason for our choice of such an output over the choice $\left[\ln m \quad \ln P\right]$ is that we are dealing with a stochastic setting, in which case we need a complete output distribution for each parameter sample rather than a single output value. While this choice would have been the correct one in a deterministic setting, our whole point in choosing the five element output vector is that when we combine inputs and outputs and then sample together as described in the previous section, we sample distributions for each input sample. If we look closely, the vector $\left[\overline{\ln m} \quad \overline{\ln P} \quad \text{Var}_{\ln m} \quad \text{Var}_{\ln P} \quad \text{Cov}_{\ln m,\ln P}\right]$ accommodates a complete distribution with mean and covariances of $\ln m$ and $\ln P$. This solves the problem when we eventually split the input and output parts and generate multiple outputs per input. The only other technicality that remains is that sometimes negative Gaussian values are sampled for the variances in the output as well. We ignore these rare samples and also use the nearest semi positive definite matrix [42] whenever numerical errors render the covariance matrix as non positive semidefinite.

Figure 4.18 shows the comparison between informations, protein entropies and noise entropies for the two methods. The results are very close to each other. Similarly, the total sensitivity indices obtained by the two methods also match very well as shown in figure 4.19. Figure 4.20 is an illustration of the summation law detailing the comparison among all orders of sensitivities for the Monte Carlo results and the unscented transform estimates. It is interesting to note that the negative fourth order interactions in the Monte Carlo case resulting from limited sampling are not encountered in case of unscented transform.

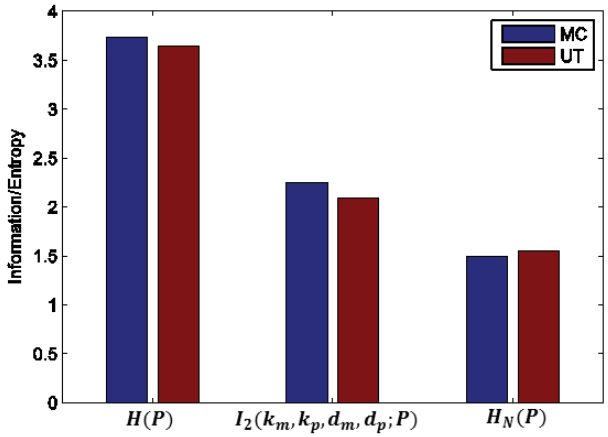

Figure 4.18: Comparisons of output entropy, mutual information, and noise entropy estimates for four parameter gene expression model. The results compared are for the full sensitivity analysis method and the corresponding approximation by unscented transform. For full Monte Carlo type analysis $N = 10^4$ and $M = 50$. Nominal parameter values are $k_m = 0.1$, $k_p = 0.1$, $d_m = 0.01$ and $d_p = 0.001$. For unscented transform version, 900 stochastic outputs were used to evaluate the nonlinear function for each of the sigma points. Also, $N = 20000$, and $M = 100$. The unscented transform parameters used are $\alpha = 0.9$, $\kappa = 0$, and $\beta = 2$

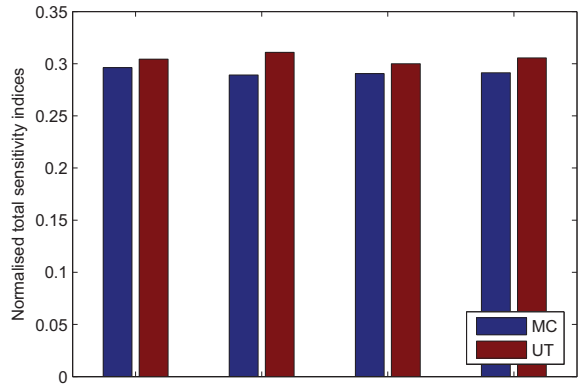

Figure 4.19: Estimates for total sensitivity indices for four parameter gene expression model. The results compared are for the full sensitivity analysis method and the corresponding approximation by unscented transform. The sensitivity indices calculated incorporate upto second order sensitivities. For full Monte Carlo type analysis $N = 10^4$ and $M = 50$. Nominal parameter values are $k_m = 0.1$, $k_p = 0.1$, $d_m = 0.01$ and $d_p = 0.001$. For unscented transform version, 900 stochastic outputs were used to evaluate the nonlinear function for each of the sigma points. Also, $N = 20000$, and $M = 100$. The unscented transform parameters used are $\alpha = 0.9$, $\kappa = 0$, and $\beta = 2$

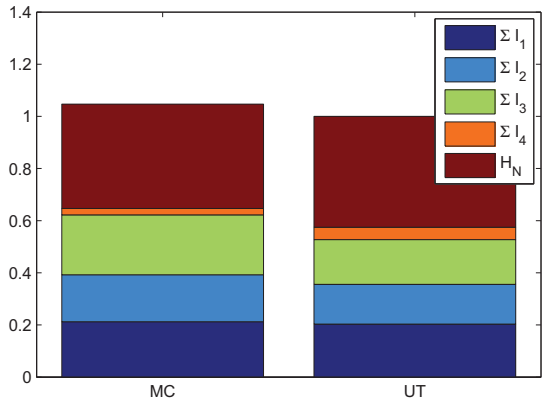

Figure 4.20: Summation theorem demonstration for four parameter gene expression model. The results compared are for the full sensitivity analysis method and the corresponding approximation by unscented transform. The sensitivity indices calculated incorporate upto second order sensitivities. For full Monte Carlo type analysis $N = 10^4$ and $M = 50$. Nominal parameter values are $k_m = 0.1$, $k_p = 0.1$, $d_m = 0.01$ and $d_p = 0.001$. For unscented transform version, 900 stochastic outputs were used to evaluate the nonlinear function for each of the sigma points. Also, $N = 20000$, and $M = 100$. The unscented transform parameters used are $\alpha = 0.9$, $\kappa = 0$, and $\beta = 2$

The setup for unscented transform application discussed in [120] is very useful when applying unscented transform for sensitivity analysis in our case. In [120], stochastic systems are analysed that are subject to both intrinsic and extrinsic variability. For the purpose, the authors design a framework where they apply linear noise approximation to analyse the stochasticity of the system. That is, the stochastic outputs corresponding to the sigma points are obtained by applying linear noise approximation. That, in effect, means two levels of approximation - replacement of Monte Carlo sampling by unscented transform and approximating exact stochastic simulations by linear noise approximation. It is noteworthy here that while we speed up our sensitivity analysis methodology by approximating with unscented transform, we do not compromise on the accuracy of the stochastic outputs. Since the simulations with unscented transform are so few in number, it is worthwhile to employ the exact stochastic simulation algorithm rather than adding another degree of approximation. This is particularly useful in nonlinear models with high stochasticity where linear noise approximation results deviate from the exact results.

To illustrate the above point, we perform sensitivity analysis for the six parameter

nonlinear gene expression model for three cases: Monte Carlo sampling combined with stochastic simulation, Monte Carlo sampling with linear noise approximation, and unscented transform approximation combined with stochastic simulations. The first case is the most computationally expensive and accurate, the second case has only one level of approximation at the level of stochastic output generation, and the third method also has one level of approximation at the sampling level. The results in figure 4.21 show that it is better to use approximation (unscented transform) at sampling level rather than trying to achieve computational efficiency by compromising on stochastic simulations. The results for linear noise approximation differ significantly from the exact results even with Monte Carlo sampling. One can only expect that further approximation by unscented transform can only worsen the accuracy of the results.

For the unscented transform results in figure 4.21, same nonlinear function output is used as for the four parameter model. The results for the unscented transform match well with the exact results, even though only 13 stochastic simulations are performed instead of the the original 15000. All the results show that the unscented transform methodology works very well and is more accurate while being efficient than the linear noise approximation approach.

4.7 Stochastic Sensitivity Analysis of Circadian Clock Model

From cyanobacteria to mammals, organisms and their biological systems use circadian rhythms to regulate their behaviour [80]. Cycles of days, months, and years appear in the normal functioning in their lives. For example, our sleep patterns are regulated everyday, and there is a typical sleep cycle we follow. Climates and temperature cycles take place repeatedly over years. This behaviour is essential to normal working of many biological systems. It is therefore desired that this periodic behaviour be as robust as possible with respect to environmental changes [3]. In this section, we shall apply the stochastic sensitivity analysis technique developed in the previous sections to a model of circadian rhythms from [126].

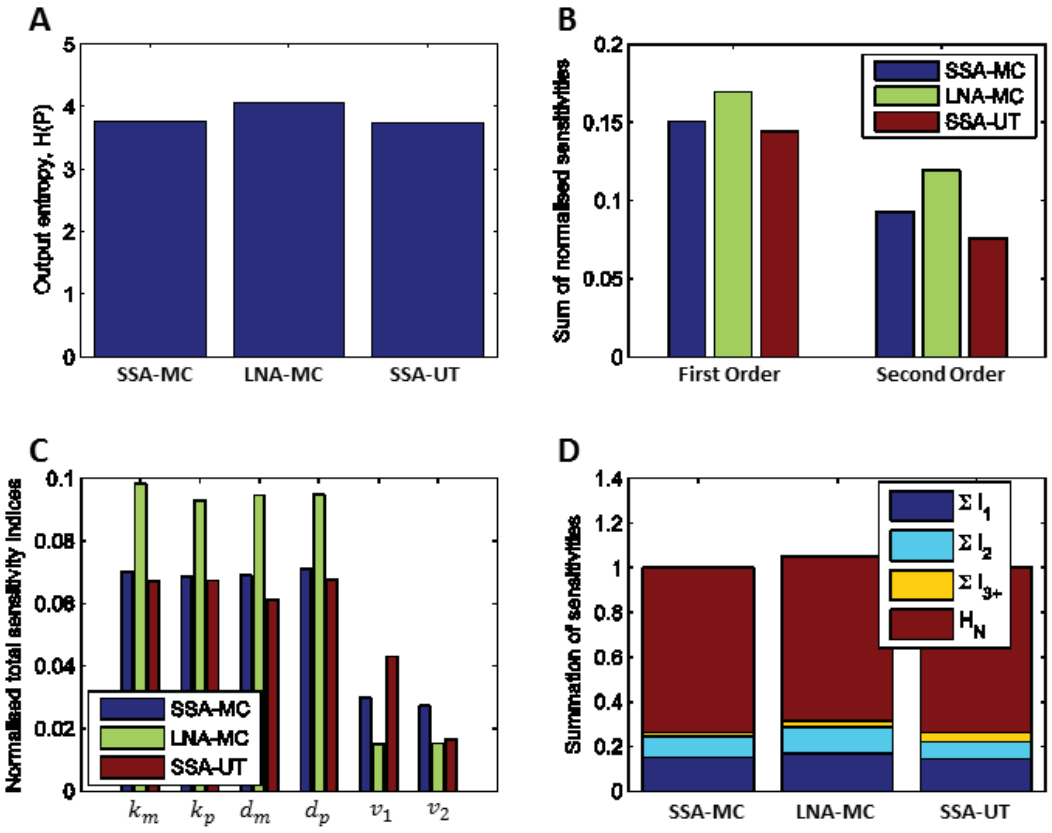

Figure 4.21: Sensitivity analysis for six parameter gene expression model with negative feedback. The results compared are for the full Monte Carlo sensitivity analysis method with SSA and the corresponding approximations by unscented transform combined with SSA and linear noise approximation combined with Monte Carlo sampling. The Unscented transform approximation combined with Gillespie simulations provides better approximation as compared to LNA with Monte Carlo sampling. For full Monte Carlo type analysis with SSA and LNA, $N = 15000$ and $M = 50$. Nominal parameter values are $k_m = 0.1$, $k_p = 0.1$, $d_m = 0.01$ and $d_p = 0.001$. For unscented transform version, 900 stochastic outputs were used to evaluate the nonlinear function for each of the sigma points. Also, $N = 20000$, and $M = 100$. The unscented transform parameters used are $\alpha = 0.9$, $\kappa = 0$, and $\beta = 2$ (A) Output entropy estimates. (B) Normalised sums of first and second order sensitivities. (C) Normalised total sensitivity indices upto second order sensitivities. (D) Summation theorem demonstration for with lognormally distributed input samples. The noise entropy obtained from SSA-MC method is used for all three illustrations The sensitivities for LNA-MC method at orders higher than 2 are negative

4.7.1 The Model

The model of circadian rhythms that we use as an application of our stochastic sensitivity analysis is discussed in [126] in detail. Figure 4.22 shows the detailed diagram of the model. The model incorporates the essential elements of gene expression that allow the organism to regulate physiological changes in order to adapt itself to different times of a day, month or other similar cycle. The model does not capture details for particular organisms, but rather provides a general setup necessary to produce circadian rhythms.

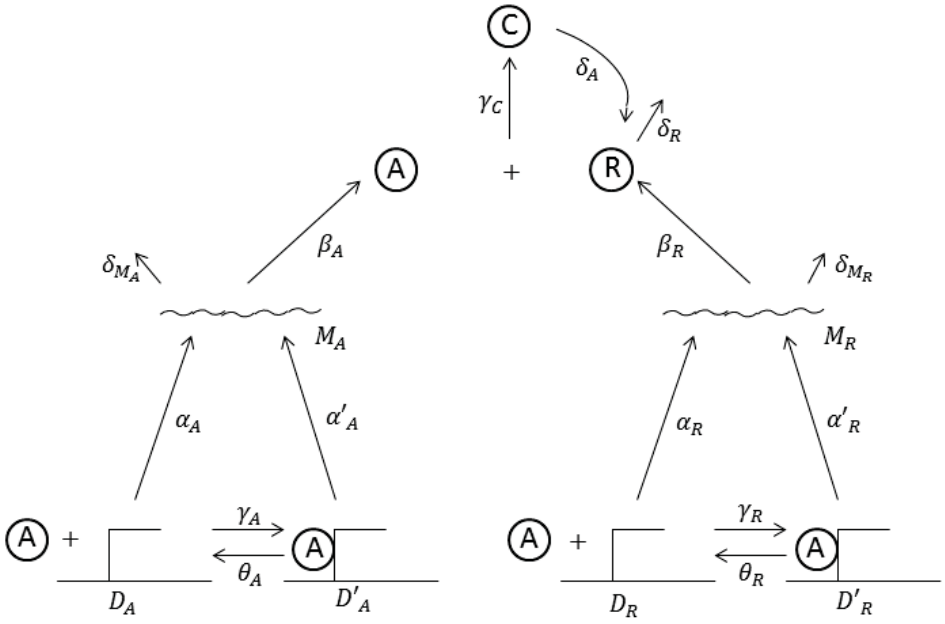

Figure 4.22: Adapted from [126], The circadian clock model that represents a general setup that produces circadian rhythms in organisms - binding and unbinding of activator and repressor genes; transcription, translation and degradation; as well as complex formation

The model consists of an activator protein A, a repressor protein R, and an inactivated complex formed by them denoted by C. At the transcription level, M_A and M_R are mRNAs of A and R respectively. Also, there are promoter states for activator genes, D'_A and D_A, denoting the states when the activator protein A is bound and not bound to its promoter respectively. Similarly, D_R and D'_R represent the repressor genes. Moreover, the complex C degrades to yield R, and A is degraded as this happens. There are 15 reaction rates involved in the whole process, where γ's and θ's represent the rates of A binding and

unbinding from others respectively, β's are translation rates, α's are transcription rates, and δ's denote spontaneous degradation. The deterministic ODE system [126] can be written as follows:

$$\begin{aligned}
\frac{dD_A}{dt} &= \theta_A D'_A - \gamma_A D_A A \\
\frac{dD_R}{dt} &= \theta_R D'_R - \gamma_R D_R A \\
\frac{dD'_A}{dt} &= \gamma_A D_A A - \theta_A D'_A \\
\frac{dD'_R}{dt} &= \gamma_R D_R A - \theta_R D'_R \\
\frac{dM_A}{dt} &= \alpha'_A D'_A + \alpha_A D_A - \delta_{M_A} M_A \\
\frac{dA}{dt} &= \beta_A M_A + \theta_A D'_A + \theta_R D_R \\
&\quad - A(\gamma_A D_A + \gamma_R D_R + \gamma_C R + \delta_A) \\
\frac{dM_R}{dt} &= \alpha'_R D'_R + \alpha_R D_R - \delta_{M_R} M_R \\
\frac{dR}{dt} &= \beta_R M_R - \gamma_C A R + \delta_A C - \delta_R R \\
\frac{dC}{dt} &= \gamma_C A R - \delta_A C
\end{aligned} \tag{4.24}$$

Figure 4.23 represents a Gillespie realisation of the model dynamics. The parameters used are given in the caption of the figure. As discussed analytically in [126], the model can be reduced to two slow variables R and C, and A is also a function of R. Therefore, these three variables capture the dynamics of the system. The Gillespie realisation of the stochastic version of the system also reveals that the numbers of the rest of the variables are very small when compared to A, R, and C. It can be easily seen that the all three of them have peaks in thousands of molecules. The other variables not shown in the figure, are at most in their tens in comparison. We shall restrict out attention to these three output variables in our analysis that follows.

In the next section we discuss the results of stochastic sensitivity analysis of the period of the circadian cycle using the unscented transform approach. We shall restrict our analysis

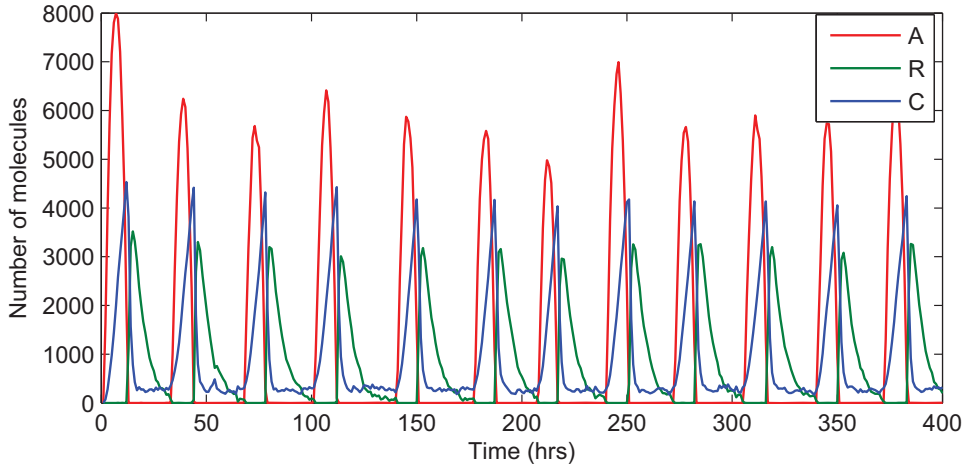

Figure 4.23: Circadian rhythms observed in A, R, and C. It can be easily seen that the periods of these three variables are the same. The model parameters used are $\gamma_A = 50M^{-1}h^{-1}$, $\theta_A = 50h^{-1}$, $\gamma_R = 1M^{-1}h^{-1}$, $\theta_R = 100h^{-1}$, $\delta_{M_A} = 10h^{-1}$, $\delta_{M_R} = 0.5h^{-1}$, $\alpha'_A = 500h^{-1}$, $\alpha_A = 50h^{-1}$, $\beta_A = 50h^{-1}$, $\beta_R = 5h^{-1}$, $\delta_A = 1h^{-1}$, $\delta_R = 0.2h^{-1}$, $\gamma_C = 2M^{-1}h^{-1}$, $\alpha_R = 0.01h^{-1}$, $\alpha'_R = 50h^{-1}$, System volume $V = 10^{-17}$L.

to one particular dynamical system behaviour where oscillations do not break down. The stability anaysis of the corresponding deterministic model can be found in [126].

4.7.2 Sensitivity Analysis Results and Discussion

For the circadian clock model under consideration, Gillespie simulations are quite time consuming. With 15 model parameters, the number of samples needed to accurately estimate entropies and mutual informations can be very large. Therefore, the stochastic sensitivity analysis with Monte Carlo type sampling can be computationally very expensive. We use the modified version of the methodology that makes use of unscented transform to capture the output distribution without the need for thousands of Gillespie simulations.

In the stochastic version of the circadian model, the proteins A, R, and the complex C are the only three slow variables. The rest of the species are produced and degraded with only a small number of molecules. So, for sensitivity analysis, we only observe and estimate the sensitivities based on the numbers of molecules of these three species. We proceed by first generating sigma points for the unscented transform methodology. To cap-

ture the input parameter distribution effectively, we assume that it is lognormal, and the sigma points then come form the associated Gaussian distribution. This ensures that we can achieve the best possible accuracy of our results with 31 sigma points for the 15 dimensional input distribution. We then feed these sigma points for Gillespie simulations. The stochastic simulations yield time series for A, R and C corresponding to each of the sigma points. For every sigma point, we calculate time periods of the cycles, and using the means and covariances of these periods as the output variables of the unscented transform, we obtain first two moments of output distributions. Assuming the responses are lognormally distributed, we sample trials from the output distribution, cross-correlated with the input distribution. We then sample further from the output distributions so as to obtain 50 stochastic realisations per input sample. The sensitivity analysis setup is now ready for entropy and mutual information estimation using bias correction.

