Inferring Rate Functions for Stochastic Models of Biological Processes

Jeremy D'Silva, Andreas Hilfinger, Nobuaki Masaki, Zhehao Zhang

Abstract

Given a markov chain modeling chemical interactions in cells whereby the birth and death rates are specified, we can obtain a stationary state distribution representing the abundance of molecules in the cell. However, we are interested in solving the inverse problem of identifying the rates from biological data representing stationary distributions. Specifically, we choose to solve for the kinetic order (n) when we specify mass action as the birth rate of one of our variables, or the Hill coefficient (n) and half-max (K) when we specify the Hill function as the birth rate. We achieve this by evaluating the 2nd and 3rd degree invariants (variance and skewness invariants) based on the data at different n and K values and numerically solving for these parameters.

I. Introduction

A. Context

Inferring interactions in biochemical reaction networks is an area of active interest. Every signal in cellular biology is transmitted through a network of interacting proteins, RNA molecules, and DNA. With the advent of genomics, transcriptomics, and proteomics, it became possible to measure these biological molecules in cells; this data is used to draw correlations and predict interactions. However, there is very limited kinetic information contained in these static measurements: there are many different interpretations of two proteins having high expression in the same sample. With the advent of single-cell technologies, we can obtain distributions of biological molecules in cells, and with multiplexing, we can obtain joint distributions, which allow us to measure the covariance between molecules of interest at the single-cell level. Although most single-cell measurements are static, these static measurements give us the stationary joint distribution of molecules in cells, which contains information about the dynamic interaction between molecules. Broadly, our goal is to exploit the dynamic information hidden in stationary joint distributions to infer which molecules interact dynamically.

More specifically, biologists can obtain joint (stationary) distributions of biochemical species in cells from multiplexed single-cell analyses such as iterated immunofluorescence, imaging mass spectrometry, or single-molecule FISH. With an assumption of ergodicity, one can get the stationary distribution from a "static" data source, like the ones listed here. We want to infer regulatory motifs from these stationary distributions. In this paper, we take a reductionist approach: although our final objective is to study larger systems, we first focus on two and three component models to characterize the applicability of our method. We then test this on a larger network to see if our method can be applied more generally.

B. Background

In a 2016 paper, Hilfinger and co-authors obtained invariant equations for birth-death processes [1]. Let $\mathbf{x} = (x_1, \cdots, x_n)$ be a list of species, and let k be the index for a set of reactions with rates $r_k(\mathbf{x})$, that change the quantities of $\mathbf{x} = (x_1, \cdots, x_n)$ by $\delta_k = (\delta_{k,1}, \cdots, \delta_{k,n})$.

$$\mathbf{x} \xrightarrow{r_k(\mathbf{x})} \mathbf{x} + \delta_k \tag{1}$$

We denote the birth flux of x_i by $R^+_{x_i}$ and the death flux of x_i by $R^-_{x_i}$, which are defined by

$$R^+_i(\mathbf{x}) = R^+_{x_i} = \sum_{\substack{k \\ \delta_{k,i}>0}} \delta_{k,i} r_k(\mathbf{x}) \qquad R^-_i(\mathbf{x}) = R^-_{x_i} = \sum_{\substack{k \\ \delta_{k,i}<0}} |\delta_{k,i}| r_k(\mathbf{x}) \tag{2}$$

Let $\langle s_i \rangle$ be the average step sizes of i^{th} component of the system. Hilfinger and colleagues showed that for each component x_i ($1 \leq i \leq n$),

$$\frac{\langle s_i \rangle}{\langle x_i \rangle} = \frac{\mathrm{Cov}(x_i,\ R^-_{x_i})}{\langle x_i \rangle \left\langle\ R^\pm_{x_i} \right\rangle} - \frac{\mathrm{Cov}(x_i,\ R^+_{x_i})}{\langle x_i \rangle \left\langle\ R^\pm_{x_i} \right\rangle} \tag{3}$$

We can specialize these results to a two-component model of the form in Equation 4. If species x_1 regulates the birth rate of x_2 by a specified function $R^+_{x_2} = f(x_1)$, then the joint distribution of x_1 and x_2 must satisfy Equation 3. Hence, if the joint distribution of x_1 and x_2 does not satisfy the invariant relations, then $R^+_{x_2} \neq f(x_1)$. This invariant provides a way to rigorously reject hypotheses about $R^+_{x_2}$.

$$\begin{aligned} x_1 &\xrightarrow{\lambda_1} x_1 + 1 \qquad & x_2 &\xrightarrow{f(x_1)} x_2 + 1 \\ x_1 &\xrightarrow{\beta_1 x_1} x_1 - 1 \qquad & x_2 &\xrightarrow{\beta_2 x_2} x_2 - 1 \end{aligned} \tag{4}$$

In this paper, we extend the approach of Hilfinger and co-authors by using invariants as a technique for parameter identification. Specifically, we assume the functional form of $R^+_{x_2}$, without assuming all parameters. Then, we simulate the distribution using at least 10^8 reactions in each variable as a "good distribution" and test whether various values of those parameters satisfy Equation 3. If the equation is not satisfied, then those parameters can be rejected. Ideally, we could use the invariant equation to determine the correct parameters: that is, if the equation is satisfied for a unique set of parameters.

We do this for two functional forms, corresponding to two reaction kinetics models: the mass-action model, given by $R^+_{x_2} = \lambda_2 x_1^n$ and the Hill function model, $R^+_{x_2} = \lambda_2 \frac{x_1^n}{x_1^n + K^n}$. The invariants do not depend explicitly on λ_i, but our goal is to use them to reject all values of n (resp. (n, K)) except for the correct parameter value (resp. values). In mass-action kinetics, n corresponds to the reaction order; commonly, this is the number of molecules of x_1 involved in the single-step reaction. In Hill kinetics, n is called the Hill coefficient, or cooperativity coefficient; it controls the "sharpness" of the response function with respect to the concentration of x_1. The parameter K represents the dissociation constant for the interaction that facilitates the reaction.

Existing work on biochemical stochastic models involve writing down the model, fitting the parameters, and comparing the distribution with known data. Numerous methods have been used in the parameter fitting process. For example, Pedraza and van Oudenaarden (2005) [2] uses moment closure schemes to approximate Hill function interactions. Similarly, equality or differential equations on moments that typically involves time series information or snapshot of expressions have been used by So, et al. (2011) [3]. Bayesian methods that sample from parameters space and update according to some distance metric is also used by Liepe et al. (2014) [4]. Our approach focuses on the covariance between pairwise expressions instead of the whole joint distribution. We do not require any temporal information. Even though we need to specify the analytical formula of the birth rate, the formula can in general be any scalar multiple of itself. We are also able to leave upstream and downstream variables unspecified and focus on

pairwise relationships in the larger network. This may allow us to easily generalize our method to large networks.

C. Summary

In Section II, we derive an invariant of skewness from the joint distribution of two variables, using a similar approach to Hilfinger et. al. In Section III, we present our method for identifying n (reaction order) in a two component model with mass-action kinetics using the 2nd invariant. Using simulations, we show that this method succeeds across parameter space and reaction orders when the true distribution is known and is relatively robust to sampling error. In section IV, we examine how we can extend this method to three component models with linear network topology and mass-action kinetics: if x_0 influences the birth rate of x_1, and x_1 influences the birth rate of x_2, we recover the correct interactions, although in some cases, we also recover spurious interactions (e.g. x_0 influencing x_2 directly). We discuss various techniques to reject the spurious interactions. In Section V, we use the second and third degree invariants to infer the correct n and K (Hill coefficient and apparent dissociation constant, respectively) for two component models where x_1 influences the birth rate of x_2 through a Hill function. In Section VI, we discuss the limitations of the method, future work, and connections to biological data.

II. Obtaining a Skewness Invariant

In this section, x is a variable governed by a birth-death process with reactions $r_k(x, y_i)$. We obtain an invariant by setting the derivative of the skewness of x to 0: at stationarity, the distribution of x must satisfy this invariant.

Lemma II.1. *Let x be a variable governed by a birth-death process with birth flux R_x^+ and death flux R_x^-. The skewness of the distribution of x is given by:*

$$\begin{aligned}\frac{d\left\langle (x-\langle x\rangle)^3\right\rangle}{dt} &= 3\mathit{Cov}\left(x^2,\ R_x^+ -\ R_x^-\right) + \left\langle \sum_k \delta_k^3 r_k(x)\right\rangle \\ &\quad - 6\left\langle x\right\rangle \mathit{Cov}\left(x,\ R_x^+ -\ R_x^-\right) + 3\mathit{Cov}\left(x, \sum_k \delta_k^2 r_k(x)\right)\end{aligned} \tag{5}$$

Proof. See Appendix 1. □

Corollary II.1. *At stationarity, by setting Equation 5 equal to zero, we obtain the following invariant:*

$$\begin{aligned}0 &= 3\mathit{Cov}\left(x^2,\ R_x^+ -\ R_x^-\right) + \left\langle \sum_k \delta_k^3 r_k(x)\right\rangle \\ &\quad - 6\left\langle x\right\rangle \mathit{Cov}\left(x,\ R_x^+ -\ R_x^-\right) + 3\mathit{Cov}\left(x, \sum_k \delta_k^2 r_k(x)\right)\end{aligned} \tag{6}$$

Proof. See Appendix 1. □

Corollary II.2. *With hypotheses as in Corollary II.1, and additionally, if $|\delta_k| = 1\ \forall k$, then the variance invariant from Equation 6 is given by*

$$\frac{\mathit{Cov}\left(x,\ R_x^+ -\ R_x^-\right)}{\left\langle x\right\rangle \left\langle\ R_{x_2}^{\pm}\right\rangle} = \frac{\mathit{Cov}\left(x^2,\ R_x^+ -\ R_x^-\right) + \mathit{Cov}\left(x,\ R_x^+ +\ R_x^-\right)}{2\left\langle x\right\rangle \left(\left\langle x\right\rangle \left\langle\ R_{x_2}^{\pm}\right\rangle\right)} \tag{7}$$

Proof. See Appendix 1. □

III. Inferring reaction order for mass-action kinetics in 2-component systems

In this section, we use the second invariant (Equation 3) to identify the correct reaction order n for a mass-action kinetics model, given the stationary probability distribution generated by the model. We show that this method works across parameter space, that it rejects the incorrect interaction in the two-component model, and that even with a small number of samples from the stationary distribution, we can obtain the invariant with high confidence.

