

Stochastic models of genetic circuits

by

Luis Alberto Gutiérrez López

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Bachelor of Science in Physics

at the

UNIVERSIDAD DE LOS ANDES

May 12, 2016

Certified by

Juan Manuel Pedraza Leal
Associate Professor
Monograph Supervisor

Stochastic models of genetic circuits

by

Luis Alberto Gutiérrez López

Submitted to the Department of Physics
on May 12, 2016 in partial fulfillment of the
requirements for the degree of
Bachelor of Science in Physics

Abstract

All living beings store their genetic information in the DNA and use similar basic mechanisms to read it and on its basis, build their structure and develop their functions. Nevertheless, the information codified on the DNA is not the only aspect that makes an organism what it is. A big and complex network of gene regulation determines which genes are read at a particular moment and the intensity of their activity making it possible, for instance, to differentiate between our muscle cells and our neurons, very different cells but with the same genetic material. Since gene expression is mediated by reactions between molecules, which at the deepest levels consist of random collisions of the reactants, gene expression and regulation is subjected to noise. A cell also regulates the expression of many genes according to the environment, which may change randomly. In response to these important sources of noise, living beings have developed their regulatory networks to work properly under its presence. This work explores the models that have been done on the last years related to stochasticity in gene expression, the insights they have given to us into the principles of biology and the design of synthetic biological circuits.

Monograph Supervisor: Juan Manuel Pedraza Leal
Title: Associate Professor

Contents

Abstract	3
Contents	6
Introduction	11
0.1 Objetivo general	12
1 Preliminary concepts	13
1.1 The central dogma of molecular biology	13
1.2 Gene regulation and biological circuits	14
1.3 Hill functions	17
1.4 Probability	20
1.5 Noise	23
1.6 Moment generating functions	24
1.7 Characteristic function	25
1.8 Stochastic processes	25
1.9 The Gillespie algorithm	26
2 Basic genetic circuits in steady state - The master equation	29
2.1 Single gene	29
2.2 Several species with linear interactions	37
2.3 Several species with non-linear interactions - Negative autorregulation . .	41

3	Cascade of regulation - The Langevin equation	43
4	Effects of bursting and senescence	51
4.1	mRNA bursts	51
4.2	Arbitrary distribution of creation times	53
4.3	Decay of molecules	60
4.4	Senescence of mRNA	64
5	Effects of cell division	67
5.1	Characterizing the noise arising from cell division	67
5.2	Independent segregation	69
5.3	Disordered segregation	70
5.3.1	General case	70
5.3.2	Random size and random accessible volume	72
5.3.3	Clustered segregation	73
5.3.4	Upper limit of the partitioning error	77
5.4	Ordered segregation	77
5.4.1	Self-volume exclusion	77
5.4.2	Binding to spindle sites	78
5.4.3	Pair formation mechanisms	81
	Bibliography	86

List of Figures

1.1	Central dogma of molecular biology	14
1.2	Transcription factors	15
1.3	Example of gene regulation	16
1.4	Parts of a gene	16
1.5	Systems biology conventions	17
1.6	Example of a biological circuit	17
1.7	Hill function for an activator	19
1.8	Hill function for a repressor	20
2.1	Model of gene expression for a single gene	30
2.2	Transitions between states for a single gene	31
2.3	Noise in proteins: comparing analytical results and simulations	36
2.4	Transitions between states in general	37
2.5	Autoregulation simulation results	41

List of Tables

Introduction

La estocasticidad o ruido en circuitos genéticos se da debido a las fluctuaciones durante la transcripción, traducción [1] y otros procesos que afectan la expresión genética. Debido a ella, células genéticamente idénticas y expuestas al mismo ambiente presentan variaciones fenotípicas notables [1] [2] [3]. Este ruido ha sido clasificado en dos grupos: intrínseco y extrínseco [2] [4]. El primero hace referencia a la variabilidad inherente a sistemas con componentes discretos y en bajos números, como lo son en este caso el ARN y las proteínas. El segundo hace referencia a factores externos como la variabilidad del ambiente, el crecimiento y la división celular.

Recientes investigaciones han mostrado la enorme importancia que tiene el ruido para los seres vivos, los cuales han adaptado sus circuitos genéticos ya sea para cumplir su función adecuadamente a pesar de la presencia de ruido (robustez) [5], o para aprovecharlo para generar variabilidad [6]. De igual manera, en el desarrollo de circuitos genéticos sintéticos es importante considerar la estocasticidad que dicho circuito podría presentar.

Lo anterior ha motivado al desarrollo de modelos estocásticos de expresión genética en los últimos años. En el trabajo pionero de Thattai y van Oudenaarden, [7] se realizó un modelo linealizado para el ruido intrínseco en la cantidad de ARN y proteínas que puede aplicarse a varios circuitos básicos. También, Pedraza y van Oudenaarden [3] desarrollaron un modelo que incluye el ruido global y mostraron cómo el ruido total se propaga a través de una cascada de regulación.

Los modelos más recientes se han centrado en considerar otros aspectos que podrían modificar las características del ruido. Entre ellos se encuentran el *bursting* en la pro-

ducción de las moléculas involucradas en la expresión, su senescencia [8] y su repartición durante la división celular [9] [10]. Una de las conclusiones más importantes que se han obtenido a partir de estos modelos es que al considerar distintos factores, el comportamiento del ruido es similar y por lo tanto lo son también los ajustes de los modelos numéricos y los experimentos. Por lo tanto a partir de las características de las fluctuaciones no se puede saber precisamente cuáles son los mecanismos que las producen.

A pesar de que se han obtenido resultados importantes, la alta no linealidad de las ecuaciones utilizadas para representar la cinética molecular ha obligado a que los modelos realizados sean en su mayoría linealizados alrededor de los puntos fijos, perdiendo así información acerca de la dinámica completa de las fluctuaciones. Sería útil entonces desarrollar modelos estocásticos que no ignoren las no-linealidades, que incluyan la dinámica temporal completa y que integren más factores como el crecimiento y la división celular.

0.1 Objetivo general

Estudiar detalladamente los principales modelos estocásticos de expresión genética.

Chapter 1

Preliminary concepts

1.1 The central dogma of molecular biology

The central dogma explains how genetic information flows within a living being. It states that DNA, the molecule that stores the genetic information, is replicated by the enzyme DNA Polymerase. Also, RNA Polymerase produces messenger RNA (mRNA) from DNA in a process called transcription. Finally, the ribosome build the proteins following the sequence of the mRNA and according to a genetic code that translates from the language of nucleotides (the structural blocks of DNA and RNA) to the language of aminoacids, the structural blocks of proteins [11].

Proteins are the structural and functional elementary units of living beings. Therefore, according to the proteins that are being produced in a certain cell it will develop certain functions. The central dogma is summarized in fig. 1.1.

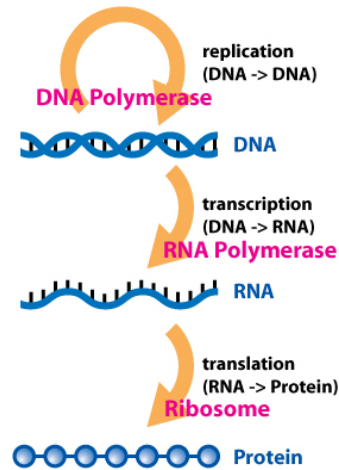


Figure 1.1: Scheme of the central dogma of molecular biology. By Dhorspool at en.wikipedia, CC BY-SA 3.0, \$3.

An important fact is that the central dogma is valid for all the living beings. The encoding of information in DNA and the mechanisms by which proteins are made according to that information, including the genetic code, are very similar between different organisms.

1.2 Gene regulation and biological circuits

(CITE SOME REVIEW)

DNA contains all the information necessary to build a living being and let him develop his functions. But genetically identical cells may differ a lot. For example, our neurons are very different in form and function than our skin cells, even though they have the same DNA and thus the same genetic information. This differentiation happens because they are expressing different sets of genes. The genetic information encoded in the DNA is called genotype, while the observable characteristics of an organism are called phenotype. In this terminology, both kinds of cells have the same genotype but differ in their phenotypes. [11]

Those differences lie on the genes that each cell is expressing at a certain time and how much they are being expressed (measured in the rate of production of proteins

corresponding to a gene). There are proteins (and even RNA molecules) that inhibit the production of other protein by stoping transcription of the corresponding gene. On the contrary, there are proteins that enhance the production of other proteins by increasing the rate of transcription. Both activators and inhibitors are called *transcription factors*, both cases can be seen on fig. 1.2.

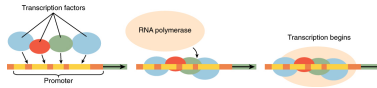


Figure 1.2: Scheme of the mechanisms of transcription factors. Retrieved from: <http://biowiki.ucdavis.edu>

From these principles of inhibition and activation, the mechanisms of gene regulation can be very complex. For example, a molecule can change the conformation of another protein that when affected by the first, inhibits the transcription of certain gene. Those molecules may be signals from the environment and with mechanisms of this type, the cell process environmental signals to express the optimal genes according to the environment. It is also important to point out that the inhibition and activation is not necessarily done individually. A certain gene may need more than one different protein to enhance its activity, or there may be genes that are activated by a protein and inhibited by others, whose production is in turn mediated by other molecules and transcription factors, a well studied case is the *lac* operon in *E. coli* whose mechanism is explained on fig. 1.3.

From the biochemical point of view, transcription factors bind specific sites on the *promoter*, a region of the DNA which is upstream the gene (or set of genes for prokaryotes), and it is where the RNA Polymerase binds to initiate transcription (see fig. ??). The binding of the transcription factor may enhance or obstruct the binding of RNA Polymerase to the promoter.

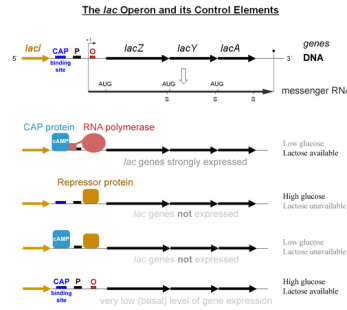


Figure 1.3: Example of gene regulation (Lac operon). Retrieved from upload.wikimedia.org/wikipedia/commons/thumb/d/d2/Lac_operon-2010-21-01.png/550px-Lac_operon-2010-21-01.png

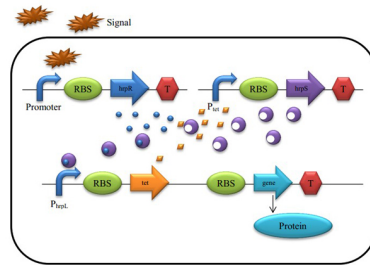


Figure 1.4: The promoter, RBS, stop codon are shown. Retrieved from http://2013.igem.org/wiki/images/c/c6/HIT-Harbin_Project_Schematic.png

Therefore, in addition to the genotype, gene expression is very important for the cells to develop properly. And, together with the genetic information, defines its structure and behavior. With this in mind, and the fact that those networks may be very large and complex, the approach that Systems Biology is proposing consists on focusing on the interactions between the different genes and components of a cell than on the details of the structure of the molecules involved. That bunch of interactions may be visualized as biological circuits, that are groups of genes that regulate each other's expression. Figure 1.5 shows some of the conventions used in the schemes of biological circuits.

$$\frac{d[nXD]}{dt} = k_+[X]^n[D] - k_-[nXD],$$

which in steady state yields

$$[X]^n[D] = \frac{k_-}{k_+}[nXD] \quad (1.1)$$

Taking the total number D_T of copies of the gene (promoter and DNA molecules) as a constant we have

$$[D_T] = [D] + [nXD]$$

Solving for the free DNA concentration $[D]$ and replacing in eq. (1.1)

$$[X]^n([D_T] - [nXD]) = \frac{k_-}{k_+}[nXD].$$

$[nXD]/[D_T]$ and $[D]/[D_T]$ are the fractions of DNA bound and unbound to the transcription factor, respectively, solving for those quantities we obtain

$$\frac{[nXD]}{[D_T]} = \frac{[X]^n}{K_d^n + [X]^n}, \quad \frac{[D]}{[D_T]} = \frac{K_d^n}{K_d^n + [X]^n} = \frac{1}{1 + \left(\frac{[X]}{K_d}\right)^n},$$

where $K_d^n := k_-/k_+$. In a timescale such that many bindings and unbindings of the transcription factor to the promoter have occurred, those fractions can be interpreted as the probability of having n bound molecules of X , and the probability for being unbound, respectively. If the promoter has a basal transcription rate a (meaning that it transcribes without induction or at full repression) OJOOO the difference between the net transcription rate $f([X])$ and the basal rate is proportional to the bound DNA for an activator, and to the unbound DNA for a repressor, the net rates are thus

$$f([X]) = a + b \frac{[X]^n}{K_d^n + [X]^n}, \quad (1.2)$$

for an activator, and

$$f([X]) = a + b \frac{1}{1 + \left(\frac{[X]}{K_d}\right)^n}. \quad (1.3)$$

for a repressor. $b+a$ is the maximum transcription rate, which happens when $[X] \rightarrow \infty$ for the case of an activator, and when $[X] = 0$ for the repressor. K_d is called the *dissociation constant*, which is the concentration of $[X]$ needed for half activation or repression. Biologically it represents OJOOO n is called the *Hill coefficient* and from the derivation can be said that it is a measure of the cooperativity of the transcription factor, being larger if the binding of a molecule of $[X]$ enhances more the binding of another one. Larger n give a more step-like Hill function. Figures 1.7 and 1.8 show the shape of typical Hill functions given by eqs. (1.2) and (??).

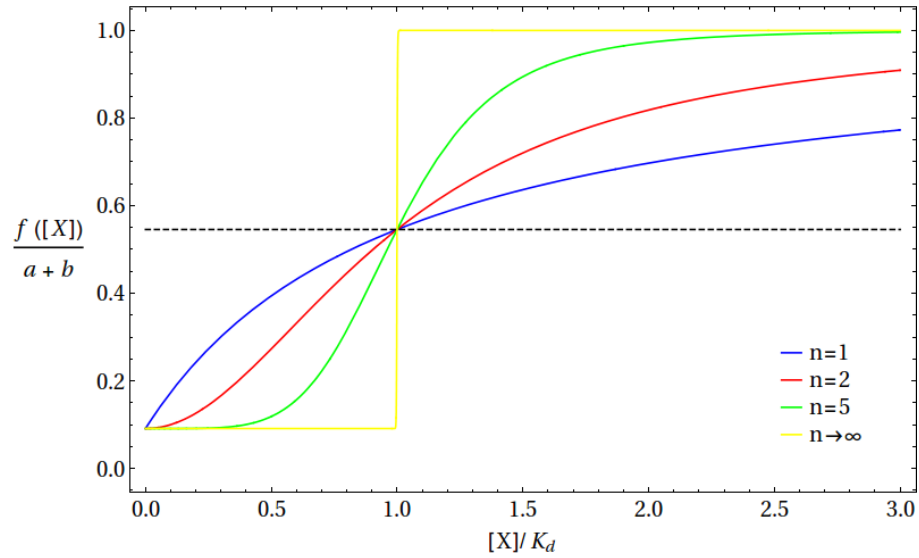


Figure 1.7: Hill functions for an activator. Various values of n are shown. The dashed line shows the point of half activation corresponding to $[X] = K_d$. All have the same value of K_d , a and b with $b/a = 10$.

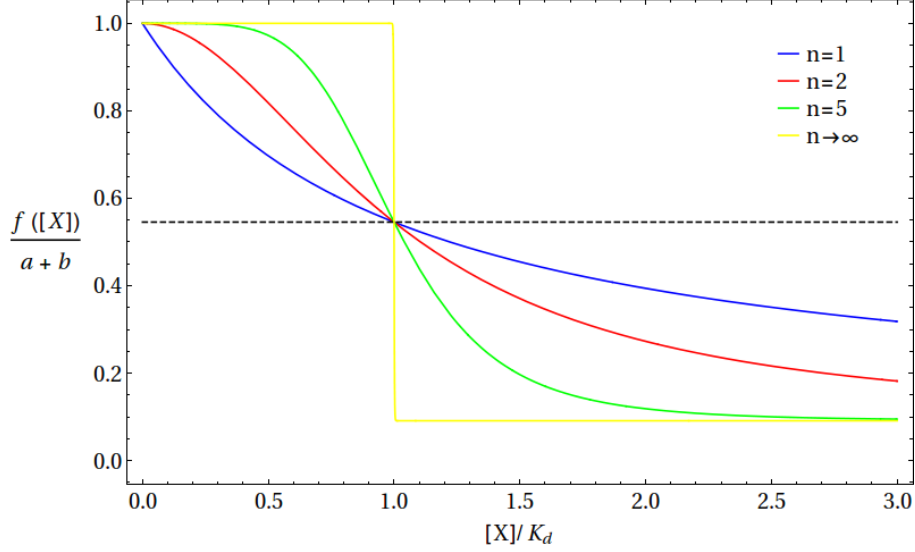


Figure 1.8: Hill functions for a repressor. With the same parameters as fig. 1.7.

