Spora: A Journal of Biomathematics

Volume 7 Article 10

2021

An Investigation of Gene Regulatory Network State Space Variability

Sara Liesman
Illinois State University, sliesma@ilstu.edu

Olcay Akman
Illinois State University, oakman@ilstu.edu

Daniel Hrozencik
Chicago State University, dhro@att.net

Intercollegiate Biomathematics Alliance

SPORA
A Journal of Biomathematics

ISSN 2473-3067 (Print)
ISSN 2473-5493 (Online)

An Investigation of Gene Regulatory Network State Space Variability

Sara Liesman[1], Olcay Akman[1,*], Daniel Hrozencik[2]

*Correspondence:
Prof. Olcay Akman,
Department of Mathematics,
Illinois State University,
100 N. University St.,
Normal, IL 61761-2467, USA
oakman@ilstu.edu

Abstract

Genes are segments of DNA that provide a blueprint for cells and organisms to effectively control processes and regulations within individuals. There have been many attempts to quantify these processes, as a greater understanding of how genes operate could have large impacts on both personalized and precision medicine. Current biological methods cannot easily reveal the details of gene interactions. Therefore, we use gene expression data to infer networks of interactions, which are called gene regulatory networks or GRNs. These methods are designed to bypass the need for large amounts of data and extensive knowledge about a network. In this work, we extend previous work by investigating additional ways to incorporate stochasticity into gene regulatory networks.

Keywords: beta distribution, gene regulatory networks, spectral density, state transition, stochasticity

1 Introduction

1.1 Biological background

Gene expression is regulated through interaction networks and a series of positive and negative feedback loops [3, 17, 20]. Similar to interaction networks on the macro-level, where organisms interact with each other in order to create an ecosystem, interaction networks also exist on the microscopic level [3, 17, 20]. On the microscopic level, cells, proteins, and molecules interact with one another.

The combination of these interactions form a network of gene, protein, and regulator reactions that we call a gene regulatory network. Current technology and experimental methods do not exist to directly reveal the intricacies of these networks. Therefore, the gene regulatory networks that we know today are networks that have been inferred from gene expression data, which are called gene regulator networks or GRNs [8]. Gene expression data provides information about the types and numbers of mRNAs, but not necessarily about binding information such as how well molecules bind to one another. Thus, GRNs are used in order to infer the interactions between genes, proteins, and other regulators [8]. This information helps us to determine when a gene is active or repressed within a system.

Although these interactions are dictated by a network, there is still some stochasticity involved among interactions [3, 4, 20, 27, 35]. This stochasticity may be due to concentration levels, binding abilities, or even distances of molecules within cells. In cells, competition between ribosomes and RNase E binding can lead to some stochasticity [27]. Mathematical models have been created in attempts to capture the details of these processes with varying degrees of success. In general, stochasticity is not captured by all inference methods, but likely plays an important role in the formation of GRNs since a large portion of cellular signals are from noise [3].

1.2 Mathematical background

For a cell to properly function, it depends on the coordination of thousands of proteins in different variations interacting at the correct time, place, and in the correct quantity [3, 20, 33]. In order to orchestrate these interactions, regulatory systems exist to help determine when mRNA is produced, how long mRNA should last, how much protein from mRNA should be created, how proteins are arranged and modified, and when they are degraded [3, 20, 33]. However, large amounts of data are needed for gene network inference [3], and limitations in experimental techniques create noisy data sets, so only a small number of interactions have been extracted [32]. Moreover, human interests have introduced a bias for networks that are related to human diseases [32]. These biological limitations have led to the need for mathematical models to help estimate biologically reasonable parameters such as inputs, time delays, and genes expressed from feedback loops [3, 17, 19, 23, 32, 33].

[1]Department of Mathematics, Illinois State University, Normal, IL, [2]Department of Mathematics and Computer Science, Chicago State University, Chicago, IL

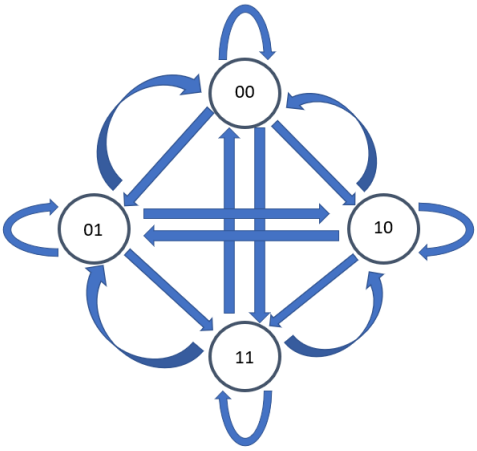

Figure 1: An example of a Boolean network represented as a directed graph.

Networks are a natural way to model biological systems with interactions [14, 23, 32]. A network is a set of objects that are connected through a set of rules. In the past, networks have been used to model a variety of interactions including protein structure networks, protein-protein interactions, transcriptional regulation, metabolism, and neuronal synaptic connections [32]. In gene regulatory networks the objects, or nodes, in a network graph represent a collection of genes and the edges represent the interactions between sets of genes. Models for GRNs can be dynamical or static, discrete or continuous, and deterministic or stochastic [3, 17]. Some models that have been studied are Bayesian networks, rule-based algorithms, ordinary differential equations, and Boolean networks [17, 23]. Table 1 describes some of the possible methods used to study GRNs. The methods discussed here are not exhaustive.

In order to use these methods large amounts of data are required. For example, differential equations can require as many experiments as there are genes in a network [35]. This likely limits the use and comparison of models to small networks. In yeast (*Saccharomyces cerevisiae*) 2355 genes have been identified contributing to regulatory networks [22], this could imply that upwards of 2355 experiments are needed in order to identify the structure and rates of these networks. In addition, transition rates are not easily obtained through experimentation because estimations often rely on linear changes, which would imply that there is no natural capacity for transition rates [35]. Combined with the fact that data size is often limited by the cost of experiments [3], it is clear that there is a need to develop more accurate GRNs that consider the effects of variation and noise.

The authors in [2] attempt to address this concern in 2018 by investigating the variation between state transitions in GRNs and noise within a network . The authors use beta distributions to estimate the propensity that state transitions occur in a GRN under noisy conditions. Through their investigation, [2] find that state transitions with high variation lead to a network with high amounts of variation [2]. This is consistent with the literature, which finds that short-term fluctuations in protein production can have larger impacts on gene expression [27]. These impacts are likely due to the cascading behavior of GRNs as well as different feedback loops within cells [4, 27]. The authors also use the assumption that propensities among state transitions are not constant and demonstrate how statistical distributions can be used to provide important insight about GRNs.

However, the process described in [2] also requires large amounts of data and becomes less stable for large networks. They also assume that an ideal network will have minimum variance among state transition propensities which may be ignoring some key ideas such as:

- Data obtained from a single cell can have large amounts of noise and variation [3]
- Noise from gene expression data can be as large as 30% [3]
- Different feedback loops can amplify or dampen variation in networks [4]
- Genetically identical cells may show cell-cell differences of more than 10% [4]

Despite some shortcomings, [2] have begun to pave the way for introducing more variability into models that will likely require stochastic properties in the near future. In this work we attempt to extend the authors' work and address concerns about noise by investigating additional ways to incorporate variation and noise into future models.

2 Methods

In this study, we look at intrinsic noise in GRNs and discuss ways to incorporate this noise into future models. During gene and protein interactions transcription factors and chemical signals come into contact with one another and bind in order to cause a reaction such as mRNA synthesis or protein assembly. Traditionally, these reactions and collisions of chemicals in a system were considered deterministic, but due to quantum indeterminacy and lack of mechanical isolation it is now argued that the processes in these systems are more likely stochastic in nature [12]. Here, we assume that we have a well-mixed system with a limited number of molecules per population of molecule so that our systems will incorporate discreteness and stochasticity. In the past, this reaction rate

was modeled by a set of ordinary differential equations, but when a system involves discreteness and stochasticity, these equations are no longer appropriate [12].

Since our system is both discrete and stochastic, it is most natural to represent our system as a Boolean network on two states with added stochasticity; this representation is similar to the ones described in [2, 28]. In Figure 1 we see an example of a Boolean network. Here, each node of the directed graph represents a state of the system. In Figure 2 we see a Boolean network with weighted edges, where each weight represents the propensity for changing to a state. Therefore, when we have a two gene system, each node will have two entries. The first entry will represent the state of the first gene, and the second will represent the state of the second gene.

These states are either on (1) or off (0). Biologically, genes may be operating at different rates and have more than two states, but mathematically it is possible to abstract information by only considering genes as active or non-active.

Let $G_1, G_2, \ldots, G_n$ represent n genes in a regulatory network. Let $x_i(t)$ be the state of G_i at time t where $i = 1, 2, \ldots, n$ and $t \in [0, T]$. The possible states of our system are 0 or 1, denoted $X_i = 0, 1$ for $i = 1, 2, \ldots, n$. If G_1 is on at time q, $0 \leq q \leq T$, then $x_1(q) = 1$. If G_1 is off at time q, $0 \leq q \leq T$, then $x_1(q) = 0$. In this study we focus on a predetermined two gene system. Therefore, the state space of this network S is the Cartesian product of each gene's state space, $S = X_1 \times X_2$.

Each edge of the directed graph represents a probability for transition to a different state. These propensities are determined by an update function involving activation and degradation propensities, $p_i^{\uparrow} \in [0, 1]$ and, $p_i^{\downarrow} \in [0, 1]$ respectively. Failure to activate and failure to degrade are represented by $1 - p_i^{\uparrow}$ and $1 - p_i^{\downarrow}$ respectively. Note that $1 - p_i^{\uparrow} \in [0, 1]$ and $1 - p_i^{\downarrow} \in [0, 1]$. These values can be seen in Table 2. Although genes are interacting with each other in a network, the propensities for activation and deactivation of genes are each independent. This occurs because we are assuming that the genes activate and deactivate independently of one another.