A. Model Structure

In this section, we consider a two component system with mass action kinetics. The variable x_1 is governed by a Poisson process; the variable x_2 is made with a birth rate that depends on x_1. In our mass-action kinetics model, x_2 has birth rate $\lambda_2 x_1^n$. Both death rates are exponential. The reactions in the model are as follows:

$$\begin{aligned} x_1 &\xrightarrow{\lambda_1} x_1 + 1 & \qquad x_2 &\xrightarrow{\lambda_2 x_1^n} x_2 + 1 \\ x_1 &\xrightarrow{\beta_1 x_1} x_1 - 1 & \qquad x_2 &\xrightarrow{\beta_2 x_2} x_2 - 1 \end{aligned} \tag{8}$$

where, with notation as in Section I, $R^+_{x_1} = \lambda_1$, $R^-_{x_1} = \beta_1 x_1$, $R^+_{x_2} = \lambda_2 x_1^n$, $R^-_{x_2} = \beta_2 x_2$. By rescaling the other parameters (dividing through by β_2), we can set $\beta_2 = 1$.

B. Invariant to infer n

Given the joint distribution of x_1, x_2, we want to infer the parameter n (reaction order), from the data. Our strategy is to substitute $\mathrm{P}(x_1, x_2)$, the joint distribution of (x_1, x_2), into the invariant from Equation 3, and evaluate the invariant at different values of n to identify the true value used to generate the data, which we denote $n^\star$.

First, we specialize Equation 3 by substituting in the birth and death fluxes of x_1 and x_2 into the equation. We note that s_i, the average step size of x_i, is 1 because $|\delta| = 1$ for all reactions.

Applying Equation 3 to x_2, we get the following equation

$$\frac{\mathrm{Var}\,(x_2)}{\langle x_2 \rangle^2} = \frac{1}{\langle x_2 \rangle} + \frac{\mathrm{Cov}\,(x_2, x_1^n)}{\langle x_2 \rangle \langle x_1^n \rangle}$$

Let $\eta_{xy} := \frac{\mathrm{Cov}(x,y)}{\langle x \rangle \langle y \rangle}$, then we can rewrite the equation as

$$\eta_{x_2,x_2} = \frac{1}{\langle x_2 \rangle} + \eta_{x_2,x_1^n} \tag{9}$$

Note that by linearity of expectation, the λ_2 coefficients cancel. This invariant can be computed from an observed joint distribution for various values of n. We can then define a relative error term comparing the two sides of the invariant equation, and compute the error for various values of n. Assuming the observed joint distribution is generated from a process of the form in Equation 8, this invariant holds for the true value of n. Hence, the error should be zero at the true value of n.

We define the error term as

$$\mathrm{Err}_2(n) = 2\frac{\frac{1}{\langle x_2 \rangle} + \eta_{x_2,x_1^n} - \eta_{x_2,x_2}}{\eta_{x_2,x_2} + \frac{1}{\langle x_2 \rangle} + \eta_{x_2,x_1^n}} \tag{10}$$

If n^* is the value of n that generated the joint stationary distribution of x_1 and x_2 via the process in Equation 8, then by the process used to obtain the invariant, it is guaranteed that $\text{Err}_2(n^*) = 0$.

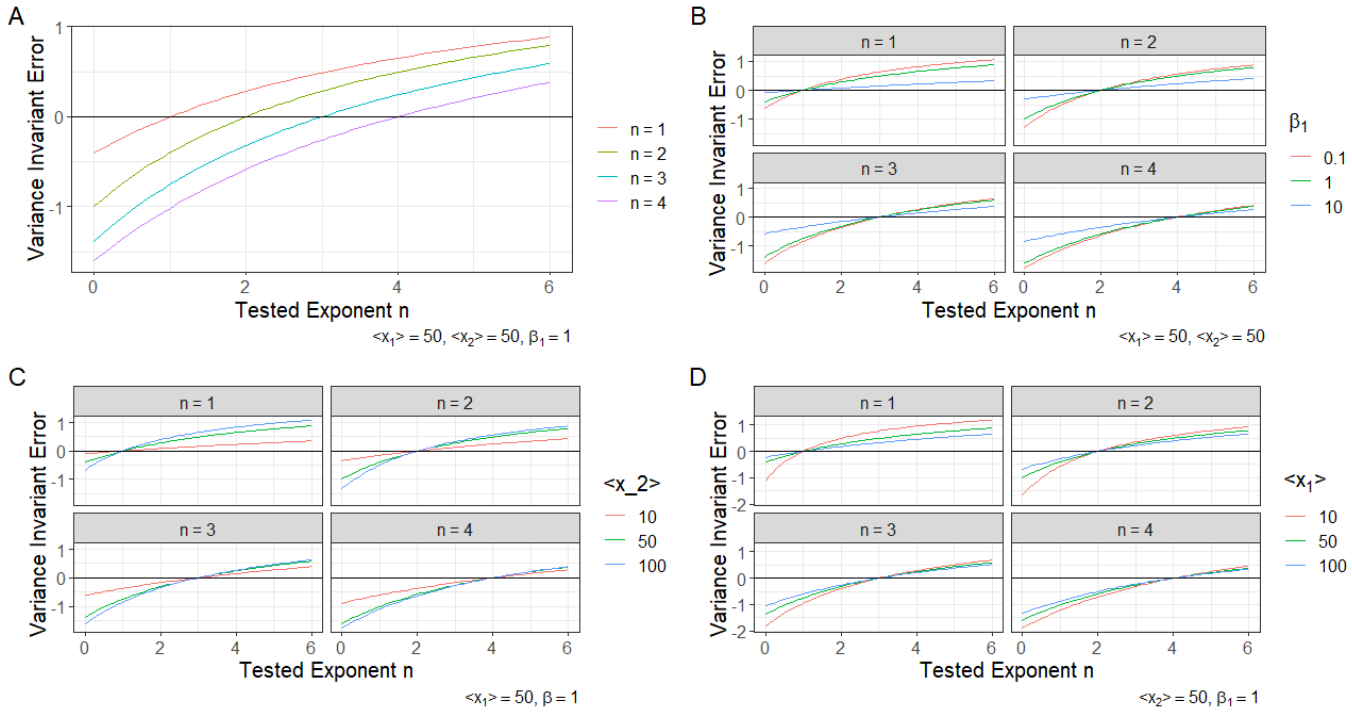

Fig. 1: $Err_2(n)$ Plotted across different n^*, β_1, $\langle x_1 \rangle$, and $\langle x_2 \rangle$

In Figure 1, we plot $Err_2(n)$ calculated across different stationary distributions generated from different parameter values. From Panel A, Figure 1, where we are keeping $\langle x_1 \rangle$, $\langle x_2 \rangle$, and β_1 fixed, we see that our method works when our data is generated for different n^*. From Panel B, where we are keeping $\langle x_1 \rangle$ and $\langle x_2 \rangle$ fixed, we see that the $Err_2(n)$ line becomes more steep as the value of β_1 decreases. Intuitively, this can be explained by the fact that the lifetime of x_1 is longer when β_1 is low, meaning that x_2 has more time to react to changes in x_1. From Panel C, where we are keeping $\langle x_1 \rangle$ and β_1 fixed, we see that the $Err_2(n)$ line becomes more steep as the mean of x_2 increases. This is expected, since when we increase x_2, $\frac{1}{\langle x_2 \rangle}$ gets smaller and a larger portion of variability is explained by the covariance term. Finally, in Panel D, we keep $\langle x_2 \rangle$ and β_1 fixed and plot changes in $Err_2(n)$ from $\langle x_1 \rangle$. We see that lower $\langle x_1 \rangle$ results in a steeper slope because x_1 follows a Possion distribution and the variability of x_1 (η_{x_1,x_1}) increases as mean of x_1 decreases

C. *Comments About* $n = 0$

When $n = 0$, both x_1 and x_2 follows independent Poisson processes. We know that $\frac{1}{\langle x_2 \rangle} = \frac{\text{Var}(x_2)}{\langle x_2 \rangle^2}$ is satisfied. However, since x_2 and x_1 are generally independent in this system, the covariance between x_2 and x_1^n should be 0 for all n theoretically. Even though numerically this could fluctuate a little bit, we would get a line of 0 or very close to 0 (See Figure 2). This is a special case for non-negative integer n.

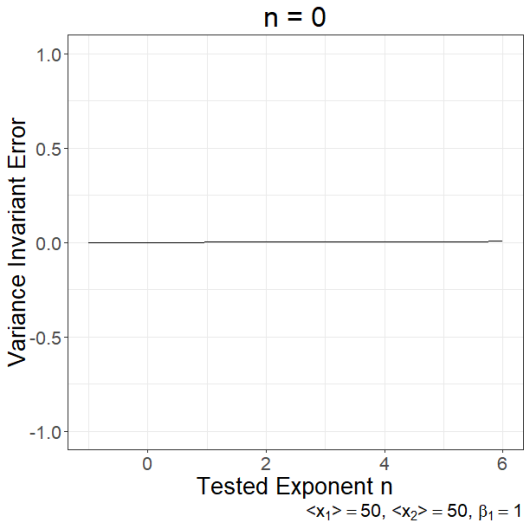

Fig. 2: When $n = 0$ in the 2 component Mass Action Model, x_1 and x_2 are independent

D. Rejecting $x_2 \to x_1$

We also want to identify the true pair of interactions in this system, that is to tell whether x_1 regulates x_2 or x_2 regulates x_1. In this case, if our model is x_1 regulates x_2 and we test on x_2 regulates x_1, we get a clear intersection at $n = 0$ (See Figure 3). This is true since indeed x_2 follows a Poisson process and x_2 is indeed made by rate $\lambda_1 x_1^0$. It is not shallow since the covariance between x_1^n and x_2 is not zero in general and even you raise x_1 to some power, the covariance term is still significant.

