## Contributions of cell growth and biochemical reactions to non-genetic variability of cells

### Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1</td>
<td>Introduction</td>
<td>182</td>
</tr>
<tr>
<td>7.2</td>
<td>Results</td>
<td>183</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Decomposition of molecule copy number variance into biochemical reaction and cell growth contributions</td>
<td>183</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Cell-division variance</td>
<td>184</td>
</tr>
<tr>
<td>7.2.3</td>
<td>Mean and variance of the copy number at a particular cell age</td>
<td>187</td>
</tr>
<tr>
<td>7.2.4</td>
<td>Variance due to the distribution of cell ages</td>
<td>188</td>
</tr>
<tr>
<td>7.2.5</td>
<td>Variance due to the biochemical reactions and (extrinsic) noise propagation</td>
<td>189</td>
</tr>
<tr>
<td>7.2.6</td>
<td>The full copy number distribution for zero-order synthesis and first-order degradation</td>
<td>190</td>
</tr>
<tr>
<td>7.2.7</td>
<td>Variance decomposition for concentrations</td>
<td>193</td>
</tr>
<tr>
<td>7.3</td>
<td>Discussion</td>
<td>194</td>
</tr>
<tr>
<td>7.4</td>
<td>Supplemental Information</td>
<td>197</td>
</tr>
<tr>
<td>7.4.1</td>
<td>Decomposition of the variance in molecule copy numbers</td>
<td>197</td>
</tr>
<tr>
<td>7.4.2</td>
<td>Variance from cell division</td>
<td>198</td>
</tr>
<tr>
<td>7.4.3</td>
<td>Variance due to partitioning for a stable molecule</td>
<td>198</td>
</tr>
<tr>
<td>7.4.4</td>
<td>Population level variance decomposition for a simple burst model</td>
<td>199</td>
</tr>
<tr>
<td>7.4.5</td>
<td>Using generating functions to calculate the cell age dependent copy number distribution</td>
<td>200</td>
</tr>
<tr>
<td>7.4.6</td>
<td>Noise in interdivision times</td>
<td>203</td>
</tr>
</tbody>
</table>

Adapted from: Biophys J. 2014 Jul 15;107(2):301-13
Cell-to-cell variability in the molecular composition of isogenic, steady-state growing cells arises spontaneously out of the inherent stochasticity of intracellular biochemical reactions and cell growth. Here, we present a general decomposition of the total variance in the copy number per cell of a particular molecule. It quantifies the individual contributions made by processes associated with cell growth, biochemical reactions, and their control. We decompose the growth contribution further into variance contributions of random partitioning of molecules at cell division, mother-cell heterogeneity, and variation in cell-cycle progression. The contribution made by biochemical reactions is expressed in variance generated by molecule synthesis, degradation and their regulation. We use this theory to study the influence of different growth and reaction-related processes, such as DNA replication, variable molecule-partitioning probability, and synthesis bursts, on stochastic cell-to-cell variability. Using simulations, we characterize the impact of noise in the generation-time and asymmetric cell division on cell-to-cell variability. This paper offers a widely-applicable theory on the influence of biochemical reactions and cellular growth on the phenotypic variability of growing, isogenic cells. The theory aids the design and interpretation of experiments involving single-molecule counting or real-time imaging of fluorescent reporter constructs.

7.1 Introduction

Single-cell experiments illustrate that isogenic cells generally differ markedly in the copy numbers of mRNA and protein molecules [Raj 2008b, Eldar 2010] and a multitude of other systems properties, such as cell volume, growth rate, and phenotypic state [Wang 2010, Stewart 2005, Koppes 1978]. These experiments exploit single-molecule counting methods [Raj 2008a] or fluorescent reporter constructs [Young 2012] to quantify the levels of specific molecules in single cells. A wide range of processes have been shown to contribute to non-genetic cell-to-cell variability [Berg 1978]: e.g. fluctuation-induced imbalances in molecule synthesis and degradation [Elowitz 2002, Ozbudak 2002], synthesis control [Ozbudak 2002], synthesis bursts [Golding 2005], binomial partitioning of molecules at cell division [Huh 2011b], bistable switching [Veening 2005], and noise propagation [Pedraza 2005].

In most single-cell studies, stochastic models are used to explain experimental findings. However, often those models highlight only a particular aspect of cellular stochasticity and they are highly simplified to overcome the problem of kinetic parameter uncertainty. As a consequence, it often remains unclear to what extent a particular stochastic phenomenon contributes to the total cell-to-cell variability as it is only one out of several and possibly many. To address this issue we present a decomposition of the total copy number variance in terms of the contributions of a wide range of cell growth and biochemical reaction processes. Such variance decomposition methods have recently been introduced to the field of stochastic cell biology [Paulsson 2004, Hilfinger 2011, Hilfinger 2012, Bowsher 2012, Swain 2002].
7.2. Results

7.2.1 Decomposition of molecule copy number variance into biochemical reaction and cell growth contributions

We shall address the variability in the copy numbers of a particular molecule across a population of isogenic cells in balanced growth. During this growth condition, individual cells progress through the cell cycle asynchronously and a probability distribution describes the fraction of cells that have reached a particular cell-cycle progression status (cell age, \( a \)). This cell age distribution, \( u(a) \), is completely determined by the interdivision time distribution [Painter 1968], which characterizes the probability for a cell to have a particular generation time. The interdivision (or generation) time, \( T \), is the time period between subsequent cell divisions and equals the duration of the cell cycle. The theory presented in this work is limited to deterministic interdivision times, some of the simulations we discuss later analyze the consequences of this assumption for cell-to-cell variability. Figure 7.1 illustrates the key processes of cell growth and biochemical reactions that contribute to cell-to-cell variability in the molecule copy number.

The variance of the copy number of a molecule \( X \), degraded by a first-order reaction with rate constant \( k_d \), in a steady-state, exponentially growing population of isogenic cells at rate \( \mu \) can be written as the following sum of variance contributions.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \langle \cdot \rangle ), ( \langle \cdot ; \cdot \rangle )</td>
<td>mean and variance ( \langle \cdot \rangle ), conditional mean and variance ( \langle \cdot ; \cdot \rangle )</td>
</tr>
<tr>
<td>( T, \sigma^2 )</td>
<td>generation time and variance</td>
</tr>
<tr>
<td>( a )</td>
<td>cell age (( 0 \leq a \leq T ))</td>
</tr>
<tr>
<td>( X_a = x[a] )</td>
<td>number of newly synthesized ( X ) molecules since the last division up to cell age ( a )</td>
</tr>
<tr>
<td>( u(a) )</td>
<td>cell age distribution; ( u(a) = \frac{\ln 2}{T} 2^{\frac{-a}{T}} ) for deterministic interdivision times</td>
</tr>
<tr>
<td>( \mu )</td>
<td>specific growth rate of the population; ( \mu = \ln 2 / T ) for deterministic interdivision times</td>
</tr>
<tr>
<td>( p(a) = e^{-k_d a} )</td>
<td>survival probability of molecule ( X )</td>
</tr>
<tr>
<td>( 1/k_d )</td>
<td>intrinsic life time of molecule ( X )</td>
</tr>
<tr>
<td>( q )</td>
<td>probability that a molecule of ( X ) at cell division will be inherited by a particular daughter cell</td>
</tr>
</tbody>
</table>

Table 7.1: Notations

They offer a powerful method to analyze the causes of cell-to-cell variability without having to consider complex mechanistic models of underlying molecular reactions and cellular growth process and as such they have great potential for the analysis of experimental data and for experiment design.
184 Chapter 7

At birth, the variance in the obtained number of molecules equals \( \langle \delta^2 x_0 \rangle \), which can be written as the sum \( \langle \delta^2 x_0 | x_T \rangle \) using the law of total variance \((x_T\) denotes the number of molecules of the mother cell at the time of division). This variance decays while the cell matures due to degradation of molecules (if \( k_d \neq 0 \)). The averages \( \langle p(a)^2 \rangle \) and \( \langle p(a)(1 - p(a)) \rangle \) can be expressed in terms of the parameter ratio \( k_d/\mu \) and describe the rate of loss of the memory of cell division variance and variance in the survived number of molecules obtained at division (section 7.4.2). If this ratio is large, the molecule lives much shorter than the generation time and \( \langle p(a)^2 \rangle \approx 0 \) and \( \langle p(a)(1 - p(a)) \rangle \approx 0 \). In this limit, the variance in \( x \) is determined only by the variance introduced during the synthesis of new \( X \) molecules and due to differences in cell age.