Thus, the probability of transitioning from one state to another can be represented as the product of two propensities. Given the independence between gene state propensities, we obtain the value of each edge propensity by finding the product of the appropriate gene propensities for activation, degradation, failure to activate, and failure to degrade. For example:

$$\begin{aligned} Pr(00 \to 01) &= (1 - p_1^{\uparrow}) \times p_2^{\uparrow}, \\ Pr(11 \to 10) &= (1 - p_1^{\downarrow}) \times p_2^{\downarrow}, \\ Pr(01 \to 11) &= p_1^{\uparrow} \times (1 - p_2^{\downarrow}), \\ Pr(10 \to 00) &= p_1^{\downarrow} \times (1 - p_2^{\uparrow}). \end{aligned}$$

These transitions can also be represented in a transition matrix. A general transition matrix is shown in Table 3. A more concrete example of a Boolean network with activation and degradation propensities is shown in Figure 2

In order to determine when a gene will transition, a transition function that has been derived and described in [28] and previously used in [2] is also used in this work:

$$f(t) = a_0 e^{-kt}, \quad t \geq 0. \tag{1}$$

The function represents the concentration of molecules in the cell as a function of time. In tandem with Equation (1), we also use a threshold m which occurs at time $t = \tau$. This threshold indicates at what concentration we would expect to see a change. A model of this process can be seen in Figure 3. For example, if gene 2 is on, $x_2 = 1$, then as soon as the concentration of molecules activating that gene decreases below m we would expect the gene to turn off, $x_2 = 0$. Likewise, if gene 2 is off, $x_2 = 0$, then as soon as the concentration of molecules repressing that gene decrease below m we would expect the gene to turn on, $x_2 = 1$.

Here, we use an exponential decay function which implies that transitions occur after a decrease in the proportion of molecules. However, biologically it is possible for transitions to occur after an increase in the proportion of molecules. Although we only use a decaying function for the transition function, we believe that any process which involves an increase in molecules could also be modeled with this exponential decay by changing the way you number the y-axis such that the $\lim_{t \to \infty} f(t) = \kappa$, where κ is the natural carrying capacity of appropriate molecules in a network's transition. In general, we would expect genes that fall below the threshold m to change states, but this is not always the case due to the stochastic nature of these processes. Equation (1) and the transition matrix together create a system that incorporates both a time delay and stochasticity.

[28] and [2] both utilize a model similar to the one described here. In [28] the edge propensities are kept constant and they work under the assumption that even if a reaction is supposed to occur, there is no guarantee that a transition will take place or even that the correct transition will take place [28]. [2] argued that the propensities discussed in [28] are not likely to remain constant and applied a beta distribution to obtain propensities for state transitions along the edges of a network [2]. These models generally focused on obtaining the propensities of state transitions and overall network variance, but paid little attention to capturing variability in individual gene to gene interactions. Here, we use the same model, but focus on gene to gene variability and propensities. First, we look at the transition function and examine how rates

Table 1: Methods for Gene Regulatory Networks [3, 17, 24].

Type	Qualities	Pros	Cons
Directed Graphs	Static and deterministic	Intuitive	Crude
Boolean Networks	Discrete, dynamic,	Simple for large data sets	Crude
Generalized Logical Networks	Discrete, dynamic, and deterministic	Accommodates asynchronous state changes and time delays	Inference is difficult
Differential Equations	Continuous, dynamic, and deterministic	Flexible	Computationally expensive
Bayesian Networks	Static, stochastic, and discrete or continuous	Stochastic	Need large amounts of data
Stochastic Master Equation	Discrete, dynamic, and stochastic	Realistic	Difficult to use

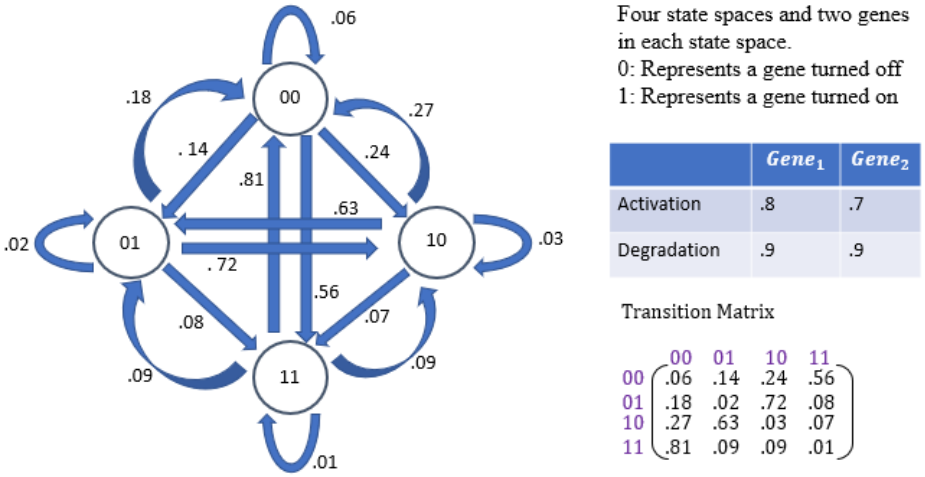

Figure 2: An example of a Boolean network with probabilities and a transition matrix.

Table 2: State Update Propensities.

Update	x_1	x_2
Activation	$p_1^{\uparrow}$	$p_2^{\uparrow}$
Degradation	$p_1^{\downarrow}$	$p_2^{\downarrow}$
Failure to Activate	$1-p_1^{\uparrow}$	$1-p_2^{\uparrow}$
Failure to Degrade	$1-p_1^{\downarrow}$	$1-p_2^{\downarrow}$

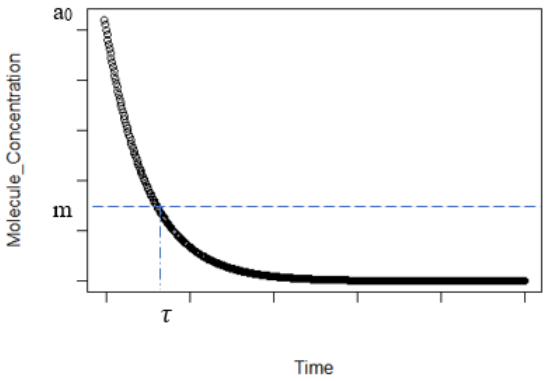

Figure 3: The transition function is shown with a threshold at m. Once the concentration of molecules dips below the threshold m we expect the gene to transition from one state to another at time τ.

of decay and thresholds impact the variation among gene transitions. Second, we revisit the beta distribution described in [2], but instead of finding transition propensities using the statistical distribution, we use the beta distribution to determine the activation and degradation propensities. Third, we explore the use of spectral density to evaluate variation and propensities in a GRN. Finally, we compare these results and discuss areas that need improvement and future directions we hope to explore.

2.1 Method 1: Transition function

Previous studies using the transition function and a Boolean network have focused on adding stochasticity into propensities and time delays between interactions [2, 28]. Variation in [2] was found based on the beta distribution of each transition variability. Here, we investigate the variation inherent to the transition function itself. Recall that the transition function represents the proportion of molecules needed in order to transition between the on and off state of a gene. Once the number of molecules reaches a threshold m at time $t = \tau$, i.e. $X(\tau) = m$, a state to state transition occurs. Arbitrarily, when $X(t) \geq m$ the gene is considered on and when $X(t) < m$ the gene is considered off. This change in state could easily be reversed without loss of generality.

In the transition function, Equation (1), we let $a_0 = 1$ so that the initial concentration of molecules, $f(0)$, can be considered as 100% and the threshold m can represent a percentage of molecules needed in order to activate or degrade a gene. Since the threshold m is determined by the strength of chemical bonds between the molecules and their binding sites, as well as the locations and number of binding sites, the threshold does not vary significantly from cell to cell [4].

In this study, we allow the rate of decay within the concentration of molecules to vary so that we can measure the effects on variance in a state to state transition. This variance is biologically relevant because some stochasticity in GRNs is likely a direct result of competition between ribosomes and RNase. Ribosomes are responsible for translating mRNA into protein synthesis. RNase E signals the degradation of mRNA after transcription and before translation, however, mRNA that binds to a ribosome will undergo translation and lead to protein production [27]. Whether mRNA degrades or leads to the production of protein is thought to be determined by its proximity to either RNase E or a ribosome [27].

To model the behavior of the transition function we use the programming language R. Let $k \sim U(0.1, 10)$, $m = 0.4$, $a_0 = 1$, and $\tau = \frac{-1}{k} * \log(\frac{m}{a_0})$. The equation for τ is derived from Equation (1), where $f(t) = m$ and $t = \tau$ such that $m = a_0 e^{-k\tau}$. Using this information we find when the transition function will cross the threshold for a given value of k. Then, we replicate this process 1000 times to examine how changes in k affect τ.

After examining the behavior of the transition function, we also investigate how stochasticity plays a role in state to state transitions. In a deterministic model, once the concentration of molecules drops below the threshold m we would expect there to be a transition between states. However, in a stochastic model we only expect a proportion of these changes to happen each time. In this study we assume that only 75% of the genes that were supposed to change states actually have changed states. The choice of 75% was predetermined and compared to other propensities without any significant changes to results. We then record every time 5% of the genes pass the threshold. This process was also simulated in R. Out of the 1000 replicates, each transition function that passed the threshold at a given time τ had a 75% probability of either changing states or remaining the same. This process was repeated for $m = \{0.1, 0.4, 0.9\}$. For each time 5% of the genes pass the threshold, we record how much variation is present in the interval and plot this variation over time. Our results are discussed in Section 4.1.

2.2 Method 2: Beta distributed propensities

The second method we applied in this work was the implementation of the beta distribution for the activation and degradation propensities. The beta distribution was utilized by [2] to find the transition propensities between the states of genes. Here, we use the beta distribution to determine the activation and degradation propensities for each gene. Therefore, we will have a transition matrix similar to Table 3 where,

$$p_i^{\uparrow} \sim Beta(\alpha_{iA}, \beta_{iA}) \qquad \text{and} \qquad p_i^{\downarrow} \sim Beta(\alpha_{iD}, \beta_{iD})$$

and where $i = 1, 2, \ldots, n$ in general, and $i = 1, 2$ for this work with only two genes.

The use of the beta distribution is appropriate for a variety of reasons. First, the beta distribution has a codomain between 0 and 1 inclusive. Second, the beta distribution is commonly used when the set of random variables are probabilities, thus the extension to propensities is not unreasonable. Last, the beta distribution has been used by Wright to model other biological phenomena such as gene frequencies in population dynamics in 1937 [40]. The beta distribution uses two parameters, a shape (α) and a rate parameter (β), but can also be understood intuitively as a number of successes α and failures β. Thus, in a series of N trials where we would expect α successes, we would also expect $N - \alpha = \beta$ failures.

Our calculations for the transition matrix make use of the gamma function, the beta function, and the Gauss

hypergeometric function which can be found in the Appendix.