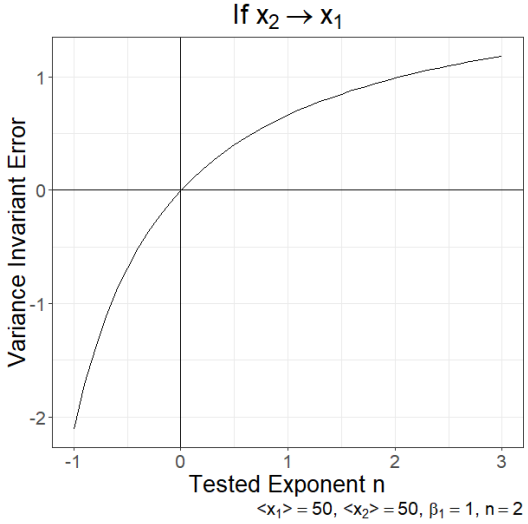

Fig. 3: Assume x_2 regulates x_1 in the model

E. Sampling

In the previous figures, we computed the error term from Equation 10 using the stationary distribution. However, in practical settings, the true stationary distribution is generally unknown:

rather, experimental data generally corresponds to sampling from an unknown, underlying stationary distribution. Using inverse sampling, we sampled from the stationary distribution (we set N, the number of samples, to be 100, 500, 1000, and 5000). We performed this sampling 1000 times for each N. For each sample, we computed the estimated error term across n; that is, we substituted the sample distribution into Equation 10 and computed it across n. For each N, we performed this process 1000 times to get 1000 estimates of the error term, and then we plotted the mean and 95% credible interval of these estimates for each N in Figure 4. As shown in the figure, the mean of the invariant error passes through the true value of n at 0, and for the number of samples $N \geq 500$, the credible interval is quite narrow around the true value of n. In a biological setting, $N = 500$ corresponds to single-cell measurements of two molecules in 500 cells, which is very plausible using fluorescence microscopy or single-cell RNA sequencing, depending on the molecules of interest.

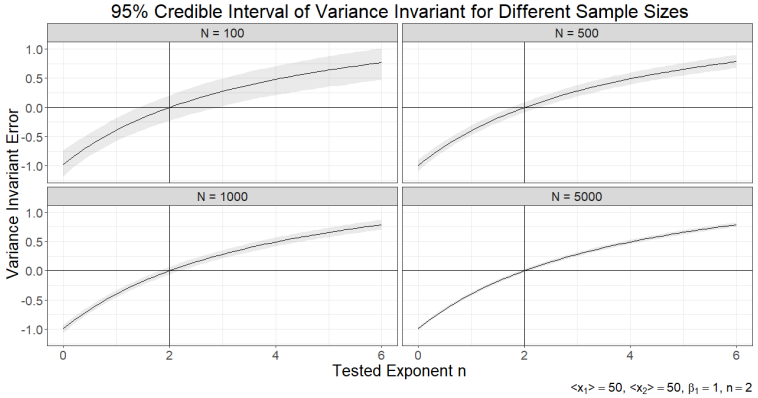

Fig. 4: 95% credible interval from different sample sizes

IV. MA 3 Components

A. Model Structure

In order to generalize this method into large networks, it is natural to ask whether it works for a chain of components in which each component regulates the birth rate of the next component by mass action kinetics. As a first response to this question, we consider a three-component system where the birth rate of x_2 depends on x_1 and the birth rate of x_3 depends on x_2. The reactions in the system are as follows:

$$\begin{aligned} x_1 &\xrightarrow{\lambda_1} x_1 + 1 & x_2 &\xrightarrow{\lambda_2 x_1^m} x_2 + 1 & x_3 &\xrightarrow{\lambda_3 x_2^n} x_3 + 1 \\ x_1 &\xrightarrow{\beta_1 x_1} x_1 - 1 & x_2 &\xrightarrow{\beta_2 x_2} x_2 - 1 & x_3 &\xrightarrow{\beta_3 x_3} x_3 - 1 \end{aligned} \tag{11}$$

B. 2nd invariant identifies spurious interactions

In the three-component case, there are two sorts of spurious interactions we might expect to identify using the invariant approach: x_3 influences x_2 or x_2 influences x_1 (interactions that reverse a correct interaction in the network) and x_1 influences the birth rate of x_3 (the "composition" of two interactions .

In the two-component model in Section III, the second invariant rejects the "reverse" spurious interaction. In particular, the variance invariant suggests that x_1 is made with order x_2^0 (i.e. at a constant rate), which is consistent with the birth rate of x_1. Intuitively, this occurs because the birth rate of x_1 is constant, and the invariant is insensitive to λ_1; Equation 3 simply witnesses that the birth rate of x_1 is constant. However, when we introduce a third component into the system, the birth rate of x_2 is no longer constant, so there might be a nonzero value of k for which $R^+_{x_2} = \lambda x_3^k$ satisfies Equation 3.

As for the composition error, it is more clear how the variance invariant might identify a spurious interaction: if x_2 has a much faster timescale than x_1, then it will reflect the amount of x_1 very faithfully; as such, it might be hard to distinguish without temporal information whether x_3 is made by x_1 or x_2. We are not so worried about this scenario but if there is a difference between x_1 and x_2 and we still get a 0 relative error from x_1 and x_3, we want to identify the correct pair.

We thought that these two problems might be ameliorated if we obtained another invariant from a higher moment of x_i, using the same approach described in the introduction. In particular, for a correct interaction, both Equation 3 and the new invariant must be zero, but for an incorrect interaction, they might not be zero at the same n, which would allow us to distinguish correct from incorrect interactions. This motivates the introduction of the skewness invariant.

C. 3rd Invariant

From Equation 7, we get the following invariant on x_2, x_3 using the 3-component system.

$$\frac{\text{Cov}\left(x_3, x_2^n\right)}{\langle x_3\rangle \langle x_2^n\rangle} - \frac{\text{Var}\left(x_3\right)}{\langle x_3\rangle^2} = \frac{\text{Cov}\left(x_3, x_2^n\right) + \text{Cov}\left(x_3, x_2^n\right)}{2\langle x_3\rangle^2 \langle x_2^n\rangle} + \frac{\text{Var}\left(x_3\right) - \text{Cov}\left(x_3^2, x_3\right)}{2\langle x_3\rangle^3} \tag{12}$$

D. Figure, Talk About Results

The time scale β_2 relative to the other time scales is very important in impacting the behavior of the method. We also need to be careful that even the intersections are at the same point, we can only conclude that the two components seems like a mass action up to "third moment". Theoretically one can use higher moments (induction formula in appendix) to examine this result. However each higher moment is "shallower" and less resistant towards the measurement error. Also, since we are only using pairwise relation here, this result could easily generalize to any chain of non-feedback mass action kinetics by induction. We need to be more careful if there exists more complex network topology in the system.

V. Hill 2 Components

In this section, we study the properties of these invariants for a higher-dimensional problem: inferring the parameters of Hill-type rate functions, that is, functions of the form

$$R^+_{x_2} = \lambda_2 \frac{x_1^n}{x_1^n + K^n} \tag{13}$$

Hill equations arise from applying the quantitative steady-state approximation to a model of enzymatic reactions; they are standard in biochemistry for modelling protein interaction kinetics. The parameter n is the Hill coefficient, a measure of cooperativity between subunits in complex formation; the parameter K represents the dissociation constant

A. *Model - Show 2nd and 3rd Invariants*

In this section, we use a similar model construction to Section III but change the birth rate of x_2 to be $\lambda_2 \frac{x_1^n}{x_1^n+K^n}$. The reaction in the models are as follows:

$$\begin{aligned} x_1 &\xrightarrow{\lambda_1} x_1+1 & x_2 &\xrightarrow{\lambda_2 \frac{x_1^n}{x_1^n+K^n}} x_2+1 \\ x_1 &\xrightarrow{\beta_1 x_1} x_1-1 & x_2 &\xrightarrow{\beta_2 x_2} x_2-1 \end{aligned} \tag{14}$$

As before, we will rescale the invariants so that they do not depend on λ_1. This problem is higher-dimensional: rather than inferring the reaction order n, as in the mass-action case, we try to infer both n and K from the joint distribution of x_1 and x_2.

Applying Equation 3, we get the following equation.

$$\frac{1}{\langle x_2 \rangle} = \frac{\mathrm{Var}\,(x_2)}{\langle x_2 \rangle^2} - \frac{\mathrm{Cov}\left(x_2, \frac{x_1^n}{x_1^n+K^n}\right)}{\langle x_2 \rangle \left\langle \frac{x_1^n}{x_1^n+K^n} \right\rangle} \tag{15}$$

As in Section III, we define an error term based on the invariant:

$$\mathrm{Err}_2(n,K) = 2\frac{\frac{1}{\langle x_2 \rangle} + \eta_{x_2,R_{x_2}^+} - \eta_{x_2,x_2}}{\eta_{x_2,x_2} + \frac{1}{\langle x_2 \rangle} + \eta_{x_2,R_{x_2}^+}} \tag{16}$$

where scalars λ_2 and β_2 cancels out using the stationary condition $\lambda_2 \left\langle \frac{x_1^n}{x_1^n+K^n} \right\rangle = \beta_2 \langle x_2 \rangle$. However, since the parameter space now has two dimensions, we would get a 1-dimensional zero stripe just based on Err_2.

