17.1 Intrinsic versus extrinsic noise sources

So far we have identified two potential sources of random fluctuations: molecular noise (due to low copy numbers) and promoter noise (due to finite rates of TF binding/unbinding). Both of these are examples of intrinsic noise, which is usually contrasted with extrinsic noise. Examples of the latter are fluctuations in the number of ribosomes in the cytoplasm, metabolic signaling pathways that regulate the gene, and stages of the cell cycle. An operational definition of gene-intrinsic noise is the difference in the expression of two almost identical genes from identical promoters in single cells averaged over a large cell population. The contribution of gene-intrinsic noise can then be investigated experimentally using two-reporter assays.

Figure 64: Measuring intrinsic and extrinsic noise in gene expression. Two almost identical genes, which encode red and green fluorescent proteins, are expressed from identical promoters, and are influenced identically by cell-specific factors, such as gene-regulatory signals. (A) Cells with equal amounts of the two proteins appear yellow (low intrinsic noise). Noise fluctuations of the two proteins in the same cell appear correlated over time. (B) If intrinsic noise is significant then the expression of the two genes becomes uncorrelated in individual cells, giving rise to a cell population in which some cells express more of one fluorescent protein than the other. (C) Plot of fluorescence in two strains of the bacterium *Escherichia coli*. Each point represents the mean fluorescence intensities from one cell. Spread of points perpendicular to the diagonal line on which the two fluorescent intensities are equal corresponds to intrinsic noise, whereas the spread parallel to this line corresponds to extrinsic noise. [Adapted from M. B. Elowitz, A. J. Levine, E. D. Siggia and P. S. Swain. Science 297, 1183–1186 (2002).]
17.2 Unregulated gene network and Poisson noise

Deterministic chemical kinetic equations based on the law of mass action assume that the concentrations of the various reactants vary continuously. For molar concentrations with molecule numbers of order $10^{23}$ this is a reasonable approximation. On the other hand, the number of mRNA and protein molecules within a cell are much smaller, and one has to take into account the discrete nature of chemical reactions (molecular noise).

Suppose that at time $t$, the number of mRNA molecules is $N(t)$. In an infinitesimal time interval $\Delta t$, there are three possible events:

1. $N(t + \Delta t) = N(t) + 1$ due to the gene transcription of one new mRNA molecule, which occurs with probability $\Omega \kappa \Delta t$.

2. $N(t + \Delta t) = N(t) - 1$ due to the degradation of one existing mRNA molecule, which occurs with probability $\gamma N(t) \Delta t$.

3. $N(t + \Delta t) = N(t)$ with probability $1 - \Omega \kappa \Delta t - \gamma N(t) \Delta t$.

The probability of two or more transitions in the small time interval is negligible. Let $P_n(t) = \mathbb{P}[N(t) = n]$ = Probability that $N(t) = n$. It follows that

$$P_n(t + \Delta t) = \Omega \kappa P_{n-1}(t) + \gamma (n+1)P_{n+1}(t) + [1 - (\Omega \kappa - \gamma n)\Delta t]P_n(t).$$

Rearranging this equation, dividing through by $\Delta t$ and taking the limit $\Delta t \rightarrow 0$ yields the equation

$$\frac{dP_n}{dt} = \Omega \kappa P_{n-1}(t) + \gamma (n+1)P_{n+1}(t) - (\Omega \kappa + \gamma n)P_n(t) \tag{17.1}$$

17.3 Mean and variance of $N(t)$

In applications one is usually more interested in moments of the distribution $P_n(t)$, in particular, the mean and variance:

$$\bar{n}(t) = \langle N(t) \rangle = \sum_{n=0}^{\infty} n P_n(t), \quad \sigma_n^2(t) = \langle (N(t) - \bar{n}(t))^2 \rangle = \sum_{n=0}^{\infty} n^2 P_n(t) - \bar{n}^2(t). \tag{17.2}$$

A classical way to calculate moments is to use a moment generating function. For the single variable process, this is defined as the discrete Laplace transform (or $z$-transform)

$$G(z, t) = \sum_{m \geq 0} z^m P_m(t), \tag{17.3}$$
Once $G(z,t)$ is known, the moments can be determined by successive derivatives:

$$G(1,t) = 1 \text{ (normalization of probability)}$$

$$\left. \frac{\partial G(z,t)}{\partial z} \right|_{z=1} = \sum_{n=0}^{\infty} n z^{n-1} P_n(t) = \bar{n}(t)$$

$$\left. \frac{\partial^2 G(z,t)}{\partial z^2} \right|_{z=1} = \sum_{n=0}^{\infty} n(n-1) z^{n-2} P_n(t) = \langle N^2(t) \rangle - \langle N(t) \rangle.$$ 