The transition matrix in Table 3 then has the following properties:

1. This matrix is stochastic and thus a transition matrix.
2. $(1-p_i^{\uparrow})$ and $(1-p_i^{\downarrow})$ each have a beta distribution.
3. The probability density function of the product of two beta distributions with shape parameters (a_1, b_1) and (a_2, b_2) for genes 1 and 2 respectively is shown in the Appendix.
4. The expected value of the matrix is the expected value of each entry.
5. The expected value of each entry is the product of the expected values of each beta distributed activation or deactivation propensity, denoted E_i for G_i.
6. The variation of each transition between the states of two beta distributions with shape parameters (a_1, b_1) and (a_2, b_2) for genes 1 and 2 respectively is

$$\frac{{E_1}^2(1+a_1){E_2}^2(1+a_2)}{(E_1+a_1)(E_2+a_2)} - (E_1E_2)^2.$$

The justification for each of these claims is discussed in the Appendix.

2.3 Method 3: Spectral density

The third and final method which we used to investigate ways to incorporate variance into gene state transitions involved using spectral density. The general idea behind spectral density is to take a finite set of static data and estimate how the total power is distributed across frequencies. Spectral analysis in particular has been used in a variety of fields such as psychology for heart rates [18], medical fields for fetal heart rates [38], geology for geological formation [5], and bioinformatics for gene prediction [25].

Let $y(t)$ for $t = 0, 1, \ldots, T$ be a discrete-time data sequence from time $t = 0$ to $t = T$, and assume that $y(t)$ has finite energy such that $\sum_{t=-\infty}^{\infty} y(t) < \infty$. Let $r(q)$ be the autocovariance sequence obtained from the lag q between sampling data such that $r(q) = E[y(t) * y(t-q)]$. Lag is the amount of time between measuring signal output. Then, the power spectral density (PSD) of $r(q)$ is

$$\phi(\omega) = \int_{-\infty}^{\infty} r(q)e^{-i\omega q}\, dq,$$

where $\phi(\omega)$ is the PSD and Fourier transform of $r(q)$, ω is the frequency of the signal, and $i = \sqrt{-1}$. Then,

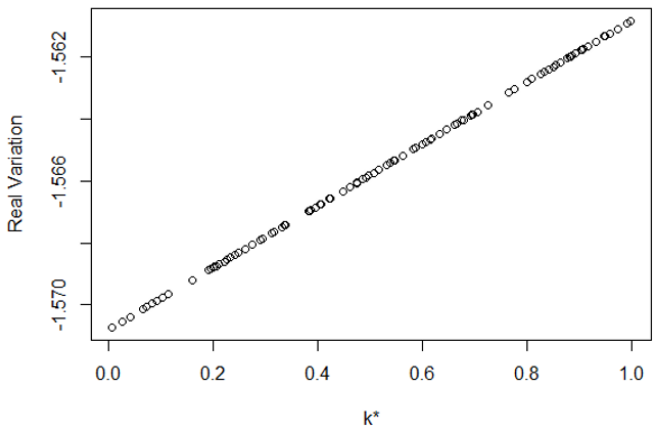

Figure 4: A plot of k_* from a uniform distribution vs. the real variation of the interaction.

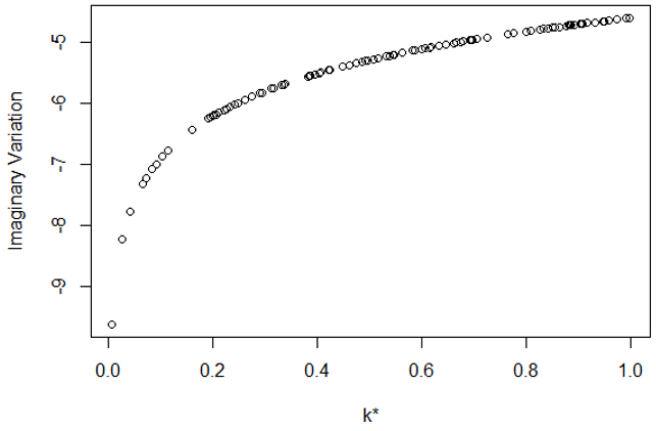

Figure 5: A plot of k_* from a uniform distribution vs. the complex variation of the interaction.

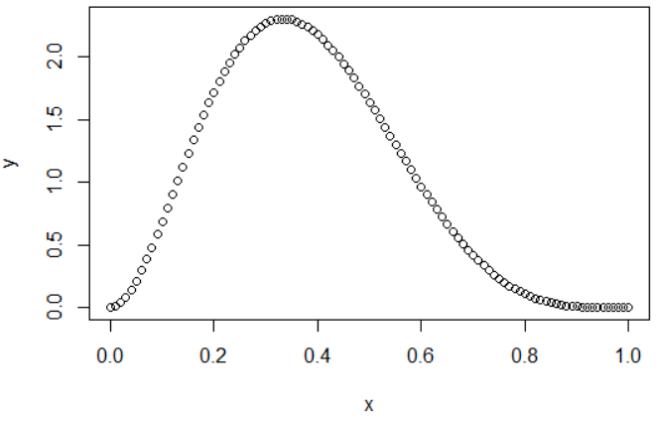

Figure 6: The beta distribution with shape parameters $\alpha = 3$ and $\beta = 5$.

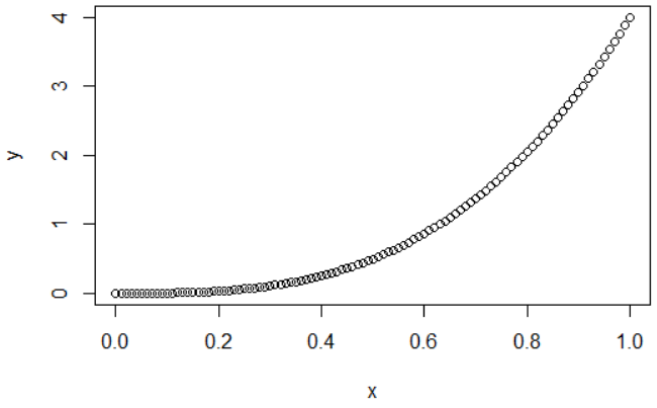

Figure 7: The beta distribution with shape parameters $\alpha = 4$ and $\beta = 1$.

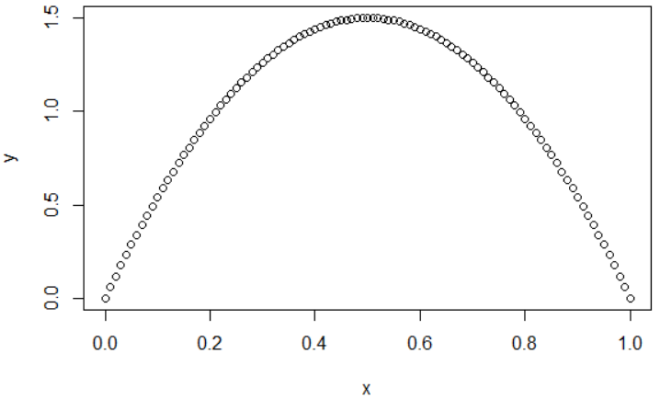

Figure 8: The beta distribution with shape parameters $\alpha = 2$ and $\beta = 2$.

$\phi(\omega)$ is the power at different bands of frequency [36]. Therefore, the integral of $\phi(\omega)$ across all possible frequencies is the total power of the signal. Furthermore, since $r(0) = E(y(t)^2)$, and if we assume that $E(y(t)) = 0$, the integral of $\phi(\omega)$ across all frequencies is also the variance [36].

Although we do not have data for the transitions of gene states, we use the transition function to demonstrate how this process would work for variance estimation. In this case, the transition function (1) is equal to $r(t)$ so that the Fourier transform of $r(t)$ will result in the PSD, and the integral of the PSD will be the variance of the system.

Biologically it is possible for the decay rate to vary [41]. If k varies, we replace it with a probability density function $g(k)$. This alters the transition function described in [28] and [12] such that the rate of transcription is $a_*(x) = kg(k)x$. Thus, $f_*(t) = a_0 e^{-k_* t}$, where $k_* = kg(k)$. Here, we allow the decay rate k to vary according to a specific statistical distribution.

In order to use a distribution for k, it is necessary to determine the variation of k. However, the decay rates of interaction molecules, like mRNA, are not well studied [41]. In this section, we explore how the variation of interactions is affected by different distributions of k.

To do this we will generate different values of k from different distributions and multiply them by their probability density function, $g(k)$, in order to obtain k_* and $f_*(t)$. The first moment of $f_*(t)$ is given by the Fourier Transform which is also the power spectral density in this case. The units of frequency that we use are cycles per sampling period. Then, we will calculate the variation of these interactions by integrating the power spectral density function.

We will include the entire time interval for these interactions, but we will only consider frequencies ranging from 0 to 100 periods per time interval. This interval has been chosen because it is unlikely that we will see frequencies greater than 100. For example, $\cos(x)$ has two frequencies because it has nonzero energy at two different amplitudes. Based on our function, it seems more likely that we only have one frequency per time interval. Either way, the interval $[0, 100]$ should include our frequency.

2.3.1 Uniform distribution

The uniform distribution describes scenarios where every value of k is equally likely to be chosen. Since k is a rate of decay, we let $k \in [0, 1]$. Then, the expected value of $g(k)$ is $E(g(k)) = 1$ and the power spectral density function is

$$\phi_*(\omega) = \frac{a_0}{i\omega + k_*}$$

Let $a_0 = 1$ and find $\int_0^{100} 1/(i\omega + k_*)\, d\omega$ to find the total variance across frequencies, $Var(\omega)$

$$Var(\omega) = -i \log\left(\tfrac{100i}{k_*} + 1\right)$$

Using polar coordinates and Euler's Theorem, we can rewrite $Var(\omega)$ as the following

$$Var(\omega) = -i \log\left(\sqrt{\left(\tfrac{100}{k_*}\right)^2 + 1}\right) - \arctan\left(\tfrac{100}{k_*}\right).$$

At 100 random values of k_* we obtain variations shown in Figures 4 and 5.