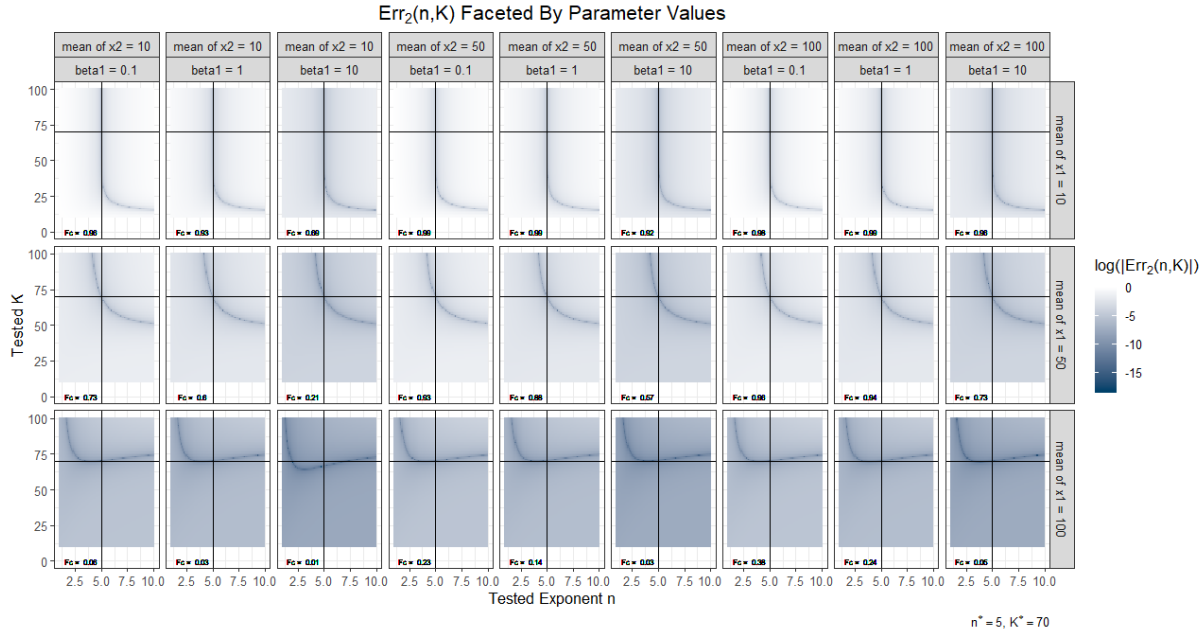

Fig. 5: Err_2 from different parameter settings

As we iterate through different settings, we can see in the plot that if we want a clear stripe for the second invariant, we want $\langle x_1 \rangle$ to be smaller than K, x_1 to be slower than x_2 and $\langle x_2 \rangle$ to be large. These follow from the variability result of the second invariant. We provide a more detailed explanation on this in F_c section and linear approximation section. One important observation is that mean of x_1 is very important in determining how confident we are to get the stripe of zeros. When $\langle x_1 \rangle > K$, the stripe is not even stable anymore as we see in the third row of the plot.

In order to further determine the (n, K) pair, we use the skewness invariant again. Applying Equation 7, we get the following:

$$\frac{\text{Cov}\left(x_2, \frac{x_1^n}{x_1^n+K^n}\right)}{\langle x_2 \rangle \left\langle \frac{x_1^n}{x_1^n+K^n} \right\rangle} - \frac{\text{Var}\left(x_2\right)}{\langle x_2 \rangle^2} = \frac{\text{Cov}\left(x_2, \frac{x_1^n}{x_1^n+K^n}\right) + \text{Cov}\left(x_2, \frac{x_1^n}{x_1^n+K^n}\right)}{2\langle x_2 \rangle^2 \left\langle \frac{x_1^n}{x_1^n+K^n} \right\rangle} + \frac{\text{Var}\left(x_2\right) - \text{Cov}\left(x_2^2, x_2\right)}{2\langle x_2 \rangle^3} \tag{17}$$

Now define

$$\begin{aligned} \text{lhs} :&= \frac{\text{Cov}\left(x_2, \frac{x_1^n}{x_1^n+K^n}\right)}{\langle x_2 \rangle \left\langle \frac{x_1^n}{x_1^n+K^n} \right\rangle} + \frac{\text{Cov}\left(x_2^2, x_2\right)}{2\langle x_2 \rangle^3} \\ \text{rhs} :&= \frac{\text{Cov}\left(x_2, \frac{x_1^n}{x_1^n+K^n}\right) + \text{Cov}\left(x_2, \frac{x_1^n}{x_1^n+K^n}\right)}{2\langle x_2 \rangle^2 \left\langle \frac{x_1^n}{x_1^n+K^n} \right\rangle} + \frac{\text{Var}\left(x_2\right)}{2\langle x_2 \rangle^3} + \frac{\text{Var}\left(x_2\right)}{\langle x_2 \rangle^2} \end{aligned} \tag{18}$$

and

$$\text{Err}_3(n, K) = \frac{2(lhs - rhs)}{lhs + rhs} \tag{19}$$

This gives another equation on (n, K). The skewness behavior follows the similar pattern as the variance invariant in Figure 9 (See appendix). If we set Err_2 and Err_3 both to be zero, we would have two equations and two unknowns. This provides a theoretical solution to infer n and K. The following plot is when we plot the zero stripes of both invariants:

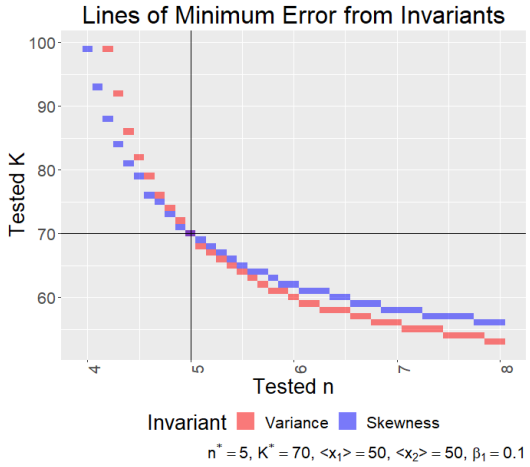

Fig. 6: Minimum Err_2 and Err_3 values from invariants

As we can see, the two lines intersect at the true value of (n, K). This is expected. However, the two intersections are not transversal. This becomes a problem when we have sampling error or measurement error in our data. The following plot is when we sample 1000 points from our good distribution and repeat 30 times:

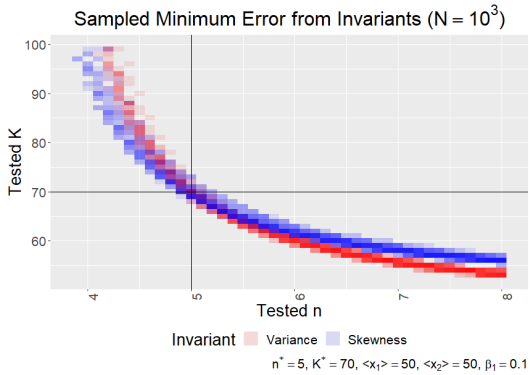

Fig. 7: Minimum Err_2 and Err_3 values from invariants in 30 times of sampling

Notice that from multiple sampling, we get an intersection area where intersections might happen due to sampling error. If we plot all the (n, K) pairs that passes through all of the intersections from our 30 trials and re-scale them (λ_2 might be different in these cases) so that they all pass through the point $(\langle y \rangle, \frac{\langle y \rangle^n}{\langle y \rangle^n + K^n})$ for true value of (n, K), we notice that they are very similar in the regime of y distribution (Figure 8). This indicates the "degenerate" case where we are not so certain about the exact parameters but we can still get the shape of Hill function in the regime where y is distributed. This is an identifiablility problem of the Hill function rather than the vulnerablity of our method.

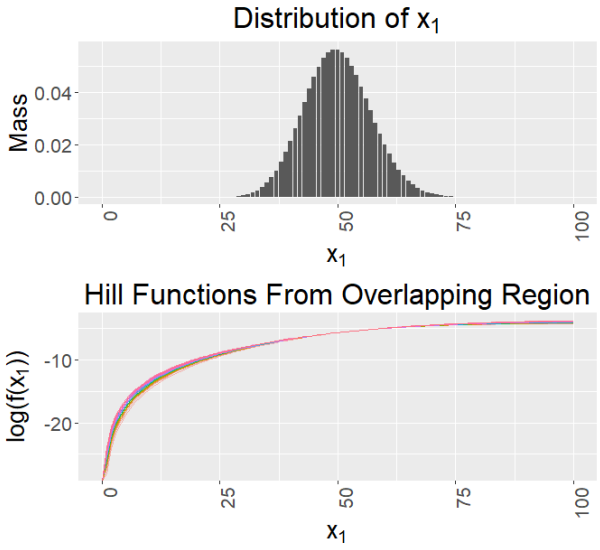

Fig. 8: Rescaled Hill funcions in intersection regions

B. Introduce F_c

In order to better understand how different settings change our result, we introduce a new statistics F_c by

$$F_c := \frac{\eta_{x_2 R_{x_2}^+}}{\frac{1}{\langle x_2 \rangle} + \eta_{x_2, R_{x_2}^+}} \tag{20}$$

Remember that $\frac{1}{\langle x_2 \rangle}$ is the intrinsic noise of x_2 and $\eta_{x_2,R_{x_2}^+}$ is the extrinsic noise of x_2 (caused by x_1) and the sum is the variability of x_2. So F_c denotes the portion of variability of x_2 explained by extrinsic noise. In Figure 5, we printed the F_c value for each subplot. In principle, as long as K larger or approximately x_2, a larger F_c would give a better intersection since it forces $\langle x_2 \rangle$ to be small and covariance to be large, we are then more confident on both variance and skewness curves. This statistics is helpful for us since in appendix we can see F_c has a monotonic relationship with Bhattacharyya distance and this allows us to get insight on how good our method works just based on the given distributions.

C. Linear Noise Approximation and Intuition

Lemma V.1. *For the model in Equation 14, by applying the linear noise approximation to the system, we obtain the following result on variability of x_2:*

$$\eta_{x_2,x_2} = \frac{1}{\langle x_2 \rangle} + \frac{\beta_1}{\beta_1 + \beta_2} \left(\frac{nK^n}{K^n + \langle x_1 \rangle^n} \right)^2 \eta_{x_1,x_1} \tag{21}$$

Proof. See Appendix 1. □

The linear noise approximation provides a good way to think about extrinsic noise across different parameter settings. In our model, $\eta_{x_1,x_1} = \frac{1}{\langle x_1 \rangle}$ as x_1 follows a Poisson process. Notice that the term $\frac{nK^n}{K^n + \langle x_1 \rangle^n}$ looks like a reflected version of hill function around half-max. So n is the asymptotic horizontal line when $\langle x_1 \rangle \to 0$ and K is the half-max. In order to have large F_c, we prefer a smaller $\langle x_1 \rangle$, larger N, larger K and relative large β_1 compared to β_2. We also want the $\langle x_2 \rangle$ to be large for intrinsic noise to be small.

