Boulder 07: Modelling Stochastic Gene Expression

Peter Swain, Physiology, McGill University
swain@cnd.mcgill.ca

The whys and wherefores of stochasticity

A system evolves stochastically if its dynamics is partly generated by a force of random strength or by a force at random times or by both. For stochastic systems, it is not possible to exactly determine the state of the system at later times given its state at the current time. Instead, to describe a stochastic system, we use the probability that the system is in a certain state and can predict exactly how this probability changes with time. Nevertheless, such a calculation is often difficult, and we usually focus on finding the moments of the probability distribution, such as the mean and variance. These two quantities are commonly measured experimentally. The level of stochasticity in a system is often referred to as its 'noise'.

Any chemical reaction is stochastic. Reactants come together by diffusion, their motion driven by rapid and frequent collisions with other molecules. Once together, these same collisions alter the internal energies of the reactants, and so their propensity to react. Both effects cause individual reaction events to occur randomly and drive the overall reaction stochastic. Is stochasticity important in biology? Intuitively, stochasticity is only significant when mean numbers of molecules are low. Then, individual reactions, which at most change the numbers of molecules by one or two, matter. Low numbers are frequent *in vivo*: gene copy number is typically one or two, and transcription factors often number in the tens, at least in bacteria [1, 2].

Unambiguously measuring stochastic gene expression, however, can be challenging [2]. Naively, we could place Green Fluorescent Protein (GFP) on a bacterial chromosome downstream of a promoter that is activated by the system of interest. By measuring the variation in fluorescence across a population of cells, we could measure the noise in the system. *Every* biochemical reaction, however, is potentially stochastic. Fluorescence variation could be because of noise in the process under study or could result from the general background 'hum' of stochasticity: stochastic effects in ribosome synthesis could lead to different numbers of ribosomes and so to differences in gene expression in each cell; stochastic effects in the cell cycle machinery may desynchronize the population; stochastic effects in signaling networks could cause each cell to respond uniquely, and so on.

Variation has then two classes: **intrinsic stochasticity** — stochasticity inherent in the dynamics of the system and that arises from fluctuations in the reactions occurring in the system — and **extrinsic stochasticity** — stochasticity originating from fluctuations in other cellular processes that interact with the system under study [3, 4]. To determine whether variation is intrinsic or extrinsic, it helps to visualize an identical second copy of the system, present in the same cell and exposed to the

same intracellular environment. For example, take a simple system like constitutive gene expression. Imagine another identical copy of the gene in each cell. Variation in the number of free ribosomes will equally affect both system copies: expression from both genes will fall if the number of free ribosomes falls and will rise if the number of free ribosomes rises — an extrinsic variation. Variation in the number of actively translating ribosomes, however, is intrinsic. It can be varied independently for each gene system. The same technique works experimentally [4]: two distinguishable alleles of GFP are placed downstream of identical promoters. The intrinsic noise is given by the variation in the difference in concentration of the two alleles, the total noise is determined from the variation in either one of the alleles, and then a simple relationship between these measurements gives extrinsic noise [3]. Stochasticity in gene expression has thus been quantified in both bacteria [4] and yeast [5].

A stochastic description of chemical reactions

For any system of chemical reactions, the ultimate level of description is the chemical master equation. This equation governs how the probability of the system being in any particular state changes with time. A system state is defined by the number of molecules present for each chemical species, and will change every time a reaction occurs. The master equation contains within it the deterministic approximation (a set of coupled differential equations) that is often used to describe system dynamics. The mean of each chemical species can be shown to obey more and more accurately these deterministic equations as the numbers of molecules of all species increase. The master equation itself is usually only solvable analytically for linear systems, i.e., systems having only first-order chemical reactions.

Nevertheless, several approximations exist, all of which exploit the tendency of fluctuations to decrease as the numbers of molecules increase. The most systematic (and complex) is the **linear noise** approach of van Kampen [6]. If the concentration of each chemical species is fixed, then changing the system size (system volume), Ω, alters the number of molecules of every chemical species. The linear noise approximation is based on a systematic expansion of the master equation in Ω^{-1}. It leads to diffusion-like equations that accurately describe small fluctuations around any stable attractor of the system. For systems that tend only to steady-state, a **Langevin** approach is also often used [7, 8, 9]. Here additive, white noise terms are included in the deterministic equations, with the magnitude of these noise terms being determined by steady-state properties of the chemical reactions. At steady-state, the Langevin and linear noise approaches are equivalent.

Unfortunately, all these methods become intractable, in general, once the number of chemical species in the system reaches more than three (we then need to analytically calculate the inverse of a 4×4 matrix or its eigenvalues). Rather than numerically solve the master equation, the **Gillespie algorithm** [10], a Monte Carlo method, is often used to simulate one sample time course from the master equation. By doing many simulations and averaging, the mean and variance for each chemical species can be calculated as a function of time.

Here we will introduce the master equation, Langevin theory, and the Gillespie algorithm.

The master equation

Once molecules start to react, the intrinsic stochasticity destroys any certainty we had of the numbers and types of molecules present. We must adopt a probabilistic description. For example, the reactions of Fig. 1 can be described by

$$\mathcal{P}(n_A \text{ molecules of } A,\ n_B \text{ molecules of } B, \text{ and } n_C \text{ molecules of } C \text{ at time } t)$$

and how this probability evolves with time. Each reaction rate is interpreted as the probability per unit time of the appropriate individual reaction.

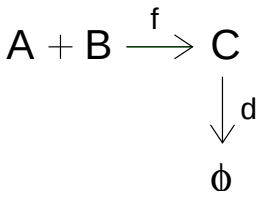

Figure 1: A simple chemical system: A and B bind irreversibly to form complex C with probability f per unit time and individual C molecules degrade with probability d per unit time.

We will write $P_{n_A,n_B,n_C}(t)$ for the probability that the system is in the state of n_A molecules of A, n_B molecules of B, and n_C molecules of C at time t. Consider a time interval δt small enough so that at most only one reaction can occur. If the system at time $t+\delta t$ has n_A, n_B, and n_C molecules of A, B, and C, then if reaction f occurred during the interval δt, the system must have been in the state n_A+1, n_B+1, and n_C-1 at time t. The probability of this reaction is

$$\mathcal{P}(f \text{ reaction}) = f(n_A+1)(n_B+1)\delta t. \tag{1}$$

Alternatively, reaction d could have occurred during δt and so the system then must have been in the state n_A, n_B, and n_C+1 at time t. Its probability is

$$\mathcal{P}(d \text{ reaction}) = d(n_C+1)\delta t. \tag{2}$$

Finally, no reaction may have occurred at all, and so the system would be unchanged at t (in the state n_A, n_B, and n_C):

$$\mathcal{P}(\text{no reaction}) = 1 - f n_A n_B \delta t - d n_C \delta t. \tag{3}$$