The first question with the sensitivity analysis of this model that arises is: how many input samples should be enough to capture the first order sensitivities well? how many for the second order sensitivities? Until now, we have only looked into the 4 and 6 parameter models. But for a 15 parameter model, we can safely assume that the number of samples needed for a decent estimate will be more than that for the earlier models. To answer this question, we systematically look into the estimates of first order mutual informations between parameters and outputs, and observe the point where they begin to asymptote along the actual value. Figure 4.24 demonstrates one such exercise. We look at three different input-output combinations and then observe the behaviour of mutual information as the number of samples for the estimates is increased. For this figure, we consider the combinations of parameter γ_A with period of A, β_A with period of A, and γ_C with period of C. We also set our test estimates with 2000, 8000, 10000, 25000, 75000 and 100000 samples and observe that for 25000 samples we start obtaining reliable estimates for mutual information. We also repeat each of the estimates 10 times for the 10000 sample case, and observe that the estimates have reasonably low errorbars. However, for accuracy up to second order, we shall use 75000 or 100000 samples for our estimates. In what follows next, we shall be denoting the input parameters as X_i and output parameters as Y_j, with the indices $i = 1, \ldots, 15$ respectively representing the parameters as in the caption of figure 4.23 and output parameters Y_1, Y_2, and Y_3 respectively representing the circadian cycle

lengths of A, R, and C.

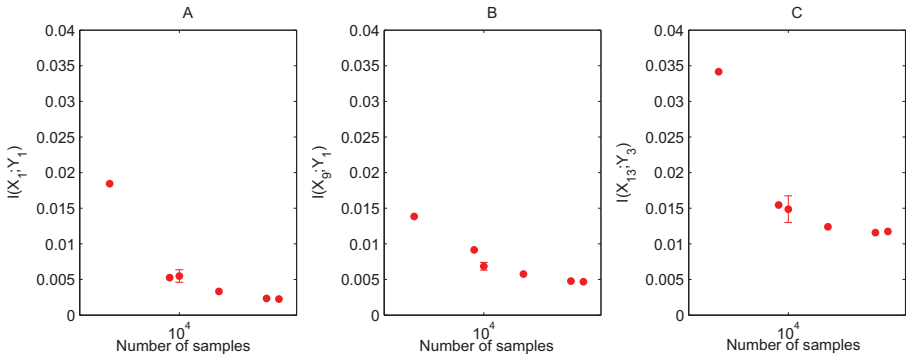

Figure 4.24: Estimates for mutual informations between input and outputs as the number of samples increase. The combinations studied are as on the y-axes. The estimates in all cases start to asymptote towards the true value for a reasonably large number of samples. Errorbars represent means and standard deviations for 10000 samples in each case. System volume used is $V = 10^{-17}$L. Unless otherwise stated, all the unscented transform analysis uses the parameter values of $\alpha = 0.9$, $\beta = 2$, and $\kappa = 0$. The sigma points are generated with variances of 10 percent of the nominal parameter values

Figure 4.25 shows the different sensitivities of circadian rhythms with respect to the parameters of the model. The sensitivities estimated are normalised first order sensitivities, where all the response entropies are 3.37 bits. From panel A, we can observe that there is a range of sensitivities exhibited for the model parameters. The orders of these normalised sensitivity estimates range from 10^{-4} to 10^{-2}. Another observation is that the sensitivity estimates for all the response variables are the same. This is expected, because all the response variables follow the same cycle. The outputs are the most sensitive to changes in parameter X_6 i.e., δ_{M_R}, the degradation rate of mRNA of the repressor protein. The least sensitivity is of the variable $\alpha_{R'}$.

Panels B and C provide a comparison of estimates of the same quantities for different levels of stochasticity. These levels are varied with the help of system volume, and as the system volume decreases, we get more and more stochasticity as we move from panels A to C. As observed in the results for gene expression model, we observe that stochasticity has the ability to change the sensitivity of the system, both in terms of sensitivity orderings and total level of sensitivity. This can also be observed in figure 4.26. As stochasticity levels change, the sensitivity orderings and distributions among various parameters change.

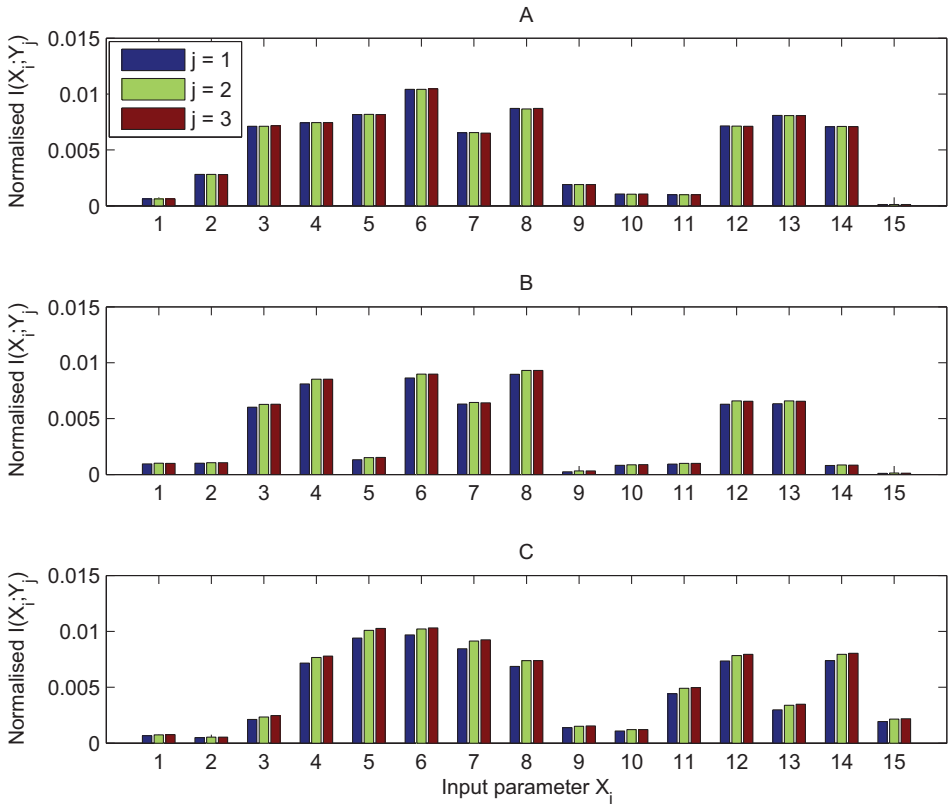

Figure 4.25: Normalised first order sensitivities of the circadian cycles of A, R, and C for different levels of stochasticity. It is observed that stochasticity can change the level of sensitivity as well as sensitivity orderings among parameters. Number of input samples used to generate these results are 10^5. The volumes used to set the different levels of stochasticity are: A) $V = 10^{-15}$L, B) $V = 10^{-16}$L, and C) $V = 10^{-17}$L

Thus, we conclude that not only can noise suppress the sensitivity of a system, at a certain level, the sensitivity can also be enhanced. Another interesting point to note is that the sensitivities for all the responses is same when the system is less noisy. But this observation changes as the system becomes more and more stochastic. This is understandable because the peaks in the circadian cycles are more noisy and the time series of one variable can potentially be more sensitive than the other as the system departs from deterministic behaviour.

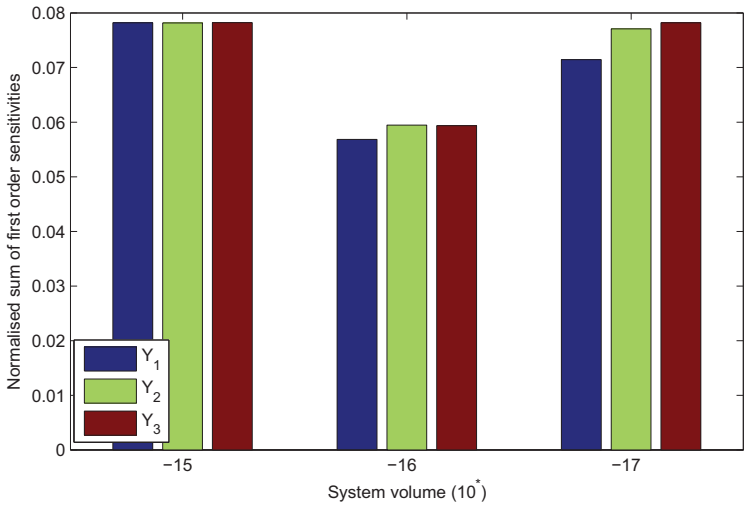

Figure 4.26: Normalised sum of first order sensitivities of the circadian cycles of A, R, and C for different levels of stochasticity. The sensitivities of all the three output variables are same for a particular level of noise, however these sensitivities change a lot with variation in noise levels. Number of input samples used are 10^5

Figure 4.27 shows the normalised second order sensitivities of the period of A with respect to the all possible two parameter combinations. Since all the response variables follow the same trend for sensitivity orderings, we only analyse one output variable in this case. All the 105 parameter combinations are shown in the figure. The most important and least important combinations are mentioned in the caption of the figure.

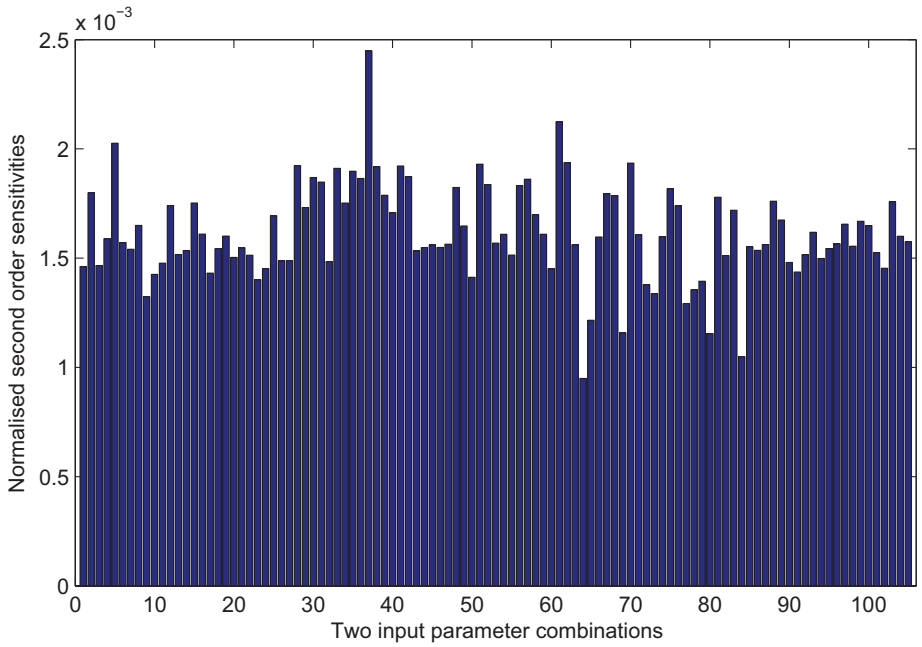

Figure 4.27: Normalised second order sensitivities of the circadian cycle of A. Number of input samples used are 75000, $V = 10^{-17}$L. The most sensitive parameter combinations are (X_3, X_{13}), (X_6, X_7), and (X_1, X_6) while the parameter combinations with least sensitivities are (X_6, X_{10}), (X_8, X_{15}), and (X_8, X_{11})

We note that X_6 exerts its influence at both orders of sensitivity, while at second order parameter pairs combined with X_{10}, X_{11}, and X_{15} are the least influential. We also note that these were the least sensitive parameters at the first order. However, we do not notice a single parameter pair that is noticeably more significant than others. This could potentially mean that no two parameter combinations are causing significantly more sensitivity than others. But at higher orders, there may be more significant parameter combinations. Accurate higher order sensitivity analysis will, however, require significantly more samples for information estimates. The summation law can also be illustrated with an estimate of noise entropy, but that analysis will require Monte Carlo parameter sampling as unscented transform methodology, as discussed before, may not yield accurate results in that case.

4.8 Concluding Remarks

In this chapter, we have developed a global stochastic sensitivity analysis methodology with the help of concepts from information theory similar to that proposed in [79] for deterministic systems. We have also discussed the different orders of parametric sensitivity in detail. We have discussed how the noise entropy evolves from deterministic discretisation entropy for stochastic systems in order to accommodate the effects of intrinsic noise. We also find that noise entropy for stochastic systems is not the sum of discretisation entropy and intrinsic noise entropy, and may even be less than the discretisation entropy for a deterministic system. We presented the results for sensitivity analysis of gene expression model with and without negative feedback for various noise levels and compared them to the corresponding deterministic results. The most interesting phenomena that we observed was that noise levels have the ability to alter the parametric sensitivity of the system, and in the cases studied for gene expression models, high noise levels implied that the protein distributions were less sensitive to parameter perturbations.

We pointed out that the methodology developed can be computationally very expensive for complex models with a high number of parameters. This means that for accurate estimation of entropies and mutual information, a lot of samples for input parameters as well as corresponding stochastic outputs are required. We proposed that this difficulty can be overcome by the use of unscented transform, wherein we use only a few number of sigma points and their corresponding outputs to capture the input and output distributions effectively. We compared the results of analysis with Monte Carlo type sampling with those from modification with unscented transform and found that they agreed well with each other.

The important advantage of using unscented transform with Gillespie for our analysis over linear noise approximation [120] is that the latter cannot deal with models with oscillations while our results are valid for systems with nonlinear effects such as systems with oscillatory dynamics. We illustrated this advantage of unscented transform by applying the sensitivity analysis methodology to a model of circadian rhythms for various noise levels. In that case, we also found that stochasticity may affect the sensitivity orderings and sensitivity levels for the parameters of the system.

The sensitivity analysis methodology proposed in this chapter can be improved further for efficiency and accuracy. In chapter 6, we discuss possible modifications and future work related to the analysis presented in this chapter.

Chapter 5

Information Theoretic Method for Specificity Quantification in a Receptor Ligand Binding System

In a receptor ligand binding system, specificity is referred to as the degree of favourability of a particular ligand to bind to a specific receptor in the presence of other available receptors and vice versa. In biological systems, a certain degree of specificity is necessary for their proper functioning [70, 123]. Real systems do not exhibit 100 percent specificity and there is always some level of cross-talk between various signaling pathways. Several studies have been done to look at specificity exhibited by certain models [45, 67, 117].

While specificity, or conversely cross-talk, can both be functionally advantageous, it is often desirable to know how specific a system is. The idea of specificity has been discussed well in literature, but there is no general way of quantifying specificity in a mathematical way. In literature, there exist some statistical mechanics approaches [55, 128, 134] as well as methodologies based on information transmission across different signaling pathways [2, 68]. In the following study, our goal is to understand and be able to quantify specificity in biological systems. In this chapter, using the ideas from information theory, we propose that specificity can be defined in terms of mutual information between a particular stimulus and output activity. We develop a setup according to which specificity is quantified, and then we discuss some of the methods for estimating specificity. We provide a general way

of estimating maximum possible specificity that can be exhibited by a system using our proposed setup. We conclude with some results for T-cell signalling experimental data from our collaborators. Also, we propose how the setup can be extended for stochastic systems.

5.1 The Receptor Ligand Binding Model

We want to determine how specific a receptor ligand system is by quantifying the mutual information in the system. The mathematical model we consider contains N receptors (A_i, where $i = 1, 2, \ldots, N$) and M ligands (B_j, where $j = 1, 2, \ldots, M$) that react according to the following scheme,

$$A_i + B_j \overset{K_D^{i,j}}{\rightleftharpoons} C_{i,j}, \tag{5.1}$$

where $C_{i,j}$ is the complex formed by the ith receptor and the jth ligand. The interaction is governed by the experimentally measured dissociation constant $(K_D^{i,j})$. At equilibrium, we have

$$A_i B_j = K_D^{i,j} C_{i,j}, \tag{5.2}$$

where A_i and B_j are concentrations of the free receptor and ligand, respectively. In addition, we assume that the total concentration of both the receptors and ligands are conserved, i.e.,

$$A_i^T = A_i + \sum_{j=1}^{M} C_{i,j}, \tag{5.3}$$

$$B_j^T = B_j + \sum_{i=1}^{N} C_{i,j}, \tag{5.4}$$

where A_i^T is the total concentration of receptor i and B_j^T is the total concentration of ligand j. Collectively, we have a set of $N \times M$ equations and the $N \times M$ unknowns (the unknowns are all the $C_{i,j}$). We assume that the system input is a stimulus vector $\mathbf{S}$, whose components

are the M ligand concentrations,

$$\mathbf{S} = \left\{B_1^T, B_2^T, B_3^T, \ldots, B_M^T\right\} \tag{5.5}$$

and the system output is a response vector, $\mathbf{R}$, whose components are the total concentration of bound receptor of each type,

$$\mathbf{R} = \left\{\sum_{j=1}^{M} C_{1,j}, \sum_{j=1}^{M} C_{2,j}, \sum_{j=1}^{M} C_{3,j}, \ldots, \sum_{j=1}^{M} C_{N,j}\right\}. \tag{5.6}$$

In what follows from here, we shall denote $\sum_{j=1}^{M} C_{i,j}$ as C_i, and use the term affinity matrix for matrix of dissociation constants, thereby implying high affinity corresponding to low numbers in the matrix. In summary, given the receptor concentrations (A_i^T) and the dissociation constant matrix $(K_D^{i,j})$, the system input is the stimulus vector $(\mathbf{S})$ which determines the system output, a response vector $(\mathbf{R})$,

$$\mathbf{R} = f(\mathbf{S}) \tag{5.7}$$

where f is a nonlinear function determined by equations (5.1) - (5.3) and equation (5.6). We mostly focus on systems with two ligands and two receptors, but the formalism presented is quite general.

5.2 Quantification of Specificity

It can sometimes be easy to qualitatively see how specific a system is by looking at the affinity matrix. For example in a 2 receptors 2 ligands model, let us consider the matrices:

$$A = \begin{bmatrix} 1 & 10 \\ 10 & 1 \end{bmatrix}, \text{ and } B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

It is easy to see that A is more specific than B. This is because for the matrix A, receptor 1 has more affinity for ligand 1 and receptor 2 has more affinity for ligand 2. So ligands 1 and 2 will *specifically* bind to receptors 1 and 2 respectively. In case of the second matrix B with all equal dissociation constants, the ligands will have equal affinity for both receptors

and their binding will not be specific in this case.

This analysis is somewhat simple in case of the matrices A and B, but in more complicated systems, one cannot determine the specificity just by looking at the affinity matrix. We need a quantification method in order to quantify specificity of a system, and be able to compare the specificity of systems relative to each other.

We propose to use concepts from information theory and use mutual information as a measure of specificity.

5.3 Mutual Information

We wish to determine how much information about $\mathbf{S}$ is retained in $\mathbf{R}$. This can be done by looking at the level of certainty that a particular stimulus $\mathbf{S}$ brings to the response $\mathbf{R}$ that is observed. In other words, the decrease in the uncertainty in $\mathbf{R}$ with the knowledge of $\mathbf{S}$can be used to determine how specific the receptor ligand binding is. Therefore, we propose that specificity of the system under consideration can be defined as the mutual information between stimulus $\mathbf{S}$ and response $\mathbf{R}$and is given by

$$I(\mathbf{R}, \mathbf{S}) = H(\mathbf{R}) - H(\mathbf{R} \mid \mathbf{S}), \tag{5.8}$$

where $H(\mathbf{R})$ is the response entropy, and is given by

$$H(\mathbf{R}) = -\sum_{\mathbf{r}} p(\mathbf{r}) \log_2 p(\mathbf{r}), \tag{5.9}$$

where $p(\mathbf{r})$ is the probability of observing a given response $\mathbf{r}$ across many trials of every possible stimulus. $H(\mathbf{R} \mid \mathbf{S})$ is the conditional entropy of $\mathbf{R}$ when $\mathbf{S}$ is fixed and is related to the noise entropy discussed in the previous chapter 4 for stochastic systems.

$$H(\mathbf{R} \mid \mathbf{S}) = \sum_{\mathbf{s}} p(\mathbf{s}) H(\mathbf{R} \mid \mathbf{S} = \mathbf{s}). \tag{5.10}$$

Given that the mathematical model is deterministic, knowing the stimulus vector $\mathbf{S}$ completely determines the response vector $\mathbf{R}$ and therefore $H(\mathbf{R} \mid \mathbf{S}) = 0$. Therefore in a

deterministic setting, evaluation of $I(\mathbf{R}, \mathbf{S})$ requires only the evaluation of the response entropy $H(\mathbf{R})$. Figure 5.1 provides a visualisation of response entropies with the same stimulus for different affinity matrices.