Notice in both graphs that as $n \rightarrow \infty$, the function appears more like a Heaviside function with the step in $[X] = K_d$. This approximation can be very useful on a first qualitative analysis of biological circuits but such high values of n are biologically unrealistic. The case $n = 1$ corresponds to the Michaelis-Menten equation.

1.4 Probability

A random variable X is a quantity that can take values randomly from a certain set. How likely is that any value x happens is determined by the probability mass function (PMF) $P(x)$ if the variable is discrete. Those functions follow the axioms of nonnegativity, additivity and normalization. [13].¹

For several random variables X_1, \dots, X_n , the joint PMF $P(x_1, \dots, x_n)$ is defined as the probability that $X_1 = x_1, \dots, X_n = x_n$. The set of random variables are **independent** if

¹We will only consider the discrete case because the models will always consist on discrete numbers of molecules. The continuous case is very similar, it reduces almost entirely to change \sum by $\int dx$ and $P(x)$ by $\rho(x)$, where $\rho(x)$ is the PDF.

$$P(x_1, \dots, x_n) = P(x_1) \cdots P(x_n) \quad (1.4)$$

The conditional probability of the r. v.s X_1, \dots, X_k given the variables X_{k+1}, \dots, X_n is denoted by $P(x_1, \dots, x_k | x_{k+1}, \dots, x_n)$ and it is defined

$$P(x_1, \dots, x_k | x_{k+1}, \dots, x_n) := \frac{P(x_1, \dots, x_n)}{P(x_{k+1}, \dots, x_n)} \quad (1.5)$$

provided that the denominator is different from 0. Notice that if all the random variables are independent it reduces to

$$P(x_1, \dots, x_k | x_{k+1}, \dots, x_n) = P(x_1, \dots, x_k) \quad (1.6)$$

the conditional and unconditional probabilities are equal, meaning that the outcome of (x_{k+1}, \dots, x_n) does not affect the outcome of (x_1, \dots, x_k) . That is a more intuitive way of interpreting independence.

To find the probability of an outcome of X , sometimes it is easier to the **total probability theorem**

$$P(x) = \sum_y P(x, y) = \sum_y P(x|y)P(y) \quad (1.7)$$

The **expected value** (also called average) of a function of a random variable $f(X)$ is defined as

$$\langle f(X) \rangle = \sum_x f(x)P(x). \quad (1.8)$$

From the definition it can be noticed that the expected value is linear, i.e., for a pair of random variables X and Y and a constant c

$$\langle X + cY \rangle = \langle X \rangle + c\langle Y \rangle. \quad (1.9)$$

The variance $\sigma^2(X)$ of X measures the dispersion of the possible outcomes of the random variable, it is defined as

$$\sigma^2(X) := \langle (X - \langle X \rangle)^2 \rangle, \quad (1.10)$$

it can be easily shown that

$$\sigma^2(X) = \langle X^2 \rangle - \langle X \rangle^2. \quad (1.11)$$

From the previous equation it can be seen that for a constant c

$$\sigma^2(cX) = c^2 \sigma^2(X) \quad (1.12)$$

If X and Y are independent, it can be shown that

$$\langle XY \rangle = \langle X \rangle \langle Y \rangle \quad \text{and} \quad \sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y) \quad (1.13)$$

The conditional expectation of X given a random variable Y , $\langle X|Y \rangle$ is itself a random variable which depends on Y , which when Y is fixed in some y is given by

$$\langle X|y \rangle := \sum_x x P(x|y), \quad (1.14)$$

and it follows that

$$\langle \langle X|Y \rangle \rangle = \langle X \rangle \quad (1.15)$$

which is the **law of total expectation**, for the variance there is an analogous theorem, called the **law of total variance**

$$\sigma^2(x) = \langle \sigma^2(x|y) \rangle + \sigma^2(\langle x|y \rangle) \quad (1.16)$$

The **covariance** of X and Y is defined as

$$\text{cov}(X, Y) := \langle (X - \langle X \rangle) (Y - \langle Y \rangle) \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle. \quad (1.17)$$

It can be thought as a measure of how correlated is their behaviour. For example, if the value of Y is known, the value of X will be more likely to be known if the covariance is high in absolute value. The covariance will be 0 if it does not give us any information. Consider the extreme case, if $Y = X$, $\text{cov}(X, X) = \sigma^2(X)$, while if X and Y are independent by eq. (1.13) we get $\text{cov}(X, Y) = 0$.

1.5 Noise

Intuitively, we may expect that a random variable is more 'random' or noisy, when the deviations relative to the expected value are bigger. With this in mind, the noise in a random variable X must increase as the variance increases and decrease as the mean increases (the same deviation from a smaller expected value contributes more to the noise than from a bigger one). The quantities that have been used to measure noise in biology are the Fano factor ν and the coefficient of variation η , which are defined by

$$\nu_X := \frac{\sigma^2(X)}{\langle X \rangle}.$$

$$\eta_X := \frac{\sigma(X)}{\langle X \rangle}.$$

The Fano factor has been used in the first studies of noise in biology since it had the particular property that for a random variable with a Poisson distribution $\nu_X = 1$ and hence it measures deviations from a Poissonian behavior. In more recent studies the coefficient of variation is being used because it is dimensionless and for that reason it does not depend on the units used for the random variable.

1.6 Moment generating functions

Let n_1, \dots, n_N be discrete random variables over \mathbb{N} and let $f(n_1, \dots, n_N)$ be the joint probability mass function. The moment generating function $F(z_1, \dots, z_N)$ is defined as

$$F(z_1, \dots, z_N) := \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} z_1^{n_1} \cdots z_N^{n_N} f(n_1, \dots, n_N). \quad (1.18)$$

Evaluating the function on $z_1 = \cdots = z_N = 1$ (denoted by $|_1$) we obtain

$$F|_1 = \sum_{n_1, \dots, n_N} f(n_1, \dots, n_N) = 1. \quad (1.19)$$

The last part stands as a consequence of the normalization of PDFs. Taking the derivative of eq. 1.18 with respect to z_i , $i = 1, \dots, N$ we get

$$\left. \frac{\partial F}{\partial z_i} \right|_1 = \sum_{n_1, \dots, n_N} n_i z_1^{n_1} \cdots z_i^{n_i-1} \cdots z_N^{n_N} f(n_1, \dots, n_N) \Big|_1 = \sum_{n_1, \dots, n_N} n_i f(n_1, \dots, n_N) = \langle n_i \rangle. \quad (1.20)$$

Taking another derivative with respect to z_j , $j = 1, \dots, N$ with $j \neq i$ we obtain

$$\left. \frac{\partial F}{\partial z_i \partial z_j} \right|_1 = \langle n_i n_j \rangle. \quad (1.21)$$

Differentiating eq. 1.18 twice with respect to z_i we obtain similarly

$$\left. \frac{\partial^2 F}{\partial z_i^2} \right|_1 = \langle n_i(n_i - 1) \rangle. \quad (1.22)$$

These properties will be very useful in the subsequent sections to find the noise of a genetic system.

1.7 Characteristic function

Another transformation of the PMF that has properties similar to the mentioned for the moment generating function is the characteristic function. It is defined for a set of random variables (X_1, \dots, X_N) as².

$$\phi(s_1, \dots, s_N) := \left\langle e^{\sum_{i=1}^N s_i x_i} \right\rangle = \sum_{x_1} \cdots \sum_{x_N} \exp \left(\sum_{i=1}^N s_i x_i \right). \quad (1.23)$$

We will denote the evaluation at $s_1 = \dots = s_N = 0$ by $|_0$. It is easy to see that by normalization

$$\phi(s)|_0 = 1. \quad (1.24)$$

Differentiating once with respect to s_i for $i = 1, \dots, N$.

$$\left. \frac{\partial \phi(s_1, \dots, s_N)}{\partial s_i} \right|_0 = \left\langle x_i e^{\sum_{i=1}^N s_i x_i} \right\rangle \Big|_0 = \langle x_i \rangle, \quad (1.25)$$

each differentiation with respect to s_i produces a factor x_i in the average, hence

$$\left. \frac{\partial^2 \phi(s_1, \dots, s_N)}{\partial s_i \partial s_j} \right|_0 = \langle x_i x_j \rangle, \quad (1.26)$$

equation that is valid for any $i, j = 1, \dots, N$, even for the case $i = j$. In that case, the left hand side becomes a second derivative and the right hand side becomes $\langle x_i^2 \rangle$. By calculating higher order derivatives we can find higher order moments in the same way.

1.8 Stochastic processes

An stochastic variable is a set of random variables indexed by another variable, which usually is taken to be the time. An outcome of the stochastic process is a function of time which varies randomly between different repetitions of the experiment. [14] [15].

²We will consider here the case with the real exponent

The **autocorrelation** C_X of a stochastic process $X(t)$ is given by

$$C_X(t, t') := \langle X(t)X(t') \rangle. \quad (1.27)$$

It tells us the degree of correlation between outcomes of the random variables at different times. If the process X is **stationary**, the autocorrelation only depends on the time difference

$$C(\tau) := \langle X(t)X(t + \tau) \rangle, \quad (1.28)$$

where $\tau := t' - t$.

The **power spectrum** of a stochastic process is defined as average of the square norm of the Fourier transform of the process

$$S_X(\omega) := \left| \langle \hat{X} \rangle \right|^2, \quad (1.29)$$

where the hat denotes Fourier transform. A mathematical tool that will be very useful in the calculations is the **Wiener-Khinchin theorem**. It states that the power spectrum and the autocorrelation are Fourier-Transform pairs, that is

$$\mathcal{F}(C_X(\tau)) = S_X(\omega), \quad \text{and} \quad \mathcal{F}^{-1}(S_X(\omega)) = C_X(\tau). \quad (1.30)$$

1.9 The Gillespie algorithm

To simulate the models we will use the Gillespie algorithm [16] which improves speed a lot with respect to brute-force stochastic algorithms. It is used to simulate events that occur with a certain probability per unit time (e.g. creation and destruction of RNA or proteins, binding of an enzyme to a substrate, etc.)

In the brute force approach we consider a fixed time interval that must be sufficiently small, and for every possible event we sample a random number that depending on

the probabilities of the events, will tell us which of the events happened or if nothing happened at that interval. This procedure is repeated for all the intervals. Since time intervals must be small and for each interval we must sample as many random numbers as events, this approach is computationally very inefficient.

In the Gillespie algorithm, we exploit the properties of the Poisson process, namely its memorylessness and its merging properties. Assuming all the events are Poissonian, the interarrival times are exponentially distributed and therefore memoryless. [13] (SAY IT MORE RIGOROUSLY)

There are not fixed time intervals in this case, a random number is sampled with an exponential distribution to find the time of occurrence of the next event and another uniform one is sampled to evaluate which of the events occurred. To sample an exponential random number X with parameter K from a uniform U between 0 and 1, the following formula can be used.

$$X = -\frac{1}{K} \ln(U) \quad (1.31)$$

Where K turns out to be the sum of the probabilities per unit time of all the events. (PROVE? MORE RIGOR?)

The Gillespie algorithm is by far more efficient. First, it does not need to have a fixed time interval, it finds it with a random number. Second, the number of random numbers sampled in the brute-force algorithm is equal to the number of events (which can be large), and some of the times there will not be an event, while in the Gillespie algorithm only two random numbers are sampled by event and the way in which the time interval is sampled allows to advance in time in many fewer steps. Finally, a very important aspect is that the Gillespie algorithm is exact, while the precision of the brute-force algorithm depends on how small is the time interval in consideration.

Chapter 2

Basic genetic circuits in steady state - The master equation

2.1 Single gene

For the process of transcription and translation the number d of DNA copies of certain gene is taken to be constant and the rate of production of RNA (n_1) is proportional to d . In the same way, the rate of production of proteins (n_2) is proportional to r . There is also a degradation rate for each molecule proportional to their concentration. The model is illustrated in fig. 2.1

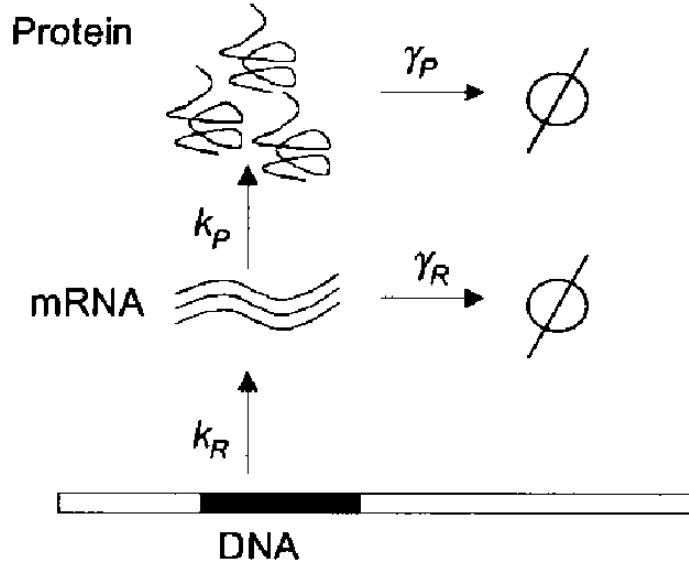


Figure 2.1: Steps of gene expression considered in the model.

The equations describing these processes are therefore given by

$$\dot{n}_1(t) = k_r d - \gamma_r n_1(t), \quad (2.1)$$

$$\dot{n}_2(t) = k_p n_1(t) - \gamma_p n_2(t). \quad (2.2)$$

In matrix notation, they can be written as

$$\dot{\mathbf{n}} = (\mathbf{A} - \mathbf{\Gamma}) \mathbf{n}, \quad (2.3)$$

where $\mathbf{n}^T = (d, n_1, n_2)$ is the vector of chemical species and the matrices \mathbf{A} and $\mathbf{\Gamma}$ are defined as

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 \\ k_r & 0 & 0 \\ 0 & k_p & 0 \end{pmatrix}, \quad \mathbf{\Gamma} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \gamma_r & 0 \\ 0 & 0 & \gamma_p \end{pmatrix}. \quad (2.4)$$

Hence, \mathbf{A} contains the creation rates and represents how each rate depends on the different species.

From eqs. 2.1 and 2.2 can be deduced that on steady state

$$\langle n_1 \rangle = \frac{k_r}{\gamma_r}, \quad (2.5)$$

$$\langle n_2 \rangle = \frac{k_p}{\gamma_p} \langle n_1 \rangle = \frac{k_p k_r}{\gamma_p \gamma_r}. \quad (2.6)$$

But these are numbers of molecules which are discrete, and on that discreteness lies part of the stochastic behaviour of those kind of systems. The molecules are created and degraded one at a time at a certain average rate but the timing between each creation or degradation should not match exactly the rates.

To model the intrinsic noise, we will consider each value of \mathbf{n} as a possible state for the system. There are transitions between the possible states which are proportional to the rates of creation and degradation as can be seen on figure 2.2.

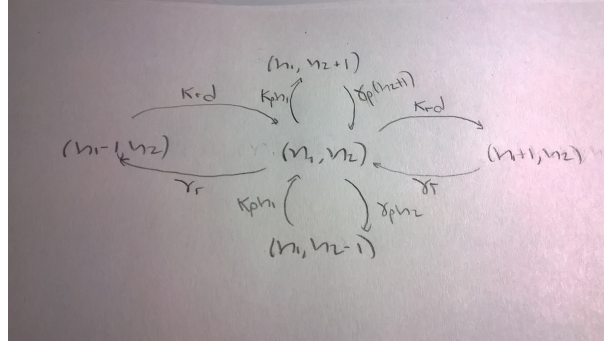


Figure 2.2: Scheme of the possible transitions involving n_1 RNA molecules and n_2 protein molecules.