D. Discussion

As we see in linear noise approximation section and plot of Error2/Error3 in different setting (see Appendix), larger n and K are helpful for us to determine the variance and skewness invariant stripe. However, to get the correct (n, K) pair from the distribution, we also want the intersection to be as transversal as possible. This is related to the shape of the hill function. In order to use the variability of x_1 to capture the dynamics of x_2, we want the birth rate of x_2 to fluctuate as much as possible. Since the hill function has a Sigmoid behavior, we want x_1 to be able to fluctuate around the half-max of the hill function, that is K. So the best scenario that we can confidently determine both n and K is when $\langle x_1 \rangle$ is slightly smaller than K and large n. When K is much larger than $\langle x_1 \rangle$, we can usually determine the right n value but it is hard to tell what K is since the intersection happens at the almost vertical part of the stripe. When K is small, in general the system is not very stable and we need to take large sample sizes in order to determine K.

VI. Future Questions

1) Compare this approach to standard parameter fitting techniques that people use (e.g. approximate Bayesian computation). In particular, look at the distribution of estimates you get for n, (n, K), and for the rate parameters (e.g. maybe you get a wide range of n but the time scales are very different and you can witness that somehow?)
2) What happens if the system is not at stationarity? First, is there a way to witness that from static data? Second, can we account for that in the invariant? (e.g. estimate the value of $\frac{d}{dt}\text{Var}(x_i)$ and keep it in the invariant.)
3) Persistent measurement error of x (e.g. undercounting by 3 molecules) or probabilistic measurement of x (e.g. measure x 70% of the time).
4) $|\delta| \neq 1$. We'd need to study the third invariant more carefully to work on this case—we've only really worked with the third invariant when $|\delta| = 1$.
5) Alternate kinetic schemes. In the two component case, it would be really nice to do something like $R^+_{x_2} = \lambda_2 x_1^m x_2^n$ (auto-catalysis?), $R^+_{x_1} = \lambda_1 x_2^m$ (classic feedback).
6) 5 component system with a mix of interactions that we expect to recover and interactions that will be more complicated (e.g. timescale spearation in the "wrong" direction). Maybe we could expand more on timescales in the preceding sections too: take the limit as β_1 gets enormous
7) Go through some of the papers we have in the google drive about "biological network motifs", make a larger list of network motifs that are interesting to study. Start with a few simple examples (phosphorylation cascades, operons, etc.).

VII. Conclusion

In summation, invariants of the stationary distribution for stochastic processes in networks not only offer the potential for rigorous rejection of models, but also for parameter estimation. These methods show great promise for recovering exact parameter values in mass-action kinetics, and for recovering rate functions, for rate functions with underlying identifiability issues.

VIII. Appendix

A. Higher Moments

In general, for birth-death processes of x_2 with all step sizes one and simple dependence birth rate $\lambda_2 f(x_1)$ and death rate $\beta_2 x_2$, we have

$$\frac{d\left\langle x^k\right\rangle}{dt} = \beta_2\left\langle (x_2-1)^k x_2\right\rangle - \beta_2\left\langle x_2^{k+1}\right\rangle + \lambda_2\left\langle (x_2+1)^k f(x_1)\right\rangle - \lambda_2\left\langle x_2^k f(x_1)\right\rangle \tag{22}$$

for any moments $k \in \mathbb{Z}^+$.

B. Skewness Invariant

Here, we prove Lemma II.1.

Proof. Next, we compute the time derivative of the the following quantity using a similar approach to the derivation of variance equation [1]. First, we rewrite skewness as:

$$\begin{aligned}\left\langle (x-\langle x\rangle)^3\right\rangle &= \left\langle x^3 - 3x^2\langle x\rangle + 3x\langle x\rangle^2 - \langle x\rangle^3\right\rangle \\ &= \left\langle x^3\right\rangle - 3\langle x\rangle\left(\left\langle x^2\right\rangle - \langle x\rangle^2\right) - \langle x\rangle^3 \\ &= \left\langle x^3\right\rangle - 3\langle x\rangle\left(\text{Var}\left(x\right)\right) - \langle x\rangle^3\end{aligned}$$

Next, we differentiate this expression:

$$\frac{d}{dt}\left\langle (x-\langle x\rangle)^3\right\rangle = \frac{d}{dt}\left\langle x^3\right\rangle - 3\left(\frac{d\langle x\rangle}{dt}\text{Var}(x) + \langle x\rangle \frac{d\text{Var}(x)}{dt}\right) - 3\langle x\rangle^2 \frac{d\langle x\rangle}{dt}$$

Now, we begin to simplify each term. For one reaction, $\frac{d}{dt}\langle x^3\rangle$ can be expressed as:

$$\begin{aligned}
\frac{d\langle x^3\rangle}{dt} &= \sum_x x^3 \frac{dP(x,t)}{dt} \\
&= \sum_x x^3(-r(x))P(x) + \sum_x x^3 r(x-\delta)P(x-\delta) \\
&= \sum_x (-x^3)r(x)P(x) + \sum_x \left(x^3 + 3x^2\delta + 3x\delta^2 + \delta^3\right) r(x)P(x) \\
&= \left\langle 3x^2\delta r(x)\right\rangle + \left\langle 3x\delta^2 r(x)\right\rangle + \left\langle \delta^3 r(x)\right\rangle.
\end{aligned}$$

For multiple reactions, this generalizes in a straightforward manner (switching summations) to

$$\frac{d\langle x^3\rangle}{dt} = \sum_k \left\langle 3x^2\delta_k r_k(x)\right\rangle + \left\langle 3x\delta_k^2 r_k(x)\right\rangle + \left\langle \delta_k^3 r_k(x)\right\rangle \tag{23}$$

$$= \left\langle 3x^2\left(\ R_x^+ - \ R_x^-\right)\right\rangle + \left\langle 3x\sum_k |\delta_k||\delta_k| r_k(x)\right\rangle + \left\langle \sum_k \delta_k^3 r_k(x)\right\rangle \tag{24}$$

The middle two terms of Equation 23 can be expressed as

$$-3\left(\frac{d\langle x\rangle}{dt}\text{Var}(x) + \langle x\rangle\frac{d\text{Var}(x)}{dt}\right) = -3\left\langle\ R_x^+ - \ R_x^-\right\rangle\left\langle x^2\right\rangle + 3\left\langle\ R_x^+ - \ R_x^-\right\rangle\langle x\rangle^2 - 3\langle x\rangle\frac{d\text{Var}(x)}{dt}.$$

The last term of Equation 23 can be expressed as:

$$-\frac{d}{dt}\langle x\rangle^3 = -3\langle x\rangle^2\frac{d\langle x\rangle}{dt} = -3\left\langle\ R_x^+ - \ R_x^-\right\rangle\langle x\rangle^2.$$

Combining these terms, we get

$$\begin{aligned}
\frac{d}{dt}\left\langle (x-\langle x\rangle)^3\right\rangle &= \left\langle 3x^2\left(\ R_x^+ - \ R_x^-\right)\right\rangle + \left\langle 3x\sum_k |\delta_k||\delta_k| r_k(x)\right\rangle + \left\langle \sum_k \delta_k^3 r_k(x)\right\rangle \\
&\quad - 3\left\langle\ R_x^+ - \ R_x^-\right\rangle\left\langle x^2\right\rangle + 3\left\langle\ R_x^+ - \ R_x^-\right\rangle\langle x\rangle^2 - 3\langle x\rangle\frac{d\text{Var}(x)}{dt} - 3\left\langle\ R_x^+ - \ R_x^-\right\rangle\langle x\rangle^2 \\
&= 3\left\langle x^2\left(\ R_x^+ - \ R_x^-\right)\right\rangle - 3\left\langle\ R_x^+ - \ R_x^-\right\rangle\left\langle x^2\right\rangle \\
&\quad + \left\langle 3x\sum_k |\delta_k||\delta_k| r_k(x)\right\rangle + \left\langle \sum_k \delta_k^3 r_k(x)\right\rangle \\
&\quad - 3\langle x\rangle\frac{d\text{Var}(x)}{dt} \\
&\quad + 3\left\langle\ R_x^+ - \ R_x^-\right\rangle\langle x\rangle^2 - 3\left\langle\ R_x^+ - \ R_x^-\right\rangle\langle x\rangle^2 \\
&= 3\text{Cov}\left(x^2, \left(\ R_x^+ - \ R_x^-\right)\right) + 3\left\langle x\sum_k |\delta_k||\delta_k| r_k(x)\right\rangle \\
&\quad + \left\langle \sum_k \delta_k^3 r_k(x)\right\rangle - 3\langle x\rangle\frac{d\text{Var}(x)}{dt}
\end{aligned}$$

Recall that

$$\frac{d\text{Var}\left(x\right)}{dt} = 2\text{Cov}\left(x,\ R_x^+ -\ R_x^-\right) + \left\langle \sum_k |\delta_k||\delta_k| r_k(x) \right\rangle.$$

Substituting this into the previous equation, we obtain

$$\begin{aligned}\frac{d}{dt}\left\langle (x - \langle x\rangle)^3 \right\rangle = 3\text{Cov}\left(x^2, \left(\ R_x^+ -\ R_x^-\right)\right) + \left\langle \sum_k \delta_k^3 r_k(x) \right\rangle \\ + 3\left\langle x \sum_k |\delta_k||\delta_k| r_k(x) \right\rangle - 3\left\langle x \right\rangle \left(2\text{Cov}\left(x,\ R_x^+ -\ R_x^-\right) + \left\langle \sum_k |\delta_k||\delta_k| r_k(x) \right\rangle \right)\end{aligned} \tag{25}$$

An alternate form of this equation is given by

$$\begin{aligned}\frac{d}{dt}\left\langle (x - \langle x\rangle)^3 \right\rangle = 3\text{Cov}\left(x^2, \left(\ R_x^+ -\ R_x^-\right)\right) + \left\langle \sum_k \delta_k^3 r_k(x) \right\rangle - 6\left\langle x \right\rangle \text{Cov}\left(x,\ R_x^+ -\ R_x^-\right) \\ + 3\text{Cov}\left(x, \sum_k \delta_k^2 r_k(x)\right)\end{aligned} \tag{26}$$

This is, in fact, the expression in Lemma II.1. □

Next, we prove Corollaries II.1 and II.2.