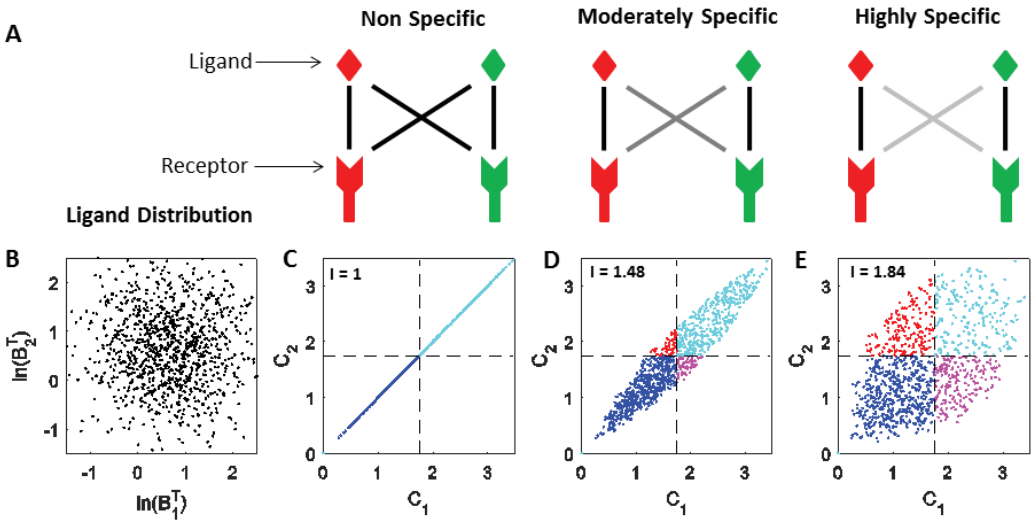

Figure 5.1: Effect of specificity on response in a two ligand two receptor system with $K_D = [1, X; X, 1]$, when larger values of X represent greater specificities. A particular stimulus is given to three different systems with different levels of specificity. The binding responses for the systems are shown to be different for different specificity levels. The response becomes more and more scattered when the specificity increases, thus motivating us to measure specificity of a system by mutual information. A) Affinity between receptors and ligands is shown on grey scale. B) Lognormal ligand distribution with $\mu_1 = \mu_2 = 0.65$ and $\sigma_1^2 = \sigma_2^2 = 0.68$. C) $X = 1$. D) $X = 2$. E) $X = 10$.

5.4 Analysis of the Model and Estimation of Mutual Information

Given an affinity matrix, we want to quantify how much maximum information can be recorded between the response $\mathbf{R}$ and stimulus $\mathbf{S}$. There is a range of questions that need to be addressed here. Some of these include, how do we generate stimulus samples to estimate mutual information, how must we bin the response, what should be the parameter(s) we maximise the mutual information against etc. In the following sections, we discuss how these various factors affect and modify our analysis.

To start with, we analyse the variation of mutual information against total concentration

of receptors. We only consider two states of the receptor, on and off. This implies that for information estimation purposes, we only bin our responses into two bins, i.e., we assume that the response is binary. The off states constitute one bin, while the on states go to the second. Off responses are those where the bound receptor concentrations are below a specific threshold. The receptor is in the on state if the bound receptor concentrations make it beyond that threshold concentration. It is interesting to note that in our case of binary response binning, the maximum response entropy that can be exhibited is 2 bits. Referring again to figure 5.1, we note that a non-specific case cannot have a response entropy greater than 1 bit.

We know that as the discretisation for information estimation becomes finer and finer, the value of $I(\mathbf{R}, \mathbf{S})$ converges to that of $H(\mathbf{R})$ because of the noise entropy. We know that in a deterministic setting, the conditional entropy $H(\mathbf{R} \mid \mathbf{S})$, is always zero. But for estimation purposes, we need to discretise the input space into a number of bins. Higher number of bins implies more accuracy. So as we accommodate finer binning of stimulus, the estimate of mutual information becomes more and more accurate. As it is easier to calculate $H(\mathbf{R})$, we shall use this value for $I(\mathbf{R} \mid \mathbf{S})$ in all the deterministic analysis to follow. In the case where the system is not purely deterministic or if the response is not binary, we need to accommodate bias correction for entropy and mutual information estimates.

An important assumption that we make for our analysis is that we are maximising mutual information given that the stimulus follows one particular distribution. In other words, we are fixing a particular type of stimulus distribution and then we want to optimise the parameters of that distribution to obtain maximum mutual information. The problem, in its generality, would be to maximise mutual information over all possible stimulus distributions, i.e., finding the *channel capacity* C of the system which is defined as:

$$C = \sup_{p_{\mathbf{S}}(\mathbf{s})} I(\mathbf{S}, \mathbf{R}).$$

This problem, however, is difficult to tackle and we therefore limit our analysis to one kind of distribution. The shape of the distributions is motivated by experimentally observed distribution of ligands as illustrated in section 5.12. This choice of distribution is studied

in the section that follows.

5.4.1 Variation of Stimulus - Logarithmic vs Lognormal

To estimate mutual information, we need to generate samples for the stimulus. Here we consider the simple 2×2 case of high specificity with matrix $\begin{bmatrix} 1 & 10 \\ 10 & 1 \end{bmatrix}$ as an example. We need to be able to cover a reasonably large range of sample ligand concentrations. We analyse the effects of different modes of stimulus variation for sampling. The obvious choices are to either sample logarithmically scaled stimulus over a range or draw sample stimuli from a lognormal distribution. Figure 5.2 shows the effects of these choices. The one with lognormal stimulus variation seems to be a better choice given that it clearly shows that the mutual concentration peaks for a certain value of A^T, and this peak is what we are interested in. Moreover, the information peak in the logarithmic case is lower than that in the lognormal case, indicating that we are prone to missing out on the maximum possible mutual information in the logarithmic case. It is worth noting that to produce this figure, we set the binary binning threshold to be absolute, at half the total receptor concentration, and we vary both A_1^T and A_2^T simultaneously. Moreover, we generate Latin hypercube samples for B_1^T and B_2^T separately but from the same values of μ and σ, μ and σ being the mean and standard deviation of the underlying normal distribution.

5.4.2 Binary Binning - Absolute vs Constant Threshold

We want to analyse how mutual information varies with the total receptor concentration. This is analysed in two different scenarios. Absolute response binning, and binning with a constant threshold. In the absolute binning case, the receptor is marked as on when the concentration of bound receptor exceeds half the total concentration. On the other hand, for the constant threshold case, a certain constant value for receptor concentration is fixed beyond which the receptor is considered on.

Figure 5.3 shows that the information gained in constant threshold case is always zero until the total receptor concentration reaches this constant threshold as expected, but remains considerably high beyond the threshold. In case of $C_T = 10$, the the maximum

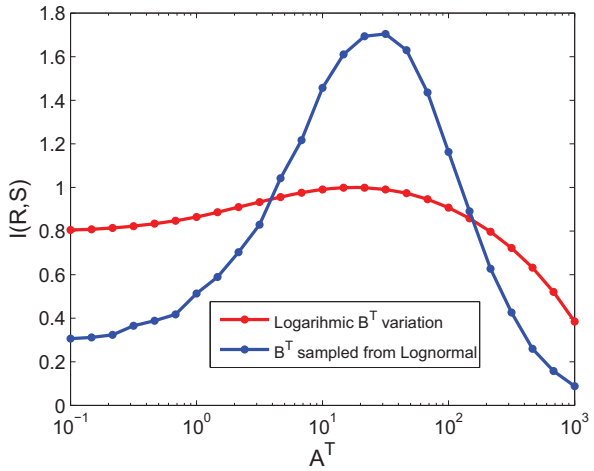

Figure 5.2: Logarithmic vs lognormal B^T variation. Number of samples = 1000 Logarithmic curve: $B^T = [10^{-1}, ..., 10^3]$, Lognormal curve: $\mu = 2.25$, $\sigma^2 = 2.5$, $K_D = [1, 10; 10, 1]$

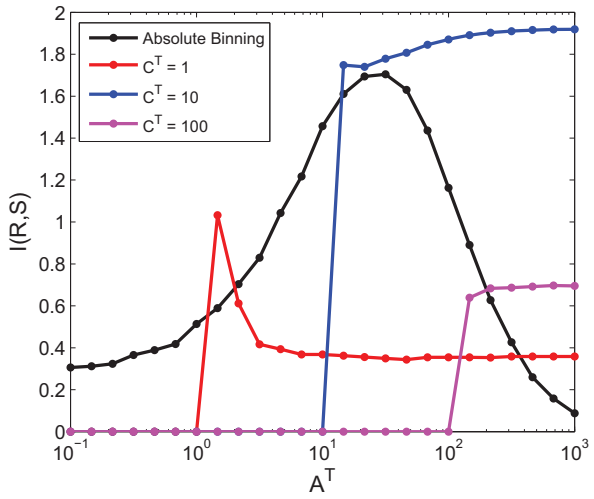

Figure 5.3: Absolute $(A^T/2)$ vs constant threshold. Number of samples = 1000, $\mu = 2.25$, $\sigma^2 = 2.5$, $K_D = [1, 10; 10, 1]$

information is even greater than the maximum information in absolute binning case. Although the constant threshold case may be biologically more relevant, the downside is that we do not know how to fix it for optimal results. The only idea we can obtain about fixing a reasonable by looking at the behaviour of $I(\mathbf{R}, \mathbf{S})$ in the absolute binning case. It is evident from the cases $C_T = 1$ and $C_T = 100$ that maximum information is significantly lower than what we obtain with absolute binning. Therefore, in our analysis that follows, we will only apply absolute response binning.

5.4.3 Changes in Stimulus Parameters

While performing analysis with lognormal B^T variation, there are two parameters involved: μ and σ, the mean and standard deviation of the underlying normal distribution. The figures 5.4 and 5.5 represent how changes in these parameters affect the value of maximum information as well as the values of A^T corresponding to which the maximum is observed. The effect of changing μ is more pronounced as shown in the figure. Both the maximum information and the corresponding A^T value change significantly as μ changes. The change in σ does not have a substantial effect on the position of optimal A^T. This means that searching for maximum information corresponding to different values of A^T does not suffice. To deal with this problem, we detail our methodology for finding the maximum information in the next section.

5.5 Optimising Mutual Information over Parameter Space

In the previous section, we observed that there are several factors involved in determining the mutual information between the stimulus and response of the receptor-ligand binding system. These parameters include the distributions of the ligand concentrations, the total concentration of receptors and also the thresholds down-stream of receptors that use the information. This can be challenging as it is not straightforward to guess or experimentally measure all of these parameters. Moreover, the effect of changing these parameters on the information is not monotone. Therefore, we use grid search for this optimal set of parameters that maximises the mutual information. This involves assigning a reasonable

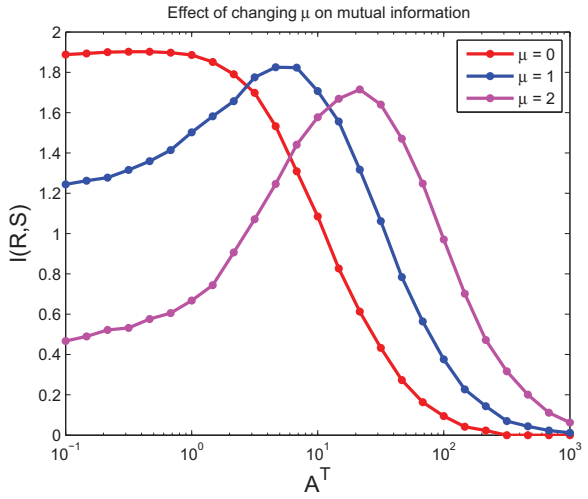

Figure 5.4: Effect of μ on maximum information. Number of samples = 1000, $\sigma^2 = 2.5$, $K_D = [1, 10; 10, 1]$

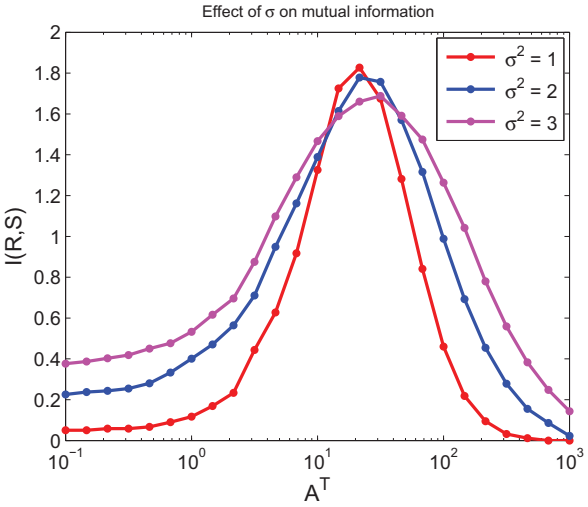

Figure 5.5: Effect of σ^2 on maximum information. Number of samples = 1000, $\mu = 2.25$, $K_D = [1, 10; 10, 1]$

search space for each parameter involved, and then evaluating the maximum entropy for each set in the parameter space.

To illustrate the grid search method, we start with bisymmetric affinity matrix for its ease in understanding. Since the matrix is bisymmetric, it is safe to assume that the optimal μ's and σ's for both ligands are the same. We also assume that the optimal total receptor concentration values for both the receptors are also equal. These assumptions conveniently reduce our parameter space from six to three dimensions. We vary each of these now three parameters logarithmically to be able to incorporate larger orders of magnitudes into the grid search.

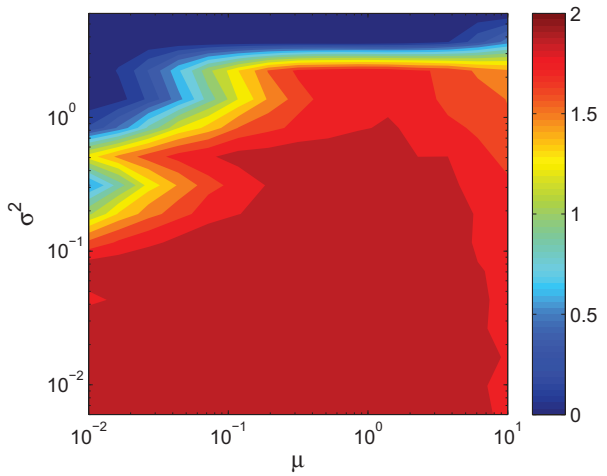

Figure 5.6: Heatmap showing maximum information obtained from each μ and σ. $K_D = [1, 10; 10, 1]$. The maxima over A^T are reported. $\mu = [10^{-2}, \ldots, 10^{1}]$, $\sigma^2 = 0.6 * [10^{-2}, \ldots, 10^{1}]$, $A^T = [10^{-1}, \ldots, 10^{2}]$. Number of elements in each parameter direction = 15. Maximum information obtained is 1.9914 bits corresponding to $\mu = 0.044$, $\sigma^2 = 0.026$ when $A^T = 0.268$

Figure 5.6 shows the results of our grid search for the very specific case. In this figure, the maximum information cannot be seen against A^T values. Only the maximum value of information among all A^T values is shown in the figure. The maximum information obtained with this affinity matrix is 1.99 bits, which is very high, as the maximum information obtainable in a 2 ligands and 2 receptors case is 2 bits. This shows that the matrix $\begin{bmatrix} 1 & 10 \\ 10 & 1 \end{bmatrix}$ is very specific.

Another factor that we are interested in is the optimality of parameter sets. Changes

in the total receptor as well as ligand concentrations can alter the information obtained. But from figure 5.7, we observe that there is no optimal set of parameter values that maximises the mutual information. There is rather a region, shown in red in the heatmaps, that corresponds to maximum mutual information.

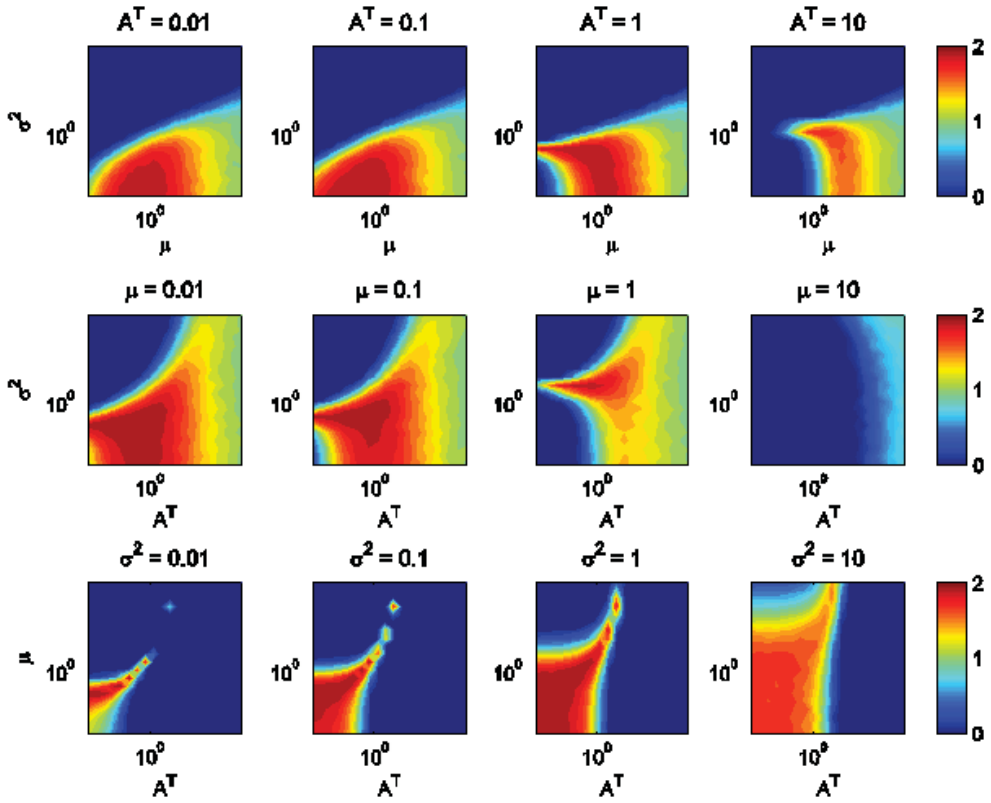

Figure 5.7: Effect of different parameters on mutual information. $K_D = [1, 10; 10, 1]$. All parameters on the axes range from 10^{-2} to 10^3. All the results reported are from grid search with Monte Carlo stimulus sampling

Figure 5.8 illustrates how the maximum mutual information obtained for a matrix $K_D = [1, X; X, 1]$ changes as X takes on different values. As expected, the mutual information increases as the X values become increasingly dominant. All the results reported in this figure are obtained from grid search methodology. Interestingly, we find that the specificity rises very quickly with X; only a factor of 2 difference in the affinities produces

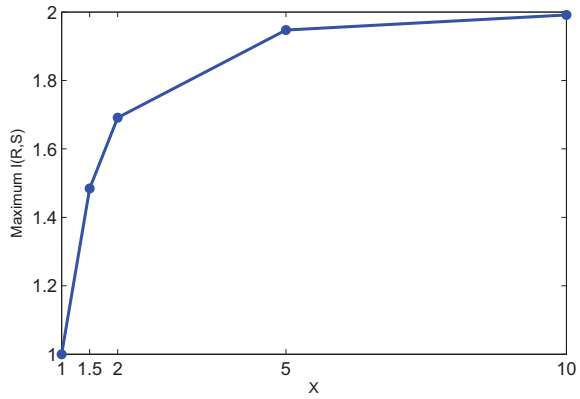

Figure 5.8: X vs maximum mutual information obtained from grid search, $K_D = [1, X; X, 1]$

a significant specificity.

The optimisation of information allows us to allocate a unique specificity measure to a K_D matrices. As we have discussed before, we already know that the matrices

$$\begin{bmatrix} 1 & 10 \\ 10 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 5 \\ 5 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

are in decreasing order of their specificities. In such cases, the relative specificities are easy to judge. However, the absolute level of specificity is not clear. Also, in case of affinity matrices where all the elements are different or where the affinity matrices are high-dimensional, it is not easy to discriminate the specificities of matrices relative to each other without specificity estimation, and the above described method can be computationally very expensive. For example in the case of figure 5.6, even with symmetry and the simplifications following it, the parameter space is 15^3 dimensional, and for each of the values in the parameter space, 1000 samples from B^T distribution are obtained and a deterministic system has to be solved for each of these samples. In summary, the heatmap in figure 5.6 required 3375000 solutions of the receptor ligand binding system. This obviously complicates further if the K_D matrix is not symmetric.

5.6 Use of Unscented Transform for Speed

As mentioned in the last section, the algorithm used for grid search can be very slow. For a larger system with multiple receptors and ligands, the performance of the algorithm can be poor. In order to overcome this problem, we use unscented transform to significantly reduce the number of calls to the function that solves the deterministic system. Since solving the biochemical system is the most time consuming step of the algorithm, the improvement due to Unscented transform can be expected to be considerable.