2.3.2 Beta distributions

We then repeat this process for the beta distribution with different shape parameters. The probability density function of the beta distribution is

$$g(k) = \frac{k^{\alpha-1} e^{-k\beta} \beta^\alpha}{\Gamma(\alpha)}.$$

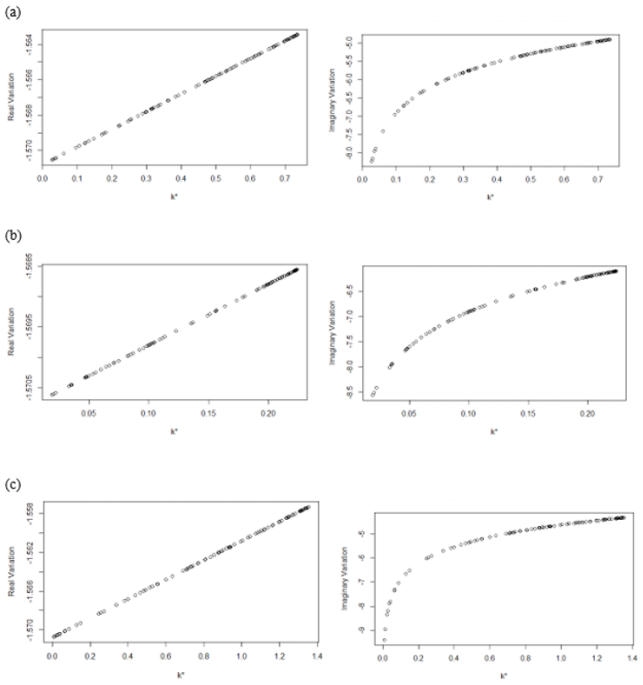

Figure 9: Variation measured from different spectral densities and beta distributions of k_*. All distributions contain real (left) and imaginary (right) components. (a) $k_* \sim Beta(\alpha = 2, \beta = 2)$; (b) $k_* \sim Beta(\alpha = 4, \beta = 1)$; (c) $k_* \sim Beta(\alpha = 3, \beta = 5)$.

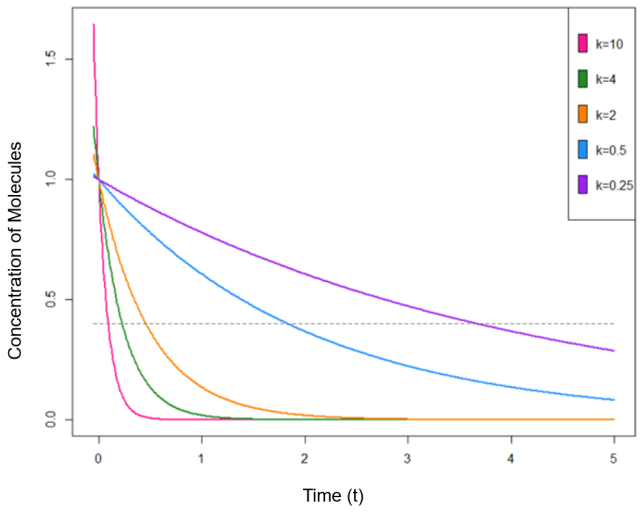

Figure 10: The transition function, $X(t) = x_0 e^{-kt}$, illustrated with different rates of decay, k, and a threshold, $m = 0.4$.

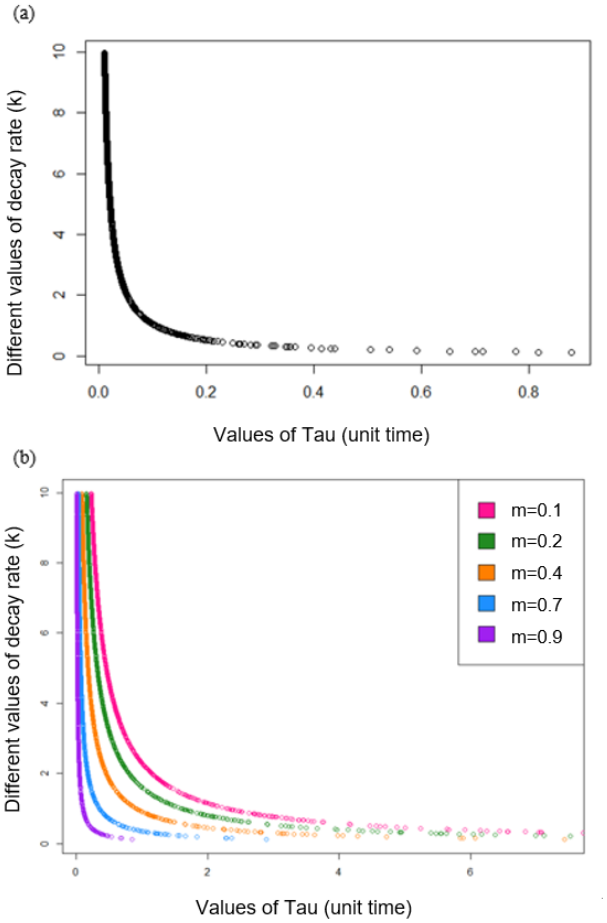

Figure 11: (a) How different decay rates, k, and time, τ, to cross the threshold, $m = 0.4$, interact. Tau is the time it takes for the proportion of molecules to cross a threshold that triggers a state change in the network. (b) How different decay rates, k, and time, τ, to cross a threshold, m, interact. This behavior is shown for different values of m to illustrate how m may impact each network.

Here, we generate three different beta distributions with shape parameters $(\alpha, \beta) = [(4,1),(2,2),(3,5)]$, which can be seen in Figures 6, 7, and 8. Notice that $g(k)$ is not a function of t, so all power spectral density functions will be in the form of

$$\phi_*(\omega) = \frac{a_0}{i\omega + k_*}.$$

Then, variation of each beta distribution, when $\omega \in (0, 100)$, will be of the form

$$-i \log\left(\frac{100i + k_*}{k_*}\right).$$

which can be rewritten as

$$Var(\omega) = -i \log\left(\sqrt{\left(\frac{100}{k_*}\right)^2 + 1}\right) - \arctan\left(\frac{100}{k_*}\right).$$

using a process similar to the one described for the uniform distribution. Due to the fact that different shape parameters for the beta distribution can result in a variety of curves, we explore three different cases. One where $\alpha > \beta$, one where $\alpha = \beta$, and one where $\alpha < \beta$. The specific shape parameters chosen visually created different curves, but we believe the general behavior of variation is still captured in Figure 9.

Case I: $\alpha = 4$, $\beta = 1$
The first case for beta distributions that we investigate is when the shape and rate parameters are $\alpha = 4$ and $\beta = 1$. This distribution can be seen in Figure 7 Then, we obtain the probability density function:

$$g(k) = \frac{k^3 e^{-4}}{6}.$$

From here, we find the expected value of k, k_*, and the PSD, $\phi_*(\omega)$:

$$k_* = \frac{k^4 e^{-4}}{6}$$
$$\phi_*(\omega) = \frac{6}{6i\omega + k^4 e^{-4}}.$$

The resulting variation can be seen in Figure 9a.

Case II: $\alpha = 2$, $\beta = 2$
The second case for beta distributions that we investigate is when the shape parameters are $\alpha = 2$ and $\beta = 2$. Figure 8 shows the behavior of this distribution. The probability density function under these parameters is

$$g(k) = 4ke^{-2k}.$$

Then, we can find the expected value of k, $E(k) = k_*$, and the PSD, $\phi_*(\omega)$:

$$k_* = 4k^2 e^{-2k}$$
$$\phi_*(\omega) = \frac{1}{i\omega + 4k^2 e^{-2k}}$$

The resulting variation can be seen in Figure 9b.

Case III: $\alpha = 3$, $\beta = 5$
The third and final case that we investigated for the beta distribution was when the shape and rate parameters were $\alpha = 3$ and $\beta = 5$. From here, we obtain the probability density function:

$$g(k) = 62.5k^2 e^{-5k}$$

Then, the expected value of k, $E(k) = k_*$, and the PSD, $\phi_*(\omega)$ are

$$k_* = 62.5k^3 e^{-5k}$$
$$\phi_*(\omega) = \frac{1}{i\omega + 62.5k^3 e^{-5k}}.$$

The resulting variation can be seen in Figure 9c.

3 Results

3.1 Results Method 1: Transition function

In Figure 10 we illustrate the behavior of the transition function, $X(t) = x_0 e^{-kt}$. The shape of the transition function and the time that the curve passes the threshold m are dependent on the value of the decay rate k. Higher values of k correlate with faster rates of decay and a shorter time τ that is needed for the curve to pass the threshold. Lower values of k correlate with slower rates of decay and a longer time period τ to cross the threshold. Concentrations of molecules that dip below the threshold are expected to change states.

In Figure 11a the relationship between the rate of decay k and the time τ that it takes to cross the threshold m is shown. A system that involves molecules with high decay rates, such as $k = 8$, will have faster response times, $\tau \leq 0.2$. In comparison, systems with molecules that take a longer time to decay $k = 1$ will take longer to change from state to state in a system, $\tau > 0.2$. In reality, the rate at which proteins can be produced is limited by the speed of transcription and translation as well as the capacity for ribosomes to assemble amino acids [4].

Although the threshold for state changes to occur does not vary significantly [4], we show how the threshold, m, can affect the time τ to change states in Figure 11b. Here, we see that when m is approximately 90% of the molecules present $\tau_{.9}$, will be much smaller than $\tau_{.1}$ when m is approximately 10%. Therefore, when m is high, the rate of decay k will have a smaller impact on the network than when m is low.

Overall, Figures 10 and 11 illustrate that GRNs with the ability to change the number of molecules quickly will have faster response times to any environmental changes.

In addition, GRNs that only require small changes in the number of molecules in order to trigger a state change will also have faster response times. Fast response times in a system lead to lower fluctuations and thus lower cell-cell variability [4]. These figures show what we expect to occur in a GRN without stochastic behavior. Since it is well known that regulatory networks have stochastic behavior [2, 4, 12, 20, 28], finding ways to incorporate stochasticity and variation within mathematical models may prove beneficial such that they may be better equipped for capturing variability in the future.

In Figures 12, 13, and 14, two sets of points are shown in each graph. In these figures we see the proportion of genes that have changed state by the indicated time step τ on the x-axis. The genes change states at different time steps because k varies uniformly from 0.1 to 10. The light gray set of points represent the deterministic outcome of a gene transition. The color coded sections of each of these figures represents 5% intervals of genes. The majority of the genes transition into a new state within the first quarter of total time steps. In Figure 12 it takes roughly 20 time steps for all of the genes to transition. This is considerably more time steps compared to the roughly 7 time steps in Figure 13 and roughly 1 time step in Figure 14. In Figure 15 we see again that when $m = 0.1$ it takes much longer for all of the genes to transition and also that there are more genes with higher variance compared to genes with a threshold of $m = 0.9$.

The variation of each of these 5% intervals is plotted in Figure 16. In Figure 16a the threshold used is $m = 0.1$ and in Figure 16b the threshold used is $m = 0.9$. This figure shows that the longer it takes for a gene to react to its surroundings, the more variation will exist during gene transitions.