The \( \langle \delta^2 (x_a) \rangle \) captures the variation in the copy numbers of \( X \) per cell due to the dependency of \( x \) on the cell age. Since \( x \) needs to double from age 0 to \( T \) (fig. 7.2), \( \langle x_a \rangle \) is an increasing function of \( a \) (e.g. linear or exponential; for experimental illustrations see [Sigal 2006, Cohen 2009, Cookson 2005, Rosenfeld 2005]), which introduces a variance in \( x \) solely due to cell age differences (as will be discussed in a later section). We emphasize that \( \langle \delta^2 (x_a) \rangle \) is not indicating cell-to-cell variability due to stochastic fluctuations, rather it arises solely from molecule synthesis accompanying the growth process.

### 7.2.2 Cell-division variance

One contribution to the variance in the copy number of a molecule derives from the variance introduced during cell division (discussed in detail by Huh and Paulsson [Huh 2011b]) and the extent by which this has decayed during a cell’s generation time. We found that this decay depends on the parameter ratio \( k_d/\mu \). Next, we will express the variance term in a newborn daughter cell, \( \langle \delta^2 x_0 \rangle \), in terms of the mother cell variance and the mechanism of cell division. Here we limit ourselves to
7.2. Results

stochastic synthesis and degradation of molecules partitioning: each molecule has a probability of $q$ to go to cell 1, and $(1-q)$ to go to cell 2. Variability in the rate of volume increase variability in the volume at which cells divide segregation into daughter cells of unequal size cell age $a$, (time since last division) number of molecules at cell birth ($a = 0$) $x_0 = 2$, $x_1 = 7$, $x_T = 6$ number of molecules at age $a$ number of molecules at cell division ($a = T$) $X_T = 7$ number of newly synthesized (and not yet degraded) molecules at age $a$ number of newly synthesized (and not yet degraded) molecules at cell division $A = 5$, $B = \ldots$, $PDF$ cell age number of molecules/cell $x_T$ PDF cell age

Figure 7.1: Overview of the biochemical reaction and cell-growth processes contributing to cell-to-cell variability. A. Heterogeneity due to (i) biochemical reactions, i.e. fluctuations in rate of molecule synthesis and its regulation, and (ii) due to binomial partitioning of molecules at cell division, i.e. variability in molecule content of mother cells, and (iii) variability in partitioning. Molecules inherited from the mother cell are shown in faded red, newly synthesized molecules in bright red. B. Heterogeneity due to the balanced growth process. At a specific cell age, the molecule copy numbers of cells follow different probability distributions. Each cell age has a certain probability of occurrence. The net distribution of molecule copy number per cell in a steady-state growing population of isogenic cells is given by marginalizing the conditional copy number distribution over the cell age distribution.
independent (binomial) partitioning of molecules but allow the partition probability, \( q \), (i.e. the probability for each molecule to end up in one daughter cell) to have a distribution by itself, denoted by \( g(q) \) (with \( q = 1/2 \)).

The condition of balanced growth together with a partitioning mechanism prescribes equations for the moments of the copy number distributions at cell birth and at cell division. For the mean we have \( \langle x|0 \rangle = \langle x|T \rangle / 2 \) and to obtain the variance, the law of total variance is applied twice (assuming that \( q \) and \( x_T \) are independent) (section 7.4.2):

\[
\langle \delta^2 x_0 \rangle = \frac{\langle \delta^2 q \rangle \langle x_T \rangle^2}{\langle \delta^2 \langle x_0 \rangle \rangle} + \frac{1}{4} \left( 1 - \langle \delta^2 q \rangle \right) \langle x_T \rangle + \frac{1}{4} \left( 1 + \langle \delta^2 q \rangle \right) \langle \delta^2 x_T \rangle
\] (7.2)

We note that rearrangement of this equation leads to the same result as given in reference [Huh 2011b]. This equation indicates that fluctuations in the partitioning probability, for instance because of different volumes of two sister cells or intracellular organisation [Huh 2011b], enhance \( \langle \delta^2 x_0 \rangle \). For a molecule that is degraded rapidly (its half-life is short in comparison to the generation time), the variance at cell division, \( \langle \delta^2 x_T \rangle \), is mainly determined by the variance of molecules that were newly synthesized during the last cell cycle. In contrast, for stable molecules (\( k_{deg} \approx 0 \)) \( \langle \delta^2 x_T \rangle = \langle \delta^2 x_0 \rangle + \langle \delta^2 X_T \rangle \), i.e. the variance at cell division has a significant contribution from the variance at cell birth and therefore also from the fluctuations in \( q \) during previous divisions. Therefore for a stable molecule, fluctuations in \( q \) are a significant contribution to the copy number noise. The population level copy number noise for a stable molecule that is synthesized by a zero-order reaction can be written as (see section 7.4.5.3 for a derivation):

\[
\frac{\langle \delta^2 x \rangle}{\langle x \rangle^2} = \frac{1}{\langle x \rangle} + \frac{(1 - 2 \ln(2))^2}{3 - 4 \langle \delta^2 q \rangle} + \frac{16 \ln(2)^2 \langle \delta^2 q \rangle}{3 - 4 \langle \delta^2 q \rangle}
\] (7.3)

For many microorganisms the coefficient of variation of \( q \) is in the range of 3–7% [Nanninga 1979, Trueba 1982] (i.e. \( \langle \delta^2 q \rangle = 0.0075 \) to 0.0175). As a result, the last term in this equation lies between 0.019 and 0.046. For a molecule with high copy numbers, where intrinsic noise (\( 1/\langle x \rangle \)) is low, this contribution can become quite significant.
7.2. Results

7.2.3 Mean and variance of the copy number at a particular cell age

The mechanism of partitioning of molecules at division, together with the condition of balanced growth, sets the boundary conditions for the distributions (and their moments) of copy numbers at birth and at division. Combining these boundary conditions with the kinetics of molecule synthesis yields a system of equations for the moments of the copy number distribution at cell age \( a \) (see also section 7.4.1, Eqs. 7.26, 7.26):

\[
\begin{align*}
\langle x_a \rangle &= \frac{p(a)\langle x_0 \rangle}{1 - p(T)\langle q \rangle} \langle X | a \rangle + \langle X \rangle \\
\langle \delta^2 x_a \rangle &= \frac{p(a)^2 \langle \delta^2 x_0 \rangle}{1 - p(T)\langle q \rangle} + p(a)(1 - p(a))\langle x_0 \rangle \langle X | a \rangle + \langle X \rangle \\
\langle \delta^2 x_0 \rangle &= \frac{\langle x_T \rangle (\langle q \rangle - \langle q \rangle^2) + \langle \delta^2 \rangle \langle x_T \rangle^2}{1 - \langle q \rangle^2 p(T)^2} + \frac{\langle q \rangle^2 (p(T) - p(T)^2) \langle x_0 \rangle + \langle q \rangle^2 \langle \delta^2 \rangle \langle x_T \rangle}{1 - \langle q \rangle^2 p(T)^2}
\end{align*}
\]

By setting \( a = T \) in eq. 7.4, 7.5 and combining with eq. 7.1, 7.2 one can express the population level variance solely in terms of the variance in the synthesis (which can include intrinsic and extrinsic components), the variance in the partition distribution, \( \langle \delta^2 q \rangle \), and the survival probability of molecules:

We show this set of equations to emphasize that the theory in this paper leads to experimentally testable relationships, in addition to understanding. Data analysis of movies resulting from real-time imaging of cell growth and dynamics of a fluorescent reporter construct [Locke 2009] allow for quantification of each of the quantities in these equations. This then allows for detailed understanding of the mechanistic origins of molecular copy variance.
7.2.4 Variance due to the distribution of cell ages

The third variance term in eq 7.1, $\langle \delta^2 \langle x_a \rangle \rangle$, quantifies the differences in the average copy numbers at different cell ages. Due to binary divisions, the average copy number per cell needs to double during a generation time (fig. 7.2). Therefore, cells in an asynchronous population have different ages and older cells will on average have a higher copy number than younger cells. The variance deriving from these cell-age differences, $\langle \delta^2 \langle x_a \rangle \rangle$ does in general not constitute a source of functional noise and derives solely from the growth process. In fact, it makes a serious contribution to copy number noise as we shall show in this section.