The transitions shown in figure 2.2 are the ways in which the vector \mathbf{n} can change from or to (d, n_1, n_2) . According to this, the so called master equation can be written

$$\begin{aligned}
\frac{dp(n_1, n_2, t)}{dt} &= k_r dp(n_1 - 1, n_2, t) - k_r dp(n_1, n_2, t) \\
&+ k_p n_1 p(n_1, n_2 - 1, t) - k_p n_1 p(n_1, n_2, t) \\
&+ \gamma_r (n_1 + 1) p(n_1 + 1, n_2, t) - \gamma_r n_1 p(n_1, n_2, t) \\
&+ \gamma_p (n_2 + 1) p(n_1, n_2 + 1, t) - \gamma_p n_2 p(n_1, n_2 + 1, t)
\end{aligned} \tag{2.7}$$

Notice that the first term refers to a transition from state $(n_1 - 1, n_2, t)$ to (n_1, n_2, t) via a creation of a RNA molecule, whereas the second term involves a transition $(n_1, n_2, t) \rightarrow (n_1 + 1, n_2, t)$ also by a creation of RNA. The third and fourth terms have a similar meaning but related to protein creation. The other terms refer to transitions due to degradation.

Multiplying by $z_1^{n_1} z_2^{n_2}$ and summing over n_1 and n_2 , both from 0 to ∞ we obtain, recalling the definition of the moment generating function (the time dependence and the upper limit of the sum are not shown for simplicity).

$$\sum_{n_1=0, n_2=0} z_1^{n_1} z_2^{n_2} f(n_1 - 1, n_2) = \sum_{n_1=-1, n_2=0} z_1^{n_1+1} z_2^{n_2} f(n_1, n_2)$$

But since n_1 represents number of molecules, it must be a positive quantity. Hence $f(-1, n_2) = 0$ and the last sum can be taken from $n_1 = 0$ getting

$$z_1 \sum_{n_1=0, n_2=0} z_1^{n_1} z_2^{n_2} f(n_1, n_2) = z_1 F(z_1, z_2). \tag{2.8}$$

For the second term of eq. 2.7 the result is trivial, for the third term we get

$$\sum_{n_1=0, n_2=0} n_1 f(n_1, n_2 - 1) = \sum_{n_1=0, n_2=-1} n_1 z_1^{n_1} z_2^{n_2+1} f(n_1, n_2).$$

Using the same argument as above, $f(n_1, -1) = 0$. Rearranging we get

$$z_1 z_2 \sum_{n_1=0, n_2=0} z_1^{n_1-1} z_2^{n_2} f(n_1, n_2) = z_1 z_2 \frac{\partial F(z_1, z_2)}{\partial z_1}. \tag{2.9}$$

For the fifth term

$$\begin{aligned}
\sum_{n_1=0, n_2=0} (n_1 + 1) z_1^{n_1} z_2^{n_2} f(n_1 + 1, n_2) &= \sum_{n_1=1, n_2=0} n_1 z_1^{n_1-1} z_2^{n_2} f(n_1, n_2) \\
&= z_1 \sum_{n_1=0, n_2=0} n_1 z_1^{n_1-1} z_2^{n_2} f(n_1, n_2) = \frac{\partial F(z_1, z_2)}{\partial z_1}.
\end{aligned} \tag{2.10}$$

The other terms are treated in a similar fashion. Putting all of this together in we obtain the equation for the moment generating function F

$$\begin{aligned}
\dot{F}(z_1, z_2, t) &= k_r d(z_1 - 1) F(z_1, z_2, t) + k_p z_1 (z_2 - 1) \frac{\partial F(z_1, z_2, t)}{\partial z_1} \\
&\quad + \gamma_r (1 - z_1) \frac{\partial F(z_1, z_2, t)}{\partial z_1} + \gamma_p (1 - z_2) \frac{\partial F(z_1, z_2, t)}{\partial z_2}.
\end{aligned} \tag{2.11}$$

Taking the derivative with respect to z_1 we obtain

$$\begin{aligned}
\frac{\partial \dot{F}}{\partial z_1} &= k_r d \left(F + (z_1 - 1) \frac{\partial F}{\partial z_1} \right) + k_p z_1 \left(\frac{\partial F}{\partial z_1} + z_1 \frac{\partial^2 F}{\partial z_1^2} \right) \\
&\quad + \gamma_r \left(-\frac{\partial F}{\partial z_1} + (1 - z_1) \frac{\partial^2 F}{\partial z_1^2} \right) + \gamma_p (1 - z_2) \frac{\partial^2 F}{\partial z_1 \partial z_2},
\end{aligned} \tag{2.12}$$

and taking the derivative of eq. 2.11 with respect to z_2

$$\begin{aligned}
\frac{\partial \dot{F}}{\partial z_2} &= k_r d(z_1 - 1) \frac{\partial F}{\partial z_2} + k_p z_1 \left(\frac{\partial F}{\partial z_1} + (z_2 - 1) \frac{\partial^2 F}{\partial z_1 \partial z_2} \right) \\
&\quad + \gamma_r (1 - z_1) \frac{\partial^2 F}{\partial z_1 \partial z_2} + \gamma_p \left(-\frac{\partial F}{\partial z_2} + (1 - z_2) \frac{\partial^2 F}{\partial z_2^2} \right).
\end{aligned} \tag{2.13}$$

Evaluating eqs. 2.12 and 2.13 at $z_1 = z_2 = 1$ and using properties (FILL) we obtain the deterministic equations

$$\begin{aligned}\langle \dot{n}_1 \rangle &= k_r d - \gamma_r \langle n_1 \rangle, \\ \langle \dot{n}_2 \rangle &= k_p \langle n_1 \rangle - \gamma_p \langle n_2 \rangle.\end{aligned}$$

Therefore, the averages follow the deterministic behavior given by eqs. 2.1 and 2.2, as expected. The steady state values are given by eqs. 2.5 and 2.6.

Differentiating eq. 2.12 with respect to z_2 , eq. 2.12 with respect to z_1 and eq. 2.13 with respect to z_2 and evaluating at $z_1 = z_2 = 1$ we obtain, respectively

$$\langle \dot{n}_1 n_2 \rangle = k_r d \langle n_2 \rangle + k_p (\langle n_1 \rangle + \langle n_1(n_1 - 1) \rangle) - (\gamma_r + \gamma_p) \langle n_1 n_2 \rangle, \quad (2.14)$$

$$\langle \dot{n}_1(n_1 - 1) \rangle = 2k_r \langle n_1 \rangle - 2\gamma_r \langle n_1(n_1 - 1) \rangle, \quad (2.15)$$

$$\langle \dot{n}_2(n_2 - 1) \rangle = 2k_p \langle n_1 n_2 \rangle - 2\gamma_p \langle n_2(n_2 - 1) \rangle. \quad (2.16)$$

We will treat the previous equations in steady state, that is, with their time derivatives equal to zero. From eq. 2.15, we get

$$0 = k_r d \langle n_1 \rangle_s - \gamma_r (\langle n_1^2 \rangle_s - \langle n_1 \rangle_s) \Rightarrow \langle n_1^2 \rangle_s = \frac{k_r d}{\gamma_r} \langle n_1 \rangle_s + \langle n_1 \rangle_s = \langle n_1 \rangle_s^2 + \langle n_1 \rangle_s. \quad (2.17)$$

Therefore, in steady state $\sigma_1^2 = \langle n_1 \rangle$ and the Fano factor and the coefficient of variation are given by (FILL)

$$\boxed{\nu_1 = 1}, \quad \boxed{\eta_1^2 = \frac{1}{\langle n_1 \rangle}}. \quad (2.18)$$

Which is characteristic of a Poisson process, as expected because ... (TODO).

From eq. 2.14 we have

$$0 = k_r d \langle n_2 \rangle_s + k_p \langle n_1^2 \rangle_s - (\gamma_p + \gamma_r) \langle n_1 n_2 \rangle_s \Rightarrow \langle n_1 n_2 \rangle_s = \frac{k_r d \langle n_2 \rangle_s + k_p \langle n_1^2 \rangle_s}{\gamma_r + \gamma_p}.$$

But from eq. 2.17 and 2.6,

$$\langle n_1^2 \rangle_s = \langle n_1 \rangle_s (\langle n_1 \rangle_s + 1) = \frac{\gamma_p}{k_p} \langle n_2 \rangle_s (\langle n_1 \rangle_s + 1), \quad (2.19)$$

thus, the covariance is given by

$$\begin{aligned} \langle n_1 n_2 \rangle_s - \langle n_1 \rangle_s \langle n_2 \rangle_s &= \langle n_2 \rangle_s \left(\frac{k_r d + \gamma_p (\langle n_1 \rangle_s + 1)}{\gamma_r + \gamma_p} - \langle n_1 \rangle_s \right) \\ &= \langle n_2 \rangle_s \frac{k_r d + \gamma_p - \gamma_r \langle n_1 \rangle_s}{\gamma_r + \gamma_p}. \end{aligned}$$

From eq. 2.5 the first and fourth term of the denominator cancel out, therefore

$$\boxed{\text{cov}(n_1, n_2)_s = \langle n_2 \rangle_s \frac{1}{1 + \frac{\gamma_r}{\gamma_p}}}. \quad (2.20)$$

From eq. 2.16 we have in steady state

$$k_p \langle n_1 n_2 \rangle_s = \gamma_p \langle n_2^2 \rangle_s - \gamma_p \langle n_2 \rangle_s$$

Replacing eq. 2.20 in the previous equation we get after rearranging

$$\begin{aligned} \langle n_2^2 \rangle_s &= \frac{k_p}{\gamma_p} \left(\langle n_1 \rangle_s \langle n_2 \rangle_s + \frac{\langle n_2 \rangle_s \gamma_p}{\gamma_r + \gamma_p} \right) + \langle n_2 \rangle_s \\ &= \langle n_2^2 \rangle_s + \frac{k_p \langle n_2 \rangle_s}{\gamma_r + \gamma_p} + \langle n_2 \rangle_s. \end{aligned}$$

Hence subtracting $\langle n_2 \rangle_s^2$ from the previous equation we obtain, in steady state,

$$\sigma_2^2 = \langle n_2 \rangle \left(\frac{k_p/\gamma_r}{1 + \gamma_p/\gamma_r} + 1 \right). \quad (2.21)$$

Therefore, the noise in steady state is given by

$$\boxed{\nu_2 = \frac{k_p/\gamma_r}{1 + \gamma_p/\gamma_r} + 1}, \quad \boxed{\eta_2^2 = \frac{1}{\langle n_2 \rangle} \left(\frac{k_p/\gamma_r}{1 + \gamma_p/\gamma_r} + 1 \right)}. \quad (2.22)$$

As can be noticed, the noise in the proteins is bigger than Poissonian. If we define $b := k_p/\gamma_r$ and taking the degradation rate for a protein to be much smaller than for RNA, we have $\gamma_p/\gamma_r \ll 1$ therefore the noise reduces to

$$\boxed{\nu_2 = b + 1}, \quad \boxed{\eta_2^2 = \frac{1}{\langle n_2 \rangle} (b + 1)}. \quad (2.23)$$

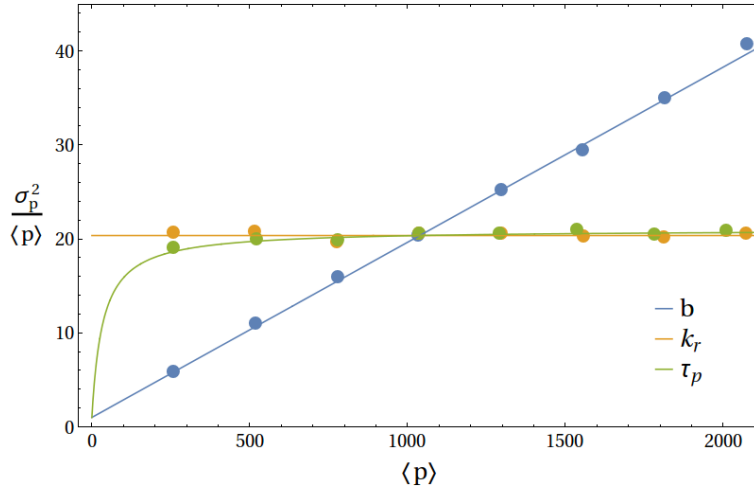


Figure 2.3: Comparison between the results of the simulations and the analytical results given by eq. 2.23. The legend indicates which parameter is varied while the others are fixed.

CONCLUSION IMPORTANT.

2.2 Several species with linear interactions

Figure 2.2 and eq. 2.7 can be generalized according to eq. 2.3 to obtain 2.4

Figure 2.4: Scheme of the possible transitions in the case of three species.

The master equation can thus be written as (NOTATION, EXPLAIN THE LIMITS OF THE SUM)

$$\dot{f}(n_i) = \sum_i \left(\sum_j A_{ij} n_j (f(n_i - 1) - f(n_i)) + \sum_j \Gamma_{ij} ((n_j + 1) f(n_i + 1) - n_j f(n_i)) \right) \quad (2.24)$$

Decaying is usually dependent on the quantity of the same molecule, assuming this the matrix Γ becomes diagonal, i.e. $\Gamma_{ij} = \delta_{ij} \Gamma_j$. With this eq. 2.24 becomes

$$\dot{f}(n_i) = \sum_i \left(\sum_j A_{ij} n_j (f(n_i - 1) - f(n_i)) + \Gamma_i ((n_i + 1) f(n_i + 1) - n_i f(n_i)) \right). \quad (2.25)$$

To get an equation for the moment generating functions, we multiply by $z_1^{n_1} \dots z_N^{n_N}$ and sum over n_1, \dots, n_N , all from 0 to ∞ . For the first term we get an expression like the following (for a fixed i) (OJO ABUSE OF NOTATION, SAY LIMITS OF SUM, DEFINE EXPLICITLY ALL Zs?)

$$\sum n_j z_1^{n_1} \cdots z_i^{n_i} \cdots z_j^{n_j} \cdots z_N^{n_N} f(n_i) = z_i z_j \sum n_j z_j^{n_j-1} z_i^{n_i} f(n_i) = z_i z_j \frac{\partial F}{\partial z_j}. \quad (2.26)$$

Where the same trick done previously on eqs. 2.8 - 2.10 were used. For the second term, similarly to the previous one

$$\sum n_j z_j^{n_j} z_i^{n_i} f(n_i) = z_j \frac{\partial F}{\partial z_j}. \quad (2.27)$$

For the third and fourth terms

$$\sum (n_i + 1) z_i^{n_i} f(n_i + 1) = \sum n_i z_i^{n_i-1} f(n_i) = \frac{\partial F}{\partial z_i}. \quad (2.28)$$

$$\sum n_i z_i^{n_i} f(n_i) = z_i \frac{\partial F}{\partial z_i}. \quad (2.29)$$

Replacing eqs. 2.26 - 2.29 in eq. 2.25 we obtain the equation for the moment generating function

$$\dot{F} = \sum_i \left(z_i \sum_j A_{ij} \frac{\partial F}{\partial z_j} - \sum_j A_{ij} z_j \frac{\partial F}{\partial z_j} + \Gamma_i \frac{\partial F}{\partial z_i} - \Gamma_i z_i \frac{\partial F}{\partial z_i} \right), \quad (2.30)$$

which after factoring becomes

$$\dot{F} = \sum_i (z_i - 1) \left(\sum_j A_{ij} z_j \frac{\partial F}{\partial z_j} - \Gamma_i \frac{\partial F}{\partial z_i} \right). \quad (2.31)$$

We have to differentiate it and use the properties (REF) to obtain equations for the moments, differentiating with respect to z_l

$$\begin{aligned} \frac{\partial \dot{F}}{\partial z_l} = \sum_i \left[(z_i - 1) \left[\sum_j A_{ij} \left(\delta_{jl} \frac{\partial F}{\partial z_j} + z_j \frac{\partial^2 F}{\partial z_j \partial z_l} \right) - \Gamma_i \frac{\partial^2 F}{\partial z_i \partial z_l} \right] \right. \\ \left. + \delta_{il} \left(\sum_j A_{ij} z_j \frac{\partial F}{\partial z_j} - \Gamma_i \frac{\partial F}{\partial z_i} \right) \right]. \end{aligned} \quad (2.32)$$

$$\begin{aligned} \frac{\partial \dot{F}}{\partial z_l} = \sum_i (z_i - 1) \left[A_{il} \frac{\partial F}{\partial z_l} + \sum_j A_{ij} z_j \frac{\partial^2 F}{\partial z_j \partial z_l} - \Gamma_i \frac{\partial^2 F}{\partial z_i \partial z_l} \right] \\ + \sum_j A_{lj} z_j \frac{\partial F}{\partial z_j} - \Gamma_l \frac{\partial F}{\partial z_l}. \end{aligned} \quad (2.33)$$