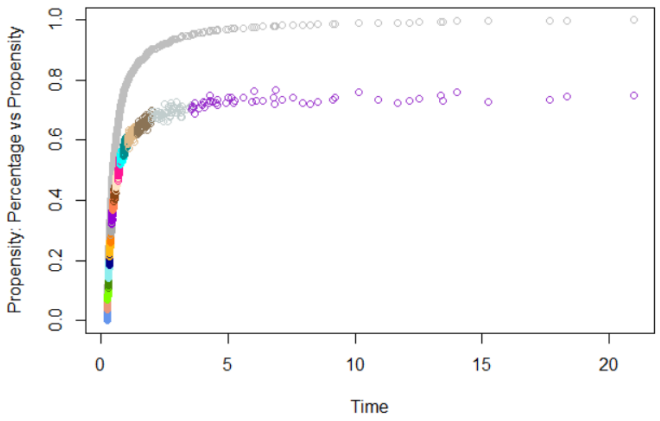

Figure 12: The proportion of genes that have changed states at time τ when $m = 0.1$.

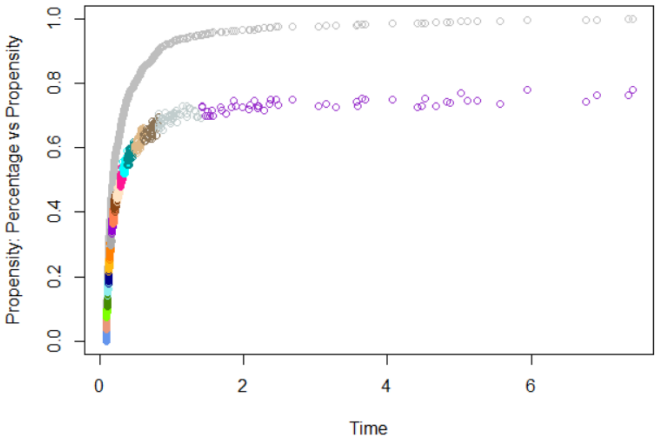

Figure 13: The proportion of genes that have changed states at time τ when $m = 0.4$.

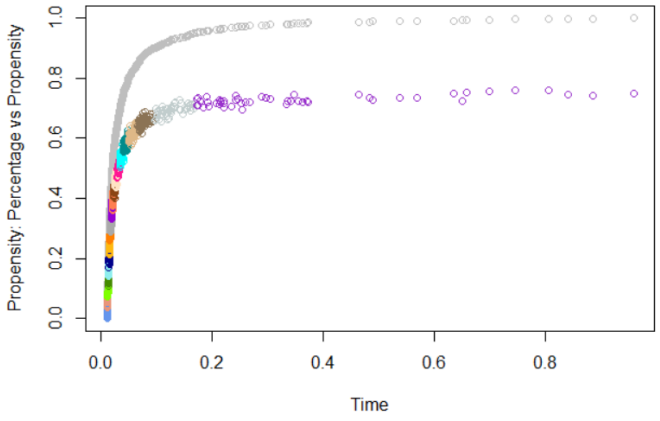

Figure 14: The proportion of genes that have changed states at time τ when $m = 0.9$.

3.2 Results Method 2: Beta distributed propensities

3.2.1 Effects on variation

In the following sections we discuss how different parameters and values effect the variation of a state transition. First we see how the two actions, one changing shape parameters of the beta distribution and two changing the expected value affect variation of transitions.

Case I: Holding Expected Value Constant
We explore how the variation of transitions is affected by different shape parameters when the expected value of an entry is held constant. In Figure 17 we see the relationship between shape parameters and variation when the entry X has 4 different expected values with corresponding gene transition probabilities (a), (b), (c), and (d). These figures reveal that high variation occurs when the shape parameter of one gene is high and the other is low.

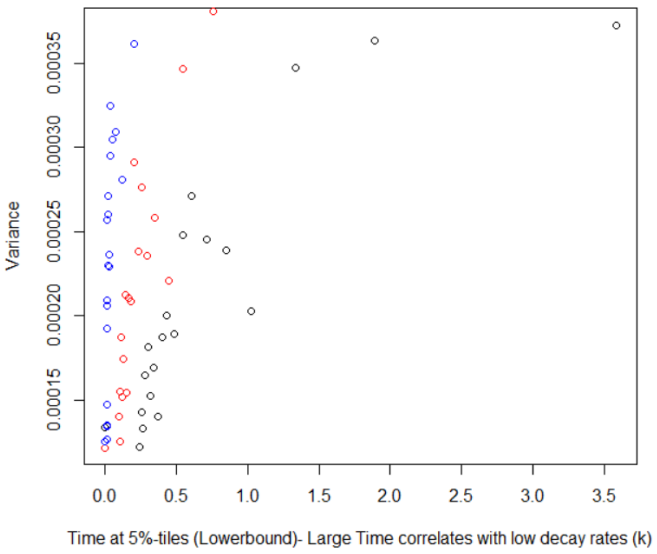

Figure 15: Graphs of the variation at every fifth percentile for $m = 0.1$ in black, $m = 0.4$ in red, and $m = 0.9$ in blue.

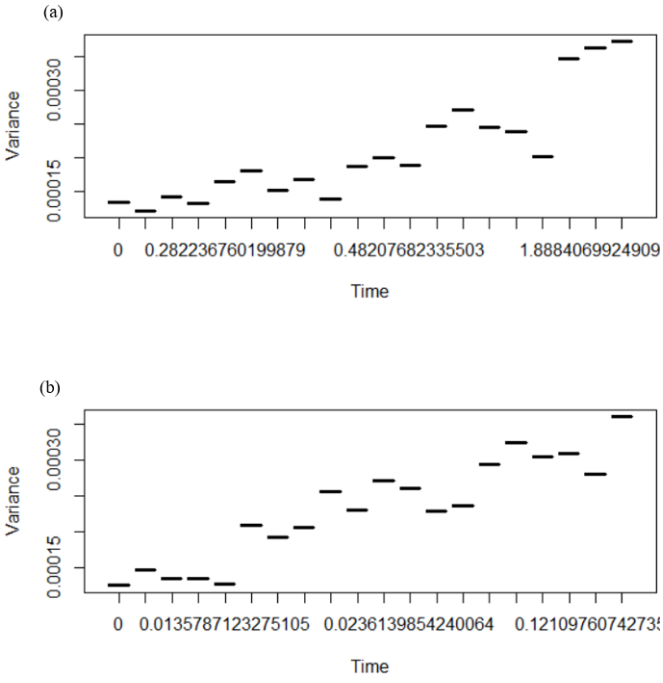

Figure 16: How variation changes with respect to the time it takes for the transition function to cross the threshold. (a) Threshold $m = 0.1$; (b) Threshold $m = 0.9$.

High variation also appears to be correlated with higher expected values (17c–17d). Since the shape of beta distributions is dependent on the shape parameters, there is no set distribution for each of the graphs being represented. The variation for each of these figures indicates that variation is highest when the shape parameter of gene 1 α_1 is high and the shape parameter of gene 2 α_2 is low. A closer look reveals that when the expected value of a transition is low for both genes, Figure 17a, the maximum variance is less than 0.1. However, when the expected value for both genes is high, Figure 17d, the maximum variance is greater than 10.

Case II: Holding Shape Parameters Constant
Here we look at how the variation is affected by changes in expected value, but when one of the shape parameters is held constant. Here, the relationship between α, β, and the expected value allow us to look at how changes in expected value affect the variation in a transition between gene states. In Figure 18 the x-axis is the expected value of gene 1 and the y-axis is the expected value of gene 2. We see that in general, when one gene has an expected value of approximately 0.5 and the other has an expected value close to 1 the variance of the transition will be high.

3.2.2 Variance of the GRN

Given independence between each of the gene transitions in the network, the variance of the entire network will have a covariance of zero and the network's variance will equal the variance of the product of the two beta distributions. The variance of the entire network is shown in Table 4.

Recall that the variation of each edge of two beta distributions with shape parameters (a_1, b_1) and (a_2, b_2) for genes 1 and 2 respectively is

$$\frac{{E_1}^2(1+a_1){E_2}^2(1+a_2)}{(E_1+a_1)(E_2+a_2)} - (E_1E_2)^2.$$

3.3 Results Method 3: Spectral density

Figures 4 and 5 show the behavior of variation for different values of k_*. The real variation is shown in Figure 4. This variation is linear and negative. The magnitude of variation is greater for lower values of k_*. In Figure 5 variation takes on a logarithmic shape and is also negative. The magnitude of variation decreases as k_* increases.

Figure 9 shows the variation from three different beta distributions, $\alpha = 4$ and $\beta = 1$, $\alpha = 2$ and $\beta = 2$, and $\alpha = 3$ and $\beta = 5$. Similar to the uniform distribution, there is a real and complex component to each of these distributions. The real component for each of these distributions is also linear, and the complex component is

logarithmic. All variation shown is negative. All variation decreases as the magnitude of k_* decreases.

4 Discussion

4.1 Method 1: Transition function

In Figure 10 we show how the rate of decay effects the time required for the curve to pass the threshold m. We see that larger values of k lead to a shorter time tau for the curve to pass the threshold. On the other hand, smaller values of k create a longer time period to cross the threshold. We also look at the relationship between the rate of decay and the time that it takes to cross the threshold in Figure 11. Again, higher rates of decay indicate a faster response time for change in gene states. We also found that higher thresholds are more responsive to change than lower thresholds and also less affected by changes in decay rates.

Later, in Figures 12, 13, and 14, we include stochasticity while using the transition function and by using a binomial distribution where there is a 75% probability that genes which are supposed to change states actually do change states. [2] argue that the use of the binomial distribution to determine the outcome of genes likely overestimates the variation of transitions [2], but this distribution was used for the sake of a preliminary investigation and could easily be changed in the future. Overtime, we see that the number of genes that transition has a propensity of 75%, which is the given probability of the binomial distribution used to determine the outcome of each gene.

In Figure 16, we see that as time increases, the variation present in transitions also increases. This is true regardless of the threshold value. This would indicate that cells with genes that can respond quickly to their surroundings are more likely to have lower variability between cells. This is consistent with literature that indicates that fast response times in a system lead to lower fluctuations and thus lower cell-cell variability [4]. It has been shown that frequent transcriptions with fewer proteins per transcription lead to networks with less variation, whereas, less frequent transcriptions and larger protein yield per transcription results in more noise [27]. Low protein yield per transcription is energy inefficient [27]. High protein yield per transcription leads to more variation but is energy efficient [27].

4.2 Method 2: Beta distributed propensities

In Figure 17 we find that large differences in shape parameters and high expected values for transitions lead to greater variation in gene state transitions. We also find that a mix of intermediate and high expected values leads to greater variation than low expected values in Figure 18. There is currently no known biological explanation for this behavior in the literature. We hypothesize that genes that are acting effectively, and thus have a high expected value or high propensity and are working appropriately, are able to handle more variation without a decrease in fitness compared to genes that are inefficiently working. It may also be unlikely to see a mixture of intermediate and high propensities in a system, since most systems were likely selected to have lower variation through evolution.