If the synthesis rate is constant throughout the cell cycle, the cell-age dependent average copy number of produced molecules can for many simple models be written as,

\[
\langle X_a \rangle = \frac{k_s}{k_d} (1 - e^{-k_d a}) \tag{7.9}
\]
\[
\langle x_a \rangle = p(a)\langle x_0 \rangle + \langle X_a \rangle \tag{7.10}
\]

where $k_s$ denotes the effective synthesis rate, which could for example include bursty production. Equation 7.10 results from rewriting equation 7.7. These equations can for instance apply to mRNA but are equally valid for most proteins provided that the lifetime of mRNA is short relative to the cell-cycle duration. Given these equation, the contribution of differences in cell age to the variance in copy numbers equals,

\[
\langle \delta^2 \langle x_a \rangle \rangle = \int_0^T w(a) (\langle x_a \rangle - \langle x \rangle)^2 da = \langle x \rangle^2 f\left(\frac{k_d}{\mu}\right) \tag{7.11}
\]

where $f(k_d/\mu)$ is a monotonically decreasing function of $k_d/\mu$. As we found earlier, the ratio $k_d/\mu$ is an essential parameter. In the limit of $k_d/\mu \to 0$, when the molecule is only diluted by growth, $\langle \delta^2 \langle x_a \rangle \rangle$ simplifies to

\[
\langle \delta^2 \langle x_a \rangle \rangle = k_s^2 \left(1 - 2 \ln(2)^2\right) \frac{1}{\mu^2} = \langle x \rangle^2 (1 - 2 \ln(2)^2) \approx 0.04 \langle x \rangle^2 \tag{7.12}
\]

with $\langle x \rangle = \frac{k_s}{\ln(2)}$ as the population average copy number. Thus, for a stable molecule differences in cell age contribute a minimum of 0.04 to the noise in copy numbers across a population of cells. This contribution can be expected to be very significant for moderately to highly expressed proteins while becoming negligible as compared to $\langle \delta^2 x | a \rangle$ for molecules with low copy numbers because the intrinsic noise contribution $1/\langle x \rangle$ dominates (see below).
7.2. Results

![Diagram](image)

Figure 7.2: **Cell-age dependent and independent noise contributions.** Average copy number as function of cell age is shown as thick gray line, results of individual stochastic simulations are shown in blue and green. The dashed black line indicates the population average copy number. Differences between the blue and gray lines at a fixed age \((x_a - \langle x_a \rangle)\) lead to functional noise, while the differences between the cell age dependent average and the total average \((\langle x_a \rangle - \langle x \rangle)\) derive from the growth process alone. The distribution of the cell age dependent averages is shown on the right.

If the synthesis rate changes during the course of the cell cycle, e.g. because the gene encoding the protein of interest is replicated, the cell-age distribution dependent contribution to the variance changes. Assuming again a stable molecule and synthesis with rate \(k_s\) up to time \(t_r\) (time of replication) and a doubled synthesis rate between \(t_r\) and division at time \(T\), the cell cycle dependent average and its variance equal:

\[
\begin{align*}
\langle x_a \rangle &= k_s(a + 2T - t_r) \quad \text{if } a < t_r \\
\langle x_a \rangle &= 2k_s(a + T - t_r) \quad \text{if } a \geq t_r \\
\langle \delta^2 \langle x_a \rangle \rangle &= \langle x \rangle^2 \left( -1 + 2 \frac{t_r}{T} (3 + \ln(4)) \right) \\
&- \langle x \rangle^2 \frac{4t_r}{T} \left( \left( \frac{t_r}{T} \right)^2 \ln(2)^2 - 2 \frac{t_r}{T} \ln(2)(1 + \ln(4)) \right) \\
&- \langle x \rangle^2 \frac{4t_r}{T} \left( 2 + 4 \ln(2)^2 + \ln(16) \right)
\end{align*}
\]

This variance depends on the ratio of \(t_r/T\) and can take values in the range of \(\approx 0.035 \langle x \rangle^2\) and \(\approx 0.046 \langle x \rangle^2\) as was determined before numerically by Marathe and Klumpp [Marathe 2010] and in the limit of \(t_r \to 0\) or \(t_r \to T\) it reduces to \(\langle x \rangle^2(1 - 2\ln(2)^2)\) as we found earlier.

### 7.2.5 Variance due to the biochemical reactions and (extrinsic) noise propagation

The term \(\langle \delta^2 X_a \rangle\) in the variance decomposition (eq. 7.1) captures contributions from the synthesis of new molecules since the last cell division event. This comprises contributions from intrinsic as well as extrinsic noise [Paulsson 2004]. We focus here on the variance decomposition for a single molecule and therefore consider all effects from upstream regulators to be part of the extrinsic noise (if multiple different...
molecule species are to be considered explicitly the approach outlined in [Huh 2011a] (eq. B1, B2) can be applied using linear noise approximation and boundary conditions for the moments. For a large number of models (considering various sources of extrinsic noise, as well as bursting) the variance of the net number of newly synthesized molecules (i.e. the number of newly made molecules that survive up to the time of interest) can be described by queueing theory. A very general bursting model with non-exponential waiting times between bursts and general distributions for burst size is described by a GI\(X\)/M/\(1\) queue for which the transient moments can be calculated from recursive equations [Holman 1982, Liu 1990]. Extrinsic noise can be modeled by using queues with input correlations [Li 1993] allowing to model (almost) any shape and time correlation for extrinsic noise. Here we illustrate the use of queueing theory to describe the variance contribution from biochemical reactions with a simple model of bursty synthesis where the times between bursts are exponentially distributed and burst sizes have a general distribution (chapter 5). For simplicity we take the molecule to be stable, i.e. \(k_d = 0\).

We denote the average burst size \(\langle b \rangle\) and its variance \(\langle \delta^2 b \rangle\). With average time between bursts equal to \(1/k_s\), the variance at cell age \(a\) is given by \(\langle \delta^2 x_a \rangle = k_s a (\langle \delta^2 b \rangle + \langle b \rangle^2)\). With this, eq 7.1 becomes (for a derivation see section 7.4.4):

\[
\langle \delta^2 x_a \rangle = \mathbf{\text{variance due to cell age distribution}} + \mathbf{\text{variance due to synthesis}}
\]

\[
\begin{align*}
\langle \delta^2 x_a \rangle &= \langle x \rangle^2 (1 - 2 \ln(2))^2 + \frac{1}{\ln 2 - 1} k_s T (\langle \delta^2 b \rangle + \langle b \rangle^2) + \\
k_s T (\langle b \rangle (2 + \langle b \rangle) + \langle \delta^2 b \rangle + 4 (\langle b \rangle (-2 + \langle b \rangle) + 4 \langle b \rangle k_s T) + \langle \delta^2 b \rangle) \langle \delta^2 q \rangle) \\
&\quad \frac{3 - 4 \langle \delta^2 q \rangle}{3 - 4 \langle \delta^2 q \rangle}
\end{align*}
\]

\(\langle \delta^2 x_0 \rangle\) variance at cell birth

(7.14)

with \(\langle x \rangle = k_s T \langle b \rangle / \ln(2)\) as the average copy number. This equation indicates that synthesis bursts enhance copy number variance and that variations in the burst size, \(\langle \delta^2 b \rangle\), and in the partition function, \(\langle \delta^2 q \rangle\), have a multiplicative contribution to overall copy number noise.

7.2.6 The full copy number distribution for zero-order synthesis and first-order degradation

If the partitioning probability \(q\) is fixed and equal to 1/2, the full copy number distribution at cell age \(a\) can be derived using probability generating functions (section 7.4.5.3). The number of molecules at a certain cell age, \(a\), equals the sum of the synthesized and survived molecules since the last division up to age \(a\) and those inherited from the mother cell. To obtain the probability distribution of such a sum of independent random variables, one needs to calculate the convolution of the two associated probability distributions. This step simplifies when instead of these distributions their probability generating functions (PGFs) are considered, as the PGF of a sum of two independent random variables equals the product of the
two associated generating functions. Moreover, the binomial partitioning at division as well as first-order degradation can be expressed in simple terms for generating functions: with \( F_x(z) \) as the generating function of some random variable \( x \), the distribution of a binomial partitioning of \( x \) over two daughter cells is given by \( F_x(z^{1/2}(1 - z/2)) \). Equivalently, the PGF for the number of molecules that remains after a period of time \( a \), where each molecule has a probability \( (1 - p(a)) \) to be degraded, equals \( F_x((1-p(a))(1-z)) \). With this, we obtain an equation for the PGF of the number of molecules at age \( a \):

\[
G(x,a)(z) = \sum_{i=0}^{\infty} F_{(X,T)}(1 - \left(\frac{1}{2} p(a) \right) \left(\frac{1}{2} p(T) \right)^i (1-z)) \]

where \( F_{(X,a)}(z) \) denotes the generating function of the number of newly synthesized molecules at age \( a \). The probability that a molecule that was present in the mother cell is inherited and not yet degraded at age \( a \) is accounted for by \( p(a)/2 \), and \((p(T)/2)^i \) reflects the probability that a molecule that was synthesized \( i \) generations ago is not yet degraded and ended up in a particular daughter cell. The population wide copy number distribution can be obtained by marginalizing out the age distribution:

\[
G_x(z) = \int_0^T u(a) G_{(x,a)}(z) da
\]