Thus we can write

$$\begin{aligned} P_{n_A,n_B,n_C}(t+\delta t) = & \\ & P_{n_A+1,n_B+1,n_C-1}(t)(n_A+1)(n_B+1)f\delta t + P_{n_A,n_B,n_C+1}(t)(n_C+1)d\delta t \\ & + P_{n_A,n_B,n_C}(t)\Big[1 - n_A n_B f\delta t - n_C d\delta t\Big] \end{aligned} \tag{4}$$

Dividing (4) through by δt and taking the limit $\delta t \to 0$ gives

$$\begin{aligned} \frac{\partial}{\partial t}P_{n_A,n_B,n_C} &= f\Big[(n_A+1)(n_B+1)P_{n_A+1,n_B+1,n_C-1} - n_A n_B P_{n_A,n_B,n_C}\Big] \\ & \quad -d\Big[n_C P_{n_A,n_B,n_C} - (n_C+1)P_{n_A,n_B,n_C+1}\Big] \end{aligned} \tag{5}$$

Eq. (5) is called a master equation (presumably because all the moments of the distribution can be derived from it) and describes how the probability of the system being in any state changes with time.

Recovering the deterministic equations

Solving the master equation is possible for linear systems, i.e. those with only first-order chemical reactions, but often only at steady-state. The standard methods of solution are given in [6] and [11]. We will use (5) to derive the equation of motion for the mean of C. The mean of C is defined as

$$\langle C(t)\rangle = \sum_{n_A,n_B,n_C} n_C P_{n_A,n_B,n_C}(t) \tag{6}$$

and is only a function of time.

Multiplying (5) by n_C and summing over n_A, n_B, and n_C gives

$$\begin{aligned}\frac{\partial}{\partial t}\langle C\rangle &= f\sum(n_C-1+1)(n_A+1)(n_B+1)P_{n_A+1,n_B+1,n_C-1}\\ &\quad -f\sum n_A n_B n_C P_{n_A,n_B,n_C} - d\sum n_C^2 P_{n_A,n_B,n_C}\\ &\quad +d\sum(n_C+1-1)(n_C+1)P_{n_A,n_B,n_C+1}\end{aligned} \tag{7}$$

where the terms in round brackets have been factored to follow the subscripts of P. Therefore, by using results such as

$$\begin{aligned}\langle ABC\rangle &= \sum_{n_A,n_B,n_C=0}^{\infty} n_A n_B n_C P_{n_A,n_B,n_C}\\ &= \sum_{n_A,n_B,n_C=0}^{\infty}(n_A+1)(n_B+1)(n_C-1)P_{n_A+1,n_B+1,n_C-1}\end{aligned} \tag{8}$$

as $P_{n_A,n_B,n_C}(t)$ is zero if any of n_A, n_B, or n_C are negative (negative numbers of molecules can not exist), we have

$$\begin{aligned}\frac{\partial}{\partial t}\langle C\rangle &= f\Big[\langle ABC\rangle + \langle AB\rangle\Big] - f\langle ABC\rangle - d\langle C^2\rangle + d\Big[\langle C^2\rangle - \langle C\rangle\Big]\\ &= f\langle AB\rangle - d\langle C\rangle\end{aligned} \tag{9}$$

Applying the law of mass action to Fig. 1, the concentration of C, $[C]$, obeys

$$\frac{d}{dt}[C] = \tilde{f}[A][B] - \tilde{d}[C] \tag{10}$$

where $\tilde{f}$ and $\tilde{d}$ are the macroscopic (deterministic) rate constants. The macroscopic concentration is related to the mean number of molecules by

$$[C] = \frac{\langle C\rangle}{V} \tag{11}$$

and so the deterministic equations are equations for the rate of change of the means of the different chemical species: using (11), (10) becomes

$$\frac{d}{dt}\langle C\rangle = \frac{\tilde{f}}{V}\langle A\rangle\langle B\rangle - \tilde{d}\langle C\rangle. \tag{12}$$

The deterministic equation, (12), should be compared with the microscopic equation for the mean derived from the master equation, (9). A relationship exists between the stochastic probabilities of reaction per unit time to the deterministic reaction rate constants:

$$\begin{aligned} \tilde{f} &= \frac{V\langle AB\rangle}{\langle A\rangle\langle B\rangle}\cdot f \qquad \text{second-order reaction} \\ \tilde{d} &= d \qquad\qquad\qquad\quad \text{first-order reaction} \end{aligned} \tag{13}$$

For first-order reactions both the rate constant and the probability are the same. The macroscopic rate $\tilde{f}$ can only have been accurately measured under conditions where the deterministic approximation holds. Then

$$\langle AB\rangle \simeq \langle A\rangle\langle B\rangle \tag{14}$$

and so

$$\tilde{f} \simeq fV \tag{15}$$

Eq. (15) is almost always used to relate the macroscopic rate and the probability of reaction for second-order reactions.

Eqs. (13) and (15) provide the inter-conversion between reaction rate constants and reaction probabilities.

An exception: homo-dimerization reactions

$$\mathsf{A} + \mathsf{A} \xrightarrow{\mathsf{f}} \mathsf{A_2}$$

Figure 2: The formation of a homo-dimer. Two A monomers combine to form an A dimer.

A homo-dimerization reaction is illustrated in Fig. 2. Two like-molecules bind to each other to form a dimer. This reaction is very common among transcription factors. The master equation is now

$$\frac{\partial P_{n_A}}{\partial t} = f\left[\binom{n_A+2}{2}P_{n_A+2} - \binom{n_A}{2}P_{n_A}\right] \tag{16}$$

where each coefficient is the number of ways of forming a dimer. Eq. (13) becomes

$$2\frac{\tilde{f}}{V}\langle A\rangle^2 = f\langle A(A-1)\rangle. \tag{17}$$

Nevertheless, consistency of the deterministic approximation implies

$$\langle A(A-1)\rangle \simeq \langle A\rangle^2 \tag{18}$$

and so to

$$\tilde{f} \simeq \frac{fV}{2} \tag{19}$$

which is the inter-conversion formula for dimerization reactions.

Converting a macroscopic kinetic rate for *Escherichia coli*

For example, a diffusion-limited reaction is expected to have

$$\tilde{f} = 10^9 \text{ M}^{-1} \text{ s}^{-1}$$

for concentrations measured in molar units [12]. The volume of a typical *E. coli* bacterium is approximately 2×10^{-15} litres [4]. Using N_{Avo} as the Avogado number, (15) implies

$$\begin{aligned} f &= \frac{\tilde{f}}{V} \\ &= \frac{10^9}{N_{\text{Avo}} V} \\ &= \frac{10^9}{6 \times 10^{23} \times 2 \times 10^{-15}} \\ &\simeq 1 \end{aligned} \tag{20}$$

i.e. a rate of 10^9 M^{-1} s^{-1} corresponds to a probability per second of almost one for a single reaction (a useful relationship to remember). Similarly, 1 molecule can be shown to have a concentration of around 1 nM.