In our methodology, we are drawing thousands of lognormally distributed Monte Carlo samples of ligand concentrations, which are then used to determine the corresponding steady state outputs of the receptor ligand system, or the response. So in turn, we obtain thousands of response samples that enable us to estimate the discrete entropy of the response, which for a deterministic system is the mutual information. The idea of using unscented transform stems from the fact that we already have lognormally distributed stimulus. With the response also following lognormal distribution for each element in the parameter space, we only need $2M + 1$ solutions of the deterministic system. This has a huge impact on the computational time of the algorithm, as thousands of system evaluations have now been reduced to only 5 for a two receptor and two ligand system.

5.6.1 Comparison between Monte Carlo and Unscented Transform Results

Figure 5.9 shows that the Unscented transform performs quite well for both specific and non-specific cases. Both the maximal mutual information value and location obtained from the unscented transform match with the results from Monte Carlo solutions. The performance of the Unscented Transform in terms of speed is significantly better.

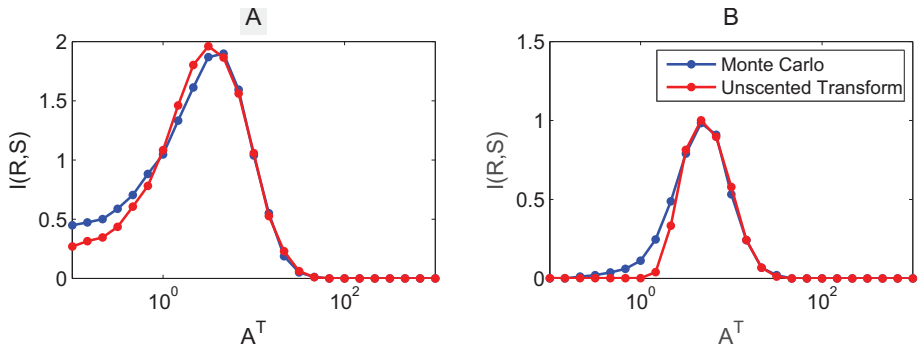

Figure 5.9: Comparison between results from Monte Carlo sampling and Unscented Transform. $\mu = 1, \sigma^2 = 0.5$. A) High specificity: $K_D = [1, 10; 10, 1]$, B) No specificity: $K_D = [1, 1; 1, 1]$

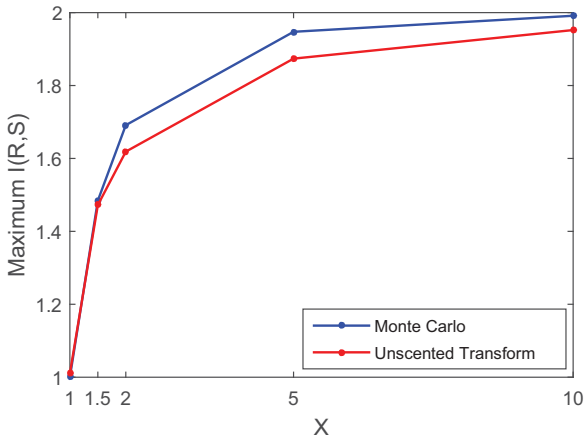

Figure 5.10: X versus maximum mutual information when $K_D = [1, X; X, 1]$. All results reported are from grid search

Figure 5.10, is an illustration of how both the methods compare in terms of the maximum mutual information values. The results from both the methods compare well against each other with the Unscented Transform almost always slightly underestimating the mutual information value.

5.6.2 Unscented Transform for Non-Bisymmetric Matrices

As discussed earlier, the grid search approach with Monte Carlo sampling is computationally very expensive even for a 2-by-2 matrix that is non bisymmetric. The bisymmetry

of K_D matrices allows us to simplify the computational efforts by assuming that the parameters for both the receptors and ligand distributions behave similarly, thus reducing the dimension of the problem by half. In case of non-bisymmetric matrices, on the other hand, we lose all these advantages and every parameter needs to be working independently of the other.

The Unscented Transform is a natural choice accompanied by grid search for non-bisymmetric matrices. But when compared for several sets of parameters, the Unscented Transform results differ significantly from those from a Monte Carlo approach. Although the Unscented Transform worked very well for matrices of type $[1, X; X, 1]$, it seems to yield incorrect mutual information values. The parameter region of high mutual information values also differed greatly from that obtained from the Monte Carlo approach.

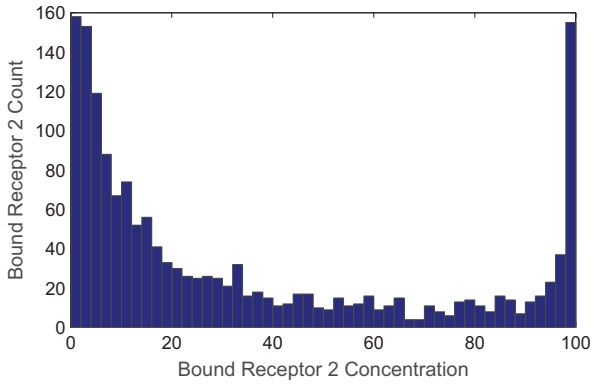

Figure 5.11: Distribution for C_2. $\mu_1 = 1$, $\mu_2 = 2$, $\sigma_1^2 = 5$, $\sigma_2^2 = 5$, $A_1^T = 10$, $A_2^T = 100$

The working of Unscented Transform ensures that we obtain the first two moments of response distribution. In our case, we generate lognormal response samples as this is more realistic. The setup of the problem stipulates that the stimulus is lognormal, but the assumption that the response is always lognormal is incorrect. Figure 5.11 shows the response distribution for second bound receptor, C_2, when $K_D = [100, 10; 10, 1]$. The distribution is bimodal which clearly shows that our assumption of lognormal response is flawed, hence unsatisfactory results.

5.7 Simulated Annealing

Studying non-bisymmetric K_D matrices with grid search approach has proved to be unsuccessful. The Unscented Transform approximation can be helpful in computationally expensive cases, but that too fails for non-bisymmetric matrices because of the response being bimodal. This problem leads us to thinking of an optimisation approach that could help in obtaining maximum mutual information.

We use simulated annealing to obtain optimal solutions for mutual information that various K_D matrices can exhibit. Table 5.1 shows simulated annealing results for some of the K_D matrices that are generally considered to be non-specific.

K_D Matrix	μ_1	μ_2	σ_1^2	σ_2^2	A_1^T	A_2^T	Maximum $I(R,S)$
$\begin{bmatrix} 10 & 10 \\ 1 & 1 \end{bmatrix}$	0.06	0.90	38.36	1.64	31.93	0.01	1.58
$\begin{bmatrix} 100 & 10 \\ 10 & 1 \end{bmatrix}$	0.56	2.48	0.06	5.52	43.37	0.19	1.58
$\begin{bmatrix} 10 & 1 \\ 10 & 1 \end{bmatrix}$	0.30	0.02	0.03	0.25	0.06	0.30	1.00

Table 5.1: Optimal parameter and specificity values for some non-bisymmetric matrices using simulated annealing

The simulated annealing methodology yields similar results as the above table when used with different initial guesses. This methodology is faster than grid search. We note that according to these results, some matrices that are generally thought to be non-specific are exhibiting a considerably high level of mutual information. For a non-specific matrix, we expect that the maximal mutual information, will not exceed the value 1 bit. To understand why this is so, we need to analyse what parameter regions, rather than parameter values, will yield high levels of mutual information. We therefore, need a faster approach to finding the maximal information as well as finding the regions in parameter space that correspond to that level of mutual information.

5.8 Markov Chain Monte Carlo Methods

Until now, we have tried different methodologies for finding the maximum mutual information between stimulus and response of our receptor ligand binding system. However, we still need to work on improving computational speed while trying to understand the relationship between optimal parameter ranges.

5.8.1 MCMC Results for Bisymmetric Matrices

Markov Chain Monte Carlo approach is applied to the problem at hand, and the computational performance is better than grid search and simulated annealing. Figure 5.12 shows the results for bisymmetric cases from MCMC approach. The results shown in the figures are based on the values of parameters corresponding to maximum 5 percent of the information values that the algorithm yields. And that also after the burning off period. The panel A in figure 5.12 confirms our earlier results from gid search. From the next panels, what we understand is that as X increases, the value of μ (note that in bisymmetric case, we take $\mu_1 = \mu_2$) somewhat decreases, while the optimal values of A^T certainly follow a decreasing trend. The role of σ^2 is even more difficult to comprehend. Figures 5.13, 5.14, and 5.15 illustrate the relationships between different parameters in the regions of high mutual information.

5.8.2 MCMC Results for Non-Bisymmetric Cases

The MCMC results for the matrices in table 5.1 referring to simulated annealing results, match exactly with those of simulated annealing. Both the simulated annealing and MCMC approaches show that the K_D matrices $\begin{bmatrix} 10 & 10 \\ 1 & 1 \end{bmatrix}$ and $\begin{bmatrix} 100 & 10 \\ 10 & 1 \end{bmatrix}$ are specific, contrary to the popular understanding. However, the matrix $\begin{bmatrix} 10 & 1 \\ 10 & 1 \end{bmatrix}$ does not exhibit any specificity. In case of $\begin{bmatrix} 10 & 1 \\ 10 & 1 \end{bmatrix}$, we note that both receptors have 10 times more affinity for ligand 2 as compared to ligand 1, but both the receptors have exactly the same characteristics.

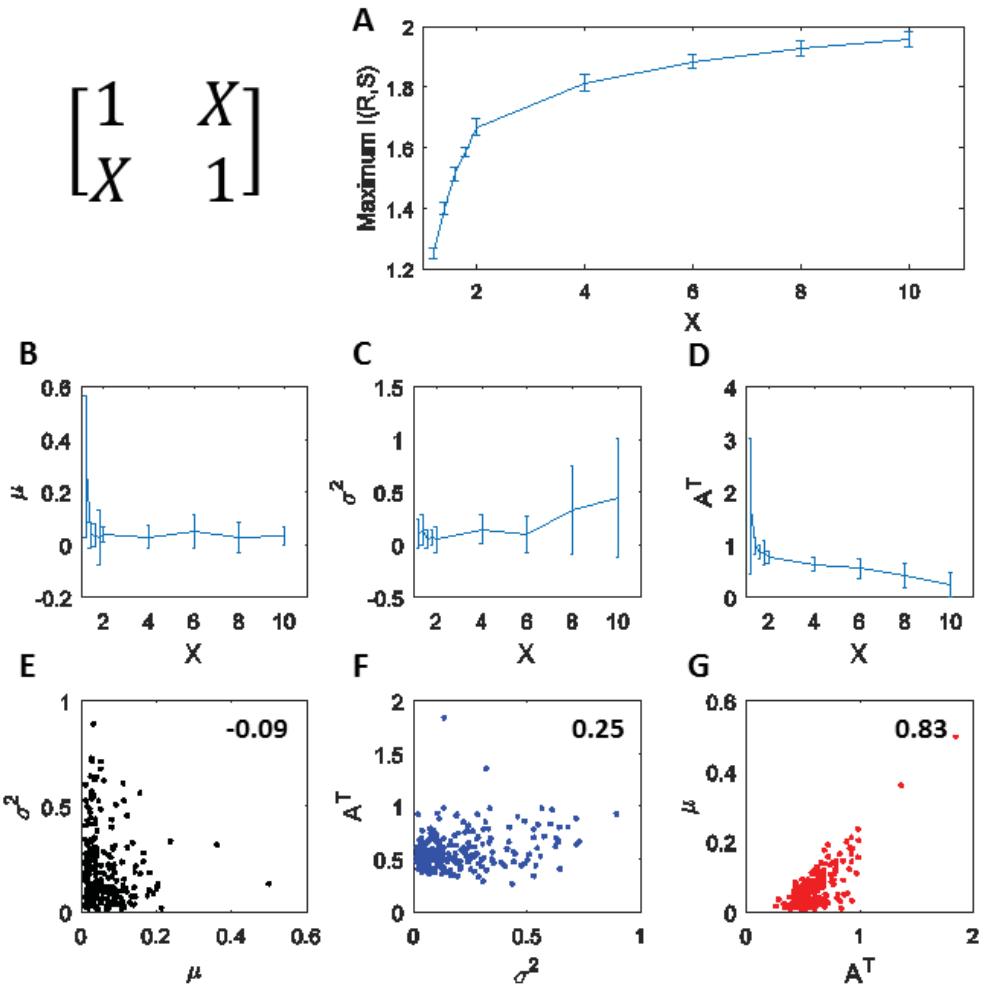

Figure 5.12: MCMC results for MI estimation for $K_D = [1, X : X, 1]$. Number of MCMC runs $= 1200$. The medians and standard deviations of those results are reported that are within the top 5 percent of the maximum values obtained from MCMC simulations. Bisymmetric K_D allows $\mu_1 = \mu_2$, $\sigma_1^2 = sigma_2^2$, and $A_1^T = A_2^T$. A) X versus the maximum MI obtained. B) X versus μ. C) X versus σ^2. D) X versus A^T. E) Relationship between μ and σ^2 when $X = 6$. Pearson correlation coefficient is also reported. F) Relationship between σ^2 and A^T when $X = 6$. G) Relationship between A^T and μ when $X = 6$.

On the other hand, for $\begin{bmatrix} 10 & 10 \\ 1 & 1 \end{bmatrix}$, the two receptors behave differently; receptor 2 likes

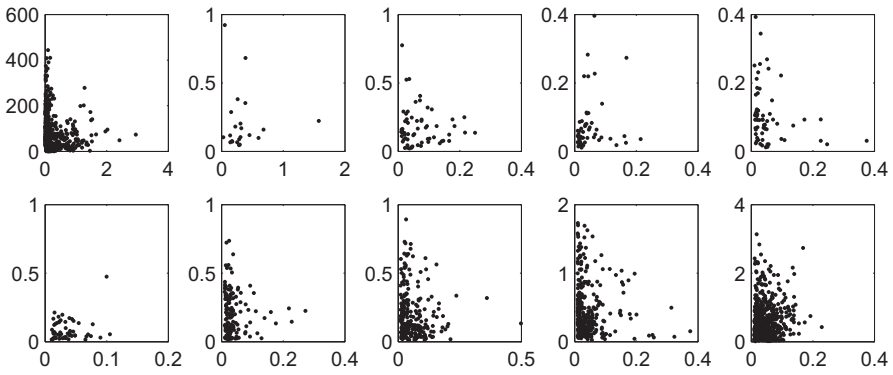

Figure 5.13: Scatterplots of μ (on x-axis) against σ^2 (on y-axis) for $X = [1, 1.2, 1.4, 1.6, 1.8, 2, 3, 4, 6, 10]$ in regions of high mutual information. Results reported are from MCMC approach

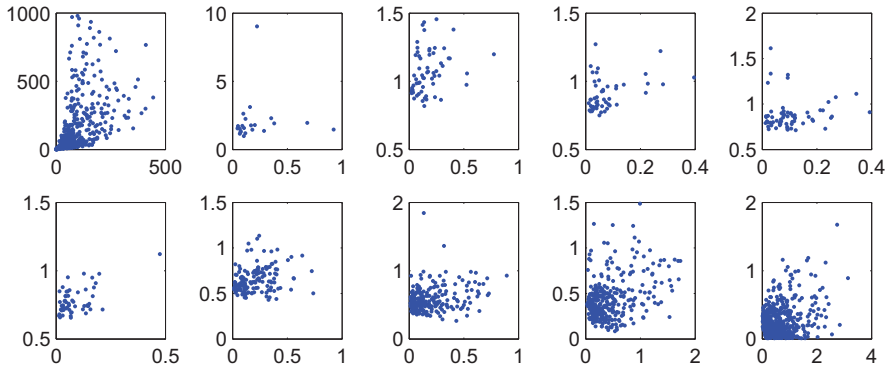

Figure 5.14: Scatterplots of σ^2 (on x-axis) against A^T (on y-axis) for $X = [1, 1.2, 1.4, 1.6, 1.8, 2, 3, 4, 6, 10]$ in regions of high mutual information. Results reported are from MCMC approach

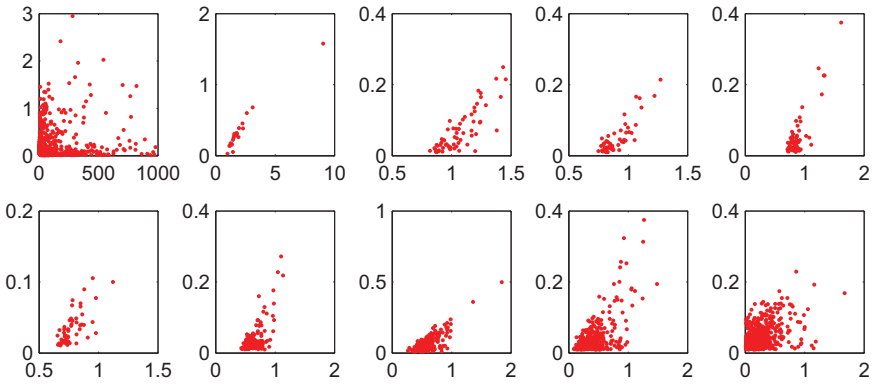

Figure 5.15: Scatterplots of A^T (on x-axis) against μ (on y-axis) for $X = [1, 1.2, 1.4, 1.6, 1.8, 2, 3, 4, 6, 10]$ in regions of high mutual information. Results reported are from MCMC approach

both the ligands 10 times more than receptor 1 does. We also note that the maximum mutual information for both the specific K_D matrices is the same, 1.58. This is somewhat unexpected as the first two matrices are seemingly very different. Furthermore, the matrices of the form $\begin{bmatrix} X & X \\ 1 & 1 \end{bmatrix}$ also tend to exhibit specificity between 1.5 and 1.6, as shown in the MCMC results in figure 5.16. Also when X is close to 1, the specificity remains close to to 1.5. Figure 5.16 suggests that a matrix of the form $\begin{bmatrix} X & X \\ 1 & 1 \end{bmatrix}$ exhibits maximum specificity of 1.58 bits, independent of the value of X. In order to gain insight into what phenomena are responsible for such estimates, we observe what is happening at the binding level. We look at questions like: Is there a specific binding pattern in the systems with these K_D matrices? How does the ratio of affinities of receptor 1 and receptor 2 affect the system? Just like bisymmetric matrices with a bound on mutual information, is the specificity of non-symmetric matrices limited?

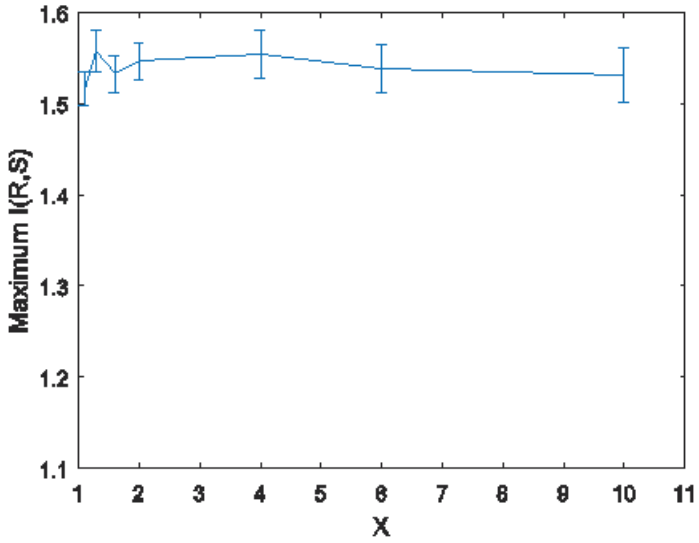

Figure 5.16: MCMC results for X versus the maximum MI obtained for $K_D = [X, X : 1, 1]$. Number of MCMC runs $= 1000$. The medians and standard deviations of those results are reported that are within the top 5 percent of the maximum values obtained from MCMC simulations

5.8.3 Theoretical Limits on Specificity for Non-Bisymmetric Matrices

Figure 5.16 illustrates the stimulus-response relationship for non-bisymmetric affinity matrices. We introduce three different ligand distributions to various non-bisymmetric matrices. We have divided the response area into four regions: I - both C_1 and C_2 are less than $\frac{A_1^T}{2}$ and $\frac{A_2^T}{2}$ respectively, II - C_1 is less than $\frac{A_1^T}{2}$, and C_2 is equal to or more than $\frac{A_2^T}{2}$, III - C_1 is equal to or more than $\frac{A_1^T}{2}$ while C_2 is less than $\frac{A_2^T}{2}$, and IV - both C_1 and C_2 are greater than $\frac{A_1^T}{2}$ and $\frac{A_2^T}{2}$ respectively. The figure shows that unlike the bisymmetric matrices studied before, these matrices have a different course of action for ligands binding of the receptors. If we look at the scatter plots, it becomes evident that rather than having a certain region in the $C_1 - C_2$ space within which binding could occur for bisymmetric matrices, some of the non-bisymmetric ones have a defined path that the binding process follows as the ligands saturate. Moreover, in the case of $\begin{bmatrix} 10 & 1 \\ 10 & 1 \end{bmatrix}$, both the receptors like ligand 2 more than ligand 1, and to the same extent. So really, the two receptors in this case

behave like one and their bound concentrations remain the same as each other as the ligand saturates. In this case, the bound receptor concentrations can lie in only two of the four regions in the $C_1 - C_2$ plane. On the other hand, receptor 2 has significantly more affinity for both ligands when compared to receptor 1 in case of $\begin{bmatrix} 10 & 10 \\ 1 & 1 \end{bmatrix}$. So the two receptors behave differently with receptor 2 binding more of the two ligands as compared to receptor 1. We therefore see incidents where receptor 2 is more than half bound while receptor 1 is not. This enables the binding path to fall into a third region, region II, as the ligands continue increasing in number.