As an illustration we consider a model with zero-order synthesis and first-order degradation: the classical "Poisson model". In order to solve the generating function given in equation 7.15 we only need to determine the generating function \( F_{(X,a)}(z) \) for the number of produced molecules that have not yet been degraded. \( F_{(X,a)}(z) \) is the generating function of the time-dependent probability distribution of the following stochastic process,

\[
\varnothing \xrightarrow{k_s} X \xrightarrow{k_d} \varnothing
\]

Starting from the initial condition with 0 molecules

The associated master equation can be solved analytically [Hemberg 2007] and gives the following distribution for the number of molecules,

\[
X_a \sim \text{Poisson}(\kappa(a)) \quad \text{with:} \quad \kappa(a) = \frac{k_s}{k_d} \left(1 - e^{-k_d a} \right)
\]

\( F_{(X,a)}(z) \) then equals \( F_{(X,a)}(z) = e^{\kappa(a)(z-1)} \). Substituting this into equation 7.15 yields:

\[
G_{(x,a)}(z) = e^{\kappa(a)(z-1)} \prod_{i=0}^{\infty} e^{\kappa(T)((1-(1-z)p(a)\frac{1}{2}p(T)^i))}\}
\]

\[
= e^{(z-1)(\kappa(a) + \frac{p(a)}{1-p(T)}\frac{1}{2}))}
\]

(7.19)
This indicates that the copy number distribution at cell age $a$ is given by a Poisson distribution with average $\kappa(a) + \frac{p(a) k_s (a + T)}{1 - p(a)}$, where the first term equals $\langle X_a \rangle$ and the second $p(a)$ (compare to eq 7.6).

For a stable molecule, i.e. $k_d = 0$, marginalizing out the age distribution and taking appropriate limits yields:

$$p(x) = \int_0^T u(a) p(x | a) da$$

$$= \int_0^T u(a) e^{-k_s (a + T)} \frac{(k_s (a + T))^x}{x!} da$$

$$= \frac{4(k_s T)^x}{x!} \Gamma[1 + x, k_s T + \ln(2)]$$

$$- \Gamma[1 + x, 2k_s T + \ln(4)] \ln(2) (k_s T + \ln(2))^{-1-x}$$

where $\Gamma[\cdot]$ denotes the incomplete gamma function.

Under these conditions the variance decomposition as given in equation 7.1 takes a simple form and gives the variance decomposition of the classical Poisson model:

$$\langle \delta^2 x \rangle = \langle (\delta^2 X_0 | x_T) \rangle + \langle (\delta^2 (x_0 | x_T) \rangle + \langle (\delta^2 X_a) \rangle$$

$$= \frac{1}{2 \ln(2) \langle x \rangle} = \frac{1}{2 \ln(2) \langle x \rangle} = \frac{1}{2 \ln(2) \langle x \rangle} = \frac{1}{2 \ln(2) \langle x \rangle}$$

$$+ \langle \delta^2 (x_0) \rangle$$

$$= \frac{1}{1 - 2 \ln(2)^2 \langle x \rangle^2}$$

(7.21)

showing that under these conditions the variance contributions from partitioning, from variance in the mother cell, and from intrinsic noise in the synthesis process are of comparable magnitude. The relative magnitude of the last term (variance deriving from the cell age distribution) in comparison to the other three depends on the average copy number. Typically we associate with the Poisson model the result that $\langle \delta^2 x \rangle = 1/\langle x \rangle$, here we see that in fact the growth process also makes a contribution to this variance due to the cell-age dependency of $x$.

For the case that active degradation of molecules is on a comparable or faster time scale as the generation time, the Poisson model gives insight into how the different variance contributions change during the cell cycle. Since the copy number distribution at any cell age is given by a Poisson distribution, the variance at age $a$ equals the average copy number at that cell age. With this we obtain:

$$\langle p(a)^2 \rangle \langle (\delta^2 x_0 | x_T) \rangle + \langle p(a)^2 \rangle \langle (\delta^2 (x_0 | x_T) \rangle + \langle p(a) (1 - p(a)) \rangle \langle x_0 \rangle$$

$$= p(a) \langle \delta^2 x_0 \rangle = p(a) \langle x_0 \rangle$$

(7.22)
Combining this with equation 7.7 yields the following expression for the variance at age $a$:

\[
\langle \delta^2 x_a \rangle = \langle \delta^2 X_a \rangle + p(a) \langle \delta^2 x_0 \rangle = \frac{k_s}{k_d} (1 - e^{-k_d a}) + e^{-k_d a} \langle \delta^2 x_0 \rangle = \langle x_a \rangle
\]  

(7.23)

This illustrates that the contribution of the variance at cell birth decays exponentially with cell age but that this decrease is balanced by the increasing variance from newly synthesized molecules.

### 7.2.7 Variance decomposition for concentrations

The results derived in the previous sections can be transferred to variance decompositions for concentrations directly if a fixed volume-age relationship is assumed. When volume, $V$, is expressed as a function of cell age, the concentration variance at age $a$ can be expressed as $\langle \delta^2 c_a \rangle = \langle \delta^2 x_a \rangle / V_a^2$ and the total variance follows as

\[
\langle \delta^2 c \rangle = \langle \langle \delta^2 c_a \rangle \rangle + \langle \delta^2 \langle c_a \rangle \rangle = \left( \frac{\langle \delta^2 x_a \rangle}{V_a^2} \right) + \left( \delta^2 \left( \frac{x_a}{V_a} \right) \right)
\]  

(7.24)

The first term can now be decomposed in an analogous manner as for copy numbers. The second term, denoting variance in concentrations from changes in concentration during the cell cycle, will in most cases be much smaller than the corresponding term for copy numbers: while the copy number inevitably doubles on average between subsequent divisions, the average concentration at age 0 must equal that at age $T$. For the case of a constant synthesis rate throughout the cell cycle and no active degradation, there is a linear increase in the average copy number with cell age. Therefore, if volumes also increase linearly with age $\langle \delta^2 \langle c_a \rangle \rangle$ will become zero, while with exponential volume growth $\langle \delta^2 \langle c_a \rangle \rangle \approx 0.0003 \langle c \rangle^2$. For the simple model with zero order synthesis and no active degradation figure 7.3 compares the resulting distributions of copy numbers and concentrations. The difference in the magnitude of the cell age distribution dependent noise is also reflected in the distributions: while concentration distributions for this model have relatively small noise and can be fit reasonably by either normal, lognormal or gamma distributions, copy number distributions have a larger spread and can be fit with neither of those distributions (especially if the copy number average is higher than $\approx 15$). The shape of the copy number distributions is different from what has been reported in the literature based on FACS experiments with cells expressing a fluorescent protein. To illustrate the effect that extrinsic noise can have on the shapes of those distributions we also show distributions for a stable protein with extrinsic noise in the synthesis rate. For
extrinsic noise with log-normal distribution [Rosenfeld 2005], squared coefficient of variation of 0.1, and a large autocorrelation time, the resulting copy number and concentration distributions are both best fit with gamma distributions in agreement with experiment [Taniguchi 2010, Cohen 2009].

Figure 7.3: **Extrinsic noise can shape the distributions of concentrations and copy numbers.** The distributions of concentrations (black) and copy numbers (gray) are shown for a stable molecule synthesized by a zero order reaction for different values of the synthesis rate $k_s$ (eq. 7.20). A. shows those distributions without extrinsic noise, B. for extrinsic noise in the synthesis rate, $k_s$. Extrinsic noise was modeled with a log-normal distribution of $k$ with squared coefficient of variation of 0.1 and an autocorrelation time that much exceeds one generation time such that the resulting distributions are mixture distributions of the distributions obtained for different values of $k_s$ drawn from the log-normal distribution. Insets show fits of the concentration and copy number distributions with intermediate $k_s$ to normal (green), log-normal (blue), and gamma(red) distributions with the same average and variance.

### 7.3 Discussion

We provided a framework theory to integrate models for intrinsic and extrinsic noise in biochemical reactions with models for noise introduced during cell division and growth. Regarding the synthesis process and its regulators, the model is completely general and results from queueing theory can be used to describe various modes of production (in bursts, or through a sequence of steps) as well as production regulated through upstream factors in an abstract and coarse grained manner where the mechanism of synthesis is characterized by the waiting time distribution between subsequent synthesis events and the burst size distribution (chapter 5), and extrinsic noise through its shape and autocorrelation function. To allow for analytical results the degradation mechanism is assumed to be first order without extrinsic noise. We note that non-exponential life times have been observed experimentally [Deneke 2013], indicating more complex mechanisms. Application of this theory would therefore require testing whether degradation is indeed first-order.