The definition of noise

Noise is typically defined as the coefficient of variation: the ratio of the standard deviation of a distribution to its mean. We will denote noise by η:

$$\eta = \frac{\sqrt{\langle N^2 \rangle - \langle N \rangle^2}}{\langle N \rangle} \tag{21}$$

for the random variable N. The noise, η, is dimensionless and measures the magnitude of a typical fluctuation as a fraction of the mean.

Example: Poisson ('birth-and-death') processes

A very simple model of gene expression can be obtained from the reaction scheme of Fig. 1 by letting n_A and n_B become constant. For example, n_A could be the number of molecules of DNA and n_B could be the number of molecules of RNA polymerase. By defining $k = f n_A n_B$, Fig. 1 collapses to the scheme of Fig. 3. Protein C is produced on average every $1/k$ seconds and degrades ('dies') every $1/d$ seconds.

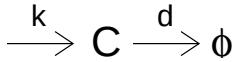

Figure 3: A simple model of gene expression.

The master equation for Fig. 3 is a simplified version of (5):

$$\frac{\partial}{\partial t} P_n = k\Big[P_{n-1} - P_n\Big] - d\Big[nP_n - (n+1)P_{n+1}\Big] \tag{22}$$

with $P_n(t)$ the probability of having n molecules of protein C at time t. Although we will not solve (22) explicitly, the steady-state solution can be found using moment generating functions [6] and is

$$P_n = \mathrm{e}^{-k/d}\frac{(k/d)^n}{n!} \tag{23}$$

which is a Poisson distribution. The first two moments are

$$\begin{aligned} \langle n \rangle &= k/d \\ \langle n^2 \rangle - \langle n \rangle^2 &= k/d \;=\; \langle n \rangle \end{aligned} \tag{24}$$

which implies that noise is

$$\eta = 1/\sqrt{\langle n \rangle}. \tag{25}$$

Eq. (25) demonstrates a general 'rule-of-thumb': noise (stochastic effects) generally become more significant as the number of molecules in the system decrease (Fig. 4).

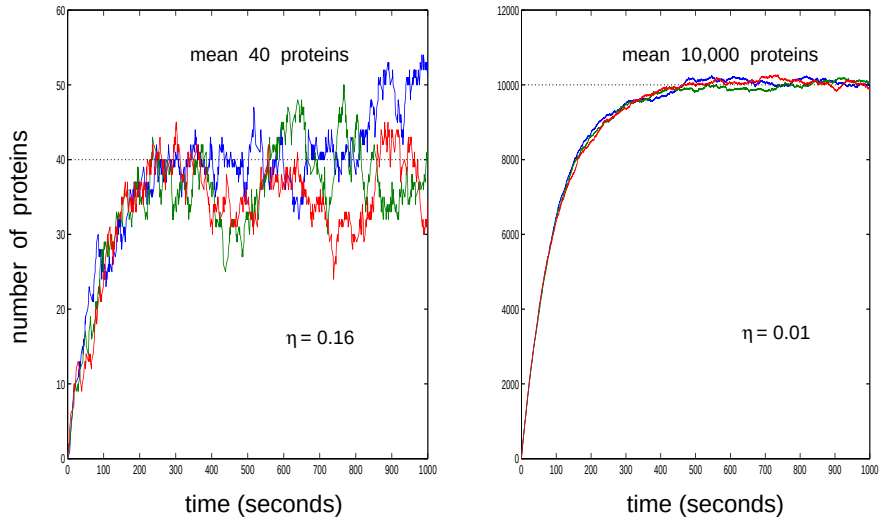

Figure 4: Three simulation runs of two simple (birth-and-death) models of gene expression. Each model involves the reactions of Fig. 3, but has different rate constants leading to different mean protein levels.

Langevin theory: an improved model of gene expression

The scheme of Fig. 3 lumps the processes of transcription and translation into one first-order reaction k. These two processes should be individually modelled. Fig. 5 makes this distinction, but is still simple enough to be exactly soluble [13]. Both mRNA, M, and protein, N, are now present and each has their own half-life (set by the inverse of their degradation rates).

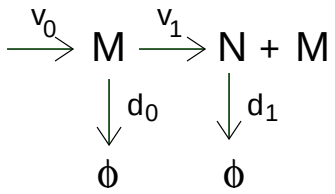

Figure 5: A model of gene expression that explicitly includes transcription (rate v_0) and translation (rate v_1) as first-order processes. mRNA is denoted by M and protein by N.

The Langevin solution

Langevin theory gives an approximation to the solution of the master equation. It is strictly only valid near steady-state and when numbers of molecules are large. Noise terms are explicitly added to the deterministic equations of the system. For the model of Fig. 5, the deterministic equations are

$$\begin{aligned} \frac{dM}{dt} &= v_0 - d_0 M \\ \frac{dN}{dt} &= v_1 M - d_1 N \end{aligned} \tag{26}$$

A Langevin model adds a stochastic variable, $\xi(t)$, to each

$$\begin{aligned} \frac{dM}{dt} &= v_0 - d_0 M + \xi_1(t) \\ \frac{dN}{dt} &= v_1 M - d_1 N + \xi_2(t) \end{aligned} \tag{27}$$

and is only fully specified when the probability distributions for the ξ_i are given. The ξ_i must be specified so that they mimic thermal fluctuations and so successfully model intrinsic noise. The solution of the Langevin equation should then be a good approximation to that of the Master equation (and an exact solution in some limit).

To define ξ, we must give its mean and variance as functions of time and its autocorrelation.

Understanding noise: autocorrelations

The autocorrelation time of a stochastic variable describes the average life-time of a typical fluctuation, as well as the average time separating such fluctuations. We will denote it by τ. Fig. 6 shows typical behaviour of a stochastic variable obeying a Poisson distribution. Time has been rescaled by the autocorrelation time. On average, the number of molecules changes significantly only over a time τ (1 in these units).

The autocorrelation time is found from the autocorrelation function. For a stochastic variable N, the autocorrelation function is

$$\begin{aligned} C_N(t_1,t_2) &= \Big\langle \Big[N(t_1) - \langle N(t_1)\rangle\Big]\Big[N(t_2) - \langle N(t_2)\rangle\Big]\Big\rangle \\ &= \Big\langle \Big\{N(t_1)N(t_2) - \langle N(t_1)\rangle N(t_2) - N(t_1)\langle N(t_2)\rangle + \langle N(t_1)\rangle\langle N(t_2)\rangle\Big\}\Big\rangle \\ &= \langle N(t_1)N(t_2)\rangle - \langle N(t_1)\rangle\langle N(t_2)\rangle. \end{aligned} \tag{28}$$