One of the benefits to using the beta distributed activation and degradation propensity method is that no simulation of the network is needed. There is also a natural extension to the beta distributed activation and degradation propensity method. When shape parameter $a_B = a_A + b_A$, the distribution is a bivariate beta distribution according to a lemma discussed in [21] and [30]. [21] also showed that this property extends to multiple independent beta random variables [21]. "If U_1, U_2, $\ldots U_p$ are independent beta random variables with shape parameters (a_i, b_i), $i = 1, 2, \ldots, p$ and if $a_{i+1} = a_i + b_i$, $i = 1, 2, \ldots, p-1$, then the product $U_1 * U_2 * \ldots U_p$ is also a beta random variable with shape parameters $(a_1,\ b_1 + \ldots + b_p)$" [21].

[30] discuss the relationships among two independent beta distributions using this lemma but leave the relationships created from multiple independent beta distributions for future work [30]. Although we limit ourselves to two genes for the majority of our work, it is more realistic to consider cases with more than 2 genes interacting and so exploring these properties between 3 and more beta distributions may yield interesting results. In fact, the bivariate beta distribution has been used to model proportions of alleles in population dynamics [40]. Therefore, future use of the bivariate beta distribution may lead to promising results and insights.

4.3 Method 3: Spectral density

In Figures 4, 5, and 9 we see the behavior of variation for a variety of distributions. In all cases, the real component of the PSD is linear and the complex component logarithmic. This indicates that the distribution of k does not have a large impact on the variation of transitions. There is some literature that agrees with the results shown here. Stochasticity in gene expression is not strongly dependent on the statistical distribution of transcription initiation [27]. Stochasticity will occur when there are a limited number of promoters for gene regulation in the cell and the time interval for transcription time is longer [27].

All of the variation represented through this method is less than zero. This is likely due to the fact that the

Fourier transform involves complex values and that $i^2 = -1$. Interpretation of negative variation is not common, but there is some evidence that the magnitude of variation is more important than the sign [9]. Likewise, complex, or imaginary, variation is also difficult to interpret. The resulting negative and complex variation for all cases is a possible downfall to this modeling process and future investigation is needed to see if this behavior is prevalent across all distributions.

5 Conclusion

Collectively, each of these modeling techniques provides different qualitative information about gene state transitions. Integrating these models and the information that they provide into future models could help to create more accurate models overall. However, these models do not capture the larger picture of interwoven networks that each impact one another. Separating networks mathematically does not provide an accurate view of an organism's complex system of networks each impacting one another.

Appendix

Justification of Method 3

Justification for (1)

A stochastic matrix is an $n \times n$ matrix such that each entry is less than or equal to 1 and each row sum is equal to 1 [7]. In the Transition Matrix, $0 \leq p_i^{\uparrow} \leq 1$ and $0 \leq p_i^{\downarrow} \leq 1$. Therefore, $0 \leq (1 - p_i^{\uparrow}) \leq 1$, and $0 \leq (1 - p_i^{\downarrow}) \leq 1$. Let x be the product of any of these terms. Then $0 \leq x \leq 1$. Furthermore,

$$(1-p_1^{\uparrow})(1-p_2^{\uparrow}) + (1-p_1^{\uparrow})p_2^{\uparrow} + p_1^{\uparrow}(1-p_2^{\uparrow}) + p_1^{\uparrow}p_2^{\uparrow} = 1,$$
$$(1-p_1^{\uparrow})p_2^{\downarrow} + (1-p_1^{\uparrow})(1-p_2^{\downarrow}) + p_1^{\uparrow}p_2^{\downarrow} + p_1^{\uparrow}(1-p_2^{\downarrow}) = 1,$$
$$p_1^{\downarrow}(1-p_2^{\uparrow}) + p_1^{\uparrow}p_2^{\downarrow} + (1-p_1^{\downarrow})(1-p_2^{\uparrow}) + (1-p_1^{\downarrow})p_2^{\uparrow} = 1,$$
$$p_1^{\downarrow}p_2^{\downarrow} + p_1^{\downarrow}(1-p_2^{\downarrow}) + (1-p_1^{\downarrow})p_2^{\downarrow} + (1-p_1^{\downarrow})(1-p_2^{\downarrow}) = 1.$$

Justification for (2)

Let $A \sim Beta(a_A, b_A)$ and $Z = 1 - A$. The probability density function of A is

$$f_A(A) = \frac{(A)^{a_A - 1} * (1-A)^{b_A - 1}}{\mathcal{B}[a_A, b_A]}.$$

The transformation of $Z = 1 - A$ creates the new pdf

$$f_Z(Z) = \frac{(1-Z)^{a_A - 1} * \big(1 - (1-Z)\big)^{b_A - 1}}{\mathcal{B}[a_A, b_A]}.$$

Which simplifies to

$$f_Z(Z) = \frac{(1-Z)^{a_A - 1} * (Z)^{b_A - 1}}{\mathcal{B}[a_A, b_A]}.$$

This implies that Z has a beta distribution with shape parameters b_A and a_A.

$$Z \sim Beta(b_A, a_A)$$

Therefore the distribution of $Z = 1 - A$, where A is a beta distribution, is also a beta distribution.

Justification of (3)

The pdf of $f(x, y)$ when $X = A * B$ and $Y = B$ where both A and B are beta distributions. $A \sim Beta(a_A, b_A)$ and $B \sim Beta(a_B, b_B)$ is shown below and obtained though a transformation of variables.

$$f(x,y) = \frac{\left(\frac{x}{y}\right)^{a_A - 1} * \left(1 - \frac{x}{y}\right)^{b_A - 1} * (y)^{a_B - 1} * (1-y)^{b_B - 1}}{\mathcal{B}[a_A, b_A] * \mathcal{B}[a_B, b_B] * y}$$

for $0 < x \leq y$ and $0 < y < 1$.

$$\int_x^1 f(x,\, y)\, dy$$
$$= \Big(\frac{x^{a_2}\Gamma[b_1]\Gamma[a_1 - a_2]\mathrm{HypG}_{2F1}[1 - b_2, 1 - a_1 - b_1 + a_2, 1 - a_1 + a_2, x]}{\Gamma[a1 + b1 - c]}$$
$$+ \frac{x^{a_1}\Gamma[b_2]\Gamma[-a_1 + a_2]\mathrm{HypG}_{2F1}[1 - b_1, 1 + a_1 - b_2 - a_2, 1 + a_1 - a_2, x]}{\Gamma[-a1 + b_2 + a_2]} \Big)$$
$$* \frac{1}{\big(x\mathcal{B}[a_1, b_1]\mathcal{B}[a_2, b_2]\big)}$$

Justification of (4)

Expected value is a positive linear operator. Results follow.

Justification of (5)

The expected value of X is dependent on the shape parameters of $A \sim Beta\,(a_A, b_A)$ and $B \sim Beta\,(a_B, b_B)$.

$$\int_0^1 \int_0^y x * f\,(x,y)\; dx\, dy = \frac{a_A \Gamma\,[1 + a_B]\,\Gamma\,[a_B + b_B]}{(a_A + b_A)\,\Gamma\,[a_B]\,\Gamma\,[1 + a_B + b_B]}$$

The expected value of X is also the same as the product of expected values of A and B. This follows naturally because A and B are independent, but this property can also be motivated by the relationship between the shape parameters.

Recall that the expected value of A is $\frac{a_A}{a_A + b_A}$ and therefore, based on a proportional relationship between a and b, the expected value E_A or E_B creates the relationship: $b_A = \frac{1 - E_A}{E_A} a_A$ and similarly $b_B = \frac{1 - E_B}{E_B} a_B$

Substituting this expression into the expected value of X for b_A or b_B leads to the expression

$$\frac{a_A \Gamma[1 + a_B]\Gamma\left[a_B + \frac{1 - E_B}{E_B} a_B\right]}{\left(a_A + \frac{1 - E_A}{E_A} a_A\right)\Gamma[a_B]\Gamma\left[1 + a_B + \frac{1 - E_B}{E_B} a_B\right]},$$