The noise decomposition given in eq. 7.1 remains valid also when the assumption of deterministic interdivision times is relaxed. However, the population copy number noise can then in general no longer be calculated from the moment equations as was done in the third section (for a more detailed discussion see section 7.4.6.1). We used
population level simulations to assess the impact of heterogeneity in interdivision times on the distributions of molecule copy numbers and concentrations. There are different mechanisms that can generate heterogeneity in the interdivision time distributions: (i) if cell divisions are imprecise, i.e. one daughter cell inherits more cell material than the other, that cell will probably divide earlier than its sister cell; (ii) the volumes at which cells divide are not deterministic but follow a distribution which for many microorganisms has a coefficient of variation of about 10-20%; (iii) the growth rate of individual cells (i.e. the rate of mass or volume increase) varies between cells which could also contribute to heterogeneity in interdivision times since cells with a higher growth rate will reach the volume at which division takes place faster. Through simulations we tested the effects of those three mechanisms separately as well as their combined effect (section 7.4.6.2) for cells that divide symmetrically (though potentially with imprecise division). The effect of heterogeneity in interdivision times on the distributions of concentrations and copy numbers is very small if interdivision time heterogeneity is caused by imprecise division or a distribution of the volume at which cells divide, but variability in the rate of volume increase can have moderate influence (fig. 7.4). At first glance, this effect seems to be smaller for molecules that have lifetimes much shorter than one generation time. However, this holds only if degradation is a true first order reaction, while degradation of molecules like mRNA and protein depends on various endo- and exonucleases or the proteasome. Degradation can then be described by quasi first order kinetics if the concentrations of these factors remain constant over time. But since those factors in most cases are stable proteins with autocorrelation times of their noise of one generation or longer one might in fact expect unstable molecules to suffer more from this type of noise.

The absolute contribution of heterogeneity in the growth rate on a molecule’s copy number distribution depends on the noise in growth rate as well as its correlation time. For E.coli and M.extorquens the correlations in rate of volume or length increase have been measured to have an autocorrelation time of about one generation or less [Strovas 2007, Wakamoto 2001, Wang 2010], therefore for those organisms the contribution of this variation to the noise in concentration can be expected to be maximally 0.5 times the noise in the distribution of rates of volume increase (for E.coli about 1% or 2% [Tsuru 2009, Wang 2010]). This by itself explains only a relatively small fraction of the extrinsic noise floor observed in a genome wide study of protein noise in E. coli [Taniguchi 2010]. However, this is a minimum level of noise from which stable proteins can only escape if their production rates were tuned to the past and present rates of volume increase. Therefore, this can also be expected to be a minimum noise floor for those molecules involved in mRNA and protein production, like polymerases, sigma factors, ribosomal proteins and RNA as well as translation initiation and elongation factors. The noise in the rate of protein production in E.coli was determined to be around 16% with a correlation time of roughly one generation and a lognormal distribution of production rates [Rosenfeld 2005]. Indeed, a positively skewed distribution for production rate is what one might expect if multiple proteins need to form a complex (and the rate
depends on the concentration of the complex) or if they are acting in a cascade (like transcription and translation) if all the distribution for the individual subunits have (positively correlated) normal or lognormal distributions themselves. The magnitude of the concentration noise is difficult to estimate since the noise in those factors can be expected to have an effect on their own production rate. Interestingly, a correlation time of the protein concentration of about one generation has been found [Wang 2010, Rosenfeld 2005], which is comparable to the correlation time of the fluctuations in the rate of volume increase.

Given the interdivision time heterogeneities observed for many microorganisms that divide symmetrically (coefficients of variation between 12 and 35% [Harvey 1967, Tsukanov 2011, Koppes 1980, Powell 1955, Powell 1963, Powell 1958, Reshes 2008a, Reshes 2008b]), the contribution of interdivision time heterogeneity to the noise in molecule concentrations is rather low. Thus, the equations given in the main text – derived under the assumption of deterministic interdivision times – can be a valuable tool in the decomposition of variance and analysis of experimental data for these types of microorganisms. This changes when asymmetric divisions are considered (section 7.4.6.3). Asymmetric divisions will in general lead to very variable interdivision times because it takes more time until the smaller daughter cell reaches the minimum size required for division. For those cases, the copy number distribution contains a large variance contribution from the unequal partitioning, but this contribution does not reflect functional noise (and its contribution to the noise in concentrations is much smaller). If partitioning is binomial with probability proportional to the volume ratio of daughter and mother cell, the average concentration of the molecule at cell birth is equal to that of the mother cell. In the simulations we vary the extent of asymmetry of the division (i.e. how much the average daughter size differs from 0.5 times the mother cell size), as well as the precision around this value. As expected, the noise in copy numbers as well as concentrations increases with increasing asymmetry. While precision has a small effect on copy number noise, it has increasing contributions to concentration noise as the asymmetry of the division increases (fig. 7.6) and can easily become the dominant contribution to concentration noise.
7.4 Supplemental Information

7.4.1 Decomposition of the variance in molecule copy numbers

As a first step we use the law of total variance to split the variance into a cell age dependent and independent part:

\[ \langle \delta^2 x \rangle = \langle \delta^2 (x|a) \rangle + \langle (\delta^2 x|a) \rangle = \langle \delta^2 \langle x_a \rangle \rangle + \langle \delta^2 x_a \rangle \]  
(7.25)

The variance at cell age \( a \) can further be split into the variance deriving from the molecules that the cell inherited from its mother and that at age \( a \) are not yet degraded, denoted \( \bar{x}_0 \), and the variance that stems from newly synthesized molecules:

\[ \langle \delta^2 x_a \rangle = \langle \delta^2 \bar{x}_0 \rangle + \langle \delta^2 X_a \rangle \]  
(7.26)

The first term can also be decomposed using the law of total variance: for a given number of molecules at cell birth the variance at age \( a \) equals \( p(a)(1-p(a)) \langle x_0 \rangle \), which is the variance of a binomial distribution. This is because each molecule has an independent survival probability, giving rise to a binomial distribution of remaining molecules for a given number of molecules to start with. Since \( x_0 \) itself is distributed, the average variance at age \( a \) must be added to this:

\[ \langle \delta^2 \bar{x}_0 \rangle = p(a)(1-p(a)) \langle x_0 \rangle + p(a)^2 \langle \delta^2 x_0 \rangle \]  
(7.27)

Combining these equations and taking the appropriate averages yields eq. 7.1 in the main text:

\[
\begin{align*}
\langle \delta^2 x \rangle &= \langle \delta^2 \bar{x}_0 \rangle + \langle \delta^2 X_a \rangle + \langle \delta^2 x_a \rangle + \langle \delta^2 x_0 \rangle + \langle \delta^2 X_a \rangle + \langle \delta^2 x_0 \rangle \\
&+ \langle \delta^2 x_a \rangle + \langle \delta^2 X_a \rangle
\end{align*}
\]  
(7.28)
With \( p(a) = e^{-k_a a} \), the averages \( \langle p(a)^2 \rangle \) and \( \langle p(a)(1 - p(a)) \rangle \) can be determined from the age distribution (eq. 7.44):

\[
\langle p(a)^2 \rangle = \int_a u(a)e^{-2k_a a}da = \frac{\left( 2 - 4\frac{k_a}{2k_d + \mu} \right) \mu}{2k_d + \mu}
\]

\[
\langle p(a)(1 - p(a)) \rangle = \int_a u(a)e^{-k_a a} \left( 1 - e^{-k_a a} \right) da
\]

\[
= \mu \left( \frac{2 - 2\frac{k_d}{k_d + \mu} - 2 + 4\frac{k_a}{2k_d + \mu}}{2k_d + \mu} \right) \quad (7.29)
\]

### 7.4.2 Variance from cell division

To calculate the variance at cell birth when the partition ratio, \( q \), is not deterministic but follows a distribution with mean \( \langle q \rangle \) and variance \( \langle \delta^2 q \rangle \), the law of total variance is applied twice under the assumption that \( x_T \) and \( q \) are independent:

\[
\langle \delta^2 x_0 \rangle = \langle \langle \delta^2 x_0 | q \rangle \rangle + \langle \delta^2 \langle x_0 | q \rangle \rangle \quad (7.30)
\]

\[
\langle \delta^2 x_0 | q \rangle = \langle \langle \delta^2 x_0 | x_T \rangle \rangle + \langle \delta^2 \langle x_0 | x_T \rangle \rangle \quad (7.31)
\]

\[
= \langle q(1 - q)x_T \rangle + \langle \delta^2 (qx_T) \rangle
\]

\[
= q(1 - q)\langle x_T \rangle + q^2 \langle \delta^2 x_T \rangle
\]