Evaluando en $z_i = 0$ para todo $i = 1, \dots, N$ obtenemos usando las propiedades (CITAR)

$$\langle \dot{n}_l \rangle = \sum_j A_{lj} \langle n_j \rangle - \Gamma_l \langle n_l \rangle. \quad (2.34)$$

Which can be written in matrix form

$$\langle \dot{\mathbf{n}} \rangle = (\mathbf{A} - \mathbf{\Gamma}) \langle \mathbf{n} \rangle. \quad (2.35)$$

Which is the same as eq. (REF), as expected. Now differentiating again with respect to z_m and some algebra

$$\begin{aligned} \frac{\partial^2 \dot{F}}{\partial z_l \partial z_m} = \sum_i (z_i - 1) \left(A_{im} \frac{\partial^2 F}{\partial z_i \partial z_m} + \sum_j A_{ij} z_j \frac{\partial^3 F}{\partial z_j \partial z_l \partial z_m} + A_{il} \frac{\partial^2 F}{\partial z_l \partial z_m} - \Gamma_i \frac{\partial^3 F}{\partial z_i \partial z_l \partial z_m} \right) \\ + \sum_j A_{mj} z_j \frac{\partial^2 F}{\partial z_j \partial z_l} + A_{ml} \frac{\partial F}{\partial z_l} - \Gamma_m \frac{\partial^2 F}{\partial z_l \partial z_m} + A_{lm} \frac{\partial F}{\partial z_m} + \sum_j A_{lj} z_j \frac{\partial^2 F}{\partial z_j \partial z_m} - \Gamma_l \frac{\partial^2 F}{\partial z_l \partial z_m}. \end{aligned} \quad (2.36)$$

Evaluating at $z_i = 1$ for all i we obtain

$$\frac{\partial^2 \dot{F}}{\partial z_l \partial z_m} = \sum_j A_{mj} z_j \frac{\partial^2 F}{\partial z_j \partial z_l} + A_{ml} \frac{\partial F}{\partial z_l} - \Gamma_m \frac{\partial^2 F}{\partial z_l \partial z_m} + A_{lm} \frac{\partial F}{\partial z_m} + \sum_j A_{lj} z_j \frac{\partial^2 F}{\partial z_j \partial z_m} - \Gamma_l \frac{\partial^2 F}{\partial z_l \partial z_m}. \quad (2.37)$$

Which can be rewritten as

$$\begin{aligned} \frac{\partial^2 \dot{F}}{\partial z_l \partial z_m} &= \sum_j (A_{mj} z_j - \Gamma_{mj}) \frac{\partial^2 F}{\partial z_j \partial z_l} + \sum_j A_{mj} \delta_{jl} \frac{\partial F}{\partial z_j} \\ &+ \sum_j (A_{lj} z_j - \Gamma_{lj}) \frac{\partial^2 F}{\partial z_j \partial z_m} + \sum_j A_{lj} \delta_{jm} \frac{\partial F}{\partial z_j}. \end{aligned} \quad (2.38)$$

Which is valid for all l and m , evaluating at $z_i = 1$ for all i we get in matrix form

$$\boxed{\nabla \nabla^T \dot{F}|_1 = ((\mathbf{\Gamma} - \mathbf{A}) \nabla \nabla^T F|_1 - \mathbf{A} \mathbf{\Theta} F|_1) + ((\mathbf{\Gamma} - \mathbf{A}) \nabla \nabla^T F|_1 - \mathbf{A} \mathbf{\Theta} F|_1)^T} \quad (2.39)$$

Where $\Theta_{ij} := \delta_{ij} \frac{\partial}{\partial z_i}$. The set of linear equations can be solved for the moments and correlation using a computer program.

2.3 Several species with non-linear interactions - Negative autorregulation

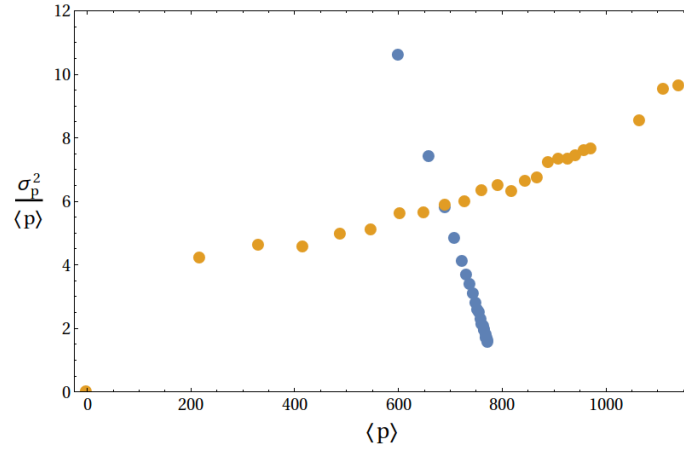


Figure 2.5: Simulation for the case of autorregulation

Chapter 3

Cascade of regulation - The Langevin equation

We will consider a set of genes whose interactions are shown on figure FILL. We will consider both intrinsic and extrinsic sources of noise. The intrinsic part refers to the inherent noise due to the low number of molecules and the nature of the reactions. This was the only source of noise consider on the previous chapter. The extrinsic part arises from another factors, such as environmental changes or sudden changes in intracellular concentrations. The fluctuations due to those sources of noise in the different genes are correlated, while the fluctuations due to intrinsic noise are not (EXPLAIN BETTER, PERHAPS ADD A SECTION ON PRELIMINARY CONCEPTS).

The differential equation for the mRNA will not be considered, we will write the equation for the proteins and include the effect of the mRNA in the rate of creation k As we have seen, the deterministic equation for the number of proteins of gene 0 is

$$\dot{x}_0(t) = k - \gamma x_0(t). \quad (3.1)$$

Where now k represents the average number of proteins created per unit time, and not the proteins per mRNA per unit time as on the previous chapter. In this approach

we add noise terms to the previous equation, one representing the intrinsic noise $\mu_0(t)$ and other representing the global noise $\xi_0(t)$. The stochastic equation then becomes

$$\dot{x}_0(t) = k - \gamma x_0(t) + \mu_0(t) + \xi_0(t). \quad (3.2)$$

Now the quantities are taken to be stochastic processes. To find the correlations and the coefficient of variation, we need some information about the noise terms. First, the average of proteins $\langle x_0 \rangle(t)$ should follow the deterministic equation 3.1, therefore

$$\langle \mu_0 \rangle = \langle \xi_0 \rangle = 0. \quad (3.3)$$

Second, we will assume that both sources are white noise, that is, the values of the noise terms at different times are uncorrelated, (EXPLAIN MORE ABOUT THE CONSTANTS AND THEIR DEFINITIONS), that is written as

$$\langle \mu_0(t) \mu_0(t + \tau) \rangle = 2\gamma(b_0 + 1)\bar{x}_0\delta(\tau), \quad (3.4)$$

$$\langle \xi_0(t) \xi_0(t + \tau) \rangle = 2\gamma\eta_G^2\bar{x}_0^2\delta(\tau). \quad (3.5)$$

where η_G is the strenght of the global noise, a parameter that is measured experimentally, and b is the average number of protein produced per mRNA. In this section the bar will denote steady state average. Also, since both sources of noise are uncorrelated

$$\langle \mu_0(t) \xi_0(t + \tau) \rangle = 0. \quad (3.6)$$

Proceeding with the calculation of the coefficient of variation, define $\delta x_0 := x_0 - \bar{x}_0$, replacing this on eq. 3.2 we get

$$\frac{d}{dt}(\delta x_0(t) + \bar{x}_0) = k - \gamma(\delta x_0(t) + \bar{x}_0) + \mu_0(t) + \xi_0(t), \quad (3.7)$$

Using the fact that $x_0 = k/\gamma$ we get

$$\delta \dot{x}_0(t) = -\gamma \delta x_0(t) + \mu_0(t) + \xi_0(t). \quad (3.8)$$

We will Fourier transform the equation, solve for δx_0 , find its square norm and use the Wiener-Khinchin theorem to find the autocorrelations in terms of the power spectrum and viceversa (EXPLAIN MORE - SEE PREVIOUS CHAPTER, ADD WK TH. ON PREL. ADD FOURIER ?).

Taking the Fourier transform of eq. 3.8 and recalling that $[\mathcal{F}(x(t))](\omega) = i\omega \mathcal{F}(x(t))(\omega)$ for a function of time x , we obtain after solving for $\delta \hat{x}_0$

$$\delta \hat{x}_0(\omega) = \frac{\hat{\mu}_0 + \hat{\xi}_0}{\gamma + i\omega}. \quad (3.9)$$

Taking the square norm and averaging we get

$$\langle |\delta \hat{x}_0|^2 \rangle = \frac{\langle |\hat{\mu}_0|^2 \rangle + \langle \hat{\mu}_0^* \hat{\xi}_0 \rangle + \langle \hat{\mu}_0 \hat{\xi}_0^* \rangle + \langle |\hat{\xi}_0|^2 \rangle}{\gamma^2 + \omega^2} \quad (3.10)$$

And using the Wiener-Khinchin theorem and eqs. 3.4 - 3.6 (EXPLAIN MORE?)

$$\begin{aligned} \langle |\delta \hat{x}_0|^2 \rangle &= \frac{(2\gamma(b_0 + 1)\bar{x}_0 + 2\gamma\eta_G^2\bar{x}_0^2) \mathcal{F}(\delta(t))}{\gamma^2 + \omega^2} \\ &= \frac{2\gamma\bar{x}_0^2 ((b_0+1)/\bar{x}_0 + \eta_G^2)}{\gamma^2 + \omega^2}, \end{aligned} \quad (3.11)$$

where the cross terms are zero by eq. 3.6. Now making the inverse Fourier transform at $t = 0$ we obtain the variance.

$$\langle \delta x_0^2 \rangle = 2\gamma\bar{x}_0^2 \left((b_0+1)/\bar{x}_0 + \eta_G^2 \right) \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{d\omega}{\omega^2 + \gamma^2}$$

The integral can be easily solved by residues and the result is π/γ , therefore

$$\langle \delta x_0^2 \rangle = \bar{x}_0^2 \left(\frac{(b_0 + 1)}{\bar{x}_0} + \eta_G^2 \right)$$

And dividing by \bar{x}_0^2 , we obtain the coefficient of variation

$$\boxed{\eta_0^2 = \frac{(b_0 + 1)}{\bar{x}_0} + \eta_G^2 = \eta_{0\text{int}}^2 + \eta_G^2}. \quad (3.12)$$

This approach enabled us to explicitly separate the total noise of gene 0 in the intrinsic and the extrinsic part. Now we will make the calculation for gene 1, which follows the equation.

$$\dot{x}_1(t) = k_1(x_{0A}) - \gamma x_1 + \mu_1 + \xi_1 \quad (3.13)$$

The decay rate γ is the same for gene 1 than for gene 0 after making the assumption that the decay is ruled by dilution due to cellular growth. The creation rate k_1 is a Hill equation for activation. The statistics for the noise terms are analogous to eqs. 3.4 - 3.6. We also need to know in this case the correlations between the noise terms corresponding to gene 0 and the ones corresponding to gene 1. As we have said, extrinsic sources of noise are uncorrelated

$$\langle \mu_0(t) \mu_1(t + \tau) \rangle = \langle \mu_0(t) \xi_1(t + \tau) \rangle = \langle \mu_1(t) \xi_0(t + \tau) \rangle = 0, \quad (3.14)$$

but the extrinsic parts of the noise of genes 0 and 1 are correlated. In analogy with eq. 3.5 (EXPLAIN) we get

$$\langle \xi_0(t) \xi_1(t + \tau) \rangle = 2\gamma \eta_G^2 \bar{x}_0 \bar{x}_1 \delta(\tau). \quad (3.15)$$

Now we will proceed in a similar way to gene 0. Defining $\delta x_1(t) := x_1(t) - \bar{x}_1$ and writing eq. 3.13 in terms of δx_1 , δx_{0A} , and making a Taylor expansion of f_1 to first order in x_{0A} we obtain.

$$\delta \dot{x}_1 = k_1(\bar{x}_{0A}) + \left. \frac{dk_1(x_{0A})}{dx_{0A}} \right|_{\bar{x}_{0A}} \delta x_{0A} - \gamma(\delta x_1 + \bar{x}_1) + \mu_1 + \xi_1, \quad (3.16)$$

but from eq. 3.13 we can see that $\bar{x}_1 = k_1(\bar{x}_{0A})/\gamma$, therefore

$$\dot{\delta x_1}(t) = c_1 \delta x_{0A} - \gamma \delta x_1 + \mu_1 \xi_1, \quad (3.17)$$

where $c_1 := dk_1(x_{0A})/dx_{0A}|_{x_{0A}^-}$ Fourier transforming and solving for $\delta \hat{x}_1$ we get

$$\delta \hat{x}_1 = \frac{c_1 \delta \hat{x}_{0A} + \hat{\mu}_1 + \hat{\xi}_1}{\gamma + i\omega}.$$

Taking the square norm and averaging

$$\begin{aligned} \langle |\delta \hat{x}_1|^2 \rangle &= \frac{1}{\omega^2 + \gamma^2} \left(c_1 \delta \hat{x}_{0A} + \hat{\mu}_1 + \hat{\xi}_1 \right) \left(c_1 \delta \hat{x}_{0A}^* + \hat{\mu}_1^* + \hat{\xi}_1^* \right) \\ &= \frac{1}{\omega^2 + \gamma^2} \left(c_1^2 \langle |\delta \hat{x}_{0A}|^2 \rangle + c_1 \left(\langle \delta \hat{x}_{0A} \hat{\xi}_1^* \rangle + \langle \delta \hat{x}_{0A}^* \hat{\xi}_1 \rangle \right) + \langle |\hat{\mu}_1|^2 \rangle + \langle |\hat{\xi}_1|^2 \rangle \right) \end{aligned} \quad (3.18)$$

Using the Wiener-Khinchine theorem and the equations for the correlations we get

$$\langle |\hat{\mu}_1|^2 \rangle = 2\gamma(b_1 + 1)\bar{x}_1, \quad (3.19)$$

$$\langle |\hat{\xi}_1|^2 \rangle = 2\gamma\eta_G^2 \bar{x}_1^2, \quad (3.20)$$

$$(3.21)$$

since the Fourier transform of the Dirac delta is 1. Also, from eqs. 3.9 and 3.11 we get

$$\langle |\delta \hat{x}_{0A}|^2 \rangle = \frac{2\gamma \bar{x}_0^2 ((b_0+1)/\bar{x}_0 + \eta_G^2)}{\gamma^2 + \omega^2}, \quad (3.22)$$

$$\langle \delta \hat{x}_{0A} \hat{\xi}_1^* \rangle = \frac{1}{\gamma + i\omega} \left(\langle \hat{\mu}_0 \hat{\xi}_1^* \rangle + \langle \hat{\xi}_0 \hat{\xi}_1^* \rangle \right) = \frac{\langle \hat{\xi}_0 \hat{\xi}_1^* \rangle}{\gamma + i\omega} \quad (3.23)$$

$$\langle \delta \hat{x}_{0A}^* \hat{\xi}_1 \rangle = \frac{1}{\gamma - i\omega} \left(\langle \hat{\mu}_0^* \hat{\xi}_1 \rangle + \langle \hat{\xi}_0^* \hat{\xi}_1 \rangle \right) = \frac{\langle \hat{\xi}_0^* \hat{\xi}_1 \rangle}{\gamma - i\omega} \quad (3.24)$$

Where the last step in the last two equations comes from the Wiener-Khinchin theorem and eq. 3.14. Replacing the previous equations in eq. 3.18 and taking the inverse transform we get for the variance

$$\begin{aligned} \langle \delta x_1^2 \rangle = & 2\gamma \bar{x}_0^2 c_1^2 \left((b_0+1)/\bar{x}_0 + \eta_G^2 \right) \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{d\omega}{(\omega^2 + \gamma^2)^2} \\ & + 2\gamma \eta_G^2 \bar{x}_0 \bar{x}_1 c_1 \frac{1}{2\pi} \left(\int_{-\infty}^{\infty} \frac{d\omega}{(\gamma + i\omega)(\omega^2 + \gamma^2)} + \int_{-\infty}^{\infty} \frac{d\omega}{(\gamma - i\omega)(\omega^2 + \gamma^2)} \right) \\ & + 2\gamma \bar{x}_1^2 \left((b_1+1)/\bar{x}_1 + \eta_G^2 \right) \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{d\omega}{\omega^2 + \gamma^2}. \end{aligned} \quad (3.25)$$

Solving the integrals in the complex plane and rearranging we get.

$$\langle \delta x_1^2 \rangle = \frac{c_1^2 \bar{x}_0^2}{2\gamma^2} \left((b_0+1)/\bar{x}_0 + \eta_G^2 \right) + \frac{c_1 \eta_G^2 \bar{x}_0 \bar{x}_1}{\gamma} + \bar{x}_1^2 \left((b_1+1)/\bar{x}_1 + \eta_G^2 \right). \quad (3.26)$$

Writing in terms of the logarithmic gain (EXPLAIN!!, ASK), $H_{10} = -\frac{c_1 \bar{x}_0}{\gamma \bar{x}_1}$, dividing by \bar{x}_1^2 and rearranging we get

$$\boxed{\eta_1^2 = \eta_{\text{1int}}^2 + \frac{1}{2} H_{10}^2 \eta_0^2 + \eta_G^2 (1 - H_{10})}. \quad (3.27)$$

Where $\eta_{\text{1int}}^2 = (b_1+1)/\bar{x}_1$ and η_0 is given by eq. 3.12.