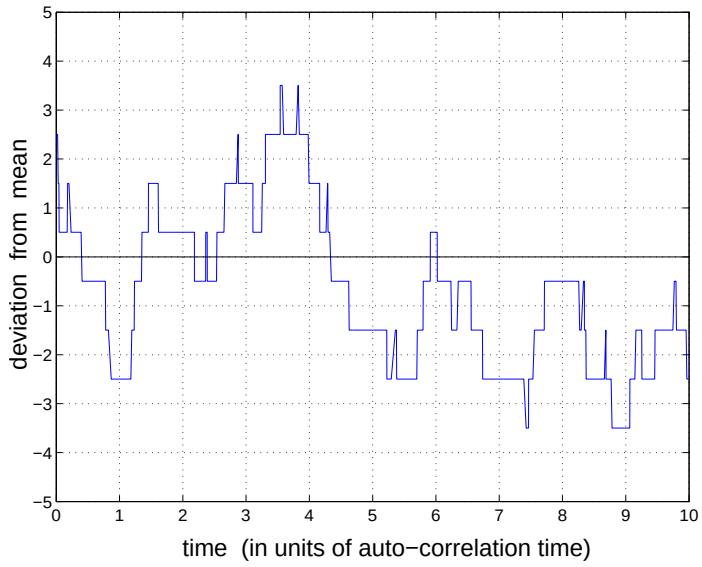

Figure 6: A time-series of the Poisson process of Fig. 3. Time has been rescaled by the autocorrelation time. The deviation from the mean, $n - \langle n \rangle$, in numbers of molecules is plotted on the y-axis.

It quantifies how a deviation of N away from its mean at time t_2 is correlated with the deviation from the mean at a later time t_1. It is determined by the typical life-time of a fluctuation. When $t_1 = t_2$, (28) is just the variance of $N(t)$.

Stationary processes are processes that are invariant under time translations and so are statistically identical at all time points. For a stationary process, such as the steady-state behaviour of a chemical system, the autocorrelation function obeys

$$C_N(t_1, t_2) = C_N(t_1 - t_2). \tag{29}$$

It is a function of one variable: the time difference between the two time points considered. Fig. 7 shows the steady-state autocorrelation function for the Poisson model of gene expression. It is normalized by the variance and is fit well by an exponential decay: $\mathrm{e}^{-t/\tau}$. A typical fluctuation only persists for the timescale τ as enough new reaction events occur during τ to significantly change the dynamics and remove any memory the system may have had of earlier behaviour.

For simple, linear systems, the time-scale associated with degradation sets the steady-state autocorrelation time. Degradation provides the restoring force that keeps the number of proteins fluctuating around their mean steady-state value. The probability of degradation in time δt, $d \times n \times \delta t$, changes as the number of proteins n changes. It increases as the number of proteins rises above the mean value, increasing the probability of degradation and of return to mean levels; it decreases as the number of proteins falls below mean levels, decreasing the probability of degradation and increasing again the probability of returning to mean values. For a linear system with multiple time-scales, the autocorrelation function is a sum of terms, each exponentially decreasing with $t_1 - t_2$ at a time-scale set by the inverse of a degradation rate.

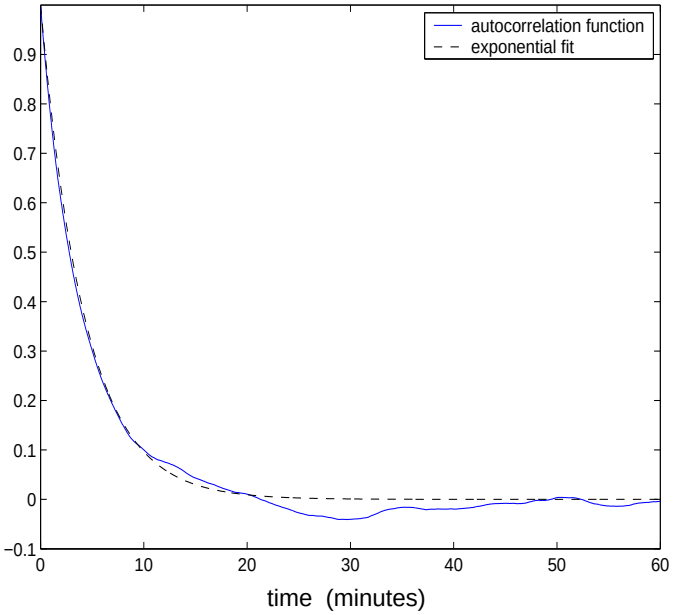

Figure 7: Auto-correlation function for the Poisson process of Fig. 3. The dotted line is an exponential fit using an autocorrelation time of $1/d \simeq 4.2$ minutes.

White noise

In Langevin theory, a stochastic variable, ξ, is added to each deterministic equation. This variable describes thermal fluctuations: those fluctuations that arise from collisions of the molecule of interest with the molecules of the surrounding gas or solvent. Such collisions can act to either increase or decrease the probability of reaction. *A priori*, there is no reason why thermal fluctuations would favour one effect over the other and so $\xi(t)$ is defined to have a mean of zero:

$$\langle \xi(t) \rangle = 0. \tag{30}$$

The time-scale associated with a collision with a solvent molecule is assumed to be much shorter than the time-scale of a typical reaction. The changes in internal energy and position of the molecule of interest because of collisions with solvent molecules are therefore uncorrelated at the reaction time-scale. Mathematically, the autocorrelation time, τ, of the autocorrelation function

$$C_\xi(t_1 - t_2) = \langle \xi(t_1)\xi(t_2) \rangle \tag{31}$$

is taken to zero. If Γ is the variance of ξ at time t, the auto-correlation function is

$$C_\xi(t_1 - t_2) = \Gamma \mathrm{e}^{-(t_1 - t_2)/\tau} \tag{32}$$

which becomes

$$\langle \xi(t_1)\xi(t_2) \rangle = \begin{cases} 0 & \text{for } t_1 \neq t_2 \\ \Gamma & \text{for } t_1 = t_2 \end{cases} \tag{33}$$

in the limit of $\tau \to 0$. Hence

$$\langle \xi(t_1)\xi(t_2) \rangle = \Gamma \delta(t_1 - t_2) \tag{34}$$

where $\delta(t)$ is the Dirac delta function. A stochastic variable that obeys (30) and (34) is referred to as 'white' noise. It is completely uncorrelated and has zero mean. Stochastic variables with zero mean and a finite auto-correlation time are considered 'coloured'. The parameter Γ is the noise strength and needs to be carefully specified (see [6] for a discussion of how Einstein famously chose Γ to appropriately model Brownian motion).

Langevin theory for stochastic gene expression

We now return to modelling the gene expression of Fig. 5. Eq. (27) is shown again below

$$\begin{aligned} \frac{dM}{dt} &= v_0 - d_0 M + \xi_1(t) \\ \frac{dN}{dt} &= v_1 M - d_1 N + \xi_2(t) \end{aligned} \tag{35}$$

and is the deterministic equations of Fig. 5 with additive, white noise terms.