Combining equations 7.30 and 7.31 yields:

\[
\langle \delta^2 x_0 \rangle = \langle q(1 - q)\langle x_T \rangle \rangle + \langle q^2 \langle \delta^2 x_T \rangle \rangle + \langle \delta^2 \langle q \langle x_T \rangle \rangle \rangle \quad (7.32)
\]

\[
= \langle q(1 - q)\langle x_T \rangle \rangle + \langle q^2 \langle \delta^2 x_T \rangle \rangle + \langle \delta^2 q \rangle \langle x_T \rangle^2
\]

\[
= \langle \delta^2 q \rangle \langle x_T \rangle^2 + \left( \frac{1}{4} - \langle \delta^2 q \rangle \right) \langle x_T \rangle + \left( \frac{1}{4} + \langle \delta^2 q \rangle \right) \langle \delta^2 x_T \rangle
\]

### 7.4.3 Variance due to partitioning for a stable molecule

Here we calculate the variance for a stable molecule \( (k_{deg} = 0) \) that is synthesized by a zero order reaction (with rate constant \( k \)), where the partitioning ratio at cell division, \( q \), fluctuates. The average and variance at cell birth are given by

\[
\langle x_0 \rangle = \langle q \rangle \langle x_T \rangle = \langle q \rangle (\langle x_0 \rangle + \langle X | T \rangle) = \langle X | T \rangle \quad (7.33)
\]

\[
\langle \delta^2 x_0 \rangle = \langle \delta^2 \rangle \langle x_T \rangle^2 + (1/4 - \langle \delta^2 q \rangle) \langle x_T \rangle
\]

\[
+ \left( 1/4 + \langle \delta^2 q \rangle \right) \langle \delta^2 x_T \rangle
\]

\[
= \langle \delta^2 x_0 \rangle + \langle \delta^2 X | T \rangle \quad (7.34)
\]

With \( \langle x_T \rangle = 2\langle x_0 \rangle = 2k_s T \) and \( \langle \delta^2 X | T \rangle = k_s T \) this yields:

\[
\langle \delta^2 x_0 \rangle = \frac{k_s T(-3 + (4 - 16k_s T)(\langle \delta^2 q \rangle))}{-3 + 4(\langle \delta^2 q \rangle)} \quad (7.35)
\]
Combining this with equations 7.25 and 7.26 yields for the population level variance:

\[
\langle \delta^2 x \rangle = \langle x \rangle^2 + \langle x \rangle^2(1 - 2 \ln(2)) + \langle x \rangle^2 \frac{16 \ln(2)^2 (\delta^2 q)}{3 - 4 (\delta^2 q)} 
\] (7.36)

### 7.4.4 Population level variance decomposition for a simple burst model

Here we apply the results from the previous section to a burst model where the times between bursts are exponentially distributed (with average time between bursts equal to \(1/k_s\)), burst sizes with a general distribution (average burst size \(\langle b \rangle\) and variance \(\langle \delta^2 b \rangle\)). For simplicity we take the molecule to be stable, i.e. \(k_d = 0\). The average and variance of newly made molecules at cell age \(a\) for this model is given by [Liu 1990]:

\[
\begin{align*}
\langle X_a \rangle &= k_s a \langle b \rangle \\
\langle \delta^2 X_a \rangle &= k_s a (\langle \delta^2 b \rangle + (\langle b \rangle)^2)
\end{align*} 
\] (7.37) (7.38)

First we solve for the mean and variance at cell birth:

\[
\begin{align*}
\langle x_0 \rangle &= \langle q \rangle \langle x_T \rangle = \langle q \rangle (\langle x_0 \rangle + k_s T \langle b \rangle) = \frac{\langle q \rangle k_s T \langle b \rangle}{1 - \langle q \rangle} = k_s T \langle b \rangle \\
\langle \delta^2 x_0 \rangle &= \frac{\langle \delta^2 q \rangle \langle x_T \rangle^2}{1 - \langle \delta^2 q \rangle} + \left( \frac{1}{4} - \langle \delta^2 q \rangle \right) \langle x_T \rangle + \left( \frac{1}{4} + \langle \delta^2 q \rangle \right) \langle \delta^2 x_T \rangle = \frac{\langle \delta^2 q \rangle (2 k_s T \langle b \rangle)^2 + \left( \frac{1}{4} - \langle \delta^2 q \rangle \right) (2 k_s T \langle b \rangle) + \left( \frac{1}{4} + \langle \delta^2 q \rangle \right) \langle \delta^2 X_T \rangle}{1 - \frac{1}{4} - \langle \delta^2 q \rangle}
\end{align*} 
\] (7.39) (7.40)

The variance due to the cell age distribution is given by

\[
\langle \delta^2 \langle x_a \rangle \rangle = \int u(a) (\langle x_a \rangle - \langle x \rangle)^2 da \\
= \int u(a) \left( x_0 + k_s a \langle b \rangle - \frac{k_s T \langle b \rangle}{\ln(2)} \right)^2 da \\
= \langle x \rangle^2 (1 - 2 \ln(2)^2)
\] (7.41)

and the average variance due to new synthesis equals:

\[
\langle \langle \delta^2 X_a \rangle \rangle = \int u(a) k_s a (\langle \delta^2 b \rangle + (\langle b \rangle)^2) da \\
= \left( \frac{1}{\ln(2)} - 1 \right) k_s T \left( \langle \delta^2 b \rangle + (\langle b \rangle)^2 \right)
\] (7.42)
Combining these equations with 7.25 yields:

\[
\langle \delta^2 x \rangle = \left( \frac{x}{\ln 2} \right)^2 (1 - 2ln(2))^2 + \frac{1}{\ln 2 - 1} k_s T \left( \langle \delta^2 b \rangle + \langle b \rangle^2 \right) + \frac{k_s T (\langle b \rangle^2 + \langle b \rangle) + \langle \delta^2 b \rangle + 4(\langle b \rangle^2 + \langle b \rangle + 4(\langle b \rangle k_s T) + \langle \delta^2 b \rangle \langle \delta^2 q \rangle)}{3 - 4(\delta^2 q)}
\]

\(\langle \delta^2 x_0 \rangle\) variance at cell birth

7.4.5 Using generating functions to calculate the cell age dependent copy number distribution

7.4.5.1 Associated cell age distribution for a discrete interdivision time distribution

We consider a discrete interdivision time distribution, \(f(t)\), which models division after deterministic time intervals, \(T\), the generation time,

\[f(t) = \delta(T - t)\]  

(7.43)

with \(\delta\) as the Dirac delta distribution such that \(f(t = T) = 1\) and \(f(t \neq T) = 0\). The cell age distribution \(u(a)\) of an exponential growing population of cells with specific growth rate \(\mu\) is related to the interdivision time distribution by (Eq 8 in reference [Painter 1968]),

\[u(a) = 2 \cdot \mu \cdot e^{-\mu a} \int_{a}^{\infty} f(t) dt\]

\[= \frac{1}{T} 2^{1 - \frac{a}{T}} \ln 2 \quad \text{for: } 0 \leq a \leq T\]  

(7.44)

With \(\mu = \frac{\ln 2}{T}\).

Note that we have the following relationships,

\[
\int_{0}^{T} u(a) da = 1
\]

\[
\langle a \rangle = \int_{0}^{T} a \cdot u(a) da = T \left( \frac{1}{\ln 2} - 1 \right) = 0.44T
\]

\[
\langle \delta^2 a \rangle = \int_{0}^{T} a^2 \cdot u(a) da - \langle a \rangle^2 = T^2 \left( \frac{1}{(\ln 2)^2} - 2 \right)
\]

\[
\frac{\langle \delta^2 a \rangle}{\langle a \rangle^2} = \frac{1 - 2(\ln 2)^2}{(\ln 2 - 1)^2} = 0.41
\]  

(7.45)
7.4. Supplemental Information

7.4.5.2 Background on generating functions

The probability generating function  Let \( X \) be a discrete random variable that takes non-negative integer values \( X \in \{0, 1, 2, \ldots\} \). The (point) probability that \( X \) takes value \( i \) is defined as \( p_i = P(X = i) \). The (probability) generating function of \( X \) denoted by \( G(z) \) is defined as,

\[
G_X(z) = \sum_{i=0}^{\infty} p_i z^i = \langle z^X \rangle \tag{7.46}
\]

The moments can be obtained from,

\[
\langle X^i \rangle = \left. \frac{1}{z!} \left( \frac{d}{dz} \right)^i G_X(z) \right|_{z=1} \tag{7.47}
\]

The generating function of a sum of independent random variables, \( X \) and \( Y \) The generating function of a sum of independent random variables, \( X \) and \( Y \), equals

\[
G_{X+Y}(z) = \langle z^{X+Y} \rangle = \langle z^X z^Y \rangle = \left( \langle z^X \rangle \langle z^Y \rangle \right) = G_X(z)G_Y(z) \tag{7.48}
\]

It is said that the probability mass function (pmf) \( P_Y(Y) \) is obtained from the convolution of the pmfs of \( X \) and \( Y \). Convolution thus means the sum of random variables and can be obtained from the product of the generating functions.