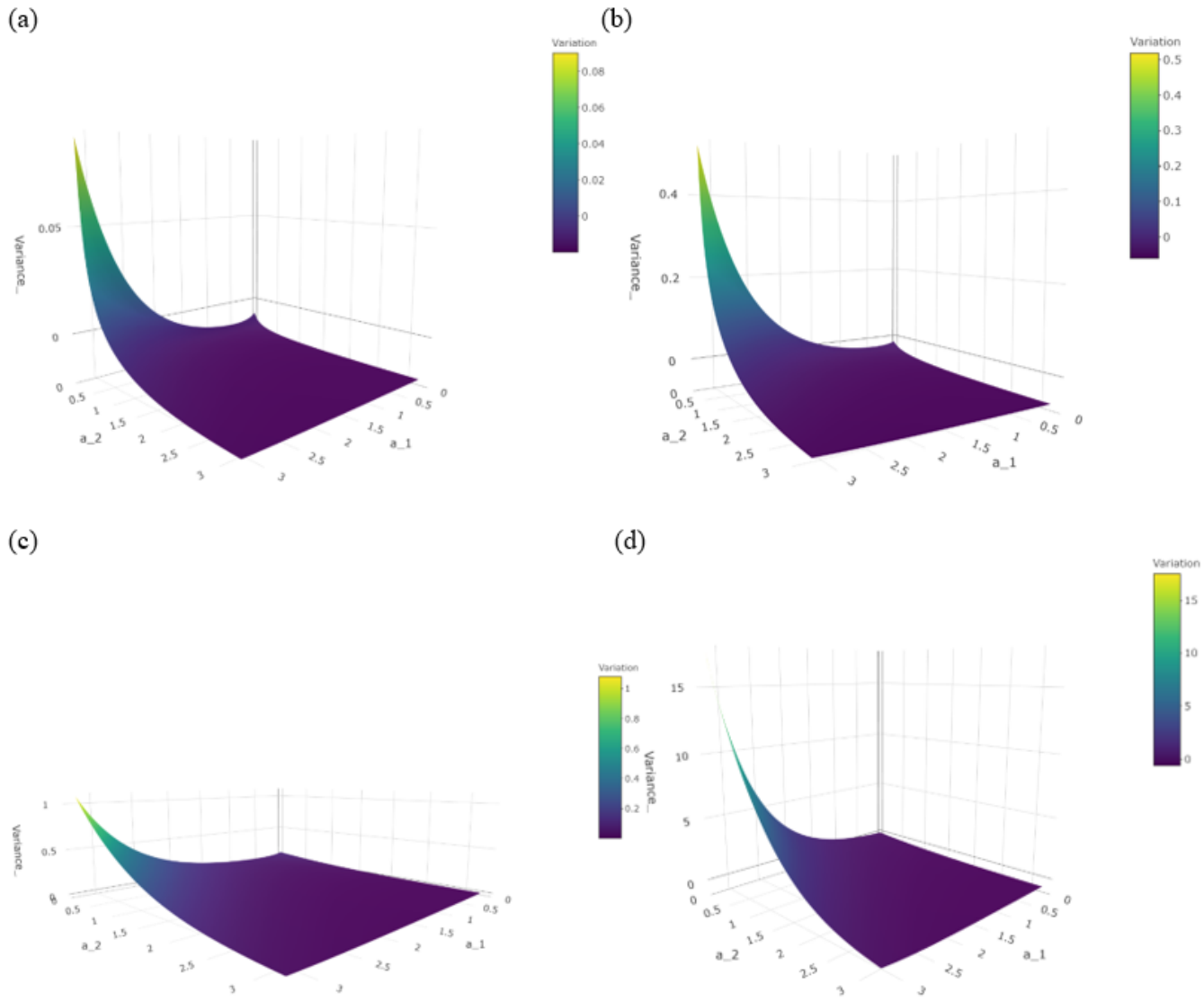

Figure 17: (a) $E(X) = 0.02$, the expected value of the transition in gene 1 is 0.2, and the expected value of the transition in gene 2 is 0.1. (b) $E(X) = 0.06$, the expected value of the transition in gene 1 is 0.3, and the expected value of the transition in gene 2 is 0.2. (c) $E(X) = 0.56$, the expected value of the transition in gene 1 is 0.8, and the expected value of the transition in gene 2 is 0.7. (d) $E(X) = 0.63$, the expected value of the transition in gene 1 is 0.9, and the expected value of the transition in gene 2 is 0.7.

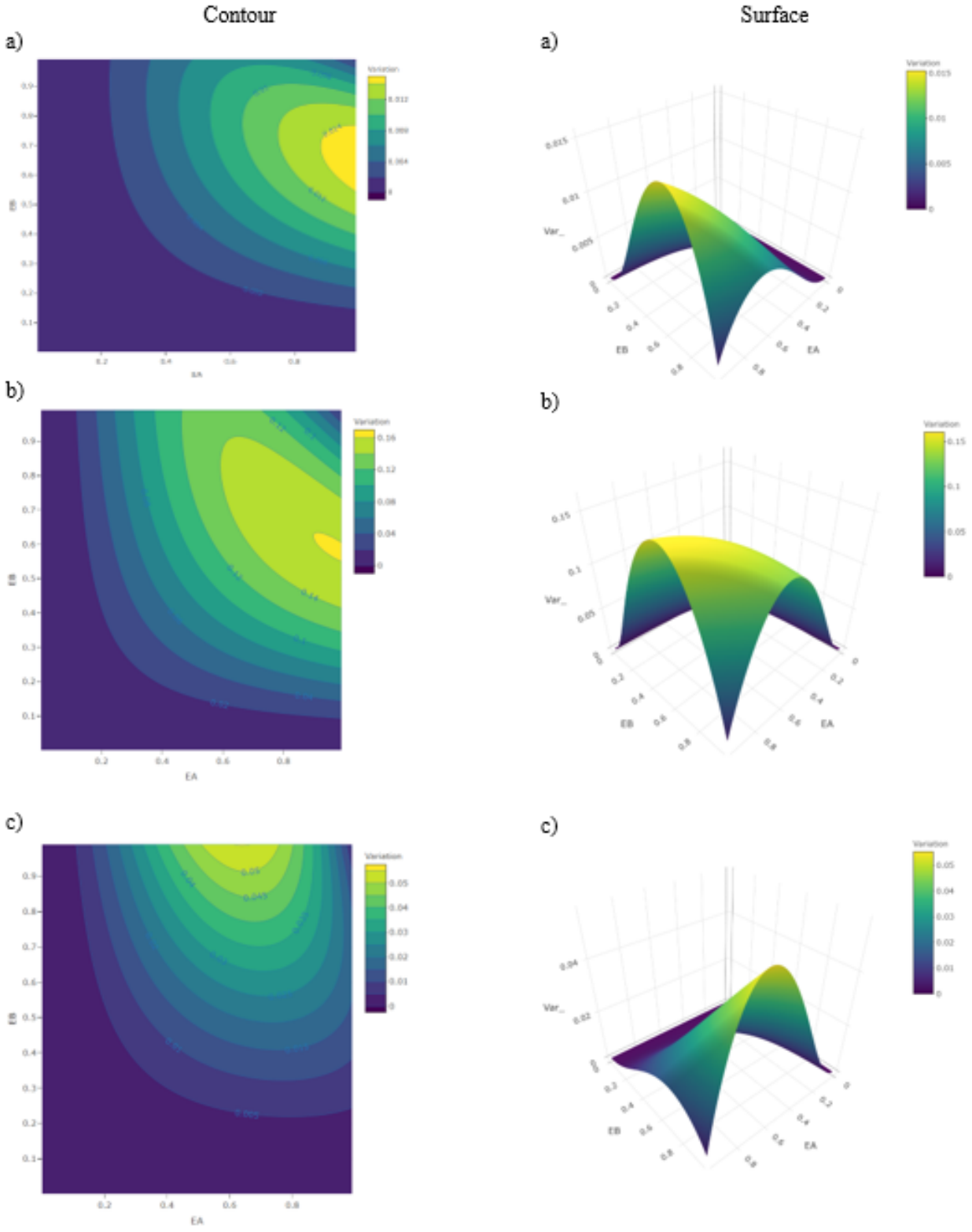

Figure 18: Three surfaces measuring variation when the expected value varies. Different shape parameters are used for each set of figures: (a) $a_a = 20$ and $a_b = 9$; (b) $a_a = 0.5$ and $a_b = 0.3$; (c) $a_a = 2$ and $a_b = 9$.

Table 3: Transition Matrix of State Update Propensities.

Input \ Output	00	01	10	11
00	$(1-p_1^{\uparrow})(1-p_2^{\uparrow})$	$(1-p_1^{\uparrow})(p_2^{\uparrow})$	$(p_1^{\uparrow})(1-p_2^{\uparrow})$	$(p_1^{\uparrow})(p_2^{\uparrow})$
01	$(1-p_1^{\uparrow})(p_2^{\downarrow})$	$(1-p_1^{\uparrow})(1-p_2^{\downarrow})$	$(p_1^{\uparrow})(p_2^{\downarrow})$	$(p_1^{\uparrow})(1-p_2^{\downarrow})$
10	$(p_1^{\downarrow})(1-p_2^{\uparrow})$	$(p_1^{\downarrow})(p_2^{\uparrow})$	$(1-p_1^{\downarrow})(1-p_2^{\uparrow})$	$(1-p_1^{\downarrow})(p_2^{\uparrow})$
11	$(p_1^{\downarrow})(p_2^{\downarrow})$	$(p_1^{\downarrow})(1-p_2^{\downarrow})$	$(1-p_1^{\downarrow})(p_2^{\downarrow})$	$(1-p_1^{\downarrow})(1-p_2^{\downarrow})$

Table 4: Variation of the entire GRN.

00	01	10	11
$\mathrm{V}((1-p_1^{\uparrow})(1-p_2^{\uparrow}))$	0	0	0
0	$\mathrm{V}((1-p_1^{\uparrow})(1-p_2^{\downarrow}))$	0	0
0	0	$\mathrm{V}((1-p_1^{\downarrow})(1-p_2^{\uparrow}))$	0
0	0	0	$\mathrm{V}((1-p_1^{\downarrow})(1-p_2^{\downarrow}))$

which can be reduced to

$$\frac{a_A a_B \Gamma[a_B]\Gamma\left[a_B+\frac{1-E_B}{E_B}a_B\right]}{a_A\left(1+\frac{1-E_A}{E_A}\right)\Gamma\left[a_B\right]\left(a_B+\frac{1-E_B}{E_B}a_B\right)\Gamma\left[a_B+\frac{1-E_B}{E_B}a_B\right]}$$
$$=\frac{a_A a_B \Gamma[a_B]\Gamma\left[a_B+\frac{1-E_B}{E_B}a_B\right]}{a_A\left(1+\frac{1-E_A}{E_A}\right)\Gamma[a_B]a_B\left(1+\frac{1-E_B}{E_B}\right)\Gamma\left[a_B+\frac{1-E_B}{E_B}a_B\right]}$$
$$=\frac{1}{\left(1+\frac{1-E_A}{E_A}\right)\left(1+\frac{1-E_B}{E_B}\right)}=\frac{E_A E_B}{1}=E_A E_B.$$

Justification for (6)

The variance of X is also dependent on the shape parameters of $A\sim Beta(a_A,b_A)$ and $B\sim Beta(a_B,b_B)$

$$Var(X)=\frac{\Gamma[2+a_A]\Gamma[b_B]\Gamma[b_A]\Gamma[2+a_B]}{\beta[a_A,b_A]\beta[a_B,b_B]\Gamma[2+a_A+b_A]\Gamma[2+a_B+b_B]}-E(X)^2$$

As before, we substitute the expressions $b_A=\frac{1-E_A}{E_A}a_A$ and $b_B=\frac{1-E_B}{E_B}a_B$ into the variance of X and we obtain the expression

$$\frac{\Gamma[2+a_A]\Gamma\left[\frac{1-E_B}{E_B}a_B\right]\Gamma\left[\frac{1-E_A}{E_A}a_A\right]\Gamma[2+a_B]}{\beta\left[a_A,\frac{1-E_A}{E_A}a_A\right]\beta\left[a_B,\frac{1-E_B}{E_B}a_B\right]\Gamma\left[2+a_A+\frac{1-E_A}{E_A}a_A\right]\Gamma\left[2+a_B+\frac{1-E_B}{E_B}a_B\right]}-(E_A E_B)^2$$

$$=\frac{(1+a_A)(1+a_B)}{\left(1+a_A+\frac{1-E_A}{E_A}a_A\right)\left(1+\frac{1-E_A}{E_A}\right)\left(1+\frac{1-E_B}{E_B}\right)\left(1+a_B+\frac{1-E_B}{E_B}a_B\right)}-(E_A E_B)^2$$

$$=\frac{{E_A}^2(1+a_A){E_B}^2(1+a_B)}{(E_A+a_A)(E_B+a_B)}-(E_A E_B)^2.$$

Equations

The gamma function

$$\Gamma(z)=\int_0^\infty x^{(z-1)}e^{-x}\,dx.$$

When z is an integer n, the gamma function is also equivalent to

$$\Gamma(n)=(n-1)!.$$

The beta function

$$\mathcal{B}(x,y)=\int_0^1 t^{x-1}(1-t)^{y-1}\,dt.$$

The beta function is also equivalent to

$$\mathcal{B}(x,y)=\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

The Gauss hypergeometric function

$$HypG_{2F1}(a,b;c;x)=\sum_{k=0}^{\infty}\frac{(a)_k(b)_k x^k}{(c)_k k!}.$$

The bivariate beta Distribution

$$f(x,y)=\frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)}x^{a-1}y^{b-1}(1-x-y)^{c-1}$$

where $a,b,c>0,\ x\geq 0,\ y\geq 0$ *and* $x+y\ \leq\ 1$.