Although we expect ξ_1 and ξ_2 to have zero mean and zero autocorrelation times, we can show that this assumptions are true explicitly by first considering the steady-state solution of (35) in the absence of the stochastic variables ξ_i:

$$M_s = \frac{v_0}{d_0} \quad ; \quad N_s = \frac{v_1}{d_1} M_s \tag{36}$$

If we assume that the system is at or very close to steady-state, and consider a time interval δt small enough such that at most only one reaction can occur, then ξ_1 and ξ_2 can only have the values

$$\xi_i \delta t = \begin{cases} +1 \\ 0 \\ -1 \end{cases} \tag{37}$$

where $i = 1$ or 2, as the number of N or M molecules can only increase or decrease by one or remain unchanged in time δt.

Define

$$P(i,j) = \mathcal{P}(\xi_1 \delta t = i, \xi_2 \delta t = j)$$

i.e. the probability that the number of mRNAs changes by an amount i and that the number of proteins changes by an amount j. Then the reaction scheme of Fig. 5 implies

$$\begin{aligned} P(+1,0) &= v_0 \delta t \\ P(+1,-1) &= 0 \\ P(+1,+1) &= 0 \end{aligned}$$

$$\begin{aligned} P(-1,0) &= d_0 M_s \delta t \\ P(-1,+1) &= 0 \\ P(-1,-1) &= 0 \end{aligned}$$

$$\begin{aligned} P(0,+1) &= v_1 M_s \delta t \\ P(0,0) &= 1 - v_0 \delta t - v_1 M_s \delta t - d_0 M_s \delta t - d_1 N_s \delta t \\ P(0,-1) &= d_1 N_s \delta t \end{aligned} \tag{38}$$

at steady-state.

We can use these probabilities to calculate the moments of the ξ_i. First,

$$\begin{aligned} \langle \xi_1 \delta t \rangle &= (+1) \times v_0 \delta t + (-1) \times d_0 M_s \delta t + (0) \times (1 - v_0 \delta t - d_0 M_s \delta t) \\ &= (v_0 - d_0 M_s) \delta t \\ &= 0 \end{aligned} \tag{39}$$

and

$$\begin{aligned} \langle \xi_2 \delta t \rangle &= (+1) \times v_1 M_s \delta t + (-1) \times d_1 N_s \delta t \\ &= (v_1 M_s - d_1 N_s) \delta t \\ &= 0 \end{aligned} \tag{40}$$

using (36). The means are both zero, as expected, and the ξ_i act to keep the system at steady-state (as they should).

For the mean square, we have

$$\begin{aligned} \langle \xi_1^2 \delta t^2 \rangle &= (+1)^2 \times v_0 \delta t + (-1)^2 \times d_0 M_s \delta t \\ &= (v_0 + d_0 M_s) \delta t \\ &= 2 d_0 M_s \delta t \end{aligned} \tag{41}$$

and, similarly,

$$\begin{aligned} \langle \xi_2^2 \delta t^2 \rangle &= 2 d_1 N_s \delta t \\ \langle \xi_1 \xi_2 \rangle &= 0 \end{aligned} \tag{42}$$

If the system is close to steady-state and the steady-state values of M_s and N_s are large enough such that

$$|M - M_s| \ll M_s \quad ; \quad |N - N_s| \ll N_s \tag{43}$$

hold, then we can assume that (38) is valid for all times. Consequently, ξ_1 at time t_1, say, is completely uncorrelated with ξ_1 at time t_2, where $|t_2 - t_1| > \delta t$ (just as the throws of a die, whose outcomes are also given by fixed probabilities, are uncorrelated). Thus, we define as white noise terms

$$\begin{aligned} \langle \xi_1(t_1) \xi_1(t_2) \rangle &= 2 d_0 M_s \delta(t_1 - t_2) \\ \langle \xi_2(t_1) \xi_2(t_2) \rangle &= 2 d_1 N_s \delta(t_1 - t_2) \\ \langle \xi_1(t_1) \xi_2(t_2) \rangle &= 0 \end{aligned} \tag{44}$$

with the noise strengths coming from (41) and (42).

This definition of ξ_1 and ξ_2 implies that the steady-state solution of (35) will have the true mean and variance of N and M obtained from the master equation, providing (43) is obeyed.

A further simplification

Although it is possible to directly solve the two coupled differential equations of (35), we can also take advantage of the different time-scales associated with mRNA and protein. Typically, mRNA life-time is of order minutes while protein life-time is of order hours. Fig. 8 shows a simulated time series of protein and mRNA. The much longer autocorrelation time of protein $(1/d_1)$ compared to mRNA $(1/d_0)$ is clearly visible.

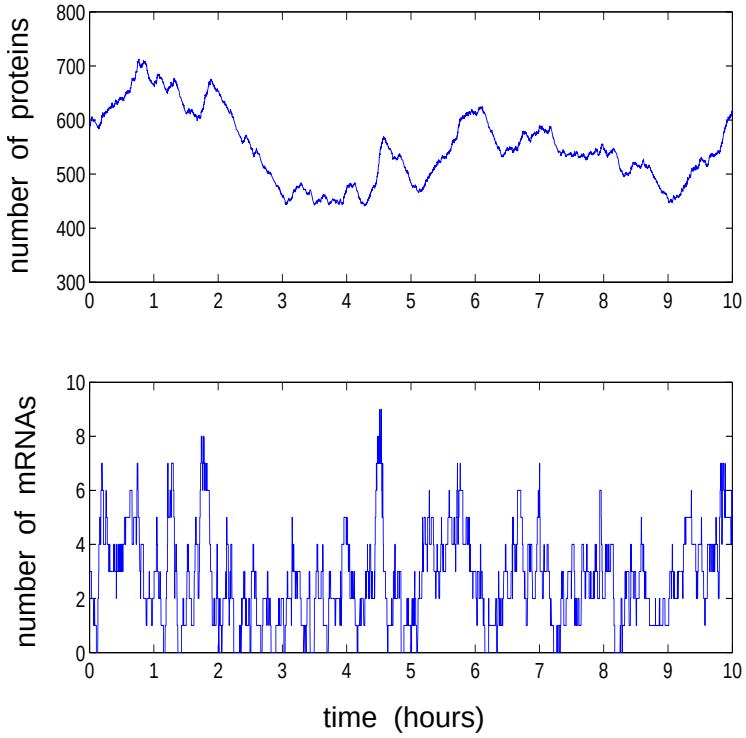

Figure 8: Protein and mRNA numbers from a simulation of the scheme of Fig. 5. Protein half-life is approximately 1 hour while that of mRNA is only 3 minutes.