The result can be interpreted as follows, the total noise in gene one is given by the intrinsic part, the noise from gene 0 that is propagated to gene 1 (including its intrinsic and global part) and the global noise that enters directly into gene 1. The factor of $1/2$ arises from the time averaging (EXPLAIN). MORE ANALYSIS, GRAPHICS, ETC.

For gene two we proceed similarly, with analogous statistics for the noise terms, the resulting noise is (CITE PAPER)

$$\boxed{\eta_2^2 = \eta_{\text{2int}}^2 + \frac{1}{2} H_{21}^2 \eta_1^2 + \frac{1}{8} H_{21}^2 H_{10}^2 \eta_0^2 + \eta_G^2 \left(1 - H_{21} - \frac{1}{4} H_{21}^2 H_{10} + \frac{1}{2} H_{21} H_{10} \right)}. \quad (3.28)$$

Which contains the intrinsic noise of gene 2, the contribution from the total noise of gene 1, the contribution from the total noise of gene 0 that is transmitted first to gene 1 and then to gene 2 and the global noise that enters directly.

The correlations can be found in a similar fashion (DO THAT?)

This approach enables us to calculate the coefficient of variation for a cascade of regulation and separate the different sources of noise. Also, it enables to write the effect of the upstream genes in terms of the logarithmic gain, making it very intuitive. The results of the theoretical model were tested with experiments where the genes that are part of the cascade are transcribed bicistronically with fluorescent reporters. The fluctuations in the intensity of the reporters was used to measure the noise in the population of cells.

Chapter 4

Effects of bursting and senescence

There are many different phenomena that could have an effect on noise. The common models have been based on assumptions and have made fits of the data recieved from fluorescent reporters according to those assumptions. Nevertheless, noise coming from different mechanisms could have the same general behaviors, making it difficult to predict characteristics of the system according to the noise.

We will use the Fluctuation-Dissipation theorem as we used it on section FILL.

4.1 mRNA bursts

Let the mRNA be produced with bursts of random size b , the degradation and protein creation is done one at a time with exponential waiting times. In this case the only modification with respect to the "standard model" is the D_{11} term of the matrix \mathbf{D} , which by definition is

$$D_{11} = \frac{1}{\langle n_1 \rangle^2} \sum_k (s_1^k)^2 r_k(\mathbf{n}). \quad (4.1)$$

All the possible k reactions include all the creation bursts and the reaction of degradation, which has rate $\langle n_1 \rangle / \tau_1$ and $s_1 = -1$, therefore

$$\sum_k (s_1^k)^2 r_k = \frac{\langle n_1 \rangle}{\tau_1} + \sum_k (s_1^k)^2 r_k. \quad (4.2)$$

Where now the index k runs over all the synthesis reactions only. We can rewrite the second term as

$$\sum_k (s_1^k)^2 r_k = \sum_k r_k \sum_k \left(\frac{r_k}{\sum_k r_k} \right) (s_1^k)^2 \quad (4.3)$$

where the sum over the term in parentheses results in 1. This term can be interpreted as the probability that the upcoming reaction turns out to be the k^{th} one. Writing it as ρ_k this yields

$$\sum_k (s_1^k)^2 r_k = \sum_k r_k \sum_k \rho_k (s_1^k)^2 \quad (4.4)$$

OJO But s_1^k is the burst size for the k^{th} synthesis reaction. Therefore the inner sum of the previous equation is actually an average over all the possible burst sizes, hence

$$\sum_k (s_1^k)^2 r_k = \sum_k r_k \langle b^2 \rangle = (\langle b \rangle^2 + \sigma_b^2) \sum_k r_k, \quad (4.5)$$

and using a similar trick we get

$$\begin{aligned} \sum_k r_k &= \sum_k r_k s_1^k \frac{\sum_k r_k}{\sum_k r_k s_1^k} = \sum_k r_k s_1^k \left(\sum_k \left(\frac{r_k}{\sum_k r_k} \right) s_1^k \right)^{-1} \\ &= \sum_k r_k s_1^k \left(\sum_k \rho_k s_1^k \right)^{-1} = \frac{1}{\langle b \rangle} \sum_k r_k s_1^k. \end{aligned} \quad (4.6)$$

And letting the system be in steady state, all the synthesis rates equal the degradation ones. Therefore

$$\sum_k r_k s_1^k = \frac{\langle n_1 \rangle}{\tau_1}$$

Obtaining

$$\sum_k r_k = \frac{\langle n_1 \rangle}{\langle b \rangle \tau_1} \quad (4.7)$$

Replacing this on eq. 4.5, and then on eq. 4.2 and 4.1 we get

$$D_{11} = \frac{1}{\langle n_1 \rangle^2} \left(\frac{\langle n_1 \rangle}{\tau_1} + \frac{\langle n_1 \rangle}{\langle b \rangle \tau_1} (\langle b \rangle^2 + \sigma_b^2) \right) = \frac{1}{\tau_1 \langle n_1 \rangle} \left(1 + \langle b \rangle \left(1 + \frac{\sigma_b^2}{\langle b \rangle^2} \right) \right) \quad (4.8)$$

The matrices then become

$$\mathbf{D} = \begin{pmatrix} D_{11} & 0 \\ 0 & \frac{2}{\tau_2 \langle n_2 \rangle} \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} \frac{1}{\tau_1} & 0 \\ -\frac{1}{\tau_2} & \frac{1}{\tau_2} \end{pmatrix}. \quad (4.9)$$

With D_{11} given by eq. 4.8 And solving the linear system $\mathbf{M}\eta + \eta\mathbf{M}^T + \mathbf{D} = 0$ using Mathematica, we obtain for the noise in the proteins

$$\boxed{\eta_{22} = \frac{1}{\langle n_2 \rangle} + \frac{1}{\langle n_1 \rangle} \frac{\tau_1}{\tau_1 + \tau_2} \frac{\langle b \rangle (1 + \sigma_b^2 / \langle b \rangle^2) + 1}{2}}. \quad (4.10)$$

Which is COMMENTS.

4.2 Arbitrary distribution of creation times

Suppose a creation event happened at $t = 0$, and let $f(t)$ be the probability density function of a creation event happening at time t (meaning a time t after the last event), i.e. $P(t \in [T, T + dt]) = f(T)dt$. According to that, the following properties are satisfied

$$P(n = 0 | t = T) = P(t > T) = 1 - P(t < T) = 1 - F(T) \quad (4.11)$$

where n stands for the number of creation events and F is the cumulative distribution function associated to $f(t)$ (CONCEPTS). Also, for one creation event to have happened

before time $t = T$, there must be a creation on a time t_1 such that $0 < t_1 < T$ and no creation events on the remaining $(T - t_1)$ time. The previous reasoning leads to the following equation

OJO NOTACION

$$P(n = 1|t = T) = \int_0^T P(t = t_1)P(t > T - t_1)dt_1 = \int_0^T f(t_1)(1 - F(T - t_1))dt_1 = f * (1 - F)|_T. \quad (4.12)$$

Where the asterisk denotes the convolution product (CONCEPTS?). Following a similar argument, we obtain for an arbitrary number of events

$$\begin{aligned} P(n = N|t = T) &= f * P(n = N - 1|t)|_T = f * f * P(n = N - 2|t)|_T \\ &= \dots = \underbrace{f * \dots * f}_{n \text{ times}} * P(n = 0, t)|_T = \underbrace{f * \dots * f}_{n \text{ times}} * (1 - F)|_T. \end{aligned} \quad (4.13)$$

Since the convolutions are difficult to deal with, we will use the Laplace transform and solve on Laplace space. The property that $\mathcal{L}(f * g) = \hat{f} \cdot \hat{g}$, where $\hat{f} := \mathcal{L}(f)$ will make the problem much simpler.

Aplying \mathcal{L} to eq. 4.13 we get

$$\hat{P}(n, s) = \hat{f}^n(s) \mathcal{L}(1 - F)(s)$$

but it can be easily shown (SHOW?) that $\hat{F} = \hat{f}/s$ and $\hat{1} = 1/s$ yielding

$$\hat{P}(n, s) = \frac{1}{s} \hat{f}^n(s) (1 - \hat{f}(s)). \quad (4.14)$$

To find the moments and the noise, we will use the moment generating function, as defined on COMPLETE. It will be denoted as $G(z, t)$. Aplying the Fourier transform we get

$$\hat{G}(z, t) = \sum_{n=0}^{\infty} z^n \hat{P}(n, s) = \frac{1}{s} (1 - \hat{f}(s)) \sum_{n=0}^{\infty} (z\hat{f}(s))^n = \frac{1 - \hat{f}(s)}{s(1 - z\hat{f}(s))}. \quad (4.15)$$

Where the geometric series converges in this case because $\hat{f}(s) \leq 1$ and we will evaluate z at 1. The first and second derivatives of G in Laplace space are given by

$$\langle \hat{n} \rangle(s) = \left. \frac{\partial \hat{G}(z, s)}{\partial z} \right|_1 = \frac{\hat{f}(s)}{s(1 - \hat{f}(s))}. \quad (4.16)$$

$$\langle \hat{n}(\hat{n} - 1) \rangle(s) = \left. \frac{\partial^2 \hat{G}(z, s)}{\partial z^2} \right|_1 = \frac{2}{s} \left(\frac{\hat{f}(s)}{1 - \hat{f}(s)} \right)^2 \quad (4.17)$$

It could also be proven that (PROVE!)

$$\hat{f}(0) = 1, \quad \left. \frac{d\hat{f}(s)}{ds} \right|_0 = -\langle t \rangle, \quad \left. \frac{d^2 \hat{f}(s)}{ds^2} \right|_0 = \langle t^2 \rangle. \quad (4.18)$$

Therefore applying the inverse Laplace transform to eqs. 4.16 and 4.17, and using properties 4.18 we could obtain the moments. (EXPLAIN BROMWICH INTEGRAL?)

$$\langle n \rangle(t) = \mathcal{L}^{-1}(\langle \hat{n} \rangle(s)) = \frac{1}{2i\pi} \oint e^{st} \frac{\hat{f}(s)}{s(1 - \hat{f}(s))} ds \quad (4.19)$$

The integral can be solved by residues. Since $\hat{f}(0) = 1$, there is a pole of order 2 in $s = 0$. The general formula for residues states that if c is a pole of order m of the function f , the residue is given by

$$\text{Res}_c(f) = \frac{1}{(m-1)!} \lim_{z \rightarrow c} \frac{d^{m-1}}{dz^{m-1}} ((z-c)^m f(z)). \quad (4.20)$$

Then

$$\begin{aligned}
\langle n \rangle(t) &= \text{Res}_0 \frac{e^{st}}{s} \frac{\hat{f}(s)}{1 - \hat{f}(s)} = \lim_{s \rightarrow 0} \frac{d}{ds} \frac{s e^{st} \hat{f}(s)}{1 - \hat{f}(s)} \\
&= \lim_{s \rightarrow 0} \frac{e^{st}}{(\hat{f}(s) - 1)^2} \left[(1 + st)(\hat{f}(s) - 1)\hat{f}(s) + s\hat{f}'(s) \right].
\end{aligned} \tag{4.21}$$

To find the limit we have to apply L'Hospital rule twice, after some algebra it yields

$$\langle n \rangle(t) = \frac{\hat{f}''(0)}{2(\hat{f}'(0))^2} - \frac{t}{\hat{f}'(0)} - 1 = \frac{t}{\langle t \rangle} + \left(\frac{\langle t^2 \rangle}{2\langle t \rangle^2} - 1 \right). \tag{4.22}$$

Now inverting eq. 4.17 we obtain

$$\begin{aligned}
\langle n(n-1) \rangle(t) &= \frac{1}{2i\pi} \oint e^{st} \frac{2}{s} \left(\frac{\hat{f}(s)}{1 - \hat{f}(s)} \right)^2 ds = \text{Res}_0 \frac{2}{s} \left(\frac{\hat{f}(s)}{1 - \hat{f}(s)} \right)^2 \\
&= \lim_{s \rightarrow 0} \frac{d^2}{ds^2} e^{st} \left(\frac{s\hat{f}(s)}{1 - \hat{f}(s)} \right)^2
\end{aligned} \tag{4.23}$$

Where we used eq. 4.20 to find the residue of a pole of order 3. After doing the necessary algebra and the L'Hospital rule three times we get (CHECK)

$$\langle n(n-1) \rangle(t) = \frac{t^2}{\langle t \rangle^2} + \frac{4t}{\langle t \rangle} \left(\frac{\langle t^2 \rangle}{2\langle t \rangle^2} - 1 \right) + 2 \left(1 - \frac{\langle t^2 \rangle}{\langle t \rangle^2} + \frac{3\langle t^2 \rangle^2}{4\langle t \rangle^4} + \frac{\langle t^3 \rangle}{3\langle t \rangle^3} \right) \tag{4.24}$$

And combining with eq. 4.22 we obtain the variance

$$\sigma_n^2(t) = \frac{t}{\langle t \rangle} \left(\frac{\langle t^2 \rangle}{\langle t \rangle^2} - 1 \right) + \left(-\frac{\langle t^2 \rangle}{2\langle t \rangle^2} + \frac{5\langle t^2 \rangle^2}{4\langle t \rangle^4} - \frac{2\langle t^3 \rangle}{3\langle t \rangle^3} \right). \tag{4.25}$$

ANALYSIS. Ignoring the second terms in parentheses in eqs. 4.22 and 4.25 we get

$$\langle n \rangle = \frac{t}{\langle t \rangle} \quad (4.26)$$

$$\sigma_n^2 = \frac{t}{\langle t \rangle} \left(\frac{\langle t^2 \rangle}{\langle t \rangle^2} - 1 \right) \quad (4.27)$$

Rearranging eq. 4.27 and using eq. 4.26 we get

$$\sigma_n^2 = \frac{t}{\langle t \rangle} \left(\frac{\langle t^2 \rangle - \langle t \rangle^2}{\langle t \rangle^2} \right) = \frac{t}{\langle t \rangle} \eta_t^2 = \langle n \rangle \eta_t^2.$$

Hence

$$\eta_n^2 = \frac{1}{\langle n \rangle} \eta_t^2 \quad (4.28)$$

Now we will include the effect of bursts of creation. Let n be the number of creation events on a given time interval (meaning the number of bursts), b_i be burst size for the i^{th} events. Both n and b_i are random variables, and each b_i follows the same probability distribution. Consider the random variable

$$x := \sum_{i=0}^n b_i \quad (4.29)$$

Which represents the total number of molecules created on the given time interval. (It is a sum of a random number of random variables). We will denote the probability mass function of x as $P_x(x)$.