Equation of the probability density function of the product of two beta distributions with shape parameters (a_1,b_1) and (a_2,b_2)

$$\left(\frac{x^{a_2}\Gamma[b_1]\Gamma[a_1-a_2]\mathrm{HypG}_{2F1}[1-b_2,1-a_1-b_1+a_2,1-a_1+a_2,x]}{\Gamma[a1+b1-c]}\right.$$
$$\left.+\frac{x^{a_1}\Gamma[b_2]\Gamma[-a_1+a_2]\mathrm{HypG}_{2F1}[1-b_1,1+a_1-b_2-a_2,1+a_1-a_2,x]}{\Gamma[-a1+b_2+a_2]}\right)$$
$$*\frac{1}{\left(x\mathcal{B}[a_1,b_1]\mathcal{B}[a_2,b_2]\right)}.$$

Author Contributions

O. A. and D. H. conceived of the presented idea. S. L. investigated the theory and developed the models. O. A., D. H., and S. L. collaborated, developed, and verified the statistical methods and models equally.

References

[1] Abou-Jaoudé, W., Ouattara, D. A., & Kaufman, M. (2009). From structure to dynamics: frequency tuning in the p53-Mdm2 network I. Logical approach. *Journal of theoretical biology 258*, 561–577.

[2] Akman, O., Comar, T., Harris, A. L., Hrozencik, D., & Li, Y. (2018). Dynamics of gene regulatory networks with stochastic propensities. *International Journal of Biomathematics*, *11*(3), 1–13. https://doi.org/10.1142/S1793524518500328

[3] Alakwaa, F. M. (2015). Modeling of Gene Regulatory Networks: A Literature Review. *Journal of Computational Systems Biology*, *1*(1), 1–8. https://doi.org/10.15744/2455-7625.1.102

[4] Alon, U. (2006). An Introduction to Systems Biology: Design Principles of Biological Circuits (1st ed.). Boca Raton, FL: Chapman & Hall/CRC.

[5] Asadzadeh, S., & Roberto de Souza Filho, C. (2016). A review on spectral processing methods for geological remote sensing. *International Journal of Applied Earth Observation and Geoinformation, 47*, 69–90.

[6] Ciliberto, A., Novak, B., & Tyson, J. J. (2005). Steady States and Oscillations in the p53/Mdm2 network. *Cell Cycle, 4*(3), e107–e112.

[7] Dobrow, R. P. (2016). Introduction to Stochastic Processes With R (1st ed.). Hoboken, New Jersey: John Wiley and Sons.

[8] Emmert-Streib, F., Dehmer, M., & Haibe-Kains, B. (2014). Gene regulatory networks and their applications: Understanding biological and medical problems in terms of networks. *Frontiers in Cell and Developmental Biology, 2*, 1–7. https://doi.org/10.3389/fcell.2014.00038

[9] Firth, B. D., & Menezes, R. de. (2004). Quasi-variances.*Biometrika, 91*(1), 65–80.

[10] Gehrmann, E., & Drossel, B. (2010). Boolean versus continuous dynamics on simple two-gene modules. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 82*(4). https://doi.org/10.1103/PhysRevE.82.046120

[11] Ghosh, I. (2019). On the Reliability for Some Bivariate Dependent Beta and Kumaraswamy Distributions: A Brief Survey. *Stochastics and Quality Control*, 1–7. https://doi.org/10.1515/eqc-2018-0029

[12] Gillespie, D. T. (2007). Stochastic Simulation of Chemical Kinetics. *Annual Review of Physical Chemistry*, *58*(1), 35–55.

[13] Ginsburg, G. S., & Phillips, K. A. (2018). Precision medicine: From science to value. *Health Affairs, 37*(5), 694–701.

[14] Gowda, T., Vrudhula, S., & Kim, S. (2009). Modeling of gene regulatory network dynamics using threshold logic. In G. Stolovitzky, P. Kahlem, & A. Califano (Eds.) *The challenges of systems biology: Community efforts to harness biological complexity*, *1158*. New York: Wiley-Blackwell.

[15] Hu, W., Feng, Z., & Levine, A. J. (2012). The Regulation of Multiple p53 Stress Responses is Mediated through MDM2. *Genes and Cancer, 3*(3–4), 199–208. https://doi.org/10.1177/1947601912454734

[16] Jambunathan, M. V. (1954). Some Properties of Beta and Gamma Distributions. *The Annals of Mathematical Statistics*, *25*(2), 401–405. https://doi.org/10.1214/aoms/1177728800

[17] Jong, H. De. (2002). Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. *Journal of Computational Biology*, *9*(1), 67–103.

[18] Jorna, P. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological Psychology*, *34*(2–3), 237–257.

[19] Kauffman, S. (1969). Homeostasis and differentiation in random genetic control networks. *Nature*, *244*, 177–178.

[20] Klymkowsky, M., & Cooper, M. (2015). Biofundamentals 2.0.

[21] Krysicki, W. (1999). On some new properties of the beta distribution. *Statistics and Probability Letters*, *42*, 131–137. https://doi.org/10.1515/dema-1999-0318

[22] Lee, W. P., & Tzou, W. S. (2009). Computational methods for discovering gene networks from expression data. *Briefings in Bioinformatics, 10*(4), 408–423. https://doi.org/10.1093/bib/bbp028

[23] Lipshtat, A, Neves,S. R., & Iyengar, R. (2009). Specification of spatial relationships in directed graphs of cell signaling networks. In G. Stolovitzky, P. Kahlem, & A. Califano (Eds.) *The challenges of systems biology: Community efforts to harness biological complexity 1158*. New York: Wiley-Blackwell.

[24] Manioudaki, M., Tzamali, E., Reczko, M., & Poirazi, P. (2009). Bioinformatics for Systems Biology (1st ed.; S. Krawetz, ed.). https://doi.org/10.1017/CBO9781107415324.004

[25] Marhon, S. M., & Kremer, S. C. (2011). Gene prediction based on DNA spectral analysis: A literature review. *Journal of Computational Biology, 18*(4).

[26] Mathur, S., & Sutton, J. (2017). Personalized medicine could transform healthcare (Review). *Biomedical Reports, 7*(1), 3–5. https://doi.org/10.3892/br.2017.922

[27] McAdams, H., & Arkin, A. (1997). Stochastic Mechanisms in Gene Expression. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(3), 814–819.

[28] Murrugarra, D., Veliz-Cuba, A., Aguilar, B., Arat, S., & Laubenbacher, R. (2012). Modeling stochasticity and variability in gene regulatory networks. *Eurasip Journal on Bioinformatics and Systems Biology*, *2012*(1), 5. https://doi.org/10.1186/1687-4153-2012-5

[29] Nadarajah, S. (2005). Reliability for some bivariate beta distributions. *Mathematical Problems in Engineering*, *2005*(1), 101–111. https://doi.org/10.1155/MPE.2005.101

[30] Nadarajah, S., & Kotz, S. (2005). Some bivariate beta distributions. *Statistics: A Journal of Theoretical and Applied Statistics*, *39*(5), 457–466. https://doi.org/10.1080/02331880500286902

[31] Olkin, I., & Trikalinos, T. A. (2015). Constructions for a bivariate beta distribution. *Statistics and Probability Letters*, *96*(January), 54–60. https://doi.org/10.1016/j.spl.2014.09.013

[32] Przulj, N. (2011). "Biological networks uncover evolution, disease, and gene functions," In P. Pevzner & R. Shamir (Eds.) *Bioinformatics for Biologists*, pp. 291–314. Cambridge: Cambridge University Press.

[33] Schwartz, R. (2011). "Regulatory network inference," In P. Pevzner & R. Shamir (Eds.) *Bioinformatics for Biologists*, pp. 315–343. Cambridge: Cambridge University Press.

[34] Song, M., Lewis, C. K., Lance, E. R., Chesler, E. J., Yordanova, R. K., Langston, M. A., Bergeson, S. E. (2009). Reconstructing generalized logical networks of transcriptional regulation in mouse brain from temporal gene expression data. *Eurasip Journal on Bioinformatics and Systems Biology*, *2009*. https://doi.org/10.1155/2009/545176

[35] Stark, J., Brewer, D., Barenco, M., Tomescu, D., Callard, R., & Hubank, M. (2003). Reconstructing gene networks: What are the limits? *Biochemical Society Transactions*,*31*(6), 1519–1525. https://doi.org/10.1042/bst0311519

[36] Stoica, P., & Moses, R. (2005). Spectral analysis of signals (T. Robbins, ed.). Upper Saddle River, New Jersey: Prentice-Hall.

[37] Thibos, L. N. (2014). Fourier analysis for beginners (6th ed.). Retrieved from http://hdl.handle.net/2022/21365

[38] Van Laar, J., Porath, M., Peters, C., & Oei, S. (2008). Spectral analysis of fetal heart rate variability for fetal surveillance: review of the literature. *Acta Obstetricia et Gynecologica Scandinavica*, *87*(3), 300–306.

[39] Waning, D. L., Lehman, J. A., Batuello, C. N., & Mayo, L. D. (2010). "Controlling the Mdm2-Mdmx-p53 circuit". *Pharmaceuticals*, *3*(5), 1576–1593. https://doi.org/10.3390/ph3051576

[40] Wright, S. (1937). The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences of the United States of America*, *23*(6), 307–320. https://doi.org/10.1126/science.85.2212.504

[41] Yang, E., van Nimwegen, E., Zavolan, M., Rajewsky, N., Schroeder, M., Magnasco, M., & Darnell, J. E. (2003). Decay rates of human mRNAs: Correlation with functional characteristics and sequence attributes. *Genome Research*, *13*(8), 1863–1872. https://doi.org/10.1101/gr.1272403

[42] Yates, C. A., Ford, M. J., & Mort, R. L. (2017). A Multi-stage Representation of Cell Proliferation as a Markov Process. *Bulletin of Mathematical Biology*, *79*(12), 2905–2928. https://doi.org/10.1007/s11538-017-0356-4