Many mRNA fluctuations occur during one protein fluctuation, and so the mean level of mRNA quickly reaches steady-state as protein fluctuates. Therefore, we can set

$$\frac{dM}{dt} \simeq 0 \tag{45}$$

which implies that

$$\begin{aligned} M &= \frac{v_0}{d_0} + \frac{\xi_1}{d_0} \\ &= M_s + \frac{\xi_1}{d_0} \end{aligned} \tag{46}$$

Consequently, the equation for protein, (35), becomes

$$\frac{dN}{dt} = v_1 M_s - d_1 N + \frac{v_1}{d_0}\xi_1 + \xi_2 \tag{47}$$

and so is a function of the two stochastic variables ξ_1 and ξ_2. To simplify (47), we define a new stochastic variable

$$\Psi = \frac{v_1}{d_0}\xi_1 + \xi_2 \tag{48}$$

which has mean

$$\langle\Psi\rangle = \frac{v_1}{d_0}\langle\xi_1\rangle + \langle\xi_2\rangle = 0 \tag{49}$$

from (39) and (40), and mean square

$$\begin{aligned}\langle\Psi(t_1)\Psi(t_2)\rangle &= \left(\frac{v_1}{d_0}\right)^2\langle\xi_1(t_1)\xi_1(t_2)\rangle + 2\left(\frac{v_1}{d_0}\right)\langle\xi_1(t_1)\xi_2(t_2)\rangle \\ &\quad +\langle\xi_2(t_1)\xi_2(t_2)\rangle\end{aligned} \tag{50}$$

From Eqs. (44), this result simplifies

$$\begin{aligned}\langle\Psi(t_1)\Psi(t_2)\rangle &= \left(\frac{v_1}{d_0}\right)^2 2d_0M_s\delta(t_1-t_2) + 2d_1N_s\delta(t_1-t_2) \\ &= 2\left[\frac{v_1^2}{d_0}M_s + d_1N_s\right]\delta(t_1-t_2) \\ &= 2d_1\left[\frac{v_1}{d_1}M_s\frac{v_1}{d_0} + N_s\right]\delta(t_1-t_2) \\ &= 2d_1N_s\left[1+\frac{v_1}{d_0}\right]\delta(t_1-t_2)\end{aligned} \tag{51}$$

and so we need only consider one equation:

$$\frac{dN}{dt} = v_1M_s - d_1N + \Psi(t) \tag{52}$$

The effects of the mRNA fluctuations have been absorbed into the protein noise term and its magnitude has increased — compare (51) and (44).

Solving the model

Eq. (52) can be written as

$$\frac{d}{dt}\left(N\mathrm{e}^{d_1t}\right) = v_1M_s\mathrm{e}^{d_1t} + \Psi\mathrm{e}^{d_1t} \tag{53}$$

and so integrated

$$N(t)\mathrm{e}^{d_1t} - N_s = \frac{v_1M_s}{d_1}\left(\mathrm{e}^{d_1t}-1\right) + \int_0^t\Psi(t')\mathrm{e}^{d_1t'}dt' \tag{54}$$

where we have assumed that $N = N_s$ when $t = 0$. Thus

$$N(t) = N_s + \mathrm{e}^{-d_1t}\int_0^t\Psi(t')\mathrm{e}^{d_1t'}dt' \tag{55}$$

Using the properties of $\Psi(t)$, (49) and (51), as well as (55), the mean protein number satisfies

$$\begin{aligned}\langle N(t)\rangle &= N_s + \mathrm{e}^{-d_1 t}\int_0^t \langle\Psi(t')\rangle \mathrm{e}^{d_1 t'}dt' \\ &= N_s \end{aligned} \tag{56}$$

and so the steady-state is stable to fluctuations (as expected).

We can also use (55) to find the autocorrelation function of the protein number:

$$\begin{aligned}&\langle N(t_1)N(t_2)\rangle \\ &\quad= \left\langle \left[N_s + \mathrm{e}^{-d_1 t_1}\int_0^{t_1}\Psi(t')\mathrm{e}^{d_1 t'}dt'\right]\times\left[N_s + \mathrm{e}^{-d_1 t_2}\int_0^{t_2}\Psi(t'')\mathrm{e}^{d_1 t''}dt''\right]\right\rangle \\ &\quad= N_s^2 + \mathrm{e}^{-d_1(t_1+t_2)}\int_0^{t_1}\mathrm{e}^{d_1 t'}dt'\int_0^{t_2}\mathrm{e}^{d_1 t''}dt''\langle\Psi(t')\Psi(t'')\rangle \end{aligned} \tag{57}$$

as $\langle\Psi\rangle = 0$. From (51), we then have

$$\langle N(t_1)N(t_2)\rangle - N_s^2 = 2d_1 N_s\left(1+\frac{v_1}{d_0}\right)\mathrm{e}^{-d_1(t_1+t_2)}\int_0^{t_1}dt'\int_0^{t_2}dt''\mathrm{e}^{d_1(t'+t'')}\delta(t'-t'') \tag{58}$$

To evaluate the double integral, we need to determine when t' is equal to t''. If $t_1 \geq t_2$, then the integral can be decomposed into

$$\begin{aligned}\int_0^{t_1}dt'\int_0^{t_2}dt'' &= \left(\int_{t_2}^{t_1}dt' + \int_0^{t_2}dt'\right)\int_0^{t_2}dt'' \\ &= \int_{t_2}^{t_1}dt'\int_0^{t_2}dt'' + \int_0^{t_2}dt'\int_0^{t_2}dt'' \end{aligned} \tag{59}$$

where we now explicitly see that $t' > t''$ for the first term (and there will be no contribution from the delta function) and t' can equal t'' for the second term (and there will be a contribution from the delta function). Therefore,

$$\begin{aligned}&\int_0^{t_1}dt'\int_0^{t_2}dt''\mathrm{e}^{d_1(t'+t'')}\delta(t'-t'') \\ &\quad= \int_{t_2}^{t_1}dt'\int_0^{t_2}dt''\mathrm{e}^{d_1(t'+t'')}\delta(t'-t'') + \int_0^{t_2}dt'\int_0^{t_2}dt''\mathrm{e}^{d_1(t'+t'')}\delta(t'-t'') \\ &\quad= \int_0^{t_2}\mathrm{e}^{2d_1 t'}dt' \\ &\quad= \frac{1}{2d_1}\left(\mathrm{e}^{2d_1 t_2}-1\right) \end{aligned} \tag{60}$$

as the first integral evaluates to zero.

Consequently, (58) becomes

$$\begin{aligned}\langle N(t_1)N(t_2)\rangle - N_s^2 &= 2d_1 N_s\left(1+\frac{v_1}{d_0}\right)\mathrm{e}^{-d_1(t_1+t_2)}\frac{1}{2d_1}\left(\mathrm{e}^{2d_1 t_2}-1\right) \\ &= N_s\left(1+\frac{v_1}{d_0}\right)\left(\mathrm{e}^{-d_1(t_1-t_2)} - \mathrm{e}^{-d_1(t_1+t_2)}\right) \end{aligned} \tag{61}$$

and we finally have

$$\langle N(t_1)N(t_2)\rangle - \langle N(t_1)\rangle\langle N(t_2)\rangle = N_s\left(1+\frac{v_1}{d_0}\right)\left(\mathrm{e}^{-d_1(t_1-t_2)} - \mathrm{e}^{-d_1(t_1+t_2)}\right) \tag{62}$$

as $\langle N(t)\rangle = N_s$.