Proof. At stationarity, we can set Equation 26 equal to 0 (Corollary II.1).

$$\begin{aligned}0 = 3\text{Cov}\left(x^2,\ R_x^+ -\ R_x^-\right) + \left\langle \sum_k \delta_k^3 r_k(x) \right\rangle \\ - 6\left\langle x \right\rangle \text{Cov}\left(x,\ R_x^+ -\ R_x^-\right) + 3\text{Cov}\left(x, \sum_k \delta_k^2 r_k(x)\right)\end{aligned}$$

If $|\delta_k| = 1\ \forall k$, then this simplifies to

$$\begin{aligned}0 &= 3\text{Cov}\left(x^2,\ R_x^+ -\ R_x^-\right) + \left\langle \sum_k \delta_k r_k(x) \right\rangle \\ &\quad - 6\left\langle x \right\rangle \text{Cov}\left(x,\ R_x^+ -\ R_x^-\right) + 3\text{Cov}\left(x, \sum_k |\delta_k| r_k(x)\right) \\ &= \text{Cov}\left(x^2,\ R_x^+ -\ R_x^-\right) - 2\left\langle x \right\rangle \text{Cov}\left(x,\ R_x^+ -\ R_x^-\right) + \text{Cov}\left(x,\ R_x^+ +\ R_x^-\right)\end{aligned}$$

(We used the fact that $\langle \sum_k \delta_k r_k(x) \rangle = \langle\ R_x^+ -\ R_x^- \rangle = \frac{d\langle x\rangle}{dt} = 0$ at stationarity). We can rearrange this and divide by $\langle x\rangle \left\langle\ R_{x_2}^{\pm}\right\rangle$ to obtain the equation given in Corollary II.2:

$$\frac{\text{Cov}\left(x,\ R_x^+ -\ R_x^-\right)}{\langle x\rangle \left\langle\ R_{x_2}^{\pm}\right\rangle} = \frac{\text{Cov}\left(x^2,\ R_x^+ -\ R_x^-\right) + \text{Cov}\left(x,\ R_x^+ +\ R_x^-\right)}{2\left\langle x\right\rangle \left(\langle x\rangle \left\langle\ R_{x_2}^{\pm}\right\rangle\right)} \tag{27}$$

Expanding this, we get

$$\frac{\text{Cov}\left(x,\ R_x^+\right)}{\langle x\rangle\left\langle\ R_{x_2}^{\pm}\right\rangle} - \frac{\text{Cov}\left(x,\ R_x^-\right)}{\langle x\rangle\left\langle\ R_{x_2}^{\pm}\right\rangle} \tag{28}$$

$$= \frac{\text{Cov}\left(x^2,\ R_x^+\right)}{2\langle x\rangle\left(\langle x\rangle\left\langle\ R_{x_2}^{\pm}\right\rangle\right)} - \frac{\text{Cov}\left(x^2,\ R_x^-\right)}{2\langle x\rangle\left(\langle x\rangle\left\langle\ R_{x_2}^{\pm}\right\rangle\right)} + \frac{\text{Cov}\left(x,\ R_x^+\right)}{2\langle x\rangle\left(\langle x\rangle\left\langle\ R_{x_2}^{\pm}\right\rangle\right)} + \frac{\text{Cov}\left(x,\ R_x^-\right)}{2\langle x\rangle\left(\langle x\rangle\left\langle\ R_{x_2}^{\pm}\right\rangle\right)} \tag{29}$$

We can also rearrange so that each term has a positive coefficient:

$$\begin{aligned}\frac{\text{Cov}\left(x,\ R_x^+\right)}{\langle x\rangle\left\langle\ R_{x_2}^{\pm}\right\rangle} + \frac{\text{Cov}\left(x^2,\ R_x^-\right)}{2\langle x\rangle\left(\langle x\rangle\left\langle\ R_{x_2}^{\pm}\right\rangle\right)} &= \frac{\text{Cov}\left(x,\ R_x^-\right)}{\langle x\rangle\left\langle\ R_{x_2}^{\pm}\right\rangle} + \frac{\text{Cov}\left(x^2,\ R_x^+\right)}{2\langle x\rangle\left(\langle x\rangle\left\langle\ R_{x_2}^{\pm}\right\rangle\right)} \\ &\quad + \frac{\text{Cov}\left(x,\ R_x^+\right)}{2\langle x\rangle\left(\langle x\rangle\left\langle\ R_{x_2}^{\pm}\right\rangle\right)} + \frac{\text{Cov}\left(x,\ R_x^-\right)}{2\langle x\rangle\left(\langle x\rangle\left\langle\ R_{x_2}^{\pm}\right\rangle\right)}\end{aligned} \tag{30}$$

□

C. Linear approximation calculation

In this section, we use the linear noise approximation to obtain expressions for η_{x_1,x_1}, η_{x_1,x_2}, and η_{x_1,x_2}. In the 2 components hill function case, first, we compute the Jacobian matrix M for the system. $H_{x_1x_1} = H_{x_2x_2} = 1$, since the death terms of each variable are linear with respect to that variable, and the birth term of each variable is independent of itself. $H_{x_1x_2}$ is zero, since the birth rate ($R_{x_1}^+$) and death rate ($R_{x_1}^-$) of x_1 are independent of x_2. The computation for $H_{x_2x_1}$ is more involved:

$$H_{x_2x_1} = \left(\frac{x_1}{R_{x_2}^-}\frac{\partial\ R_{x_2}^-}{\partial x_1} - \frac{x_1}{R_{x_2}^+}\frac{\partial\ R_{x_2}^+}{\partial x_1}\right)\Bigg|_{x_2=\langle x_2\rangle, x_1=\langle x_1\rangle} \tag{31}$$

$$= 0 - \frac{\langle x_1\rangle}{\lambda_1\frac{\langle x_1\rangle^n}{\langle x_1\rangle^n+K^n}}\lambda_1\frac{nK^n\langle x_1\rangle^{n-1}}{\left(\langle x_1\rangle^n+K^n\right)^2} \tag{32}$$

$$= -\frac{nK^n\langle x_1\rangle^n\left(\langle x_1\rangle^n+K^n\right)}{\langle x_1\rangle^n\left(\langle x_1\rangle^n+K^n\right)^2} \tag{33}$$

$$= -\frac{nK^n}{K^n+\langle x_1\rangle^n} \tag{34}$$

Hence,

$$H = \begin{bmatrix} 1 & 0 \\ -\frac{nK^n}{K^n+\langle x_1\rangle^n} & 1 \end{bmatrix} \tag{35}$$

Recall that $M_{ij} = \frac{H_{ij}}{\tau_i}$; that is, the entries of the ith row of M are the entries of the ith row of H divided by the lifetime $\tau_i = \frac{\langle x_i\rangle}{\left\langle\ R_{x_i}^-\right\rangle}$. For exponential death processes, the lifetime is simply the inverse of the decay rate, so we multiply the ith row by β_i Hence,

$$M = \begin{bmatrix} \beta_1 & 0 \\ -\beta_2\frac{nK^n}{K^n+\langle x_1\rangle^n} & \beta_2 \end{bmatrix} \tag{36}$$

We now solve for the diffusion terms: the diffusion matrix is symmetric, with entries expressed in terms of the $\langle s_{ij} \rangle$. Since all reactions involve only one type of species, $s_{ij} = 0$ for $i \neq j$; hence, the diffusion matrix is diagonal, with diagonal entries $s_{ii} = \frac{2}{\tau_i} \frac{\langle s_{ii} \rangle}{\langle x_i \rangle}$. Thus,

$$D = \begin{bmatrix} \frac{2\beta_1}{\langle x_1 \rangle} & 0 \\ 0 & \frac{2\beta_2}{\langle x_2 \rangle} \end{bmatrix} \tag{37}$$

Now, we write the matrix equation

$$\begin{bmatrix} \beta_1 & 0 \\ -\beta_2 \frac{nK^n}{K^n + \langle x_1 \rangle^n} & \beta_2 \end{bmatrix} \begin{bmatrix} \eta_{x_1,x_1} & \eta_{x_2,x_1} \\ \eta_{x_2,x_1} & \eta_{x_2,x_2} \end{bmatrix} + \begin{bmatrix} \eta_{x_1,x_1} & \eta_{x_2,x_1} \\ \eta_{x_2,x_1} & \eta_{x_2,x_2} \end{bmatrix} \begin{bmatrix} \beta_1 & 0 \\ -\beta_2 \frac{nK^n}{K^n + \langle x_1 \rangle^n} & \beta_2 \end{bmatrix} = \begin{bmatrix} \frac{2\beta_1}{\langle x_1 \rangle} & 0 \\ 0 & \frac{2\beta_2}{\langle x_2 \rangle} \end{bmatrix}$$

This gives us three equations. For the diagonal equations, we divide through by $2\beta_i$ to write

$$\eta_{x_2,x_2} = \frac{1}{\langle x_2 \rangle} + \frac{nK^n}{K^n + \langle x_1 \rangle^n} \eta_{x_2,x_1}$$
$$\eta_{x_1,x_1} = \frac{1}{\langle x_1 \rangle}$$

For the off-diagonal equation, we obtain

$$(\beta_2 + \beta_1)\, \eta_{x_2,x_1} - \beta_2 \frac{nK^n}{K^n + \langle x_1 \rangle^n} \eta_{x_1,x_1} = 0 \tag{38}$$