Substituting equation (17.3) into (17.1) gives

$$\frac{\partial G}{\partial t} + \gamma (z-1) \frac{\partial G}{\partial z} = \Omega \kappa (z-1) G. \quad (17.4)$$

Equation (17.4) is a linear first-order partial differential equation (PDE) with non-constant coefficients, which can be solved using what is called the method of characteristics, see below. However, if one is only interested in steady-state moment, then one can set time derivatives to zero and obtain the ODE

$$\frac{dG}{dz} = \frac{\Omega \kappa}{\gamma} G.$$

This has the solution

$$G(z) = \exp \left( \frac{\Omega \kappa (z-1)}{\gamma} \right), \quad G(0) = 1.$$ 

It immediately follows that

$$\langle N \rangle = \left. \frac{\partial G(z)}{\partial z} \right|_{z=1} = \frac{\Omega \kappa}{\gamma},$$

and

$$\langle N^2 \rangle - \langle N \rangle = \left. \frac{\partial^2 G(z)}{\partial z^2} \right|_{z=1} = \left( \frac{\Omega \kappa}{\gamma} \right)^2.$$ 

Hence, the variance $\sigma^2 = \langle N \rangle$, as expected for a Poisson process.

One commonly used measure of the level of noise in regulatory networks is the so-called Fano factor:

$$\text{Fano factor} = \frac{\sigma^2}{\bar{n}}. \quad (17.5)$$

For the unregulated process, the Fano factor is one. This is a baseline value for quantifying the effects of gene regulation on the level of noise.

**Time-dependent solution.** The basic idea of the method of characteristics is to assume $z = z(t)$, with the function $z(t)$ to be determined, and set $G(t) \equiv G(z(t), t)$. From the chain rule of calculus,

$$\frac{dG}{dt} = \frac{\partial G}{\partial t} + \frac{dz}{dt} \frac{\partial G}{\partial z}.$$ 

This is consistent with equation (17.4) provided that

$$\frac{dz}{dt} = \gamma (z-1), \quad \frac{dG}{dt} = \Omega \kappa (z-1) G.$$
Solving for \( z(t) \), we have

\[
z(t) = 1 + se^{\gamma t}
\]

where \( s \) is some parameter. Then

\[
dG = \Omega \kappa se^{\gamma t} G, \quad G(t) = F(s) \exp \left( \Omega \kappa se^{\gamma t}/\gamma \right)
\]

for some function \( F \) determined by the initial condition. For concreteness, we take \( N(0) = n_0 \) so that \( P_n(0) = \delta_{n,n_0} \).

In order to determine the solution \( G(z,t) \) we eliminate \( s \) in terms of \( z \), which gives

\[
G(z,t) = F\left( [z - 1]e^{-\gamma t} \right) \exp \left( \Omega \kappa (z - 1)/\gamma \right).
\] (17.6)

Since \( G(1,t) = 1 \), we require \( F(0) = 1 \). Moreover, given the initial condition \( P_n(0) = \delta_{n,n_0} \), we have \( G(z,0) = z^n \) and \( F(z) = (1 + z)^{n_0}e^{-\Omega \kappa z/\gamma} \). It follows that

\[
G(z,t) = \left[ 1 + e^{-\gamma t}(z - 1) \right]^{n_0}e^{\Omega \kappa (1 - e^{-\gamma t})(z - 1)/\gamma},
\] (17.7)

Calculating derivatives and setting \( z = 1 \) thus yields

\[
\hat{n}(t) = (n_0 - \Omega \kappa /\gamma)e^{-\gamma t} + \Omega \kappa /\gamma, \quad \sigma_n^2(t) = \hat{n}(t) - n_0e^{-2\gamma t}.
\] (17.8)

Note that in the limit \( t \to \infty \), we recover the steady-state moments.

The analysis simplifies considerably if \( n_0 = 0 \). For then

\[
G(z,t) = e^{\lambda(t)(z-1)/\gamma} = e^{-\lambda(t)}e^{z\lambda(t)} = e^{-\lambda(t)} \sum_{n=0}^{\infty} \frac{\lambda(t)^n}{n!} z^n,
\]

where

\[
\lambda(t) = \frac{\Omega \kappa}{\gamma} (1 - e^{-\gamma t}),
\]

and we have Taylor expanded in powers of \( z \). Comparison with equation (17.3) shows that \( P_n(t) \) is given by a Poisson distribution with time-dependent rate \( \lambda(t) \):

\[
\hat{P}_n(t) = \frac{\lambda(t)^n}{n!} e^{-\lambda(t)}.
\] (17.9)

At large times, the rate approaches a constant \( \lambda(t) \to \lambda_0 = \Omega \kappa /\gamma \), and we have constant mean and variance \( \hat{n} = \lambda_0 = \sigma_n^2 \).