Eq. (62) is the autocorrelation function for protein number and becomes

$$C_N = N_s \left(1 + \frac{v_1}{d_0}\right) \mathrm{e}^{-d_1(t_1-t_2)} \tag{63}$$

after long times $t_1 > t_2 \gg 1$. The protein autocorrelation time is $1/d_1$.

Eq. (52) has the same structure as the equation for mRNA

$$\frac{dM}{dt} = v_0 - d_0 M + \xi_1(t) \tag{64}$$

i.e. a constant rate of production and first-order degradation. The solution of (64) will therefore be of the same form as (63), but with d_1 replaced by d_0 and the magnitude of the noise term coming from (44) rather than (51). This substitution gives

$$C_M = M_s \mathrm{e}^{-d_0(t_1-t_2)} \tag{65}$$

so that the autocorrelation time of the mRNA is $1/d_0$.

When $t_1 = t_2$, the autocorrelation becomes the variance. We calculate the noise in mRNA levels as

$$\begin{aligned} \eta_M^2 &= \frac{\langle M(t)^2\rangle - \langle M(t)\rangle^2}{\langle M(t)\rangle^2} \\ &= \frac{M_s}{M_s^2} \\ &= \frac{1}{\langle M\rangle} \end{aligned} \tag{66}$$

Eqs. (65) and (66) are the solutions to any simple birth-and-death model and correspond to the expressions given in (24) and (25).

The protein noise is a little more complicated. It satisfies

$$\begin{aligned} \eta_N^2 &= \frac{1}{N_s} + \frac{v_1}{d_0}\frac{1}{N_s} \\ &= \frac{1}{N_s} + \frac{d_1}{d_0}\frac{1}{M_s} \\ &= \frac{1}{\langle N\rangle} + \frac{d_1}{d_0}\frac{1}{\langle M\rangle} \end{aligned} \tag{67}$$

which should be compared with (25) for the simple model of Fig. 3. The mRNA acts as a fluctuating source of proteins and increases the noise above the Poisson value. Eq. (67) can be described as

$$(\text{protein noise})^2 = (\text{Poisson noise})^2 + \frac{\text{mRNA lifetime}}{\text{protein lifetime}} \times (\text{mRNA noise})^2 \tag{68}$$

The Poisson noise is augmented by a time average of the mRNA noise. As the protein life-time increases compared to the mRNA life-time, each protein averages over more mRNA fluctuations and the overall protein noise decreases. Ultimately, η_N approaches the Poisson result as $d_1/d_0 \to 0$.

Typical numbers for constitutive expression

Some typical numbers for constitutive (unregulated) expression in *E. coli* are

$$\begin{aligned} d_1 = 1/\text{hour} \quad &; \quad d_0 = 1/3 \text{ minutes} \\ \langle N \rangle = 10^3 \quad &; \quad \langle M \rangle = 5 \end{aligned} \tag{69}$$

and so (67) becomes

$$\begin{aligned} \eta_N^2 &= 1/1000 + 3/60 \times 1/5 \\ &= 0.001 + 0.01 \end{aligned} \tag{70}$$

The mRNA term determines the overall magnitude of the noise.

Using (36), Eq. (67) can be re-written as

$$\eta_N^2 = \frac{d_1}{v_1 M_s} + \frac{d_1}{d_0} \cdot \frac{1}{M_s} \tag{71}$$

Only the first term contains the translation rate v_1. Therefore, transcription dominates translation, and determines protein noise, if

$$\frac{d_1}{d_0} \cdot \frac{1}{M_s} \gg \frac{d_1}{v_1 M_s} \tag{72}$$

which simplifies to

$$v_1 \gg d_0. \tag{73}$$

Ribosomes are believed to translate at a rate of around 40 nt s^{-1} [14]. For a 1000 nt protein, v_1 satisfies

$$\frac{1}{v_1} = \frac{1000 \text{ nt}}{40 \text{ nt s}^{-1}} \tag{74}$$

and so $v_1 \simeq 0.04 \text{ s}^{-1}$. Eq. (73) then becomes

$$0.04 \gg \frac{1}{3 \times 60} \simeq 0.006 \tag{75}$$

which certainly holds. Transcription, rather than translation, is often the likely source of gene expression noise [3, 15]. More recently, it has been shown that including transitions in the state of the DNA between forms capable and incapable of initiating transcription better fits experimental data [5, 16]. Physically, this additional process may correspond to re-modelling of the secondary structure of chromosomes.

Simulating stochastic biochemical reactions

The Gillespie algorithm [10] is most commonly used to stochastically simulate biochemical systems. The equivalent of two dice are rolled on the computer: one to choose which reaction will occur next and the other to choose when that reaction will occur. Assume that we have a system in which n different reactions are possible, then the probability that starting from time t a reaction only occurs between $t + \tau$

and $t + \tau + \delta\tau$ must be calculated for each reaction. Let this probability be $P_i(\tau)\delta\tau$ for reaction i, say.

For example, if reaction i corresponds to the second-order reaction of Fig. 1, then

$$\begin{aligned} \mathcal{P}(\text{reaction } i \text{ in time } \delta\tau) &= n_A n_B f \delta\tau \\ &= a_i \delta\tau \end{aligned} \tag{76}$$

where a_i is referred to as the propensity of reaction i. Therefore,

$$\begin{aligned} P_i(\tau)\delta\tau &= \mathcal{P}(\text{no reaction for time } \tau) \\ &\quad \times \mathcal{P}(\text{reaction } i \text{ happens in time } \delta\tau) \\ &\equiv P_0(\tau) a_i \delta\tau \end{aligned} \tag{77}$$

with $P_0(\tau)$ the probability that no reaction occurs during the interval τ. This probability satisfies

$$P_0(\tau + \delta\tau) = P_0(\tau)\Big[1 - \sum_{j=1}^{n} a_j \delta\tau\Big] \tag{78}$$

which implies

$$\frac{dP_0}{d\tau} = -P_0 \sum_{j=1}^{n} a_j \tag{79}$$

and so

$$P_0(\tau) = \exp\left(-\tau \sum a_j\right). \tag{80}$$

Thus we have

$$P_i(\tau) = a_i \mathrm{e}^{-\tau \sum a_j} \tag{81}$$

from (80).

To choose which reaction to simulate, an n-sided die is rolled with each side corresponding to a reaction and weighted by the reaction's propensity. A second die is then used to determine the time when the reaction occurs by sampling from (80). All the chemical species and the time variable are updated to reflect the occurrence of the reaction, and the process is then repeated. See [10] for more details.