We will use the properties of the characteristic function $\phi(s)$ and find the moments in a similar way that was done on previous chapter using the moment generating function. From its definition (CONCEPTS!!).

$$\phi(s) := \langle e^{xs} \rangle_{tot} = \sum_{a=0}^{\infty} e^{xs} P_x(x = a).$$

Using the total probability theorem, we can write it as

$$\begin{aligned}
\phi(s) &= \sum_{a=0}^{\infty} e^{xs} \sum_{n=0}^{\infty} P_x(x=a|n) P(n) = \sum_{n=0}^{\infty} \left(\sum_{a=0}^{\infty} e^{xs} P_x(x=a|n) \right) P(n) \\
&= \sum_{n=0}^{\infty} \langle e^{xs} \rangle_{x|n} P(n)
\end{aligned} \tag{4.30}$$

Where $\langle \rangle_x$ denotes average with respect to the distribution of x . Now using eq. 4.29 we get for $\langle e^{xs} \rangle_{x|n}$

$$\langle e^{xs} \rangle_{x|n} = \langle e^{s \sum_{i=0}^n b_i} \rangle_b = \langle \prod_{i=0}^n e^{sb_i} \rangle_b,$$

Notice that the average is now taken with respect to the distribution of burst sizes (denoted by $\langle \rangle_b$ since we wrote the function in terms of that variable. Now assuming independence of the burst sizes we get

$$\langle e^{xs} \rangle_{x|n} = \prod_{i=0}^n \langle e^{sb_i} \rangle_b,$$

and since all the b_i s follow the same distribution, the product is in fact independent of i , yielding

$$\langle e^{xs} \rangle_{x|n} = \prod_{i=0}^n \langle e^{sb} \rangle_b = \langle e^{sb} \rangle_b^n$$

Replacing this result in eq. 4.30 we obtain

$$\phi(s) = \sum_{n=0}^{\infty} \langle e^{sb} \rangle_b^n P(n) = \langle \langle e^{sb} \rangle_b^n \rangle_n \tag{4.31}$$

Where $\langle \rangle_n$ denotes average with respect to the distribution of events P_n .

We proceed to obtain the moments, using the properties of the characteristic function

CITE CONCEPTS.

$$\langle x \rangle = \left. \frac{d\phi(s)}{ds} \right|_0 = \left. \frac{d}{ds} \langle \langle e^{bs} \rangle_b^n \rangle_n \right|_0 = \langle n \langle e^{bs} \rangle_b^{n-1} \langle b e^{bs} \rangle_b \rangle_n \Big|_0 = \langle n \langle b \rangle_b \rangle_n = \langle n \rangle_n \langle b \rangle_b \quad (4.32)$$

which makes a lot of sense: the average number of molecules produced is the average number of bursts times the average burst size. For the second moment we have

$$\begin{aligned} \langle x^2 \rangle &= \left. \frac{d^2\phi(s)}{ds^2} \right|_0 = \left. \frac{d\phi(s)}{ds} \langle n \langle e^{bs} \rangle_b^{n-1} \langle b e^{bs} \rangle_b \rangle_n \right|_0 \\ &= \langle n(n-1) \langle e^{bs} \rangle_b^{n-2} \langle b e^{bs} \rangle_b^2 + n \langle e^{bs} \rangle_b^{n-1} \langle b^2 e^{bs} \rangle_b \rangle_n \Big|_0 \\ &= \langle n^2 \rangle_n \langle b \rangle_b^2 - \langle n \rangle_n \langle b \rangle_b^2 + \langle n \rangle_n \langle b^2 \rangle_b \end{aligned} \quad (4.33)$$

Using the previous result with eq. 4.32 to find the variance we get

$$\begin{aligned} \sigma_x^2 &= \langle n^2 \rangle_n \langle b \rangle_b^2 - \langle n \rangle_n \langle b \rangle_b^2 + \langle n \rangle_n \langle b^2 \rangle_b - \langle n \rangle_n^2 \langle b \rangle_b^2 \\ &= \langle b \rangle_b^2 (\langle n^2 \rangle_n - \langle n \rangle_n^2) + \langle n \rangle_n (\langle b^2 \rangle_b - \langle b \rangle_b^2) \\ &= \langle b \rangle_b^2 \sigma_n^2 + \langle n \rangle_n \sigma_b^2. \end{aligned} \quad (4.34)$$

And dividing by $\langle x \rangle^2 = \langle n \rangle_n^2 \langle b \rangle_b^2$ we get

$$\eta_x^2 = \frac{\sigma_n^2}{\langle n \rangle_n^2} + \frac{\sigma_b^2}{\langle n \rangle_n \langle b \rangle_b^2} = \eta_n^2 + \frac{1}{\langle n \rangle_n} \eta_b^2.$$

Using eq. 4.28 yields

$$\eta_x^2 = \frac{1}{\langle n \rangle} (\eta_t^2 + \eta_n^2) = \frac{\langle b \rangle (\eta_t^2 + \eta_n^2)}{\langle x \rangle}, \quad (4.35)$$

where

$$\langle x \rangle = \langle b \rangle \frac{t}{\langle t \rangle}. \quad (4.36)$$

Result that holds for a pure birth process.

4.3 Decay of molecules

We include decay of molecules considering a binomial partitioning during cell division. Let $P_{\text{Dr}}(l|m)$ be the probability of finding l molecules in volume fraction r given that there are m molecules before division.

Assuming that each molecule segregates independently (we will see next that this does not necessarily happens and that introduces additional uncertainties), and that the probability of arriving at a volume is proportional to it we obtain a binomial distribution

$$P_{\text{Dr}}(l|m) = \binom{m}{l} r^l (1-r)^{m-l}. \quad (4.37)$$

Let $P_{\text{Br}}(m)$ and $P_{\text{Ar}}(m)$ be the probabilities of having m molecules before and after division, respectively, for a fixed volume fraction r . We have then

$$P_{\text{Ar}}(l) = \sum_{m=0}^{\infty} P_{\text{Dr}}(l|m) P_{\text{Br}}(m) = \sum_{m=0}^{\infty} \binom{m}{l} r^l (1-r)^{m-l} P_{\text{Br}}(m). \quad (4.38)$$

Multiplying by z^l and summing we get the moment generating function $G_{\text{Ar}}(z)$

$$\begin{aligned} G_{\text{Ar}}(z) &= \sum_{l=0}^{\infty} z^l \sum_{m=0}^{\infty} \binom{m}{l} r^l (1-r)^{m-l} P_{\text{Br}}(m) \\ &= \sum_{m=0}^{\infty} \left(\sum_{l=0}^{\infty} (zr)^l (1-r)^{m-l} \right) P_{\text{Br}}(m). \\ &= \sum_{m=0}^{\infty} (zr + 1 - r)^m P_{\text{Br}}(m), \end{aligned} \quad (4.39)$$

where we used the binomial expansion formula on the last step.

The number of molecules of at the end of a growth stage (following P_{Br}) equals the number of molecules at the beginning (following P_{Ar}) plus the number of molecules

created during the cycle (following $P_{x,\tau} := P_x|_{t=\tau}$. Assuming both random variables as independent and recalling that the probability mass function of the sum of random variables is the convolution of the individual PMFs we have (CONCEPTS)

$$P_{\text{Br}}(m) = P_{\text{Ar}} * P_{x,\tau}(m)$$

Therefore (CONCEPTS),

$$G_{\text{Br}}(z) = G_{\text{Ar}}(z)G_{x,\tau}(z). \quad (4.40)$$

Finding the moments by using the properties of G and the previous equation we obtain

$$\langle n \rangle_{\text{Br}} = \left. \frac{\partial G_{\text{Br}}(z)}{\partial z} \right|_1 = G_{\text{Ar}}(1) \left. \frac{\partial G_{x,\tau}(z)}{\partial z} \right|_1 + \left. \frac{\partial G_{\text{Ar}}(z)}{\partial z} \right|_1 G_{x,\tau}(1) = \langle m \rangle_{x,\tau} + \langle m \rangle_{\text{Ar}},$$

from eq. 4.39

$$\begin{aligned} \left. \frac{\partial G_{\text{Ar}}(z)}{\partial z} \right|_1 &= \langle m \rangle_{\text{Ar}} = \left. \frac{\partial}{\partial z} \sum_{m=0}^{\infty} (zr + 1 - r)^m P_{\text{Br}}(m) \right|_1 \\ &= \left. \sum_{m=0}^{\infty} m(zr + 1 - r)^{m-1} r P_{\text{Br}}(m) \right|_1 \\ &= r \sum_{m=0}^{\infty} m P_{\text{Br}}(m) = r \langle m \rangle_{\text{Br}} \end{aligned}$$

Therefore

$$\langle m \rangle_{\text{Br}} = \frac{1}{1-r} \langle m \rangle_{x,\tau} \quad \langle m \rangle_{\text{Ar}} = r \langle n \rangle_{\text{Br}} = \frac{r}{1-r} \langle m \rangle_{x,\tau}. \quad (4.41)$$

The variances are obtained by taking the second derivative

$$\begin{aligned}
\langle m(m-1) \rangle_{\text{Br}} &= \left. \frac{\partial^2 G_{\text{Br}}(z)}{\partial z^2} \right|_1 \\
&= G_{\text{Ar}}(1) \left. \frac{\partial^2 G_{x,\tau}(z)}{\partial z^2} \right|_1 + \left. \frac{\partial^2 G_{\text{Ar}}(z)}{\partial z^2} \right|_1 G_{x,\tau}(1) + 2 \left. \frac{\partial G_{\text{Ar}}(z)}{\partial z} \right|_1 \left. \frac{\partial G_{x,\tau}(z)}{\partial z^2} \right|_1 \\
&= \langle m(m-1) \rangle_{x,\tau} + \langle m(m-1) \rangle_{\text{Ar}} + 2 \langle m \rangle_{x,\tau} \langle m \rangle_{\text{Ar}}.
\end{aligned} \tag{4.42}$$

but from eq. 4.39

$$\begin{aligned}
\langle m(m-1) \rangle_{\text{Ar}} &= \left. \frac{\partial^2}{\partial z^2} \sum_{m=0}^{\infty} (zr + 1 - r)^m P_{\text{Br}}(n) \right|_1 \\
&= \left. \sum_{m=0}^{\infty} m(m-1) (zr + 1 - r)^{m-2} r^2 P_{\text{Br}}(n) \right|_1 = r^2 \langle n(n-1) \rangle_{\text{Br}}.
\end{aligned} \tag{4.43}$$

For any random variable x , we can write $\langle x(x-1) \rangle = \sigma_x^2 - \langle x \rangle + \langle x \rangle^2$. Using this on eqs. 4.42 and 4.43 we get (MORE DETAILED?)

$$\begin{aligned}
\sigma_{\text{Br}}^2 - \langle m \rangle_{\text{Br}} + \langle m \rangle_{\text{Br}}^2 &= (\sigma_{x,\tau}^2 - \langle m \rangle_{x,\tau} + \langle m \rangle_{x,\tau}^2) \\
&\quad + 2(r \langle m \rangle_{\text{Br}}) [(1-r) \langle m \rangle_{\text{Br}}] + r^2 (\sigma_{\text{Br}}^2 - \langle m \rangle_{\text{Br}} + \langle m \rangle_{\text{Br}}^2)
\end{aligned}$$

After some algebra we obtain (MORE DETAIL?)

$$\sigma_{\text{Br}}^2 = \frac{1}{1-r^2} \sigma_{x,\tau}^2 + \frac{r}{1+r} \langle m \rangle_{\text{Br}} \tag{4.44}$$

Dividing by $\langle m \rangle_{\text{Br}}^2$ and using eq. 4.41 we get

$$\begin{aligned}
\eta_{\text{Br}}^2 &= \frac{1}{1-r^2} \sigma_{x,\tau}^2 \frac{(1-r)^2}{\langle m \rangle_{x,\tau}^2} + \frac{r}{1+r} \langle m \rangle_{\text{Br}} \\
&= \frac{1-r}{1+r} \eta_{x,\tau}^2 + \frac{r}{1+r} \frac{1}{\langle n \rangle_{\text{Br}}}
\end{aligned} \tag{4.45}$$

Also, from eqs. 4.43 and 4.44 we get

$$\begin{aligned}
\sigma_{\text{Ar}}^2 - \langle m \rangle_{\text{Ar}} + \langle m \rangle_{\text{Ar}}^2 &= r^2 (\sigma_{\text{Br}}^2 - \langle m \rangle_{\text{Br}} + \langle m \rangle_{\text{Br}}^2) \\
&= r^2 \left(\frac{1}{1-r^2} \sigma_{x,\tau}^2 + \frac{r}{1+r} \langle m \rangle_{\text{Br}} \right) - r^2 \langle m \rangle_{\text{Br}} + r^2 \langle m \rangle_{\text{Br}}^2
\end{aligned} \tag{4.46}$$

Using eq. 4.41 and after a little algebra (MORE?) we get

$$\sigma_{\text{Ar}}^2 = \frac{r^2}{1-r^2} \sigma_{x,\tau}^2 + \frac{1}{1+r} \langle m \rangle_{\text{Ar}}, \tag{4.47}$$

hence, dividing by $\langle m \rangle_{\text{Ar}}^2$ and using eq. 4.41 we get

$$\begin{aligned}
\eta_{\text{Ar}}^2 &= \frac{r^2}{1-r^2} \sigma_{x,\tau}^2 \left(\frac{1-r}{r \langle m \rangle_{x,\tau}} \right)^2 + \frac{1}{1+r} \frac{1}{\langle m \rangle_{\text{Ar}}} \\
&= \frac{1-r}{1+r} \eta_{x,\tau}^2 + \frac{1}{1+r} \frac{1}{\langle m \rangle_{\text{Ar}}}.
\end{aligned} \tag{4.48}$$

EXPLAIN THE CASE OF NORMAL DILUTION BY DIVISION. AND THE LIMIT

$$\boxed{\eta^2 = \frac{\langle b \rangle (\eta_t^2 + \eta_b^2) + 1}{2 \langle n \rangle}} \tag{4.49}$$

Corresponding to the factor, CHECK NOTATION, COMPLETE, ANALYSIS.

4.4 Senescence of mRNA

Suppose that mRNAs are created at a constant rate λ_1 and that they senesce through a sequence of N steps labeled as X_1, X_2, \dots, X_N . Thus, the states and their transitions are

$$\text{Transcription} \rightarrow X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_N \rightarrow \text{Degradation} \quad (4.50)$$

The number of mRNA molecules in each step is labeled as x_i for $1 \leq i \leq N$. Also, assume that the rates of transcription per mRNA between the states are the same, we will denote it as β_1 .

$$\begin{aligned} x_1 &\xrightarrow{\lambda_1} x_1 + 1 \\ \{x_i, x_{i+1}\} &\xrightarrow{\beta_1 x_i} \{x_i - 1, x_{i+1} + 1\}, \quad \text{for } 1 \leq i \leq N - 1 \\ x_N &\xrightarrow{\beta_1 x_N} x_N - 1 \end{aligned} \quad (4.51)$$

Where the first line denotes transcription, the second denotes transitions between states and the third stands for degradation. Notice that the same constant β_1 is used both for transitions and for degradation as assumed. Now, denote as x_{N+1} the number of proteins in a cell and suppose that independently of the current state of the mRNA molecules, the translate with the same rate λ_2 per molecule. Also, denoting the degradation rate per protein as β_2 the possible transitions for the proteins are

$$\begin{aligned} x_{N+1} &\xrightarrow{\lambda_2 \sum_{i=1}^N x_i} x_{N+1} + 1 \\ x_{N+1} &\xrightarrow{\beta_2 x_{N+1}} x_{N+1} - 1 \end{aligned} \quad (4.52)$$

The average dynamics thus follows

$$\begin{aligned}
\langle \dot{x}_1 \rangle &= \lambda - 1 - \beta_1 \langle x_1 \rangle \\
\langle \dot{x}_{i+1} \rangle &= \beta_1 (\langle x_i \rangle - \langle x_{i+1} \rangle) \quad \text{for } 1 \leq i \leq N \\
\langle \dot{x}_{N+1} \rangle &= \lambda_2 \sum_{i=1}^N \langle x_i \rangle - \beta_2 \langle x_{N+1} \rangle
\end{aligned} \tag{4.53}$$

At steady state we get

$$\begin{aligned}
\langle x_1 \rangle &= \frac{\lambda_1}{\beta_1} \\
\langle x_{i+1} \rangle &= \langle x_i \rangle \quad \text{for } 1 \leq i \leq N-1 \\
\langle x_{N+1} \rangle &= \frac{\lambda_2}{\beta_2} \sum_{i=1}^N \langle x_i \rangle
\end{aligned} \tag{4.54}$$

Therefore,

$$\langle x_i \rangle = \lambda_1 \tau_1 \quad \text{for } 1 \leq i \leq N, \tag{4.55}$$

where $\tau_1 := 1/\beta_1$.