\section{Chemical master equation}

Consider a set of chemical species \( i = 1, \ldots, K \) with concentrations \( x_i \) undergoing \( R \) single-step reactions labeled \( a = 1, \ldots, R \). Recall from lecture 1 that the kinetic law of mass-action for the
concentrations $x_i$ takes the general form

$$\frac{dx_i}{dt} = \sum_{a=1}^{R} S_{ia} f_a(x), \quad i = 1, \ldots, K, \quad f_a(x) = \kappa_a \prod_{j=1}^{K} x_j^{S_{aj}}, \quad (17.10)$$

where $S$ is the $K \times R$ stoichiometric matrix and the functions $f_a$ are the propensities. In the case of the autoregulatory network (15.2) we have $K = 2$ with $x_1$ and $x_2$ the concentrations of mRNA and protein, respectively. There are $R = 4$ reactions. For $a = 1, 2$ (mRNA production and degradation), we have

$$S_{i,1} = \delta_{i,1}, \quad S_{i,2} = -\delta_{i,1}, \quad f_1(x) = \kappa g(x_1), \quad f_2(x) = \gamma x_1.$$ 

Similarly, for $a = 3, 4$ (protein production and degradation), we have

$$S_{i,3} = \delta_{i,2}, \quad S_{i,4} = -\delta_{i,2}, \quad f_3(x) = \kappa_p x_1, \quad f_4(x) = \gamma_p x_2.$$ 

Here $x = (x_1, x_2)^\top$.

As in the case of the simple unregulated gene network, we expect the various chemical species to have low copy numbers. This means that one can no longer model the dynamics in terms of a deterministic system of ODEs. Instead, one needs to keep track of the stochastic changes in the number of molecules of each species. The possible transitions and reaction rates are determined by the stoichiometric matrix and propensities introduced above.

Let $N_i(t)$ denote the stochastic number of molecules of species $i$ at time $t$ and introduce the probability distribution for $n = (n_1, \ldots, n_K)$:

$$P(n, t) = \mathbb{P}[N_1(t) = n_1, \ldots, N_K(t) = n_K].$$

In an infinitesimal time interval $\Delta$, one of the following events can occur

1. $N_i(t + \Delta t) = N_i(t) + S_{ia}$ for all $i = 1, \ldots, K$ and a single reaction $a = 1, \ldots, R$, which occurs with probability $\Omega f_a(N(t)/\Omega) \Delta t$.

2. $N_i(t + \Delta t) = N_i(t)$ for all $i = 1, \ldots, K$, with probability $1 - \Omega \sum_a f_a(N(t)/\Omega) \Delta t$.

The distribution $P(n, t)$ thus evolves according to the following chemical master equation, which is a generalization of the birth-death master equation of lecture 3.

The chemical master equation for the reaction scheme of equation (17.10) is

$$\frac{dP(n, t)}{dt} = \Omega \sum_{a=1}^{R} \left( \prod_{i=1}^{K} E^{-S_{ia}} - 1 \right) f_a(n/\Omega) P(n, t), \quad (17.11)$$

where $\Omega$ denotes the system size, $n = (n_1, \ldots, n_K)^\top$ and $E^{-S_{ia}}$ is a step or ladder operator such that for any function $g(n)$,

$$E^{-S_{ia}} g(n_1, \ldots, n_i, \ldots, n_N) = g(n_1, \ldots, n_i - S_{ia}, \ldots, n_N). \quad (17.12)$$
As an example, we can rewrite the birth-death master equation (9.14) as

$$\frac{dp(n,t)}{dt} = \Omega \left[ (E^{-1} - 1) \hat{\omega}_+(n/\Omega)p(n,t) + (E - 1) \hat{\omega}_-(n/\Omega)p(n,t) \right].$$

One additional point to note is that when the number of molecules is sufficiently small, the characteristic form of a propensity function $f(x)$ has to be modified:

$$\left( \frac{n_j}{\Omega} \right)^{s_j} \to \frac{1}{\Omega^{s_j}} \frac{n_j!}{(n_j - s_j)!}.$$ 

In general, it is not possible to obtain exact solutions of the master equation (17.11) even in the case of a stationary solution. Therefore, one usually resorts to some form of approximation scheme. The most common is the diffusion approximation, which is obtained by carrying out a system-size expansion along analogous lines to the approximation of birth-death processes (see lecture 8).