Solving for η_{x_2,x_1}, we obtain

$$\eta_{x_2,x_1} = \left(\frac{\beta_2}{\beta_1 + \beta_2} \right) n \left(\frac{K^n}{K^n + \langle x_1 \rangle^n} \right) \eta_{x_1,x_1} \tag{39}$$

Using the diagonal equations, we could rewrite this as

$$\eta_{x_2,x_1} = \left(\frac{\beta_2}{\beta_1 + \beta_2} \right) \left(\frac{K^n}{K^n + \langle x_1 \rangle^n} \right) \frac{n}{\langle x_1 \rangle} \tag{40}$$

Substitute in the previous equation, we get

$$\eta_{x_2,x_2} = \frac{1}{\langle x_2 \rangle} + \left(\frac{\beta_2}{\beta_1 + \beta_2} \right) \left(\frac{nK^n}{K^n + \langle x_1 \rangle^n} \right)^2 \frac{1}{\langle x_1 \rangle} \tag{41}$$

D. Linear noise approximation for Mass action

Similarly, for linear noise approximation in mass action case, we construct three matrices.

$$H = \begin{bmatrix} 1 & 0 \\ -n & 1 \end{bmatrix} \tag{42}$$

$$M = \begin{bmatrix} \beta_1 & 0 \\ -\beta_2 n & \beta_2 \end{bmatrix} \tag{43}$$

$$D = \begin{bmatrix} \frac{2\beta_1}{\langle x_1 \rangle} & 0 \\ 0 & \frac{2\beta_2}{\langle x_2 \rangle} \end{bmatrix} \tag{44}$$

This gives us three equations. For the diagonal equations, we divide through by $2\beta_i$ to write

$$\eta_{x_2,x_2} = \frac{1}{\langle x_2 \rangle} + n\eta_{x_1,x_2}$$
$$\eta_{x_1x_1} = \frac{1}{\langle x_1 \rangle}$$

For the off-diagonal equation, we obtain

$$(\beta_1 + \beta_2)\,\eta_{12} - \beta_2 n \eta_{x_1x_1} = 0 \tag{45}$$

Solving for η_{12}, we obtain

$$\eta_{x_1,x_2} = \left(\frac{\beta_2}{\beta_1 + \beta_2}\right) n^2 \eta_{x_1x_1} \tag{46}$$

Using the diagonal equations, we could rewrite this as

$$\eta_{x_1,x_2} = \left(\frac{\beta_2}{\beta_1 + \beta_2}\right) \frac{n^2}{\langle x_1 \rangle} \tag{47}$$

And substitute into previous equation, we get

$$\eta_{x_2,x_2} = \frac{1}{\langle x_2 \rangle} + \left(\frac{\beta_2}{\beta_1 + \beta_2}\right) \frac{n^2}{\langle x_1 \rangle} \tag{48}$$

E. Error2 and Error3 Plots over different parameter settings

Like Figure 5, we can plot Err_3 under different parameter settings. It follows a similar pattern as we can see that when mean of x_1 is small, β_1 is small and mean of x_2 large, we get a large F_c and larger contrast between the third invariant line and the background, meaning we are more confident for the line of Err_3.

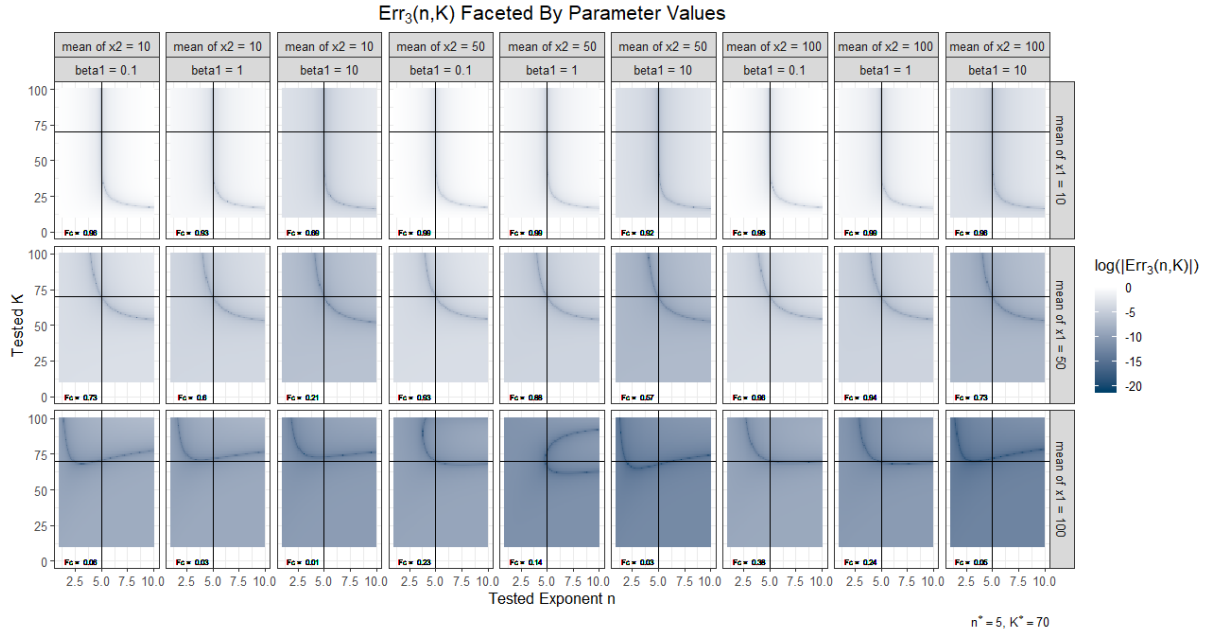

Fig. 9: Err_3 from different parameter settings

Previously we are fixing the true (n, K) pair and look at different parameters. Alternatively, we can also fix all other parameters and change n, K. As we see in the following figure, as n increases and K increases, we are more confident in the second invariant curve we get.

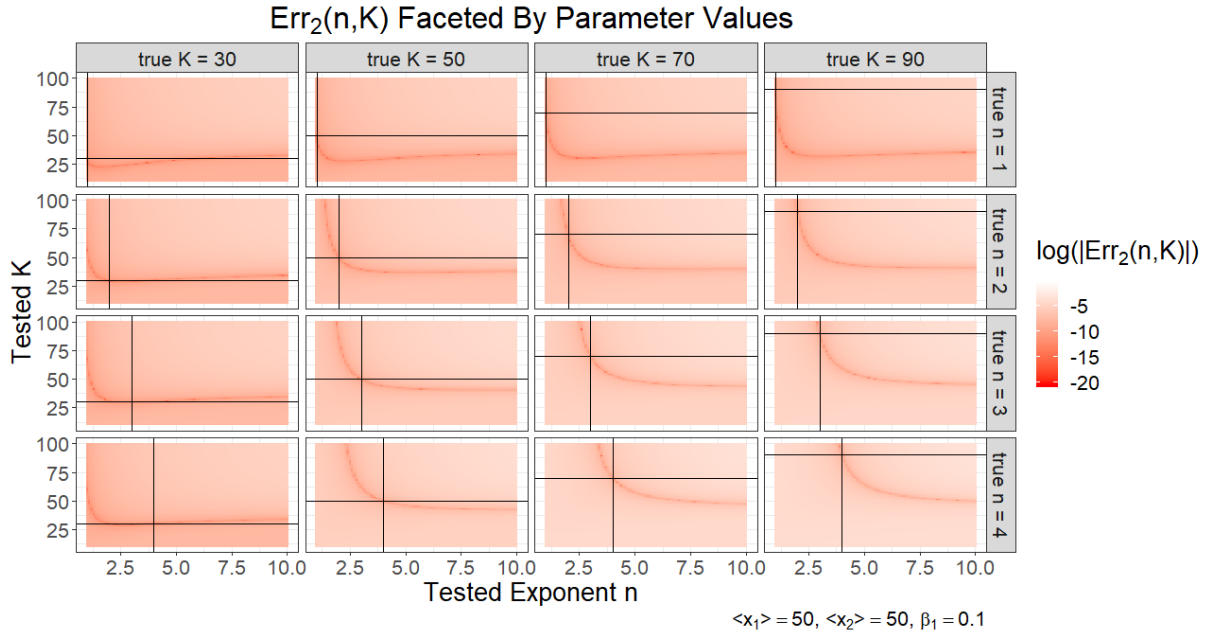

Fig. 10: Err_2 from different (n, K) pairs

Similar story for the third invariant:

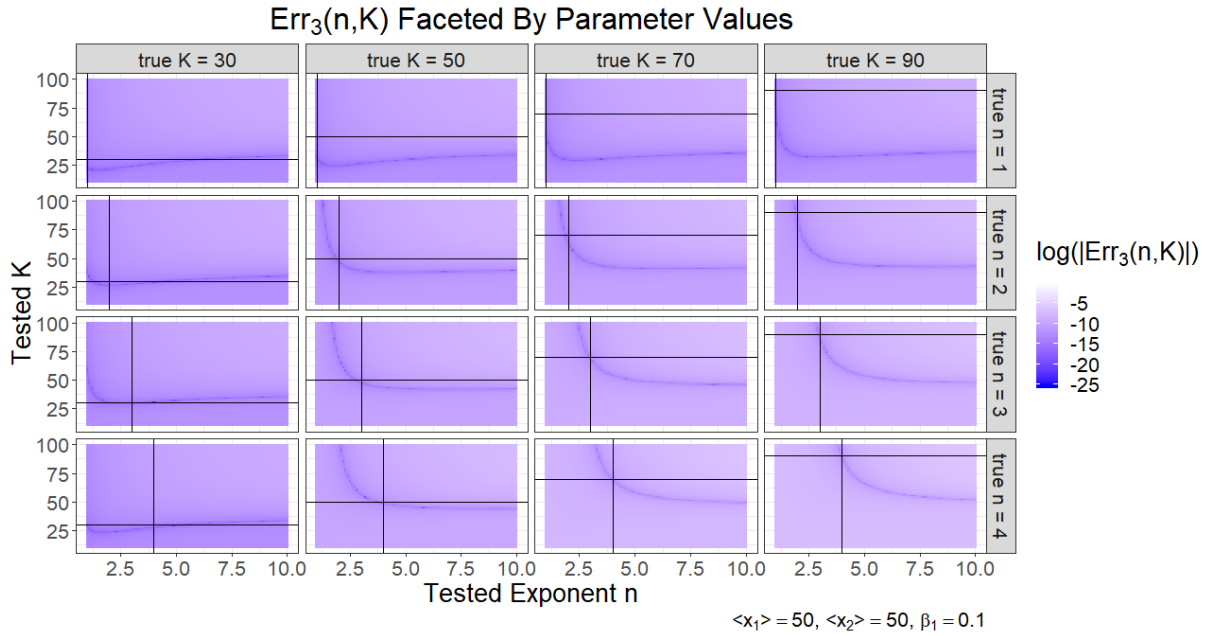

Fig. 11: Err_3 from different (n, K) pairs

Since the intrinsic noise $\frac{1}{\langle x_2 \rangle}$ is usually much smaller than the extrinsic term, we find that

change of $\langle x_2 \rangle$ does not affect the result very significantly. Thus, we could take a closer look just at how β_1 and $\langle x_1 \rangle$ affect our result. The following two plots shows how confident we are in Err2 and Err3 under different β_1 and $\langle x_1 \rangle$.