Denote the total mRNA as m , then $m = \sum_{i=1}^N x_i$ and taking the average

Chapter 5

Effects of cell division

When the cell divides, all the components (organelles, proteins, genetic material, etc.) must be distributed among the daughter cells. Nevertheless, this partition is not even and this unevenness could be a great source of noise even for components present at high numbers. In this chapter we will explore some general mechanisms of partition of molecules during cell division and how they can affect noise statistics.

5.1 Characterizing the noise arising from cell division

Let $x = l + r$ be the number of copies of some component (e.g. a certain protein) for a dividing cell, with l and r being the number of copies each daughter cell receives. Also, let v be the number of molecules of some component that affects the partition such as vacuoles or spindles. On average, we expect the molecules to distribute symmetrically. Therefore

$$\langle l \rangle = \langle r \rangle = \frac{\langle x \rangle}{2}. \quad (5.1)$$

We will find the statistics for l , from them the statistics for r are trivially derived

(EXPLAIN MORE). Using the law of total variance (CONCEPTS), the variance of l is given by

$$\sigma^2(l) = \sigma^2(\langle l|x, v \rangle) + \langle \sigma^2(l|x, v) \rangle. \quad (5.2)$$

Where the first and second terms can be viewed as the contributions to randomness from prior noise on x and v and the noise due solely to cell division (EXPLAIN MORE?), respectively. From eq. (5.1), $\langle L|x, v \rangle = x/2$, using this and dividing by $\langle l \rangle^2 = \langle x \rangle^2/4$ we get

$$\begin{aligned} \eta^2(l) &= 4 \frac{\sigma^2(x/2)}{\langle x \rangle^2} + \frac{\langle \sigma^2(L|x, v) \rangle}{\langle L \rangle^2} \\ &= \eta^2(x) + Q_x^2, \end{aligned} \quad (5.3)$$

(CONCEPTS: $\text{VAR}(\text{ax}) = \text{aavar}(\text{x})$). where

$$Q_x^2 := \frac{\langle \sigma^2(L|x, v) \rangle}{\langle L \rangle^2}, \quad (5.4)$$

is the noise arising from cell division. Eq. (5.3) states more clearly what was said previously: the noise after cell division (measured with the coefficient of variation), is the sum of the noise before division and the noise arising at the division process.

The term Q_x can be interpreted in another way. From the definition of variance and eq. (5.1)

$$Q_x^2 = \frac{1}{\langle l \rangle^2} \langle (l - \langle l \rangle)^2 | x, v \rangle = \frac{4}{\langle x \rangle^2} \left\langle \left(l - \frac{x}{2} \right)^2 \middle| x, v \right\rangle \quad (5.5)$$

but $l - x/2 = 1/2(2l - (l - r)) = l - r/2$, then

$$Q_x^2 = \frac{4}{\langle x \rangle^2} \left\langle \left(\frac{l - r}{2} \right)^2 \middle| x, v \right\rangle = \frac{1}{\langle x \rangle^2} \langle (l - r)^2 \rangle. \quad (5.6)$$

OJO EN LA EC. ANTERIOR ULT. PASO. Therefore the term Q_x^2 can be interpreted as the average square deviation between the quantities of the molecules of each daughter

cell. For a perfect division $l = r$, making $Q_x = 0$. And the most noisy case is when one receives x molecules and the other 0 molecules, which makes intuitive sense.

EXPLAIN GENERAL FRAMEWORK?

5.2 Independent segregation

In the case of independent segregation, each molecule has an equal probability per unit time to switch from a cell half to another. Assuming there are l and $x - l$ molecules in each half, a process that can describe this statistic is given by

$$\begin{aligned} l &\xrightarrow{x-l} l+1 \\ l &\xrightarrow{l} l-1 \end{aligned} \tag{5.7}$$

Where the jacobian and diffusion (1×1) matrices are given by

$$\begin{aligned} \mathbf{A} &= \partial_l ((x - l) - l) = -2, \\ \mathbf{B} &= (x - l) + l = x \end{aligned} \tag{5.8}$$

which solving for the variance (covariance matrix) in steady state gives

$$\sigma^2(l|x) = \frac{x}{4}, \tag{5.9}$$

and averaging and using eq. (5.4) we get

$$Q_x^2 = \frac{4}{\langle x \rangle^2} \langle \sigma^2(l|x) \rangle = \frac{4}{\langle x \rangle^2} \frac{\langle x \rangle}{4} = \frac{1}{\langle x \rangle}. \tag{5.10}$$

In the following sections, we will find the noise for some common division mechanisms and compare it to the case of independent segregation.

5.3 Disordered segregation

5.3.1 General case

First we will consider a general case in which the rate with which each molecules goes from a cell half to the other is proportional to the available space generated by the upstream component. For a fixed number v of molecules of the upstream component, there are n and $v - n$ available spaces in a cell independently of x . Therefore, the process can be written as

$$\begin{aligned} l &\xrightarrow{n(x-l)} l + 1 \\ l &\xrightarrow{(v-n)l} l - 1 \end{aligned} \tag{5.11}$$

From the law of total variance we have

$$\sigma^2(l|x, v) = \langle \sigma^2(l|x, v, n) \rangle_{(n|v)} + \sigma^2(\langle l|x, v, n \rangle)_{(n|v)} \tag{5.12}$$

Where the subscript $(n|v)$ denotes that averages and variances are evaluated over the conditional PDF of n given x . Notice that taking averages over $(n|v)$ and over $(n|x, v)$ is the same in this case by the assumption that n is independent of x .

And by symmetry we have $\langle n|v \rangle = \frac{v}{2}$.

Therefore, by finding the first and second moments for $P(l|x, v, n)$, we can use eq. (5.12) to find the variance for l given x and v . We will use the method of the moment generating function on $P(l|x, v, n)$. The master equation for this PDF is given by.

$$\partial_t P(l|x, v, n) = n(x - (l - 1))P(l - 1|x, v, n) - (v - n)lP(l|x, v, n). \tag{5.13}$$

Writing the moment generating function (MGF) as

$$G(z) := \sum_{l=0}^x z^l P(l|x, v, n) \quad (5.14)$$

the master eq. in terms of G is given by (COMPELTE PROCEDURE)

$$\partial_t G(z) = nxzG(z) - (v - n + nz)z\partial_z G(z)$$

COMPLETE SOLUTION!!. At steady state we have

$$\partial_z G(z) = \frac{nxz}{(v - z + nz)z} G(z) \quad (5.15)$$

Solving with the boundary condition $G(1) = 1$, which follows from the normalization of the PDF. We find

$$G(z) = \left(1 + \frac{n}{v}(s - 1)\right)^x = \sum_{l=0}^x \binom{x}{l} \left(\frac{n}{v}\right)^l \left(1 - \frac{n}{v}\right)^{x-l} z^l. \quad (5.16)$$

Comparing with eq. (5.14) we get

$$P(l|x, v, n) = \binom{x}{l} \left(\frac{n}{v}\right)^l \left(1 - \frac{n}{v}\right)^{x-l}, \quad (5.17)$$

which is a binomial distribution as expected (COMMENT). From the CONCEPTS section, the average and variance are given by

$$\langle l|x, v, n \rangle = \frac{n}{v}x, \quad \sigma^2(L|x, v, n) = \frac{n}{v} \left(1 - \frac{n}{v}\right)x \quad (5.18)$$

We can also have been used the properties of G to find the first two moments. (DO IT?).

Therefore, taking the average of the conditional variance we get

$$\begin{aligned}\langle \sigma^2(l|x, v, n) \rangle_{(n|v)} &= \left\langle \frac{n}{v} \left(1 - \frac{n}{v}\right) x \middle| x, v, n \right\rangle_{(n|v)} = \left\langle \left(\frac{n}{v} - \frac{n^2}{v^2}\right) x \middle| x, v, n \right\rangle_{(n|v)} \\ &= \left(\frac{\langle n|v \rangle}{v} - \frac{\sigma^2(n|v) + \langle n|v \rangle^2}{v^2} \right) x\end{aligned}\quad (5.19)$$

Where we replaced $\langle n^2|v \rangle = \sigma^2(n|v) + \langle n \rangle^2$. Now since $\langle n|v \rangle = v/2$ we get

$$\langle \sigma^2(l|x, v, n) \rangle_{(n|v)} = \left(\frac{1}{2} - \frac{\sigma^2(n|v)}{v^2} + \frac{1}{4} \right) x = \frac{x}{4} (1 - Q_v^2), \quad (5.20)$$

where $Q_v^2 := 4\sigma^2(n|v)/v^2$. On the other hand we have

$$\sigma^2(\langle l|x, v, n \rangle)_{(n|v)} = \sigma^2\left(\frac{n}{v}x\right)_{(n|v)} = \frac{x^2}{v^2}\sigma^2(n|v) = \frac{x^2}{4}Q_v^2. \quad (5.21)$$

Replacing eqs. (5.20) and (5.21) on eq. (5.12) we get

$$\sigma^2(l|x, v) = \frac{x}{4}(1 - Q_v^2) + \frac{x^2}{4}Q_v^2,$$

averaging and multiplying by $4/\langle x \rangle^2$ we get

$$Q_x^2 = \frac{4}{\langle x \rangle^2} \langle \sigma^2(l|x, v) \rangle = \frac{4}{\langle x \rangle^2} \frac{1}{4} \langle x(1 - Q_v^2) + x^2 Q_v^2 \rangle = \frac{1}{\langle x \rangle} - \frac{\langle Q_v^2 x \rangle}{\langle x \rangle^2} + \frac{\langle Q_v^2 x^2 \rangle}{\langle x \rangle^2}. \quad (5.22)$$

Now, we will use this equation to calculate the partitioning error Q_x^2 at different scenarios.

5.3.2 Random size and random accessible volume

Now we will consider an available molecule for the molecules that varies randomly. Let n be the fraction of available volume in one of the daughter cells and assume that each molecule is equally likely to occupy any volume unit. Hence, the probability per unit time of each molecule leaving its cell half is proportional to the available volume in the

other cell half. Therefore the process is

$$\begin{aligned} l &\xrightarrow{n(x-l)} l+1 \\ l &\xrightarrow{(1-n)l} l-1 \end{aligned} \quad (5.23)$$

Which is the general case with $v = 1$. Assuming the volume variance is independent of x , eq. (5.22) becomes

$$Q_x^2 = \frac{1}{\langle x \rangle} - \frac{Q_v^2}{\langle x \rangle^2} (\langle x \rangle - \langle x^2 \rangle) = \frac{1}{\langle x \rangle} (1 - \langle Q_v^2 \rangle) + \langle Q_v^2 \rangle \frac{\langle x^2 \rangle}{\langle x \rangle^2}, \quad (5.24)$$

but $\langle x^2 \rangle / \langle x \rangle^2 = \sigma^2(x) / \langle x \rangle^2 = \eta_x^2 + 1$. Also, recalling the definition $Q_v^2 := 4\sigma^2(n|v)/v^2$. In this case $v = 1$ and it is fixed. Denoting it as Q_{vol}^2 we have $Q_{\text{vol}}^2 = Q_v^2 = \langle Q_v^2 \rangle = \sigma^2(n) / \langle n \rangle^2$. Therefore

$$\boxed{Q_x^2 = \frac{1 - Q_{\text{vol}}^2}{\langle x \rangle} + Q_{\text{vol}}^2(\eta_x^2 + 1)} \quad (5.25)$$

5.3.3 Clustered segregation

The clustering of molecules into vesicles could increase randomness in cell division. Let x and v be the total number of molecules and vesicles in a cell before division, respectively. Let x_i be the number of molecules in vesicle i , then $\sum_{i=0}^v x_i = x$. Two processes add randomness: the migration of the molecules between vesicles and the partition of the vesicles into each daughter cells.

In the first part, a vesicle loses a molecule with a probability proportional to its number of molecules.

$$(x_1, \dots, x_i, \dots, x_j, \dots, x_v) \xrightarrow{x_i} (x_1, \dots, x_i - 1, \dots, x_j + 1, \dots, x_v), \quad \text{for all } j \neq i. \quad (5.26)$$

In the second part, let n be the number of vesicles in one of the daughters, then similarly to the previous sections.

$$\begin{aligned} n &\xrightarrow{v-n} n+1 \\ n &\xrightarrow{n} n-1 \end{aligned} \tag{5.27}$$

If we assume both processes are independent, they could be done in any order to calculate the analytical expressions. i.e. it is the same to first distribute the molecules in each vesicle and then distribute the vesicles into each cell than to first distribute the empty vesicles between cells and then distribute the molecules. We will follow the second approach.

Let x_1, \dots, x_n be the number of molecules in each of the vesicles in one of the daughter cells and x_{n+1}, \dots, x_v be the number of molecules in the vesicles of the other daughter cell. As usual, let l be the number of molecules in one of the cells, then $l = \sum_{i=1}^n x_i$, and $x - l = \sum_{i=n+1}^v x_i$. Therefore, among all the possible transitions of eq. (5.26), the transitions coming from x_i and entering into x_j , for $i, j = 1, \dots, n$, or $i, j = n+1, \dots, v$, both with $i \neq j$ does not change the number of molecules. The net effect on the number of molecules in one of the daughter cells is

$$\begin{aligned} l &\xrightarrow{n(x_{n+1}+\dots+x_v)} l+1 \\ l &\xrightarrow{(v-n)(x_1+\dots+x_n)} l-1 \end{aligned} \tag{5.28}$$

EXPLAIN. Which can be simplified to

$$\begin{aligned} l &\xrightarrow{n(x-l)} l+1 \\ l &\xrightarrow{(v-n)l} l-1, \end{aligned} \tag{5.29}$$

which is the same as eq. (5.11), thus, from eq. (5.22)

$$Q_x^2 = \frac{1}{\langle x \rangle} - \frac{\langle Q_v^2 x \rangle}{\langle x \rangle^2} + \frac{\langle Q_v^2 x^2 \rangle}{\langle x \rangle^2} \quad (5.30)$$

Also, a correspondence can be made between eq. (5.7) and eq. (??) by switching $l \leftrightarrow n$ and $x \leftrightarrow v$. Then, using this correspondence on eq. (5.9) we get

$$\sigma^2(n|v) = \frac{v}{4} \quad (5.31)$$

And recalling that $Q_v^2 := 4\sigma^2(n|v)/v^2$ we have in this case

$$Q_v^2 = \frac{4}{v^2} \frac{v}{4} = \frac{1}{v}, \quad (5.32)$$

and replacing on eq. (5.30) we get

$$Q_x^2 = \frac{1}{\langle x \rangle} + \frac{1}{\langle x \rangle^2} \left(\left\langle \frac{x^2}{v} \right\rangle - \left\langle \frac{x}{v} \right\rangle \right) \quad (5.33)$$

EXPLAIN TERMS. If x and v are independent we can write it as

$$\begin{aligned} Q_x^2 &= \frac{1}{\langle x \rangle} - \frac{1}{\langle x \rangle} \left\langle \frac{1}{v} \right\rangle + \frac{\langle x^2 \rangle}{\langle x \rangle^2} \left\langle \frac{1}{v} \right\rangle = \frac{1}{\langle x \rangle} \left(1 - \left\langle \frac{1}{v} \right\rangle \right) + \frac{\langle x^2 \rangle}{\langle x \rangle^2} \left\langle \frac{1}{v} \right\rangle \\ &\approx \frac{1}{\langle x \rangle} + \frac{\langle x^2 \rangle}{\langle x \rangle^2} \left\langle \frac{1}{v} \right\rangle = \frac{1}{\langle x \rangle} + (1 + \eta_x^2) \left\langle \frac{1}{v} \right\rangle, \end{aligned} \quad (5.34)$$

under the assumption that $\langle 1/v \rangle \ll 1$, which also allow us to making a Taylor expansion of $\langle 1/v \rangle$ around $\langle v \rangle$ obtaining (????)

$$\begin{aligned} \left\langle \frac{1}{v} \right\rangle &\approx \left\langle \frac{1}{\langle v \rangle} - \frac{v - \langle v \rangle}{\langle v \rangle^2} + \frac{(v - \langle v \rangle)^2}{\langle v \rangle^3} \right\rangle \\ &= \frac{1}{\langle v \rangle} + \frac{\langle (v - \langle v \rangle)^2 \rangle}{\langle v \rangle^3} = \frac{1}{\langle v \rangle} (1 + \eta_v^2). \end{aligned} \quad (5.35)$$

After replacing we obtain

$$Q_x^2 \approx \frac{1}{\langle x \rangle} + \frac{(1 + \eta_x^2)(1 + \eta_v^2)}{\langle v \rangle} \quad (5.36)$$

When $x \ll 1$, $Q_x^2 \approx (1 + \eta_v^2)/\langle v \rangle$, (????) meaning that the segregation of clusters is the more significant factor on partitioning errors.