A compound distribution and its generating function  Let \( Y \) be a sum of independent and identically distributed random variables \( X_i \) then

\[
Y = X_1 + X_2 + \ldots + X_N \tag{7.49}
\]

and let \( N \) be a positive integer-valued random variable. We define \( G_X(z) \) and \( G_N(z) \) and \( G_Y(z) \) now equals,

\[
G_Y(z) = \langle z^Y \rangle = \langle z^{X_1+X_2+\ldots+X_N} \rangle = \langle z^{X_1}z^{X_2}z^{X_3}\ldots z^{X_N} \rangle = \langle G_X(z)^N \rangle = G_N(G_X(z)) \tag{7.50}
\]

Bernoulli distribution and its generating function  A random variable \( X \) is Bernoulli distributed if,

\[
X = \begin{cases} 
1 & \text{when the event is successful, probability } p, \\
0 & \text{when the event is successful, probability } q = 1 - p
\end{cases} \tag{7.51}
\]
The generating function is then 
\[ G_X(z) = p_0z^0 + p_1z^1 = q + pz = 1 - p + pz. \]

We are going to need this generating function to calculate the probability for the number of molecules obtained at cell division and to calculate the probability for the number of molecules that have not been degraded within a certain time interval.

### 7.4.5.3 The generating function for the number of molecules in a population of cells engaged in balanced growth

Denote by \( G(x,a)(z) \) the probability generating function (pgf) for the number of molecules, \( x \), at cell age \( a \). \( H(x_0,a)(z) \) is the pgf for the distribution of molecules obtained at cell birth, \( x_0 \), that have survived until age \( a \). \( F(x,a)(z) \) is the pgf of the number of molecules newly produced that have not yet been degraded, \( X \), during time \( a \).

At any cell age \( a \) the number of molecules equals the sum of the number of molecules obtained from birth that have not yet been degraded and those have been newly produced and not yet been degraded. Thus, the pgf of the number of molecules at cell age \( a \) can be obtained from the convolution,

\[
G(x,a)(z) = H(x_0,a)(z)F(x,a)(z)
\]

\( H(x_0,a)(z) \) is a compound generating function for the probability of the number of molecules that have not been degraded until age \( a \) of the \( x_0 \) molecules obtained at birth. With first order independent degradation of molecules the survival probability for each molecule equals \( p(t) = e^{-kt} \). With this \( H(x_0,a)(z) \) can be expressed as

\[
H(x_0,a)(z) = G(x_0)(1 - p(a) + p(a)z)
\]

We assume independent (binomial) partitioning at division where one daughter cell receives molecules with probability \( q \) (the other daughter with \( 1 - q \)). We consider that \( q \) follows a probability density function \( g(q) \). The "partitioning probability" \( q \) can be influenced by the volume ratio of the two daughter cells or binding of molecules to intracellular compartments. The copy number probabilities at age 0 in a daughter cell is related to the copy number probability at the time of division, \( T \), in the mother cell,

\[
H(x_0,0) = G(x,0) = \langle G(x,T) \left(1 - q + qz\right) \rangle_{\text{binomial partitioning of the } x \text{ molecules at time } T \text{ in the mother}}
\]

Combining equations 7.52 to 7.54 yields

\[
H(x_0,a)(z) = G(x_0)(1 - p(a) + p(a)z)
\]

\[
= \langle G(x,T)(1 - q + q(1 - p(a) + p(a)z)) \rangle
\]

\[
= \langle G(x,T)(1 - qp(a)(1 - z)) \rangle
\]

\[
G(x,a)(z) = \langle G(x,T)(1 - qp(a)(1 - z)) \rangle F(x,a)(z)
\]
If $q = 1/2$ (i.e. $g(q = 1/2) = 1$) eq. 7.56 can be solved by iteration for $G_{(x,T)}$ (if $q$ follows a distribution, the moments of the copy number distribution can still be obtained from eq. 7.56):

$$G_{(x,T)}(z) = G_{(x,T)}(1 - \left(\frac{1}{2}p(T)(1-z)\right))F_{(x,T)}(z)$$  \hspace{1cm} (7.57)

$$= G_{(x,T)}(1 - \left(\frac{1}{2}p(T)^2(1-z)\right))$$

$$\times F_{(x,T)}(1 - \left(\frac{1}{2}p(T)(1-z)\right))F_{(x,T)}(z)$$

$$= G_{(x,T)}(1 - \left(\frac{1}{2}p(T)^n(1-z)\right))$$

$$\times \prod_{i=0}^{n} F_{(x,T)}(1 - \left(\frac{1}{2}p(T)^i(1-z)\right))$$  \hspace{1cm} (7.58)

In the limit of $n \to \infty$ the first term in equation 7.58 becomes one resulting in:

$$G_{(x,T)}(z) = \prod_{i=0}^{\infty} F_{(x,T)}(1 - \left(\frac{1}{2}p(T)^i(1-z)\right))$$  \hspace{1cm} (7.59)

$$G_{(x,0)}(z) = \prod_{i=0}^{\infty} F_{(x,T)}(1 - \left(\frac{1}{2}p(T)^i(1-z)\right))$$  \hspace{1cm} (7.60)

$$G_{(x,a)}(z) = F_{(x,a)}(z) \prod_{i=0}^{\infty}$$

$$\times F_{(x,T)}(1 - \left(\frac{1}{2}p(a)(\frac{1}{2}p(T))^i(1-z)\right)) \hspace{1cm} (7.61)$$

### 7.4.6 Noise in interdivision times

#### 7.4.6.1 Interdivision time heterogeneity as additional noise source

Throughout the main text we have assumed interdivision times to be deterministic, but the times between subsequent divisions would better also be described by a random variable. For a number of symmetrically dividing microorganisms the distributions of interdivision times have been measured and were found to have coefficients of variation between 12 and 35\% [Harvey 1967, Tsukanov 2011, Koppes 1980, Powell 1955, Powell 1963, Powell 1958, Reshes 2008a, Reshes 2008b, Reshes 2008c, Tsuru 2009, Marathe 2010, Canela-Xandri 2010, Wu 2012c]. With interdivision time heterogeneity the analytical approach taken in the previous sections has only limited applicability. The reason for this is that three different types of population samples now need to be distinguished: extant cells (all cells that exist at a given moment in time), baby cells (all cells that were incepted during the same time interval), and mother cells (all cells that divide within the same interval). Denoting
those different types of samples with subscripts of e, b, and m, respectively we obtain:

\[
\langle \delta^2 x_{a(e)} \rangle = p(a)^2 \langle \delta^2 x_{0(e)} \rangle + p(a)(1 - p(a)) \langle x_{0(e)} \rangle + \langle \delta^2 X_a \rangle
\]

\[
\langle \delta^2 x_{0(b)} \rangle = (\delta^2 q) \langle x_{T(m)} \rangle^2 + \left( \frac{1}{4} - (\delta^2 q) \right) \langle x_{T(m)} \rangle
\]

\[
+ \left( \frac{1}{4} + (\delta^2 q) \right) \langle \delta^2 x_{T(m)} \rangle
\]

This means that the variance decomposition remains the same (as applied to a sample of extant cells) but the variance at cell birth of a sample of extant cells can no longer in a straightforward way be expressed as a function of the copy number variance at the time of division. An alternative approach to calculate the copy number noise for a model with interdivision time heterogeneity is by use of cell population balance models [Wu 2012c, Spetsieris 2012, Spetsieris 2009, Friedlander 2008]. The disadvantage of the population balance equation is that it doesn’t allow decomposition of variance into its different sources and that it can usually only be solved numerically.