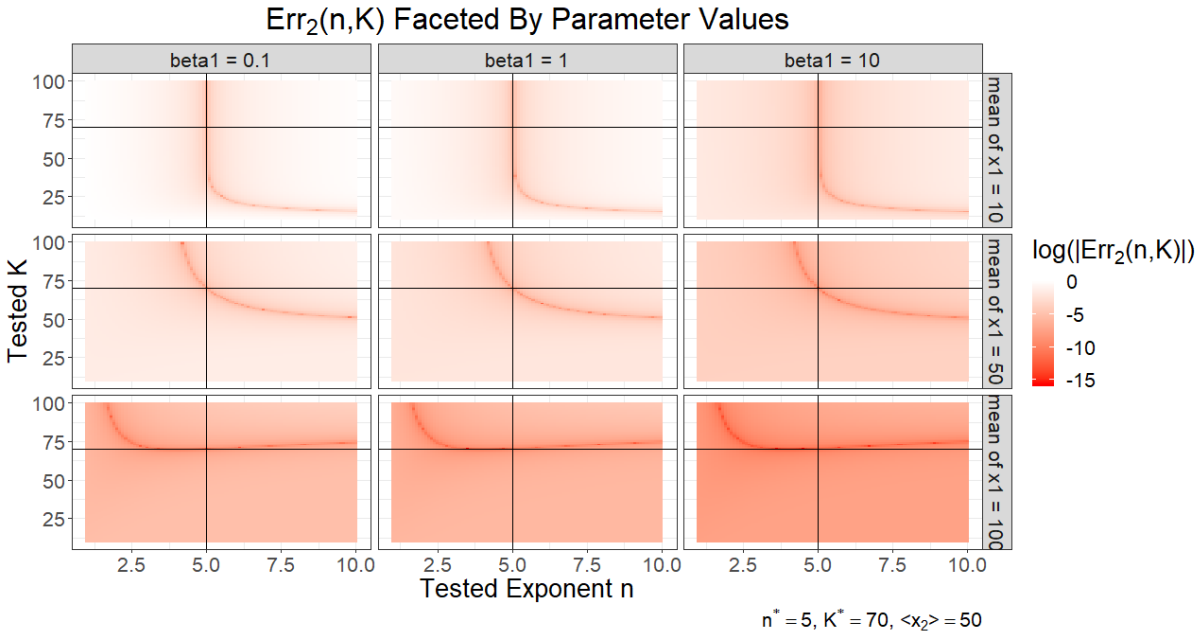

Fig. 12: Err_2 from different β_1 and $\langle x_1 \rangle$

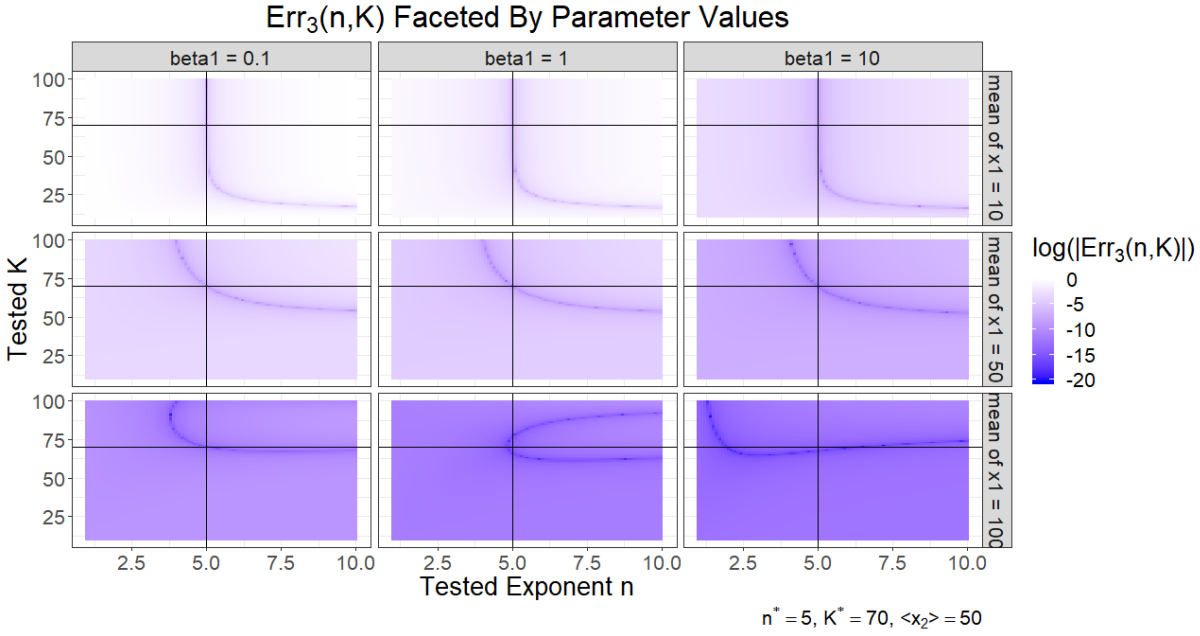

Fig. 13: Err_3 from different β_1 and $\langle x_1 \rangle$

F. 2 component Mass Action derivative

For the model in Section III, we can compute the numerical derivative under different settings. The result is similar as we see in the panel of Figure 1. In order to get a large derivative, we want $\langle x_2 \rangle$ to be large, $\langle x_1 \rangle$ to be small and β_1 to be small.

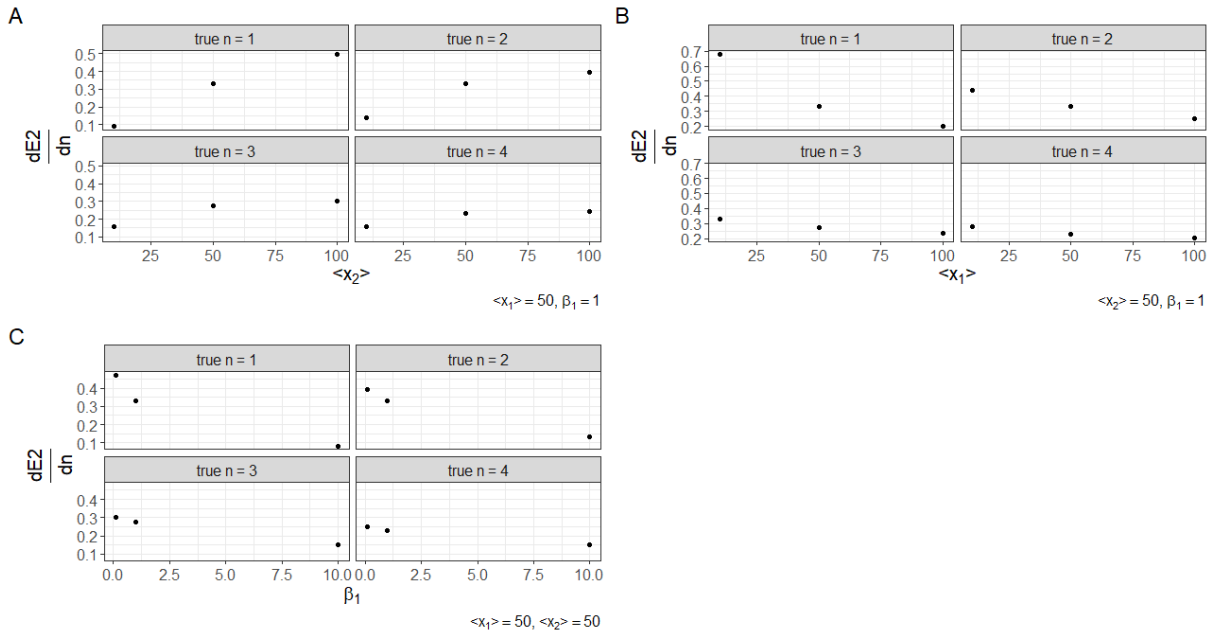

Fig. 14: Derivatives of Err_2 in Mass Action Case

REFERENCES

[1] A. Hilfinger, T. M. Norman, and J. Paulsson, "Exploiting natural fluctuations to identify kinetic mechanisms in sparsely characterized systems," *Cell systems*, vol. 2, no. 4, pp. 251–259, 2016.

[2] A. v. O. Juan M Pedraza, "Noise propagation in gene networks," *Science*, vol. 307, no. 5717, pp. 1965–1969, 2005.

[3] L.-h. So, A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding, "General properties of transcriptional time series in escherichia coli," *Nature genetics*, vol. 43, no. 6, p. 554, 2011.

[4] J. Liepe, P. Kirk, S. Filippi, T. Toni, C. P. Barnes, and M. P. Stumpf, "A framework for parameter estimation and model selection from experimental data in systems biology using approximate bayesian computation," *Nature protocols*, vol. 9, no. 2, p. 439, 2014.