Now assume that $\langle x|v \rangle = sv$ where s is a constant representing the average number of molecules per vesicle. The term in parentheses of eq. becomes by definition of expected value

$$\begin{aligned} \left\langle \frac{x^2}{v} \right\rangle - \left\langle \frac{x}{v} \right\rangle &= \sum_{x,v} \left(\frac{x^2}{v} - \frac{x}{v} \right) P(x, v) = \sum_v \frac{1}{v} \left[\sum_x (x^2 - x) P(x|v) \right] P(v) \\ &= \sum_v \frac{1}{v} (\langle x^2|v \rangle - \langle x|v \rangle) P(v) = \sum_v \frac{1}{v} (\sigma^2(x|v) + \langle x|v \rangle^2 - \langle x|v \rangle) P(v) \\ &= \sum_v \frac{1}{v} \left(\frac{sv\sigma^2(x|v)}{\langle x|v \rangle} + s^2v^2 - sv \right) P(v) = s \left\langle \frac{\sigma^2(x|v)}{\langle x|v \rangle} \right\rangle + s^2\langle v \rangle - s \end{aligned} \quad (5.37)$$

Where we used the conditional probability theorem (??) to write $P(x, v) = P(x|v)P(v)$. Defining $q := \sigma^2(x|v)/\langle x|v \rangle$ and recalling that by the law of iterated expectations $\langle x \rangle = \langle \langle x|v \rangle \rangle = s\langle v \rangle$ we get

$$\left\langle \frac{x^2}{v} \right\rangle - \left\langle \frac{x}{v} \right\rangle = s\langle x \rangle + s(q - 1). \quad (5.38)$$

Replacing in eq. we get

$$Q_x^2 = \frac{1}{\langle x \rangle} + \frac{1}{\langle x \rangle^2} (s\langle x \rangle + s(q - 1)) \quad (5.39)$$

And if $(1 - q)$ is very small (or 0 as in the case of a Poissonian), we can approximate it as

$$Q_x^2 \approx \frac{1}{\langle x \rangle} + \frac{s}{\langle x \rangle} \quad (5.40)$$

5.3.4 Upper limit of the partitioning error

PUT IT IN ANOTHER PART There is an upper bound for the partitioning error corresponding to the case when one daughter cells keeps all the molecules. There is an equal probability of each daughter to keep all of them, hence

$$\sigma^2(L|x) = \left\langle \left(l - \frac{x}{2} \right)^2 \right\rangle = \frac{1}{2} \left(0 - \frac{x}{2} \right)^2 + \frac{1}{2} \left(x - \frac{x}{2} \right)^2 = \frac{x^2}{4} \quad (5.41)$$

Therefore

$$Q_x^2 = \frac{4}{\langle x \rangle^2} \langle \sigma^2(l|x) \rangle = \frac{4}{\langle x \rangle^2} \langle x^2 \rangle 4 = \eta_x^2 + 1. \quad (5.42)$$

It only depends on the prior heterogeneity of the mother cells.

5.4 Ordered segregation

5.4.1 Self-volume exclusion

By analogy to eq. FILL making the correspondence FILL. If a molecule occupy a fixed volume fraction k of the total cell volume, we have

UNDERSTAND

$$\sigma^2(l|x) = \frac{1}{4} x(1 - kx), \quad (5.43)$$

so that

$$Q_x^2 = \frac{4}{\langle x \rangle^2} \langle \sigma^2(l|x) \rangle = \frac{4}{\langle x \rangle^2} \frac{1}{4} (\langle x \rangle - k \langle x^2 \rangle) = \frac{1}{\langle x \rangle} - k \frac{\langle x^2 \rangle}{\langle x \rangle^2} = \frac{1}{\langle x \rangle} - k(\eta_x^2 + 1). \quad (5.44)$$

COMMENTS. It can be noticed that the reduction with respect to independent segregation gets bigger when the volume fraction occupied by each molecule is bigger. This makes sense because it makes the exclusion bigger when there are more molecules, having the net effect of reducing the unevenness.

5.4.2 Binding to spindle sites

Suppose each dividing cell has a random number of binding sites v which are also distributed randomly between both daughter cells. Letting x be the (random) total number of molecules of some type which are going to bind the sites before division. We will separate cells in which $v < x$ and $v \geq x$. Assume also that the binding is such that all possible molecules of x are bound, that is, at equilibrium if $v < x$ all binding sites are occupied, and if $v > x$, all molecules are bound to sites.

Let n be the number of binding sites on a cell half, and suppose that it increases with a rate dependent on the number of binding sites in the other cell half, then

$$\begin{aligned} n &\xrightarrow{f(v-n)} n+1 \\ n &\xrightarrow{f(n)} n-1 \end{aligned} \tag{5.45}$$

where f is some function. Also, the rate at which a molecule leaves a cell half is proportional to the number of molecules in its cell half and the number of free sites in the other cell half, obtaining

$$\begin{aligned} l &\xrightarrow{\lambda(n-l)(x-l)} l+1 \\ l &\xrightarrow{\lambda(v-n-x+l)l} l-1 \end{aligned} \tag{5.46}$$

UNDERSTAND AND EXPLAIN. SOLVE. Solving the FDR we get

$$\sigma^2(l|x, v) = \frac{1}{4} \left(x - \frac{x^2}{v} + Q_v^2 x^2 \right), \quad \text{for } v \geq x, \quad (5.47)$$

with $Q_v^2 := 4\sigma^2(n|v)/v^2$ as before. In the case $v < x$, the v copies of the molecule that are bound segregate along with n , and for the remaining copies suppose they segregate independently. The result is (COMPARE?)

$$\sigma^2(l|x, v) = \frac{1}{4}(x - v) + \sigma^2(n|v), \quad \text{for } v < x. \quad (5.48)$$

OJO, ERROR AQUI (CREO QUE YA CORREGIDO) Putting together both cases we get by definition of expectations

$$\begin{aligned} Q_x^2 &= \frac{4}{\langle x \rangle^2} \langle \sigma^2(l|x) \rangle = \frac{1}{\langle x \rangle^2} \sum_{x,v} \sigma^2(l|x) P(x, v) \\ &= \frac{1}{\langle x \rangle^2} \left[\sum_{v \geq x} \left(x - \frac{x^2}{v} + Q_v^2 x^2 \right) P(x, v) + \sum_{v < x} ((x - v) + 4\sigma^2(n|v)) P(x, v) \right] \\ &, \end{aligned}$$

notice that there is an x in both sums than can be taken out as a $\langle x \rangle$, also, by replacing $4\sigma^2(n|v) = v^2 Q_v^2$ and separating the sums by first summing x and then over all vs we obtain

$$\begin{aligned} Q_x^2 &= \frac{1}{\langle x \rangle} - \frac{1}{\langle x \rangle^2} \sum_{v=0}^{\infty} \left[\sum_{x=0}^v \left(\frac{1}{v} - Q_v^2 \right) x^2 P(x, v) + \sum_{x=v+1}^{\infty} (v - v^2 Q_v^2) P(x, v) \right] \\ &= \frac{1}{\langle x \rangle} - \frac{1}{\langle x \rangle^2} \sum_{v=0}^{\infty} \left[\left(\frac{1}{v} - Q_v^2 \right) \sum_{x=0}^v x^2 P(x, v) + (v - v^2 Q_v^2) \sum_{x=v+1}^{\infty} P(x, v) \right]. \end{aligned} \quad (5.49)$$

To make interpretations easier, consider a special case in which v is fixed, each daughter cell receives exactly $v/2$ binding sites, $\langle x \rangle = v$, and $P(x)$ is symmetric. With these assumptions, the previous eq. can be reduced

$$Q_x^2 = \frac{1}{\langle x \rangle} - \frac{1}{\langle x \rangle^2} \left[\frac{1}{v} \sum_{x=0}^v x^2 P(x) + v \sum_{x=v+1}^{2v} P(x) \right]$$

where we made $Q_v^2 = 0$ because there is no uncertainty on n since each cell receives exactly $v/2$ sites. Also, for the sum over v only survives the term corresponding to the fixed number v of binding sites. Writing $x^2 = (x - \langle x \rangle)^2 + 2x\langle x \rangle - \langle x \rangle^2$ on the first sum we get

$$Q_x^2 = \frac{1}{\langle x \rangle} - \frac{1}{\langle x \rangle^2} \left[\frac{\sigma^2(x)}{v} \sum_{x=0}^v P(x) + \frac{2\langle x \rangle}{v} \sum_{x=0}^v xP(x) - \frac{\langle x \rangle^2}{v} \sum_{x=0}^v P(x) + v \sum_{x=v+1}^{2v} P(x) \right]$$

evaluating $\langle x \rangle = v$ we get

$$Q_x^2 = \frac{1}{\langle x \rangle} - \frac{1}{\langle x \rangle^2} \left[\frac{\sigma^2(x)}{v} \sum_{x=0}^v P(x) + 2 \sum_{x=0}^v xP(x) - v \sum_{x=0}^v P(x) + v \sum_{x=v+1}^{2v} P(x) \right],$$

and since $P(x)$ is symmetric about $x = v$ we have $\sum_{x=0}^v P(x) = \sum_{x=v+1}^{2v} P(x) = 1/2$

$$Q_x^2 = \frac{1}{\langle x \rangle} - \frac{1}{\langle x \rangle^2} \left[\frac{\sigma^2(x)}{2v} + 2 \sum_{x=0}^v xP(x) \right]. \quad (5.50)$$

The absolute deviation $\langle |x - \langle x \rangle| \rangle$ can be written using the symmetry of $P(x)$ as

$$\langle |x - \langle x \rangle| \rangle = \sum_{x=0}^{\infty} |x - \langle x \rangle| P(x) = 2 \sum_{x=0}^v (\langle x \rangle - x) P(x) = \langle x \rangle - 2 \sum_{x=0}^v xP(x). \quad (5.51)$$

Thus, solving for the sum and replacing we get

$$Q_x^2 = \frac{1}{\langle x \rangle} - \frac{1}{\langle x \rangle^2} \left[\frac{\sigma^2(x)}{2v} + \langle x \rangle - \langle |x - \langle x \rangle| \rangle \right] = \frac{\langle |x - \langle x \rangle| \rangle}{\langle x \rangle^2} - \frac{\eta_x^2}{2v}$$

And using $v = \langle x \rangle$

$$Q_x^2 = \frac{1}{\langle x \rangle} \left(\frac{\langle |x - \langle x \rangle| \rangle}{\langle x \rangle} - \frac{\eta_x^2}{2} \right) \quad (5.52)$$

ANALYSIS, η_x cannot exceed 1 by the symmetry of $P(x)$?

5.4.3 Pair formation mechanisms

A mechanism of ordered segregation consists in the pair formation of the molecules to be segregated and then spindles are formed which separates each molecule forming the pair into each cell half.

Assume that in each cell there are z pairs of molecules and m independent molecules i.e. $x = m + 2z$. Suppose also that the paired molecules do not interact with the unpaired ones, then

$$\begin{aligned} \langle \sigma^2(l|x) \rangle &= \langle \sigma_p^2(L|2z) \rangle + \langle \sigma_u^2(l|m) \rangle = \sum_{x,z} [\sigma_p^2(l|2z) + \sigma_u^2(l|m)] P(2z, x) \\ &= \sum_{x,z} [\sigma_p^2(l|2z) + \sigma_u^2(l|x - 2z)] P(2z|x) P(x) \end{aligned} \quad (5.53)$$

where the subscripts 'p' and 'u' stand for 'paired' and 'unpaired' and $P(2z, x)$ is the PMF of having z pairs and a total of x molecules before division. The variances can be added in that way because for independent random variables, the variance of a sum is the sum of the variances (IS THIS THE REASON?).

Now We will proceed to find each one of the variances. The unpaired molecules segregate independently, therefore, by comparison with eq. (5.9) we get

$$\sigma_u^2(l|x - 2z) = \frac{x - 2z}{4}. \quad (5.54)$$

For the paired molecules, assume that each pair is split to separate daughters with probability p and to the same daughter with probability $1 - p$. MORE COMMENTS

$$\begin{aligned}
(M, l, r) &\xrightarrow{pM} (M - 2, l + 1, r + 1) \\
(M, l, r) &\xrightarrow{(1-p)M/2} (M - 2, l + 2, r) \\
(M, l, r) &\xrightarrow{(1-p)M/2} (M - 2, l, r + 2)
\end{aligned} \tag{5.55}$$

where the first line represents a succesfull split and the other two unsuccessful ones.

EXPLAIN ALL THE FDR, AND THE MATRICES, ETC. —————

we get

$$\sigma_p^2 = (1 - p)z. \tag{5.56}$$

Replacing eqs. (5.54) and (5.56) in eq. (5.53) we get

$$\begin{aligned}
\langle \sigma^2(l|x) \rangle &= \sum_x \left[\sum_z \left((1-p)z + \frac{x-2z}{4} \right) P(2z|x) \right] P(x) = \sum_x \left(\frac{1-p}{2} \langle 2z|x \rangle + \frac{x - \langle 2z|x \rangle}{4} \right) P(x) \\
&= \frac{1}{4} \sum_x (2(1-p)\langle 2z|x \rangle + x - \langle 2z|x \rangle) P(x) = \frac{1}{4} \sum_x (x - (2p-1)\langle 2z|x \rangle) P(x) \\
&= \frac{1}{4} (\langle x \rangle - (2p-1)\langle 2z \rangle).
\end{aligned} \tag{5.57}$$

Hence, (POR QUE EL SOBRE 2, SUSTITUCION?)

$$\boxed{Q_x^2 = \frac{1 - (2p-1)k}{\langle x \rangle}} \tag{5.58}$$

where $k := \langle 2z \rangle / \langle x \rangle$ is the average fraction of molecules that are in pairs. If $k = 0$ there is independent segregation and $Q_x = 1/\langle x \rangle$ on the previous equation. For the segregation into pairs to be 'ordered', p must be greater than $1/2$, in the opposite case, the paired molecules have a higher chance of not being split, increasing segregation error with respect to the independent case.

[7] [1] [2] [3] [4] [5] [6] [8] [9] [10] [16] [17] [13] [14] [15]

Bibliography

1. Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464 (2005).
2. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 183–186 (2002).
3. Pedraza, J. M. & van Oudenaarden, A. Noise Propagation in Gene Networks. *Science* **307**, 2005 (1965–1969).
4. Paulsson, J. Models of stochastic gene expression. *Phys. Life Rev.* **2**, 157–175 (2005).
5. Alon, U., Surette, M. G., Barkai, N. & Leibler, S. Robustness in bacterial chemotaxis. *Nature* **397**, 168–171 (1999).
6. Arkin, A., Ross, J. & McAdams, H. H. Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage λ -Infected Escherichia coli Cells. *Genetics* **149**, 1633–1648 (1998).
7. Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8614–8619 (2001).
8. Pedraza, J. M. & Paulsson, J. Effects of Molecular Memory and Bursting on Fluctuations in Gene Expression. *Science* **319**, 339–343 (2008).
9. Huh, D. & Paulsson, J. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat. Genet.* **43**, 95–100 (2011).

10. Huh, D. & Paulsson, J. Random partitioning of molecules at cell division. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15004–15009 (2011).
11. Alberts, B. *et al.* *Essential Cell Biology* (Garland Science, New York, 2013).
12. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman and Hall/CRC, Boca Raton, 2006).
13. Bertsekas, D. P. & Tsitsiklis, J. N. *Introduction to Probability* (Athena Scientific, Belmont, 2008).
14. Van Kampen, N. G. *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992).
15. Gardiner, C. W. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (Springer, Berlin, 2003).
16. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).
17. Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004).