In case of $\begin{bmatrix} 10 & 10 \\ 1 & 1 \end{bmatrix}$, where the binding scheme can lie in three regions, the maximum mutual information, which is the maximum output entropy for deterministic setting, is 1.58. This can be connected to the fact that if a response has a $\frac{1}{3}$ probability of lying in any of the three regions and a zero probability in the fourth, then the entropy becomes $\log_2 3$ which equals 1.58. Therefore, just as the bisymmetric matrices have an upper bound of 2 for mutual information, non-bisymmetric matrices can exhibit a maximum mutual information of 1.58 bits.

5.8.4 Some More Affinity Matrices and Their Scatter Plots

Although we have been able to successfully optimise the parameters using MCMC methods: ligand means and standard deviations, and total receptor concentrations to obtain the maximum possible specificity of the system, we are unsure of the numbers this analysis gives for specificity. For example in case of bisymmetric matrices, the magic number is 2 - no bisymmetric matrix can exhibit a maximum specificity of more than 2 bits. We also studied the special case of bisymmetric matrices of the form $\begin{bmatrix} 1 & X \\ X & 1 \end{bmatrix}$ and found that the specificity increases quite rapidly as soon as X starts taking values greater than 1. On the other hand, the number 1.58 comes up repeatedly for non-bisymmetric cases. Moreover, some matrices like $\begin{bmatrix} 10 & 1 \\ 10 & 1 \end{bmatrix}$, which may not look non-specific, turned out to be so with maximum possible specificity of 1 bit. To understand the stimulus to response mapping in more detail, we study the scatterplots for some more affinity matrices.

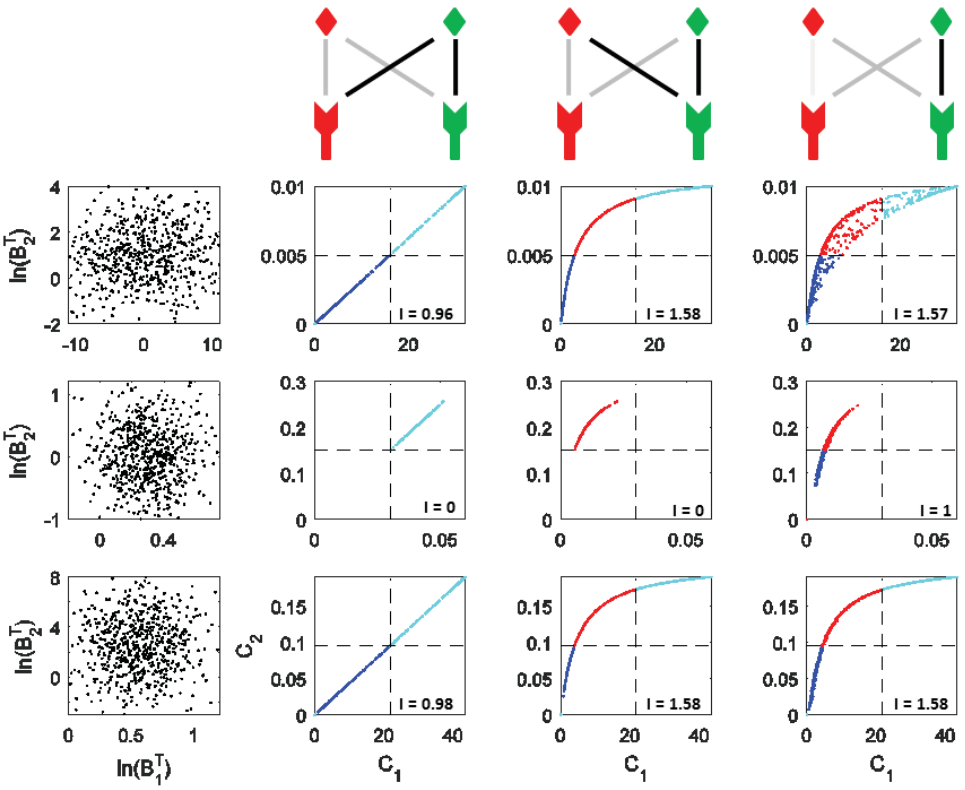

Figure 5.17: Responses for non-bisymmetric K_D matrices. Three different stimuli and receptor concentrations are given to systems with three different non-bisymmetric K_D matrices. The analysis shows that the comparative affinities from the K_D matrices for different ligands determine the shape of the region which the response covers. The rows represent different stimuli and A^T values. The values of parameters used are $\mu_1 = \{0.06, 0.3, 0.56\}$, $\mu_2 = \{1, 0.02, 2.5\}$, $\sigma_1^2 = \{38, 0.03, 0.06\}$, $\sigma_2^2 = \{1.6, 0.25, 5.52\}$, $A_1^T = \{32, 0.06, 43.4\}$, and $A_2^T = \{0.01, 0.3, 0.19\}$ in respective order. The columns are for different K_D matrices. The matrices used are $[10, 1; 10, 1]$, $[10, 10; 1, 1]$, and $[30, 10; 10, 1]$

We fix a stimulus, that is, ligand means and variances, as in the caption of figure 5.18. We then look at the scatter plots for both the bound receptors corresponding to the given stimulus samples. We observe that the K_D matrix in panel A effectively behaves as a bisymmetric matrix $\begin{bmatrix} 1 & X \\ X & 1 \end{bmatrix}$, with some value of X. This is because the scattering region from this matrix is similar in shape to that in 5.1. The region is also symmetric as C_1 and C_2 increase. We observe similar behaviour in panel D, though the symmetry is lost here. There are outputs in all four regions as in panel A but the shape of the boundary is different for the scattering region. We can attribute this to the fact that for the matrix in panel A, both the ligands had affinity for corresponding receptor to an equal extent, which is not the case in the matrix of panel D. If we use $\sim$ to denote the equivalence of two matrices in the way that the maximum specificity exhibited by both is equal, then we conclude from this comparison that

$$\begin{bmatrix} A & B \\ B & A \end{bmatrix} \sim \begin{bmatrix} 1 & X \\ X & 1 \end{bmatrix}$$

for some X, and

$$\begin{bmatrix} A & B \\ C & A \end{bmatrix} \sim \begin{bmatrix} 1 & X \\ Y & 1 \end{bmatrix}$$

for some X and Y.

For the matrix in panel B, we see that there is no region where scattering takes place, the region is rather just reduces to a line, whereas in panel E, although one would think it is a similar case, there is region formed by the outputs. For the matrix in panel A, the second receptor is 10 times more likely to bind than receptor one, to both the ligands. That is, the level of preference for receptor 2 over 1 is equal for both ligands. In the matrix of panel E, the ligand 2 favours receptor 2 10 times more than receptor 1, but ligand 1 favours receptor 2 only twice as much as 1. Therefore, this level of likeliness shows in the scatter plots. The more a ligand favours one receptor over another, the sharper the curve of the boundary, and if both ligands have different preference levels for receptors, this shows as the area between the two curves. This conclusion is confirmed by the results in panel C and F.

Now that we know the response scatterplots have these diverse behaviours, we would like to point out that for the present setup with binary binning and the threshold at half

the receptor concentration level, the scatter plots determine the maximum specificity a matrix can exhibit. A matrix with response in all the four binning regions can exhibit more specificity than that with response in three regions. For example, the matrices in panel A and D can exhibit a maximum specificity (response entropy) of 2 bits, whereas for matrices similar to the matrix in panel F, this maximum cannot be achieved because one of the binning regions never has a response in it. The maximum that such a matrix can achieve is therefore 1.58, when all the responses are equally distributed among three regions. Similarly a matrix with straight response line without any curve can only exhibit 1 bit of maximum specificity.

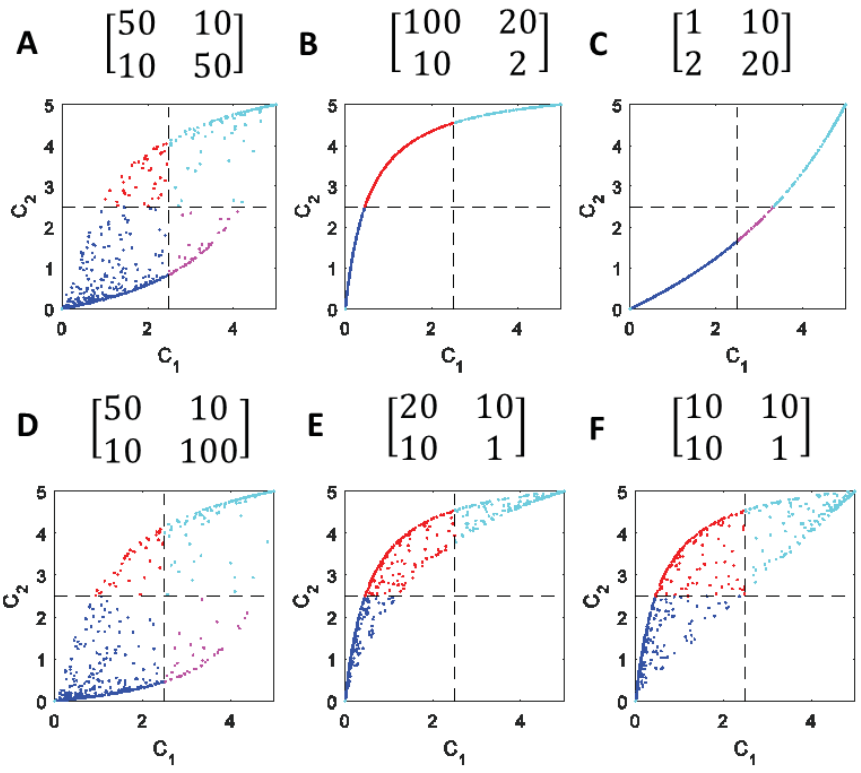

Figure 5.18: Responses for non-bisymmetric K_D matrices. The ratios between the column elements of the K_D matrices determine the shape and scattering of the response, which in turn determines the bound on the maximum possible specificity that can be exhibited by that matrix.The stimulus used here is $\mu_1 = 0.06$, $\mu_2 = 1$, $\sigma_1^2 = 38$, $\sigma_2^2 = 1.6$, $A_1^T = 5$, $A_2^T = 5$

5.9 Overall Summary of Results for Maximum Specificity of 2×2 Affinity Matrix Case

The observations in the previous section give us an idea of the specificity limit of an affinity matrix. We already know that for each affinity matrix, there is an optimal region of parameters consisting of ligand distributions and total receptor concentrations. We have shown that MCMC methods are the most efficient way of finding such an optimal parameter combination. We also know now that by looking at an affinity matrix, we can tell the upper limit of the specificity that it can exhibit. It is worth noting again that although the specificity is dependent in ligand distributions and receptor concentrations, the maximum specificity we are referring to is the optimal specificity corresponding to the parameters from the optimal region. A matrix that can exhibit a maximum specificity of two bits may not always exhibit that much specificity, depending upon the particular stimulus used and total receptor concentrations. We also assert that these maxima are only valid for our current specificity estimation setup, i.e., a 2×2 case with binary binning of bound receptors at thresholds of $A^T/2$.

We now summarise the findings for maximum specificity values for a specificity matrix. Let us assume that the K_D matrix is of the form

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

We denote the ratios of the dissociation constants for the two receptors as:

$$r_1 = \frac{a_{11}}{a_{21}}, \qquad r_2 = \frac{a_{12}}{a_{22}}.$$

Then we have the following findings:

(i) $r_1 < 1$ *and* $r_2 < 1$: Maximum specificity $= 1.58$ bits

(ii) $r_1 > 1$ *and* $r_2 > 1$: Maximum specificity $= 1.58$ bits

(iii) $r_1 > 1$ *and* $r_2 < 1$: Maximum specificity $= 2$ bits

(iv) $r_1 < 1$ *and* $r_2 > 1$: Maximum specificity $= 2$ bits

(v) $r_1 = 1$ *or* $r_2 = 1, \quad r_1 \neq r_2$: Maximum specificity $= 1.58$ bits

(vi) $r_1 = r_2 = 1$: Maximum specificity $= 1$ bit

5.10 Specificity Comparison when One of the Ligands is Absent

As discussed before in chapter 1, it is commonly believed that in some signaling systems where both the receptors bind both the ligands, the role of one of the ligands becomes redundant because the two are performing the same task [56]. The analysis of this unexplained redundancy is particularly important for the ligands CD80 and CD86 in the CD28-CTLA4 signaling system because they both bind to receptors CD28 and CTLA4 with a similar fold difference in affinity. Their absolute affinities are different so the redundancy is not entirely obvious. With our specificity estimation setup, we can show that the two ligands in such a case are completely redundant. The functionality of both is important to the system because the specificity when both the ligands are expressed is more than that of the system when only one ligand is expressed.

Figure 5.19 shows the comparison among mutual information estimates when one of the two ligands is absent and when both the ligands are present. The comparison shows that the presence of second ligand brings in more information to the system and the resulting system is more specific. We can therefore conclude, contrary to common understanding, that the second ligand is not necessarily redundant in such signaling systems. This is a a redundancy estimating setup that can then be used by biologists and experimentalists by measuring ligand and receptor concentrations in real systems.

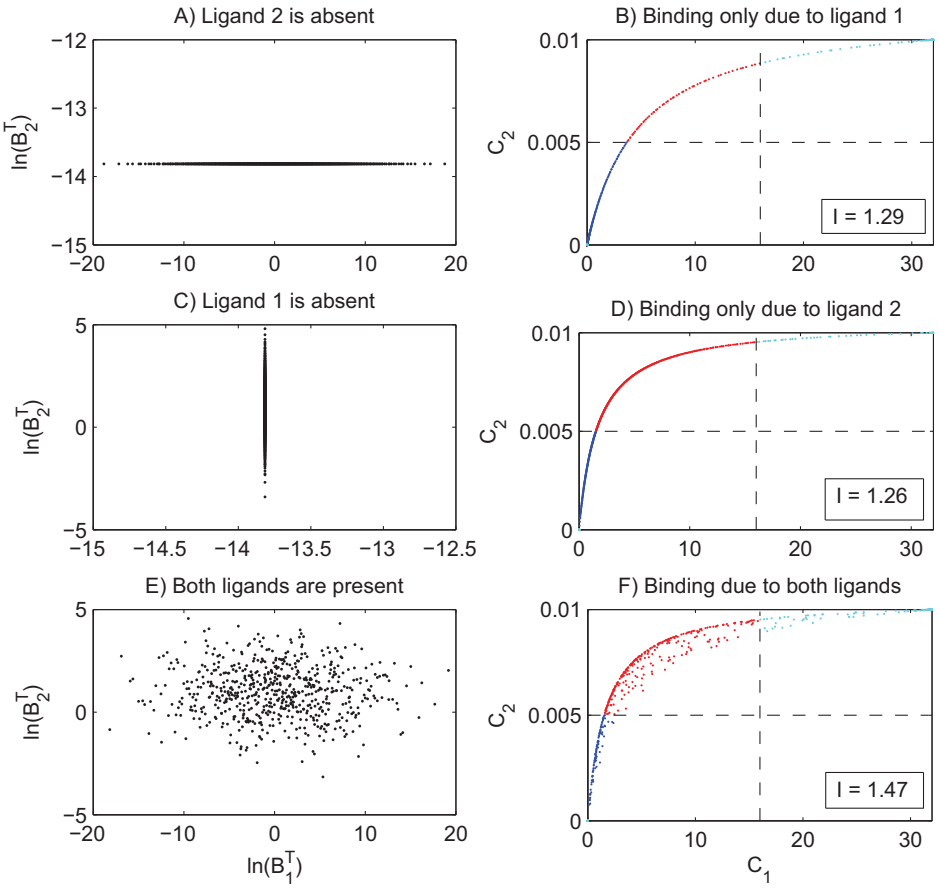

Figure 5.19: The difference in specificity and mutual informations is analysed when when one of the ligands in absent and then compared to the case when both the ligands are present. The specificity in case of both ligands being present is more than the cases where one of the ligands is absent. This shows that there is additional information that the presence of a second ligand brings into the system. For all stimuli, $\mu_1 = 0.06$, $\mu_2 = 1$, $\sigma_1 = 38$, $\sigma_2 = 1.6$, $A_1^T = 32$, $A_2^T = 0.01$, $K_D = [0.2, 0.04; 0.026, 0.002]$

5.11 Generalising the Specificity Estimation Setup Further - Binning Thresholds as Optimisation Parameters

With our current specificity estimation and optimisation setup, we have kept the binary binning thresholds fixed at half the total concentration of receptors. But from the scatter

plots is figure 5.18, we can see that the specificity limits mentioned in the last section can change with different placements of binning thresholds. For example the matrix in panel E from figure 5.18 has a specificity upper limit of 1.58. However, if the thresholds are placed such that the response is in all the four regions for bound receptor concentrations, the specificity might exceed our current limit of 1.58 bits. This gives us another insight into ways of enhancing specificity. If we treat the binning thresholds as optimisation parameters, we can optimise the specificity more effectively.

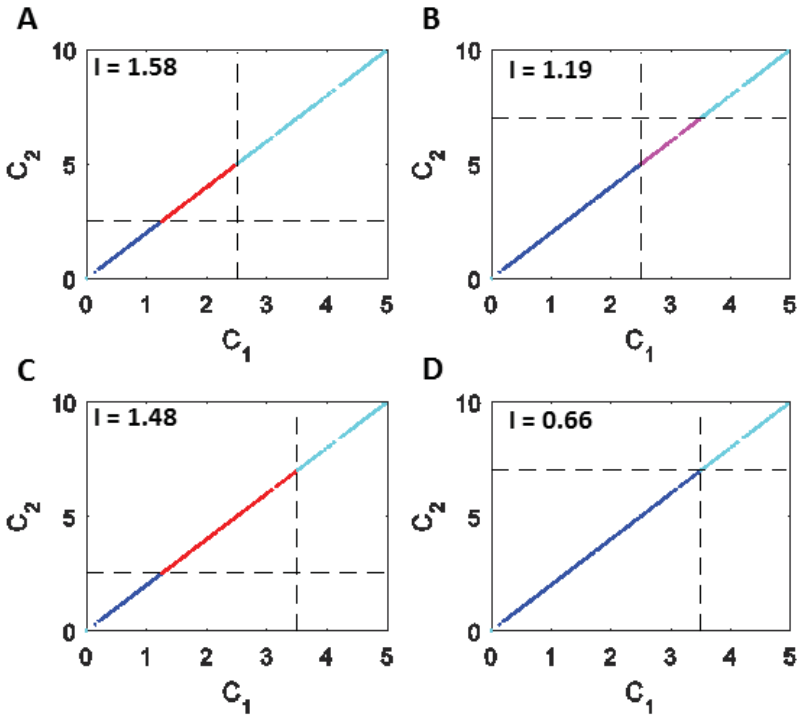

Figure 5.20: Mutual information estimates vary with different threshold settings for $K_D = [1, 1; 1, 1]$. The same stimulus is presented in each case and only the threshold settings are varied. The parameters used are $\mu_1 = 0.5$, $\mu_2 = 1$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1.5$, $A_1^T = 5$, $A_2^T = 10$

Figure 5.20 is the case of non-specific matrix, a matrix of all ones. This figure shows that different placements of thresholds result in different response entropies, and therefore, different specificities. We conclude that the previously thought of as a non-specific matrix can still be specific, based on the binning setup. Unfortunately, it is not always clear how

to place these thresholds, and sometimes the best we can do is to treat them as optimisation parameters, for example, in the MCMC optimisation. This approach is used in the following section that deals with a specific biological example.