### 7.4.6.2 Heterogeneity in the rate of volume increase has the largest effect on concentration noise

There are different mechanisms that can cause heterogeneities in interdivision times:

i) asymmetric or imprecise division, ii) a distribution of volumes at which cells divide, iii) variation in the rate of volume increase between cells. Although living organisms will have contributions to interdivision time heterogeneity from all three sources we first explore their effects separately by simulating production and degradation of a molecule for a population of cells that have only one of the above mentioned sources of heterogeneity in interdivision time. To make those results comparable we determine distributions of division fraction, volume at division, and rate of volume increase such that the resulting interdivision time distribution remains approximately the same. For the first mechanism we choose a distribution of the division fraction that is peaked at \(q = 1/2\) (symmetric but imprecise division) and analyze the effects of asymmetric division in the following section. Simulation of these models (figure 7.4; description of simulation algorithm in section 7.4.6.4) show that the effects on copy number and concentration distributions are rather small for the first two mechanisms while heterogeneity in the rate of volume increase changes the distribution of concentrations appreciably. This effect becomes more pronounced for high average concentrations when intrinsic noise is smaller.

If contributions of these different mechanisms that generate heterogeneity in interdivision times were independent one would expect the squared coefficient of
Figure 7.4: Variations in interdivision times have small effects on variability of copy numbers while fluctuations in rate of volume increase do change distributions of concentrations. A) Schematic of the simulated scenarios for cell divisions: deterministic cell cycle (gray), variability in the division fraction (magenta), variability in the volume at which a cell divides (green), or variability in the rate of volume increase (blue). For the last three scenarios, the distributions of division fraction, volume at division, and rate of volume increase are depicted at the right and were chosen to yield the approximately same interdivision time distribution which is depicted at the bottom panel together with the resulting cell age distribution. For comparison, the cell age distribution for the scenario with deterministic interdivision times is shown in dashed gray. B) to E) show distributions of copy numbers and concentrations for molecule X produced through a zero order and degraded with first order reaction for the four scenarios with the production rate set to give averages of 50 molecules per cell. B) and D) show simulations where $k_d = 0$, C) and E) where $k_d = 20 \ln(2)/\mu$.

The variation of the molecule concentrations to be a sum of noise terms:

$$\frac{\langle \delta^2 c \rangle}{\langle c \rangle^2} = \frac{\langle \delta^2 c \rangle}{\langle c \rangle^2} \text{ deterministic cell cycle} + \sum_{i=1}^{n} \frac{\langle \delta^2 c \rangle}{\langle c \rangle^2} \text{ noise caused by mechanism } i$$

(7.64)

However, simulations with all three mechanisms combined show that the total noise is lower than this sum. This is not too surprising since also the noise in the interdivision time distribution is less than three times that of the distribution used for the simulations with only one source of IDT heterogeneity.

If variations in the rate of volume increase persist over multiple generations the squared coefficient of variation in concentrations can be expected to be the sum of ‘intrinsic’ terms and the noise in the distribution of interdivision times: for the case that heterogeneity in the rates of volume increase is the only source of variation in interdivision times, the distribution of interdivision times can be derived from the distribution of rates of volume increase using the change-of-variable technique: with $g(k)$ as the distribution of rates and $f(\tau)$ the distribution of times and $\tau = 1/k$

$$f(\tau) = g(1/\tau) \frac{1}{\tau^2}$$

(7.65)
For deterministic interdivision times the average protein concentration depends linearly on the interdivision time \( (c = k_s(\tau)) \); according to the law of total variance for long autocorrelation times of the fluctuations in \( k \) (and therefore also in \( \tau \)) the total noise in protein concentration equals the sum of noise in interdivision times and the average noise for a model with deterministic interdivision times. That this is indeed the case is shown in fig. 7.5 for a gaussian distribution of rates of volume increase, correlated over time with an exponentially decaying autocorrelation function (since the gaussian distribution of rates in volume increase was applied to a sample of baby cells, enforcing the correlation leads to a slight increase in specific growth rate with increasing autocorrelation time; however the effect of this on the squared coefficient of variation is small).

**Figure 7.5:** The squared coefficient of variation for copy numbers (A) and concentrations (B) increases with the correlation time of the rate of volume increase. Simulations were run for a stable protein (no degradation), average concentration of 50 molecules per average cell volume, with rates of volume increase sampled from a gaussian distribution with \( CV = 0.15 \) and an exponentially decaying autocorrelation function. The specific growth rate of the population increases with increasing correlation because faster growth rates are overrepresented in a populations of extant cells.

### 7.4.6.3 Asymmetric cell division leads to skewed copy number distributions

Fig. 7.6 shows the results of simulations of populations of asymmetrically dividing cells where production of molecules is again described by zero and degradation by first order kinetics. We explore the effects of asymmetry as well as precision of the division fraction by using a distribution of division fractions described by a mixture of two gaussian distributions centered at \( 0.5 - f \) and \( 0.5 + f \) with standard deviation \( \sigma \) (fig. 7.6(A)). The higher the \( f \)-value the more asymmetric are divisions on average while \( \sigma \) is a measure of the precision. These simulations were performed either with a fixed volume at which cells divide (the distribution of division fractions completely determines the interdivision time distribution) or with volumes at division distributed according to a normal distribution with \( CV = 10\% \) (leading to a more realistic distribution of volumes for the total population). Both types of simulations give almost indistinguishable results. As expected, the squared coefficient of variation is an increasing function of \( f \) for both copy numbers and
concentrations, but while the precision ($\sigma$) adds a constant term to the noise in copy numbers for all values of $f$, for concentrations the magnitude of the noise contributed by $\sigma$ increases strongly with $f$. This effect is also seen on the level of copy number and concentration distributions: while the precision has relatively small effects on the distribution of copy numbers, increased $\sigma$ for high $f$ changes the shape of the distribution of molecule concentrations quite considerably. Both types of distributions become broader and more tailed for increasing asymmetry.

Figure 7.6: Asymmetry in cell division introduces strong skew to distributions of copy numbers and concentrations. A) PDF of the division fraction used for simulations: a mixture of two gaussians with centers at $(0.5 - f)$ and $(0.5 + f)$ and standard deviation $\sigma$, so that $f$ is a measure for the asymmetry of the division and $\sigma$ for its precision. B) Scaling of the squared coefficient of variation with molecule abundance for $f = 0.25$ and $\sigma = 0.01$ (black) or 0.07 (dashed gray); dashed lines are used for concentrations, solid lines for copy numbers. C and D) show the increase of the squared coefficient of variation as a function of $f$ for $\sigma = 0.01$ (black) and 0.07 (gray) for a stable molecule with zero order production and first order degradation ($k_d = 0.01 T$, $(x) = 50$). Insets show distribution of copy numbers and concentrations for $f = 0$ and 0.3 and $\sigma = 0.01$ (black) and 0.07 (dashed gray).

It has been suggested that asymmetry in the division of budding yeast could explain the noise floor observed for many yeast proteins on the level of copy numbers (or total cell fluorescence) [Volson 2006]. With $f = 0.25$ the noise in copy numbers equals about the minimum noise level observed in a genome wide survey [Newman 2006]. In that study it was shown that this noise reduces strongly when considering subsamples of cells gated with narrow radius on forward and side scattering (i.e. selecting cells with roughly the same volume). Fig 7.6(B) shows the scaling of the squared coefficient of variation (for both concentrations and copy numbers) with molecule abundance for the simple model described above. Both curves initially decrease linearly (when noise is dominated by its intrinsic part) and
level off to a constant level (when the extrinsic noise from the asymmetric division becomes dominant), but the endpoints differ: for copy numbers, where noise is insensitive to the precision \( \sigma \) of the division, the endpoint depends only on the level of asymmetry \( f \). In contrast, precision has a strong effect on the noise in concentrations (when divisions are asymmetric) and hence the noise floor reached for high abundances depends on both \( f \) and \( \sigma \). For low values of \( \sigma \) (0.01), the curves for both concentrations and copy numbers follow the same trends as observed in [Newman 2006] where concentration is taken as a proxy for the noise in gated samples of cells (fig. 7.7 shows the results for the coefficient of variation as function of molecule abundance for different subsamples with narrow volume ranges - the general trends are similar, but the coefficient of variation for comparable molecule abundance is in general lower and the curves differ between different selected volumes - something that could easily be tested by FACS measurements).

### 7.4.6.4 Simulation of cell populations in balanced growth

Reactions were simulated using the next reaction method with extension to volume dependent reactions [Gibson 2000]. The time of cell division (calculated from the volume at cell birth, the volume at division, and the cell’s rate of volume increase) was added to the list of reaction times. Upon division, all molecules were distributed binomially over the two daughter cells with a probability equal to the ratio of daughter to mother volume. The complete lineage tree was simulated, either starting with a single cell (cell divisions with a distributed division fraction and with distributed rate of volume increase) or with a collection of cells where the remaining life length was calculated from the theoretically calculated interdivision time distribution according to [Painter 1968] (cell divisions with distribution of volumes at division). The reason for this is that the latter mechanism on its own doesn’t lead to a time invariant cell age distribution when starting with a single cell.
Simulations were run until all distributions (cell age, copy number, volume) became stationary.