5.12 Specificity Estimates for the CD28-CTLA4 System

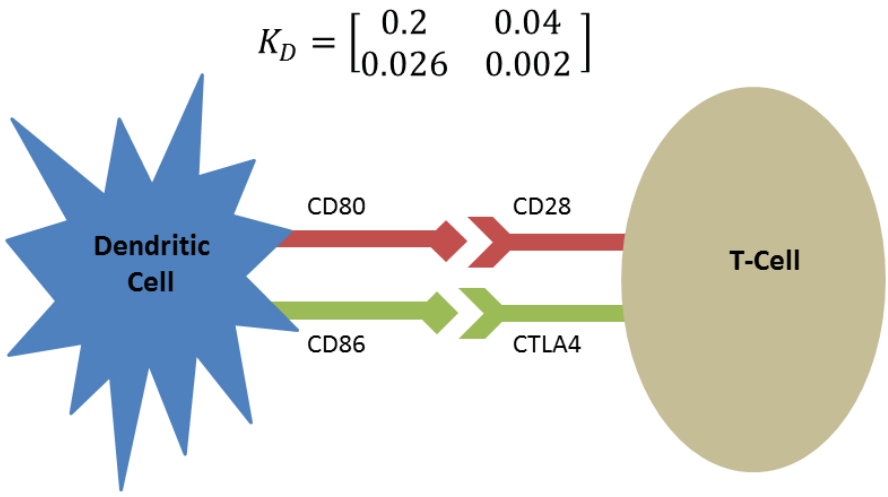

Figure 5.21: A cartoon for CD28-CTLA4 signaling system. The two ligands CD80 and CD86 are present on antigen presenting dendritic cells. The receptors CD28 and CTLA4 are proteins expressed on the surface of T-cells

T-cells are white blood cells, that are important for proper functioning of immune responses against antigens, which carry disease and infection. T-cells circulate in blood looking for antigen presenting cells, and get activated upon encounter. The appropriate activation is important for correct immune responses and for encountering the threats due to antigens. An inappropriate activation may result in allergic reactions and autoimmune disorders.While

the primary signalling route in application is through the T-cell receptors, there is also an important role through signaling by the CD28 and CTLA4 receptors in the T-cell receptors. The ligands for these receptors that are found on dendritic cells (DCs) are CD80 and CD86. CD80 and CD86 are proteins expressed on antigen presenting DC cells, while the receptor proteins CD28 and CTLA4 are expressed on the surface of T-cells. Their binding mechanism determines immune responses. The 2×2 affinity matrix for this system is obtained previously in [10]. Our collaborators Enas Abu Shah and Omer Dushek at Oxford have also measured the distributions of the ligands and receptors in single T-cells and dendritic cells by FACS. In this section we describe the use of our specificity estimation setup for this system. It is often desirable that the receptor ligand binding takes place is a specific way, thus exhibiting maximum specificity.

5.12.1 Data Calibration and Specificity Estimation Setup

The data we use are obtained from experiments using FACS technique. Fluorescent proteins are used to measure the level of expression of the proteins. The experiments show that the ligands concentrations are lognormally distributed. From the experimental measurements of raw intensities, we have means and variances from lognormal distributions for autofluorescence (unstained background) and total fluorescence (total fluorescence meaning the signal fluorescence in addition to the autofluorescence) for both ligands and receptors for two different types of T-cells: CD4 and CD8. With the experimental observations that all the raw fluorescence intensities follow lognormal distributions, we use the means and standard deviations from the data to obtain raw intensity distribution for the signal i.e, fluorescence purely due to the ligand and receptor proteins. We use the following convolution here

$$p(F_T) = g(F_{AF}) * h(F_S) \tag{5.11}$$

where F_T,F_{AF}, and F_S denote total fluorescence, autofluorescence, and signal fluorescence respectively, and p, g, and h their respective probability distributions. With means and variances for F_T and F_{AF}, given that they follow lognormal distributions, we sample form the distributions and bin the samples to obtain p and g. We then apply deconvolution using inverse Fourier transform to h. With h and binning scheme in hand, we generate samples

for F_S. These raw intensity samples for F_S are then transformed into calibrated data, and the required statistics calculated.

To be able to apply our specificity estimation setup to the experimental setup, we need to be able to produce calibrated data for ligand distribution statistics, i.e., means and variances of the underlying normal distribution, as well as total receptor concentration. Although we have a distribution for receptors, here we use the mean of the receptor concentration as the total available receptor. We scale the experimental affinity matrix, and all the calibrated data for comparable units and then apply our specificity estimation setup to calculate specificities. From the experimental data we have, we calculate how much specificity the system actually exhibits in our case for variable binning thresholds. Note that as here ligand distributions and receptor levels are measured and we only do the MCMC optimisation over thresholds.

5.12.2 Results and Discussion

The measurements we have of the raw intensities of the receptors and ligands are in two states: resting (immature) and activated (mature). Therefore, we look at all the various combinations of states for ligands and receptors to analyse and compare the specificities in these cases.

Figure 5.22 shows how the stimulus from our experiments is mapped onto the response. The figure shows the results for two cases: when both the DCs and CD4s are resting, and when the DCs are resting while CD4 cells are in the activated state. The figure shows a sample case when both the thresholds are set at half the total receptor concentration. The specificities in these cases are 1.39 and 1.32 bits which is quite high, although this can be improved with better placements of binning thresholds.

Figure 5.23 is the case where we use the stimulus statistics of the experimental data, and then vary the threshold placements. The maximum specificities that we obtain are 1.47 and 1.58 bits for resting and activated CD4 cells respectively. It is worth noting that the thresholds are bound by total receptor concentrations that we set to be the mean receptor concentration values from the data.

We perform similar analyses of other combinations for the states of dendritic as well

as CD4 and CD8 cells. For examples, Figures 5.24 and 5.25 provide the same analysis as before in case of activated or mature DCs. These cases exhibit zero specificity, regardless of the state of the CD4 cells. From the scatterplots in figure 5.24, it is clear that for the stimulus presented by the distribution of dendritic cells, both CD28 and CTLA4 are in on state.

Figures 5.26 and 5.27 provide scatter plots for stimulus and response and threshold variation as before for the case where DCs are resting and the receptors CD28 and CTLA4 are on the CD8 cells. The results are similar to the case of CD T-cells, and exhibit some specificity, more so in the activated CD4 case.

Figures 5.28 and 5.29 again show that when the DCs are in the mature state, the system is non-specific for the CD8 cells as well. As in the case of CD4 cells, both the receptors are almost completely bound when the DCs are in activated state.

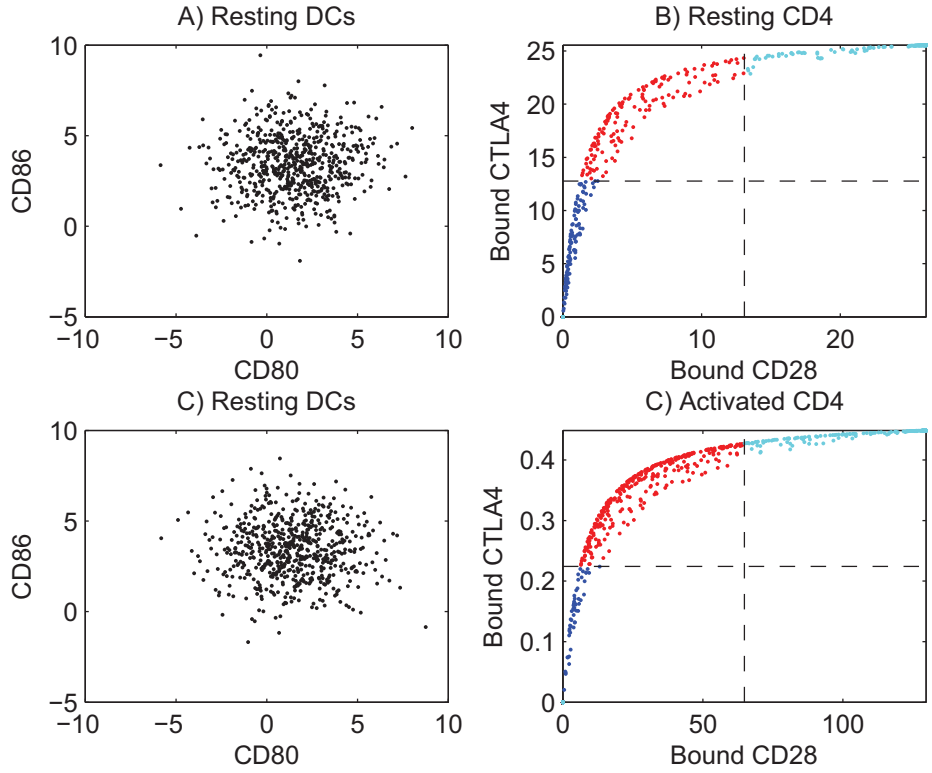

Figure 5.22: Stimulus and response for the experimental data when the DCs and CD4 cells are resting (Panels A and B), and when DCs are resting and CD4 cells are activated (Panels C and D). The response mutual informations in the two cases are 1.39 and 1.32 bits respectively

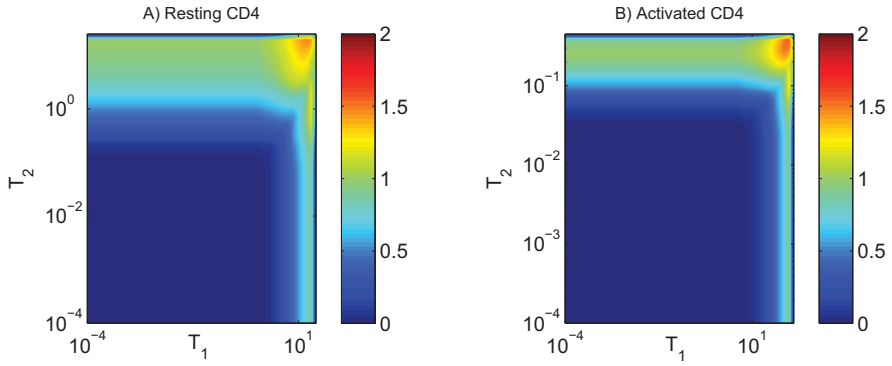

Figure 5.23: Heatmaps for response specificities as the response thresholds are varied for the cases when DCs and CD4 cells are resting (Panel A), and when DCs are resting and CD4 cells are activated (Panel B). The maximum specificities observed in the two cases are 1.47 and 1.58 bits respectively

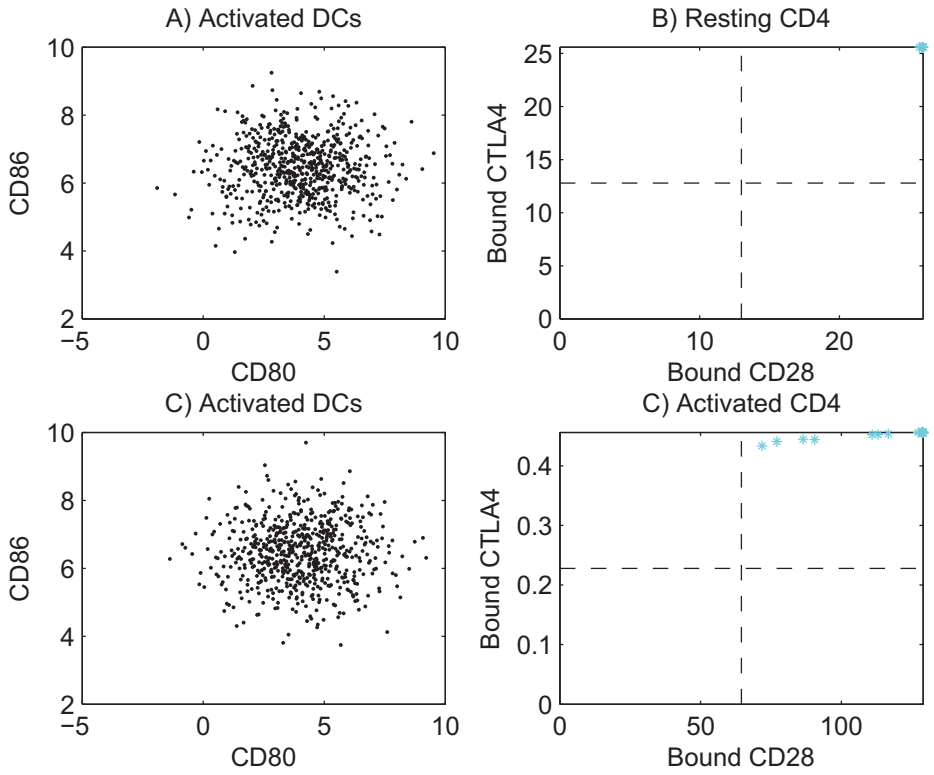

Figure 5.24: Stimulus and response for the experimental data when the DCs are activated and CD4 cells are resting (Panels A and B), and when both DCs and CD4 cells are activated (Panels C and D). The response mutual informations in the two cases are 0 bits

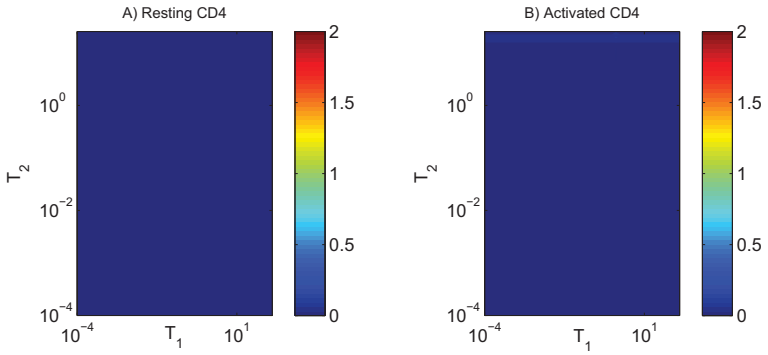

Figure 5.25: Heatmaps for response specificities as the response thresholds are varied for the cases when DCs are activated and CD4 cells are resting (Panel A), and when both DCs and CD4 cells are activated (Panel B). The maximum specificities observed in the two cases are 0 and 0.07 bits respectively

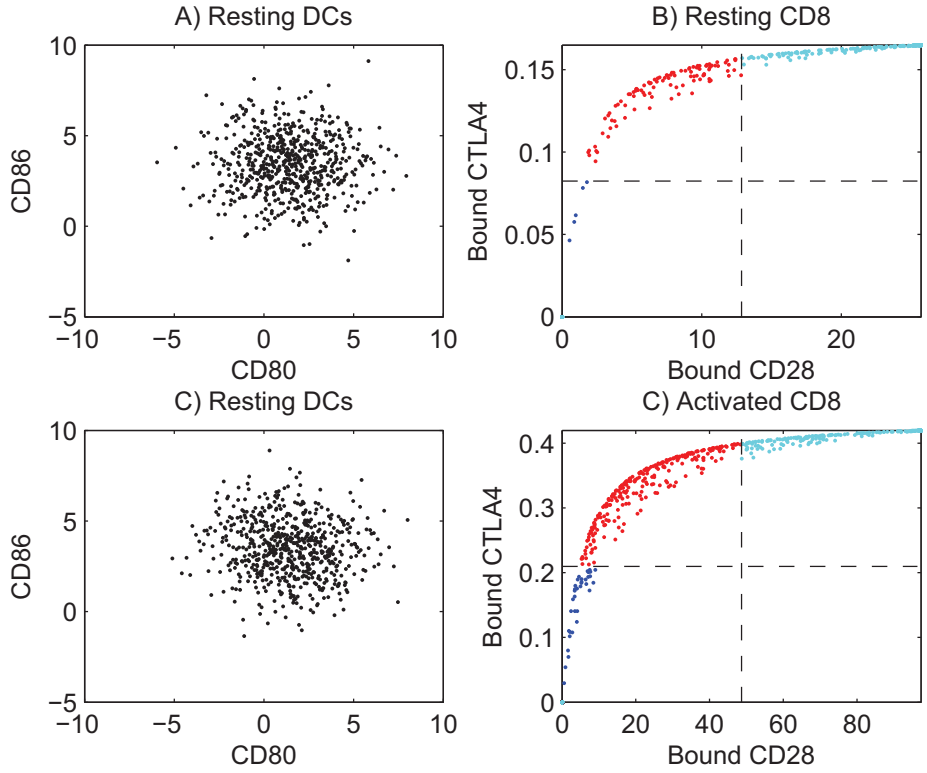

Figure 5.26: Stimulus and response for the experimental data when the DCs and CD8 cells are resting (Panels A and B), and when DCs are resting and CD8 cells are activated (Panels C and D). The response mutual informations in the two cases are 0.71 and 1.24 bits respectively

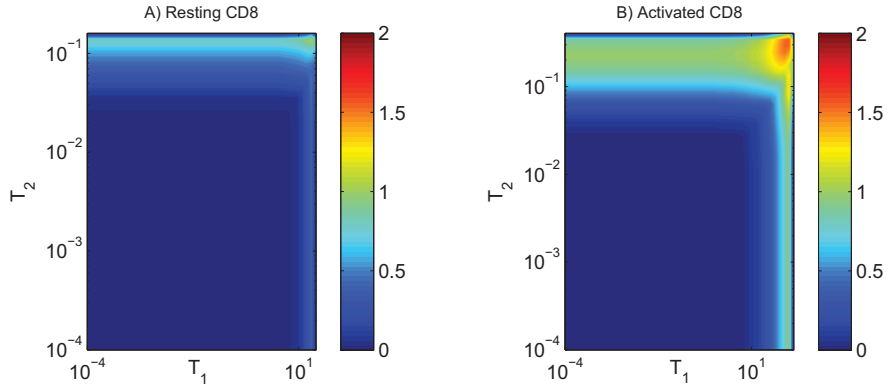

Figure 5.27: Heatmaps for response specificities as the response thresholds are varied for the cases when DCs and CD8 cells are resting (Panel A), and when DCs are resting and CD8 cells are activated (Panel B). The maximum specificities observed in the two cases are 1.04 and 1.57 bits respectively

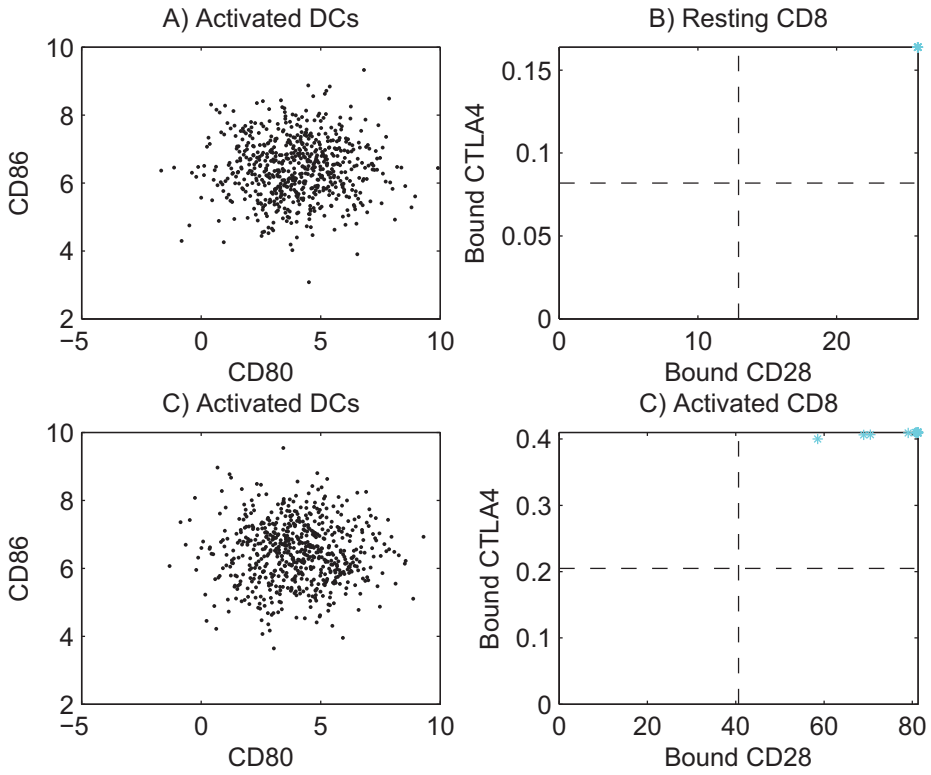

Figure 5.28: Stimulus and response for the experimental data when the DCs are activated and CD8 cells are resting (Panels A and B), and when both DCs and CD8 cells are activated (Panels C and D). The response mutual informations in the two cases are 0 bits

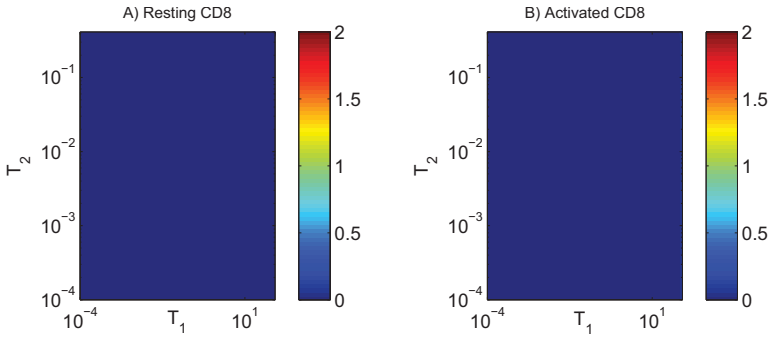

Figure 5.29: Heatmaps for response specificities as the response thresholds are varied for the cases when DCs are activated and CD8 cells are resting (Panel A), and when both DCs and CD8 cells are activated (Panel B). The maximum specificities observed in the two cases are 0 and 0.07 bits respectively

5.13 Specificity in a Stochastic Setting

In the previous sections, we have looked at specificity of a purely deterministic system. But we know that biochemical systems are noisy and it is more realistic to study them from a stochastic point of view. In a stochastic system the receptor ligand binding will be noisy. But our definition of specificity in terms of mutual information remains the same in a stochastic setup. Mutual information is versatile in the sense that it easily accommodates for stochasticity. When switching from a purely deterministic system to a stochastic one, we are replacing a single response value to a distribution. Therefore, accommodating several values of response per stimulus value is the key to mutual information estimation.

The idea of extension of specificity quantification for stochastic systems is, however, not completely straightforward. There are a few technicalities that need to be dealt with. The first difficulty that arises is that Gillespie simulations are required for each solution to a stochastic system. This means that, if previously we were generating 1000 samples to capture a stimulus distribution and solving a deterministic system for fixed points for each of these samples for responses, we now need to run Gillespie simulations 1000 times to generate this many response distributions, which then yield the response entropies and mutual informations.

The second complication comes from the fact that we cannot use response entropy as mutual information. In a deterministic system, there is no noise entropy and for fine enough stimulus binning scheme, the value of mutual information converges to that of response entropy. However, finer binning does not help for stochastic case as there is additional uncertainty due to intrinsic noise of the system. For binary responses, the output entropy estimation directly does not produce much bias, so bias correction is not required. However, while computing mutual information, the stimulus need to be discretised, and there can only be a finite number of stimulus bins. Also, the entropy of one variable for N samples is always less biased than mutual information between two variables with N realisations each. Therefore, the estimate of mutual information needs to be bias corrected as well.

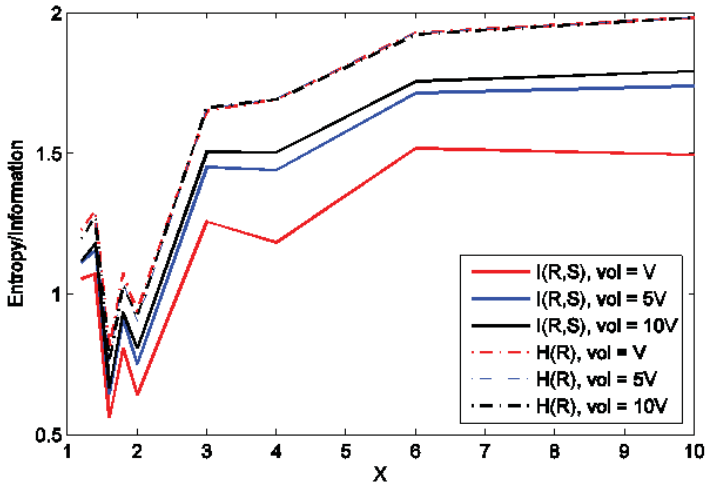

Figure 5.30: Change in mutual information as X varies in the K_D matrix $[1, X; X, 1]$ and the system is stochastic. All the levels of stochasticity are set using different system volumes where $V = 10^{-15}$. The same stimulus is used for all Gillespie simulations, except when volumes are used for binary reaction rates. All the parameters are set at the same values that provide optimal specificities in the corresponding deterministic case. The estimates of mutual information and entropies are bias corrected

Figure 5.30 provides a glimpse of what stochastic specificity looks like in the receptor ligand binding system. We have limited our analysis to bisymmetric K_D matrix of the form $\begin{bmatrix} 1 & X \\ X & 1 \end{bmatrix}$, and estimated specificities using the same parameters that provide the results of figure 5.8. The parameters for binary reactions were changed to accommodate for the different levels of stochasticity of the system. The results show that the specificity of the system decreases as noise is increased. This is expected because when the system gets more and more stochastic, it is difficult for ligands to bind to specific receptors as noise increases fidelity and cross-talk.

Ideally, with the accommodation of a few extra steps of Gillespie simulations and bias correction, the MCMC methodology provides best solution for finding the optimal parameter region that maximises the stochastic specificity.

5.14 Concluding Remarks

In this chapter, we have proposed a novel method for quantification of biological specificity in a receptor ligand binding system using entropy and mutual information estimates. We developed a specificity estimation setup in terms of mutual information estimates between stimulus and response, where the stimuli and responses are based on the concentrations of ligands and receptors in the system. We worked within a deterministic setting with a 2×2 affinity matrix and solved for response distribution corresponding to a particular stimulus distribution. The response was binned into two bins, representing only on and off states of the receptors, while the stimulus followed a lognormal distribution. Primarily working with bisymmetric affinity matrices, we showed how parameters of the specificity estimation setup influence the specificity exhibited by a system. We also observed that there were optimal ranges corresponding to maximum specificity that a system can exhibit. The optimal parameter solution were sought by grid search, simulated annealing and MCMC methods. MCMC methods proved to be most efficient for finding regions of parameters where specificity is maximal.

We extended our analysis of specificity quantification to non-bisymmetric cases also, and showed that there were bounds on maximum specificities for different kinds of affinity matrices. Next, we discussed different types of affinity matrices and reported their specificity upper bounds with our quantification methods. We then introduced the idea of considering the response binning thresholds as optimisation parameters and proposed that the specificity bounds could be pushed upwards with variable binning thresholds. Finally, we applied our specificity quantification methodology to experimental data for T-cells and showed that, against common understanding, the specificity estimates of more than 1 bit ensure that the on-off and off-on states for receptors are also identifiable in addition to the on-on and off-off states, and this contributes towards the specificity of the system. At the end of this chapter, we proposed how this setup can be extended for a stochastic setting and illustrated that specificity in a stochastic setting will be less than that of the corresponding deterministic system.

Chapter 6

Conclusions and Future Work

In this thesis, we have studied two main problems in mathematical biology: stochastic sensitivity analysis and specificity quantification for biochemical networks. We have detailed the motivation behind the study of these problems in chapter 1, wherein we discussed that biological systems are stochastic in nature, and this stochasticity becomes very important at low copy numbers of the species involved. We have also elaborated that stochasticity has the capability to change system dynamics, and because of this, there is a need to study biological systems in a stochastic setup. Biological systems are complex as they are, but the added element of stochasticity makes their study even more complicated and highly nontrivial.

We have explained in chapter 1 that the general phenomena of robustness of stochastic systems is closely related to sensitivity analysis. There are many tools and methodologies in literature to deal with analysis of sensitivity of a system. We have presented a review of these methods in chapter 3. In chapter 4, we developed a global stochastic sensitivity analysis methodology using tools from information theory. The methodology developed here is a generalisation of the information theoretic sensitivity analysis setup [79] for stochastic systems. Information theory setup is a useful approach for such a study in the sense that it accommodates for stochasticity relatively easily, and it is straightforward to estimate quantities like entropies and mutual informations. We described how we define sensitivities of outputs with respect to input parameters, and how second and higher order parameter sensitivities are defined and estimated. We also introduced the idea of intrinsic noise entropy

and how discretisation entropy [79] evolves as noise entropy for stochastic systems. We also showed that the presence of both intrinsic noise entropy and discretisation entropy can alter the distribution of sensitivities dictated by sensitivity summation laws.

The implementation of the method, as in chapter 4, is not straightforward for stochastic systems. However, the method itself is versatile in the sense that it takes care of both stochasticity and global characteristics of sensitivities at the same time. So one can expect that these advantages come at the cost of computational efficiency. Monte Carlo sampling needs to be performed to capture the input parameter space. Gillespie simulations are then performed for each of these parameter samples, which is the most computationally expensive step of this sensitivity analysis setup. Although bias correction is applied to the entropy and mutual information estimates but most of the bias correction techniques already require a lot of samples, which further require performing that many Gillespie simulations. The Gillespie outputs are used for bias corrected entropy and mutual information estimates that define the sensitivity estimates corresponding to various parameters of the system.

We studied a simple four parameter gene expression model with the help of stochastic sensitivity analysis setup that we developed in chapter 4. We repeated the analysis for different levels of stochasticity of the system and observed the interesting phenomena that as noise level of the system changes, the parametric sensitivity changes. For this model, in particular, the sensitivities decreased as the system became more and more noisy. We applied the same methodology to a six parameter gene expression model with negative feedback and observed the same trend.

Although the sensitivity analysis methodology provides some very useful and interesting insights into the systems studied, the methodology can prove to be computationally expensive in its complete generality for complex systems. This motivated us to use the concept of unscented transform to replace Monte Carlo sampling. The idea of unscented transform is to capture the essence of input distribution with a carefully calculated set of sigma points and obtain the output distribution by applying an appropriate nonlinear function to only that set of sigma points. The size of the sigma point set is only $2L + 1$ for an L parameter system. This provides a huge advantage for the sensitivity analysis methodology as thousands of Gillespie simulations are replaced by only $2L + 1$. In the later part of chapter 4, we applied this modified and efficient sensitivity analysis methodology to

the gene expression models and observed a good comparison between the results from full method with Monte Carlo sampling and the modified version with unscented transform. The new version with unscented transform is suitable for use in more complex systems, and we demonstrated this advantage with the application on stochastic model of circadian rhythms. It is worth noting that oscillatory behaviour cannot be analysed with linear noise approximation, and Gillespie simulations are expensive for such analysis. Therefore, for oscillatory systems like the circadian clock example, the unscented transform aided sensitivity analysis is an efficient way of analysis. Other approaches that use linear noise approximation [69, 120] will not be suitable for the study of highly nonlinear and oscillatory systems.

In the sensitivity analysis applications presented in chapter 4, we have focused on regions of parameter space within which system dynamics do not change. Although the gene expression model has the ability to exhibit bimodality in a stochastic setup, but that only happens when the strength of the negative feedback is extremely low. Our studies have been limited to systems (self-repressor gene expression model as well as circadian clock model) that did not exhibit bifurcations for the parameters ranges analysed. We would like to point out here that it will be interesting to study a system like the self-activator gene expression model that can potentially exhibit bistability with this sensitivity analysis technique, and this can be a future area of research. We can expect in such systems, the sensitivity orderings for parameters will change dramatically when the critical values at which bifurcations take place are crossed. Although the unscented transform application does not provide good results for bimodal output variable cases, the sensitivity analysis technique remains applicable in its generality. A detailed study of sensitivity orderings and related parameter values could potentially signify bifurcations. Furthermore, sensitivity analysis of parameters that undergo extrinsic fluctuations can also be helpful in detecting bifurcations in system dynamics induced by extrinsic noise.

The setup developed in chapter 4 provides the necessary equipment for sensitivity analysis. But there are many avenues and several modifications regarding this methodology that need to be explored. One such modification can be how we define sensitivities. In the present setup, we define first order sensitivity to be mutual information between an input parameter and an output variable, which is then normalised by output entropy. If we look

at the definition of mutual information, we observe that this mutual information is the difference in the uncertainties of output with and without the knowledge of input. Another similar way of approaching the idea is to measure some kind of distance between the distributions of output itself and the distribution of output conditioned on the input variable. Kullback Leibler divergence [71] can be one such measure.

The framework for sensitivity analysis presented in chapter 4 has many computational components. One of the component is bias correction of entropy and mutual information estimates. The examples and applications we have looked at use the Panzeri-Treves (PT) method [92]. There are many other bias correction techniques that can be applied for the purpose and may result in more accurate results for less samples. The most interesting methods can include Panzeri-Treves method coupled with shuffled estimators and Nemenman-Shafee-Bialek (NSB) method. We expect that for NSB method, the number of samples required for accurate bias corrected estimates will be considerably less as compared to PT method. However, it is known [49] that NSB method is computationally expensive and time consuming. It will therefore be interesting to study how the efficiency compares for the two methods. Another potentially interesting idea would be to analyse what effects, if any, does Latin hypercube sampling yield on the bias of estimates.

Another modification of the methodology, which is perhaps less intuitive and difficult to interpret, is the use of continuous differential entropies and mutual informations. This is especially useful when the inputs and outputs follow a particular distribution. For example, there are analytical formulas that provide continuous mutual information and entropy values for Gaussian variables. The problem here is that these values for differential entropies can be negative, which hinders the complete understanding of the consequences. If this difficulty can be overcome, it potentially spares us all the discretisation and subsequent bias corrections, and can therefore be of huge advantage.

A drawback of the sensitivity analysis technique studied is that sometimes sensitivities can be negative, especially at higher order. While this again poses the question of how we actually interpret these, the actual inefficiency while calculating multivariate mutual informations lies in the way we decompose them. It is well established that sometimes redundant terms are taken into account [132]. To deal with this problem, we need to modify the mutual information decomposition and ensure that all components are non-negative.

The ideas presented in [102, 132], for example of further decomposing Shannon entropy, can be helpful while exploring this avenue towards a more efficient sensitivity analysis method.

Chapter 5 attempts to answer the problem of specificity quantification in a receptor ligand binding system. Specificity is a well recognised and well developed area of research [45, 67, 117]. The importance of specificity for an organised functioning of biological systems is also well known. Unfortunately, little has been done to develop mathematical methods for quantification of specificity. There are some statistical mechanics approaches [128, 134] that try to address this need. Information theory, however, provides a strong tool to quantify specificity of a system. In chapter 5, we developed a setup in which application of the concepts of entropy and mutual information becomes more effective and provides meaningful results.

Our specificity quantification methodology relies on a stimulus response system, where we set the stimulus to be ligand concentrations, and the responses to be the concentrations of bound receptors. This enables us to quantify specificity in the form of mutual information between stimulus and response. Therefore, we can easily quantify the specificity of a ligand-receptor binding system given the ligand distribution and total receptor concentration. In chapter 5, we also discussed in detail, how we can estimate the maximum possible specificity exhibited by a system with a given affinity matrix. For this reason, we further detailed our setup and described how to actually estimate the entropies and mutual informations using binary response binning i.e., for on and off states of the receptors. We also showed how different parameters of the system affect the maximal specificity of the system. We observed that there is a parameter region, rather than a specific parameter value that optimises specificity. We tried different methodologies ranging from grid search, simulated annealing, to MCMC methods to search for this optimal parameter region and found that MCMC methods were most helpful and least expensive way to do this. This analysis provided us with the ranges of parameters to set in order for a system with a certain affinity matrix to exhibit maximum specificity.

In the later half of chapter 5, we discussed the bounds on specificity of the system given the affinity matrix. We found that for our specificity estimation setup, just the affinity matrix could lead us to guessing the maximum possible specificity of the system. This can

be interesting, as one can identify a real system based on how much specificity the system actually exhibits. We also pointed out that these limits are specific to the estimation setup. With another free variable like the on-off switching threshold, these systems can potentially exhibit even more specificity. In addition to this we also showed that our specificty estimation setup is useful from a biological point of view in that it helps us quantify the so-called redundancy of ligands in signaling systems. We predicted that the presence of a second ligand, as in the case of CD28-CTLA4 systems, is not necessarily unimportant as this can effect the overall specificity of the system.

We applied our specificity estimation methodology to the experimental data provided to us by our collaborators at Oxford university. The data are for antigen presenting dendritic cells with ligands CD80 and CD86, and for two types of T-cells, CD4 and CD8 both with receptor proteins CD28 and CTLA4. With the help of these data, we estimated the specificities of the system, and found that these systems are more specific for resting DCs. The results also showed the interesting fact that more than 1 bit of specificity implies that the cells do not only identify the states when both the receptors are simultaneously on and simultaneously off, rather the states where one receptor is on and the other off, are also identifiable and contribute towards specificity of the system. This suggests that the ligands CD80 and CD86 are not completely redundant and they potentially encode different biological information.

We concluded chapter 5 with the idea of how the analysis can be extended for stochastic systems. It is pointed out that the stochastic specificity analysis becomes computationally expensive with the need to use Gillespie simulations, and bias correction methods for mutual information estimation. We showed some basic results for stochastic system to compare with the deterministic results and found that, as expected, noise suppresses specificity of a system.

The study in chapter 5 is a basic setup for specificity quantification. Moreover, the analysis we have presented here is focused at a two ligands and two receptors system. Future extensions of this methodology can be aimed in different directions. Even in a 2×2 system, there is plenty of room for further exploration. For example, in our setup we are assuming that the stimulus is lognormally distributed, mainly because this brings us closer to the specificity peak we are targeting. In a more general setup, this assumption

can be relaxed and we can look for the distribution that maximises mutual information. In other words, we can aim our study at channel capacity of the system, wherein the stimulus distribution is not limited to lognormal. Another possible extension of our study in a 2×2 system can focus on non-binary binning of response. Rather than assuming that each response component is either in on or off state, we can consider a finite number of response states. The challenge in that case arises from the fact that the response entropy will not be limited to 2 bits, which makes interpretation of the results more difficult.

The obvious generalisation of our specificity quantification setup is for more than two receptors and ligands. In general for N receptors with binary outputs, maximum specificity can be N bits. For example, we can say that the affinity matrix of the form $\begin{bmatrix} X & 1 & 1 \\ 1 & X & 1 \\ 1 & 1 & X \end{bmatrix}$ for a large enough X should technically exhibit this maximum specificity of 3 bits, but unlike in the 2×2 case, finding the specificity bound for a general affinity matrix is computationally highly demanding.

The most interesting future direction for this specificity quantification setup is for stochastic systems. We already know that no real system exhibits complete specificity. Intrinsic noise present in these systems is one of the reasons why this could be so. Our specificity estimation setup lends itself well to generalisation for stochastic systems as we have discussed in the last section of chapter 5. We have presented that with the help of Gillespie simulations, the analysis can be done for stochastic systems. However, it still needs to be researched whether the optimal parameter regions for maximum stochasticity in deterministic systems and stochastic systems are the same. We expect that MCMC methods work equally well for stochastic systems in search for the optimal parameter regions. Our preliminary results for comparison showed that in all cases studied for particular parameter values, the specificity of the stochastic system is less than that for the corresponding deterministic system. We expect that this observation is true in general. A detailed analysis of specificity quantification in stochastic systems is a direction for our future study.

References

[1] Sébastien Aumaître, Kirone Mallick, and François Pétrélis. Noise-induced bifurcations, multiscaling and on–off intermittency. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(07):P07016, 2007.

[2] Lee Bardwell, Xiufen Zou, Qing Nie, and Natalia L Komarova. Mathematical models of specificity in cell signaling. *Biophysical Journal*, 92(10):3425–3441, 2007.

[3] Naama Barkai and Stanislas Leibler. Biological rhythms: Circadian clocks limited by noise. *Nature*, 403(6767):267–268, 2000.

[4] M Bentele, I Lavrik, M Ulrich, S Stösser, DW Heermann, H Kalthoff, PH Krammer, and R Eils. Mathematical modeling reveals threshold mechanism in cd95-induced apoptosis. *The Journal of Cell Biology*, 166(6):839–851, 2004.

[5] M Bentele, I Lavrik, M Ulrich, S Stösser, DW Heermann, H Kalthoff, PH Krammer, and R Eils. Mathematical modeling reveals threshold mechanism in cd95-induced apoptosis. *The Journal of Cell Biology*, 166(6):839–851, 2004.

[6] Alexander Borst and Frédéric E Theunissen. Information theory and neural coding. *Nature Neuroscience*, 2(11):947–957, 1999.

[7] JA Burns, A Cornish-Bowden, AK Groen, R Heinrich, H Kacser, JW Porteous, SM Rapoport, TA Rapoport, JW Stucki, JM Tager, et al. Control analysis of metabolic systems. *Trends in Biochemical Sciences*, 10(1):16, 1985.

[8] Jean M Carlson and John Doyle. Complexity and robustness. *Proceedings of the National Academy of Sciences*, 99(suppl 1):2538–2545, 2002.

[9] Vladimír Černỳ. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications*, 45(1):41–51, 1985.

[10] Alison V Collins, Douglas W Brodie, Robert JC Gilbert, Andrea Iaboni, Raquel Manso-Sancho, Björn Walse, David I Stuart, P Anton van der Merwe, and Simon J Davis. The interaction properties of costimulatory molecules revisited. *Immunity*, 17(2):201–210, 2002.

[11] Begoña Comín-Anduix, Joan Boren, Sonia Martinez, Cristina Moro, Josep J Centelles, Raisa Trebukhina, Nataly Petushok, Wai-Nang Paul Lee, Laszlo G Boros, and Marta Cascante. The effect of thiamine supplementation on tumour proliferation. *European Journal of Biochemistry*, 268(15):4177–4182, 2001.

[12] Marie E Csete and John C Doyle. Reverse engineering of biological complexity. *Science*, 295(5560):1664–1669, 2002.

[13] RI Cukier, CM Fortuin, Kurt E Shuler, AG Petschek, and JH Schaibly. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I theory. *The Journal of Chemical Physics*, 59(8):3873–3878, 1973.

[14] RI Cukier, JH Schaibly, and Kurt E Shuler. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. III. analysis of the approximations. *The Journal of Chemical Physics*, 63(3):1140–1149, 1975.

[15] Bryan C Daniels, Yan-Jiun Chen, James P Sethna, Ryan N Gutenkunst, and Christopher R Myers. Sloppiness, robustness, and evolvability in systems biology. *Current Opinion in Biotechnology*, 19(4):389–395, 2008.

[16] Wiet de Ronde, Filipe Tostevin, and Pieter Rein Ten Wolde. Multiplexing biochemical signals. *Physical Review Letters*, 107(4):048101, 2011.

[17] Wiet Hendrik de Ronde, Filipe Tostevin, and Pieter Rein Ten Wolde. Effect of feedback on the fidelity of information transmission of time-varying signals. *Physical Review E*, 82(3):031914, 2010.

[18] Andrea Degasperi and Stephen Gilmore. Sensitivity analysis of stochastic models of bistable biochemical reactions. In *Formal Methods for Computational Systems Biology*, pages 1–20. Springer, 2008.

[19] Robert P Dickinson and Robert J Gelinas. Sensitivity analysis of ordinary differential equation systemsa direct method. *Journal of Computational Physics*, 21(2):123–143, 1976.

[20] Eugene P Dougherty, Jenn-Tai Hwang, and Herschel Rabitz. Further developments and applications of the greens function method of sensitivity analysis in chemical kinetics. *The Journal of Chemical Physics*, 71(4):1794–1808, 1979.

[21] Johan Elf and Måns Ehrenberg. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Research*, 13(11):2475–2484, 2003.

[22] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

[23] David A Fell. Metabolic control analysis: a survey of its theoretical and experimental development. *Biochemical Journal*, 286(Pt 2):313, 1992.

[24] Michael A Gibson and Jehoshua Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry A*, 104(9):1876–1889, 2000.

[25] Jeff Gill. *Bayesian methods: A social and behavioral sciences approach*, volume 20. CRC press, 2014.

[26] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

[27] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[28] Daniel T Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001.

[29] Daniel T Gillespie. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, 58:35–55, 2007.

[30] Didier Gonze and Albert Goldbeter. Circadian rhythms and molecular noise. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 16(2):026110, 2006.

[31] Didier Gonze, José Halloy, and Albert Goldbeter. Robustness of circadian rhythms with respect to molecular noise. *Proceedings of the National Academy of Sciences*, 99(2):673–678, 2002.

[32] Ramon Grima. An effective rate equation approach to reaction kinetics in small volumes: Theory and application to biochemical reactions in nonequilibrium steady-state conditions. *The Journal of Chemical Physics*, 133(3):035101, 2010.

[33] Rudiyanto Gunawan, Yang Cao, Linda Petzold, and Francis J Doyle. Sensitivity analysis of discrete stochastic systems. *Biophysical Journal*, 88(4):2530–2540, 2005.

[34] Ankit Gupta and Mustafa Khammash. An efficient and unbiased method for sensitivity analysis of stochastic reaction networks. *Journal of The Royal Society Interface*, 11(101):20140979, 2014.

[35] Ryan N Gutenkunst, Joshua J Waterfall, Fergal P Casey, Kevin S Brown, Christopher R Myers, and James P Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*, 3(10):e189, 2007.

[36] Heikki Haario, Marko Laine, Antonietta Mira, and Eero Saksman. DRAM: efficient adaptive MCMC. *Statistics and Computing*, 16(4):339–354, 2006.

[37] Heikki Haario, Eero Saksman, and Johanna Tamminen. An adaptive metropolis algorithm. *Bernoulli*, pages 223–242, 2001.

[38] Marc Hafner, Heinz Koeppl, Martin Hasler, and Andreas Wagner. glocalrobustness analysis and model discrimination for circadian oscillators. *PLoS Comput Biol*, 5(10):e1000534, 2009.

[39] Jean Hausser and Korbinian Strimmer. Entropy inference and the james-stein estimator, with application to nonlinear gene association networks. *The Journal of Machine Learning Research*, 10:1469–1484, 2009.

[40] Reinhart Heinrich and Tom A Rapoport. A linear steady-state treatment of enzymatic chains. *European Journal of Biochemistry*, 42(1):97–105, 1974.

[41] Desmond J. Higham. Modeling and simulating chemical reactions. *SIAM Rev.*, 50(2):347–368, May 2008.

[42] Nicholas J Higham. Computing the nearest correlation matrixa problem from finance. *IMA Journal of Numerical Analysis*, 22(3):329–343, 2002.

[43] George M Hornberger and RC Spear. Approach to the preliminary analysis of environmental systems. *J. Environ. Manage.;(United States)*, 12(1), 1981.

[44] George M Hornberger and RC Spear. Approach to the preliminary analysis of environmental systems. *J. Environ. Manage.;(United States)*, 12(1), 1981.

[45] Eun-Mi Hur and Kyong-Tai Kim. G protein-coupled receptor signalling and crosstalk: achieving rapidity and specificity. *Cellular Signalling*, 14(5):397–405, 2002.

[46] Jenn-Tai Hwang, Eugene P Dougherty, Suzanne Rabitz, and Herschel Rabitz. The greens function method of sensitivity analysis in chemical kinetics. *The Journal of Chemical Physics*, 69(11):5180–5191, 1978.

[47] Pablo A Iglesias and Brian P Ingalls. *Control theory and systems biology*. MIT Press, 2010.

[48] AEC Ihekwaba, DS Broomhead, RL Grimley, N Benson, and DB Kell. Sensitivity analysis of parameters controlling oscillatory signalling in the nf-kb pathway: the roles of ikk and ikba. *Syst Biol*, 1:93–103, 2004.

[49] Robin AA Ince, Rasmus S Petersen, Daniel C Swan, and Stefano Panzeri. Python for information theoretic analysis of neural data. *Python in Neuroscience*, page 133, 2015.

[50] Sagar Indurkhya and Jacob Beal. Reaction factoring and bipartite update graphs accelerate the gillespie algorithm for large-scale biochemical systems. *PloS One*, 5(1):e8125, 2010.

[51] Brian Ingalls. Sensitivity analysis: from model parameters to system behaviour. *Essays in Biochemistry*, 45:177–194, 2008.

[52] Brian P. Ingalls. A frequency domain approach to sensitivity analysis of biochemical systems. *Journal of Physical Chemistry B*, 108:2004, 2004.

[53] Brian P. Ingalls and Herbert M. Sauro. Sensitivity analysis of stoichiometric networks: an extension of metabolic control analysis to non-equilibrium trajectories. *Journal of Theoretical Biology*, 222:23–36, 2003.

[54] AD Irving. Stochastic sensitivity analysis. *Applied Mathematical Modelling*, 16(1):3–15, 1992.

[55] Joël Janin. Quantifying biological specificity: the statistical mechanics of molecular recognition. *Proteins: Structure, Function, and Bioinformatics*, 25(4):438–445, 1996.

[56] Andreas Jansson, Eleanor Barnes, Paul Klenerman, Mikael Harln, Poul Srensen, Simon J. Davis, and Patric Nilsson. A theoretical framework for quantitative analysis of the molecular basis of costimulation. *The journal of Immunology*, 175(3):1575–1585, 2005.

[57] Jaewook Joo, Steven J Plimpton, and Jean-Loup Faulon. Noise-induced oscillatory shuttling of nf-{\ kappa} b in a two compartment ikk-nf-{\ kappa} bi {\ kappa} b-a20 signaling model. *arXiv preprint arXiv:1010.0888*, 2010.

[58] Simon J Julier. The scaled unscented transformation. In *American Control Conference, 2002. Proceedings of the 2002*, volume 6, pages 4555–4559. IEEE, 2002.

[59] Simon J Julier and Jeffrey K Uhlmann. Consistent debiased method for converting between polar and cartesian coordinate systems. In *AeroSense'97*, pages 110–121. International Society for Optics and Photonics, 1997.

[60] Simon J Julier and Jeffrey K Uhlmann. New extension of the kalman filter to nonlinear systems. In *AeroSense'97*, pages 182–193. International Society for Optics and Photonics, 1997.

[61] H Kacser, , and JA34 Burns. The control of flux. In *Symp. Soc. Exp. Biol.*, volume 27, pages 65–104, 1973.

[62] Kyung Hyuk Kim and Herbert M Sauro. Sensitivity summation theorems for stochastic biochemical reaction systems. *Mathematical Biosciences*, 226(2):109–119, 2010.

[63] Scott Kirkpatrick. Optimization by simulated annealing: quantitative studies. *Journal of Statistical Physics*, 34(5-6):975–986, 1984.

[64] Hiroaki Kitano. Towards a theory of biological robustness. *Molecular Systems Biology*, 3(1):137, 2007.

[65] Hiroaki Kitano, Kanae Oda, Tomomi Kimura, Yukiko Matsuoka, Marie Csete, John Doyle, and Masaaki Muramatsu. Metabolic syndrome and robustness tradeoffs. *Diabetes*, 53(suppl 3):S6–S15, 2004.

[66] Edda Klipp, Wolfram Liebermeister, Christoph Wierling, Axel Kowald, Hans Lehrach, and Ralf Herwig. " systems biology: A textbook, 2009.

[67] Heather Knight and Marc R Knight. Abiotic stress signalling pathways: specificity and cross-talk. *Trends in Plant Science*, 6(6):262–267, 2001.

[68] Natalia L Komarova, Xiufen Zou, Qing Nie, and Lee Bardwell. A theoretical framework for specificity in cell signaling. *Molecular Systems Biology*, 1(1), 2005.

[69] Michał Komorowski, Maria J Costa, David A Rand, and Michael PH Stumpf. Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. *Proceedings of the National Academy of Sciences*, 108(21):8645–8650, 2011.

[70] Daniel E Koshland. The seven pillars of life. *Science*, 295(5563):2215–2216, 2002.

[71] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[72] Edo Kussell and Stanislas Leibler. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309(5743):2075–2078, 2005.

[73] Marko Laine. MCMC toolbox for MATLAB. `http://helios.fmi.fi/~lainema/mcmc/`, 2013. [Online; accessed 6-October-2015].

[74] Ioannis Lestas, Glenn Vinnicombe, and Johan Paulsson. Fundamental limits on the suppression of molecular fluctuations. *Nature*, 467(7312):174–178, 2010.

[75] Difei Li and Chunguang Li. Noise-induced dynamics in the mixed-feedback-loop network motif. *Physical Review E*, 77(1):011903, 2008.

[76] Genyuan Li, Jishan Hu, Sheng-Wei Wang, Panos G Georgopoulos, Jacqueline Schoendorf, and Herschel Rabitz. Random sampling-high dimensional model representation (rs-hdmr) and orthogonality of its different order component functions. *The Journal of Physical Chemistry A*, 110(7):2474–2485, 2006.

[77] Genyuan Li, Sheng-Wei Wang, Herschel Rabitz, Sookyun Wang, and Peter Jaffé. Global uncertainty assessments by high dimensional model representations (hdmr). *Chemical Engineering Science*, 57(21):4445–4460, 2002.

[78] Wolfram Liebermeister. Response to temporal parameter fluctuations in biochemical networks. *Journal of Theoretical Biology*, 234(3):423–438, 2005.

[79] Niklas Lüdtke, Stefano Panzeri, Martin Brown, David S Broomhead, Joshua Knowles, Marcelo A Montemurro, and Douglas B Kell. Information-theoretic sensitivity analysis: a general method for credit assignment in complex networks. *Journal of The Royal Society Interface*, 5(19):223–235, 2008.

[80] Biliana Marcheva, Kathryn M Ramsey, Clara B Peek, Alison Affinati, Eleonore Maury, and Joseph Bass. Circadian clocks and metabolism. In *Circadian clocks*, pages 127–155. Springer, 2013.

[81] Simeone Marino, Ian B Hogue, Christian J Ray, and Denise E Kirschner. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology*, 254(1):178–196, 2008.

[82] Sergei Maslov and Kim Sneppen. Specificity and stability in topology of protein networks. *Science*, 296(5569):910–913, 2002.

[83] Siobhan S Mc Mahon, Aaron Sim, Sarah Filippi, Robert Johnson, Juliane Liepe, Dominic Smith, and Michael PH Stumpf. Information theory and signal transduction systems: from molecular information processing to network inference. In *Seminars in Cell & Developmental Biology*, volume 35, pages 98–108. Elsevier, 2014.

[84] Harley H McAdams and Adam Arkin. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences*, 94(3):814–819, 1997.

[85] Marcelo A Montemurro, Riccardo Senatore, and Stefano Panzeri. Tight data-robust bounds to mutual information combining shuffling and model selection techniques. *Neural Computation*, 19(11):2913–2957, 2007.

[86] Mineo Morohashi, Amanda E Winn, Mark T Borisuk, Hamid Bolouri, John Doyle, and Hiroaki Kitano. Robustness as a measure of plausibility in models of biochemical networks. *Journal of Theoretical Biology*, 216(1):19–30, 2002.

[87] Max D Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.

[88] Ilya Nemenman, Fariel Shafee, and William Bialek. Entropy and inference, revisited. *arXiv preprint physics/0108025*, 2001.

[89] Jürgen Pahle, Joseph D Challenger, Pedro Mendes, and Alan J McKane. Biochemical fluctuations, optimisation and the linear noise approximation. *BMC Systems Biology*, 6(1):86, 2012.

[90] David J Pannell. Sensitivity analysis of normative economic models: theoretical framework and practical strategies. *Agricultural Economics*, 16(2):139–152, 1997.

[91] Stefano Panzeri, Riccardo Senatore, Marcelo A Montemurro, and Rasmus S Petersen. Correcting for the sampling bias problem in spike train information measures. *Journal of Neurophysiology*, 98(3):1064–1072, 2007.

[92] Stefano Panzeri and Alessandro Treves. Analytical estimates of limited sampling biases in different information measures. *Network - Comp. Neural.*, 7:87–107.

[93] Vladimir Privman, Jian Zhou, Jan Halámek, and Evgeny Katz. Realization and properties of biochemical-computing biocatalytic xor gate based on signal change. *The Journal of Physical Chemistry B*, 114(42):13601–13608, 2010.

[94] Herschel Rabitz, Mark Kramer, and D Dacol. Sensitivity analysis in chemical kinetics. *Annual Review of Physical Chemistry*, 34(1):419–461, 1983.

[95] Rajesh Ramaswamy, Nélido González-Segredo, and Ivo F Sbalzarini. A new class of highly efficient exact stochastic simulation algorithms for chemical reaction networks. *The Journal of Chemical Physics*, 130(24):244104, 2009.

[96] David A Rand. Mapping global sensitivity of cellular network dynamics: sensitivity heat maps and a global summation law. *Journal of The Royal Society Interface*, 5(Suppl 1):S59–S69, 2008.

[97] Muruhan Rathinam, Patrick W Sheppard, and Mustafa Khammash. Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks. *The Journal of Chemical Physics*, 132(3):034103, 2010.

[98] A Raue, C Kreutz, T Maiwald, U Klingmuller, and J Timmer. Addressing parameter identifiability by model-based experimentation. *Systems Biology, IET*, 5(2):120–130, 2011.

[99] Christine Reder. Metabolic control theory: a structural approach. *Journal of Theoretical Biology*, 135(2):175–201, 1988.

[100] Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

[101] Andrea Rocco. Stochastic control of metabolic pathways. *Physical Biology*, 6(1):016002, 2009.

[102] Fernando Rosas, Vasilis Ntranos, Christopher J Ellison, Sofie Pollin, and Marian Verhelst. Understanding interdependency through complex information sharing. *arXiv preprint arXiv:1509.04555*, 2015.

[103] Andrea Saltelli, Marco Ratto, Stefano Tarantola, and Francesca Campolongo. Sensitivity analysis for chemical models. *Chemical Reviews*, 105(7):2811–2828, 2005.

[104] John H Schaibly and Kurt E Shuler. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. II applications. *The Journal of Chemical Physics*, 59(8):3879–3888, 1973.

[105] Vahid Shahrezaei, Julien F Ollivier, and Peter S Swain. Colored extrinsic fluctuations and stochastic gene expression. *Molecular Systems Biology*, 4(1), 2008.

[106] Vahid Shahrezaei and Peter S Swain. Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences*, 105(45):17256–17261, 2008.

[107] Vahid Shahrezaei and Peter S Swain. The stochastic nature of biochemical networks. *Current Opinion in Biotechnology*, 19(4):369–374, 2008.

[108] Fernando Siso-Nadal, Julien F Ollivier, and Peter S Swain. Facile: a command-line network compiler for systems biology. *BMC systems Biology*, 1(1):36, 2007.

[109] Ilya M Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and Computers in Simulation*, 55(1):271–280, 2001.

[110] Sunil Srinivasa. A review on multivariate mutual information. *Univ. of Notre Dame, Notre Dame, Indiana*, 2:1–6, 2005.

[111] Jörg Stelling, Uwe Sauer, Zoltan Szallasi, Francis J Doyle, and John Doyle. Robustness of cellular functions. *Cell*, 118(6):675–685, 2004.

[112] Steven P Strong, Roland Koberle, Rob R de Ruyter van Steveninck, and William Bialek. Entropy and information in neural spike trains. *Physical Review Letters*, 80(1):197, 1998.

[113] Peter Swain. Modelling stochastic gene expression.

[114] Peter S Swain. Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *Journal of Molecular Biology*, 344(4):965–976, 2004.

[115] Peter S Swain, Michael B Elowitz, and Eric D Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, 99(20):12795–12800, 2002.

[116] I Swameye, TG Müller, J t Timmer, O Sandra, and U Klingmüller. Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proceedings of the National Academy of Sciences*, 100(3):1028–1033, 2003.

[117] Yurong Tang, Xia Sheng, and Michael PH Stumpf. The roles of contact residue disorder and domain composition in characterizing protein–ligand binding specificity and promiscuity. *Molecular BioSystems*, 7(12):3280–3286, 2011.

[118] Gašper Tkačik, Curtis G Callan Jr, and William Bialek. Information capacity of genetic regulatory elements. *Physical Review E*, 78(1):011910, 2008.

[119] Gašper Tkačik, Aleksandra M Walczak, and William Bialek. *Physical Review E*, 80(3):031920, 2009.

[120] Tina Toni and Bruce Tidor. Combined model of intrinsic and extrinsic variability for computational network design with application to synthetic biology. *PLoS Comput Biol*, 9(3):e1002960, 2013.

[121] Filipe Tostevin and Pieter Rein Ten Wolde. *Physical Review Letters*, 102(21):218101, 2009.

[122] Mark K Transtrum, Benjamin B Machta, Kevin S Brown, Bryan C Daniels, Christopher R Myers, and James P Sethna. Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *The Journal of Chemical Physics*, 143(1):010901, 2015.

[123] Jeffrey A Ubersax and James E Ferrell Jr. Mechanisms of specificity in protein phosphorylation. *Nature Reviews Molecular cell biology*, 8(7):530–541, 2007.

[124] Rudolph Van Der Merwe. *Sigma-point Kalman filters for probabilistic inference in dynamic state-space models*. PhD thesis, Oregon Health & Science University, 2004.

[125] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.

[126] José MG Vilar, Hao Yuan Kueh, Naama Barkai, and Stanislas Leibler. Mechanisms of noise-resistance in genetic oscillators. *Proceedings of the National Academy of Sciences*, 99(9):5988–5992, 2002.

[127] Christian Waltermann and Edda Klipp. Information theory based approaches to cellular signaling. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1810(10):924–932, 2011.

[128] Jin Wang, Xiliang Zheng, Yongliang Yang, Dale Drueckhammer, Wei Yang, Gennardy Verkhivker, and Erkang Wang. Quantifying intrinsic specificity: A potential complement to affinity in drug screening. *Physical Review Letters*, 99(19):198101, 2007.

[129] Hans V Westerhoff, Albert K Groen, and Ronald JA Wanders. Modern theories of metabolic control and their applications. *Bioscience Reports*, 4(1):1–22, 1984.

[130] Wikipedia. Simulated annealing - wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Simulated_annealing&oldid=711274215`, 2016. [Online; accessed 22-March-2016].

[131] Darren J Wilkinson. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics*, 10(2):122–133, 2009.

[132] Paul L Williams and Randall D Beer. Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515*, 2010.

[133] Elizabeth Skubak Wolf and David F Anderson. A finite difference method for estimating second order parameter sensitivities of discrete stochastic chemical reaction networks. *The Journal of Chemical Physics*, 137(22):224112, 2012.

[134] Zhiqiang Yan and Jin Wang. Specificity quantification of biomolecular recognition and its implication for drug discovery. *Scientific Reports*, 2, 2012.

[135] Y Zheng and A Rundell. Comparative study of parameter sensitivity analyses of the tcr-activated erk-mapk signalling pathway. *IEE Proceedings-Systems Biology*, 153(4):201–211, 2006.

[136] Zhike Zi. Sensitivity analysis approaches applied to systems biology models. *Systems Biology, IET*, 5(6):336–346, 2011.

[137] Zhike Zi, Yanan Zheng, Ann E Rundell, and Edda Klipp. Sbml-sat: a systems biology markup language (sbml) based sensitivity analysis tool. *BMC Bioinformatics*, 9(1):1, 2008.