Appendix 1: Dirac delta function

The Dirac delta function can be considered the limit of a zero mean normal distribution as its standard deviation tends to zero:

$$\delta(x) = \lim_{n \to \infty} \frac{n}{\sqrt{\pi}} \mathrm{e}^{-n^2 x^2} \tag{A1}$$

This limit gives a function whose integral over all x is one, but that becomes increasingly more and more spiked at zero (Fig. 9). Ultimately

$$\delta(x) = 0 \text{ for all } x \neq 0 \tag{A2}$$

and is not strictly defined at $x = 0$, but does retain the property

$$\int_{-\infty}^{\infty} \delta(x) dx = 1. \tag{A3}$$

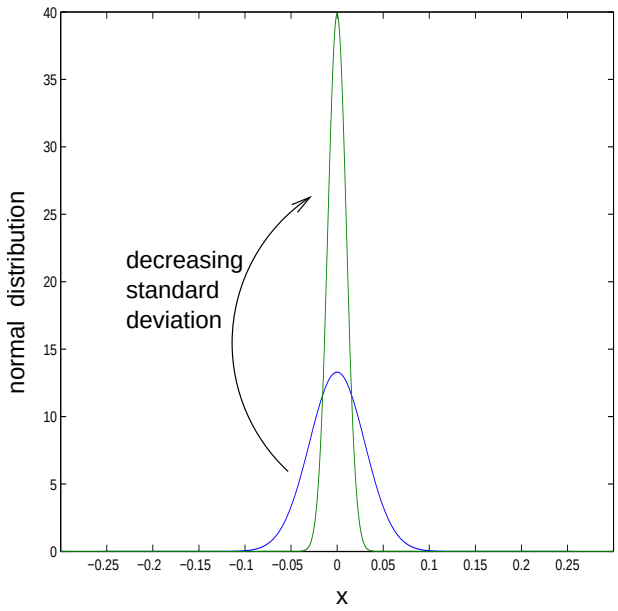

Figure 9: The Dirac delta function is the 'spike' limit of a normal distribution as its standard deviation tends to zero.

These two characteristics imply that the integral of a product of a delta function and another function, $f(x)$, will only give a non-zero result at $x = 0$. The delta function effectively selects the value $f(0)$ from the integral:

$$\int_{-\infty}^{\infty} f(x)\delta(x)dx = f(0) \tag{A4}$$

More generally,

$$\int_{-\infty}^{\infty} f(x)\delta(x-y)dx = f(y). \tag{A5}$$

Appendix 2: Sampling from a probability distribution

Often we wish to sample from a particular probability distribution, $P(x)$, say. The cumulative distribution of $P(x)$ is

$$F(x) = \int_{x_{\min}}^{x} P(x')dx' \tag{A6}$$

and

$$\begin{aligned} \mathcal{P}(x \leq x_0) &= \int_{x_{\min}}^{x_0} P(x')dx' \\ &= F(x_0) \end{aligned} \tag{A7}$$

A sketch of the typical behaviour of $F(x)$ is shown in Fig. 10. If $x \leq x_0$, then $F(x) \leq F(x_0)$ because $F(x)$ is a monotonic function (by definition).

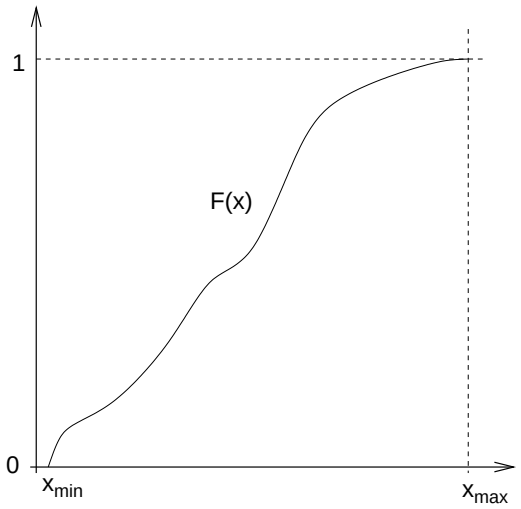

Figure 10: A typical plot of cumulative frequency versus x.

To sample from $P(x)$, first let y be a uniform random number with $0 \leq y \leq 1$ (easily obtained on a computer), then

$$\mathcal{P}(y \leq y_0) = \int_0^{y_0} dy' = y_0 \tag{A8}$$

for some $0 \leq y_0 \leq 1$. Define

$$x = F^{-1}(y) \tag{A9}$$

where $F(x)$ is the cumulative frequency of $P(x)$. Consequently,

$$\begin{aligned} \mathcal{P}(x \leq x_0) &= \mathcal{P}(F^{-1}(y) \leq x_0) \\ &= \mathcal{P}(F.F^{-1}(y) \leq F(x_0)) \end{aligned} \tag{A10}$$

given that $F(x)$ is monotonic. As $F.F^{-1}(y) = y$, we have

$$\begin{aligned} \mathcal{P}(x \leq x_0) &= \mathcal{P}(y \leq F(x_0)) \\ &= F(x_0) \end{aligned} \tag{A11}$$

as y is a sample between 0 and 1 from the uniform distribution: see (A8). Thus the x of (A9) obeys (A7) and so is a sample from $P(x)$.

If we can calculate the inverse function of the cumulative frequency of a distribution $P(x)$, then applying this inverse function to a sample from the uniform distribution gives a sample from $P(x)$.

References

[1] C. V. Rao, D. M. Wolf, and A. P. Arkin. Control, exploitation and tolerance of intracellular noise. *Nature*, 420:231–237, 2002.

[2] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.*, 6:451–464, 2005.

[3] P. S. Swain, M. B. Elowitz, and E. D. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U. S. A.*, 99:12795–12800, 2002.

[4] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297:1183–1186, 2002.

[5] J. M. Raser and E. K. O'Shea. Control of stochasticity in eukaryotic gene expression. *Science*, 304:1811–1814, 2004.

[6] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, New York, New York, 1990.

[7] E. Renshaw. *Modelling biological populations in space and time*. Cambridge University Press, New York, New York, 1991.

[8] J. Hasty, J. Pradines, M. Dolnik, and J. J. Collins. Noise-based switches and amplifiers for gene expression. *Proc. Natl. Acad. Sci. U. S. A.*, 97:2075–2080, 2000.

[9] P. S. Swain. Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *J. Mol. Biol.*, 344:965–976, 2004.

[10] D.T. Gillespie. Exact stochastic simulation of coupled chemical-reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.

[11] C. W. Gardiner. *Handbook of stochastic processes*. Springer, New York, New York, 2004.

[12] D. A. Lauffenburger and J. Lindermann. *Receptors: Models for Binding, Trafficking, and Signaling*. Oxford University Press, New York, New York, 2003.

[13] M. Thattai and A. van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U. S. A.*, 98:8614–8619, 2001.

[14] H. Bremer and P. P. Dennis. Modulation of chemical composition and other parameters of the cell by growth rate. In F. C. Neidhardt, editor, *Escherichia coli and Salmonella: cellular and molecular biology*, pages 1553–1569. A. S. M. Press, Washington, D. C., 1996.

[15] M. Kollmann, L. Lovdok, K. Bartholome, J. Timmer, and V. Sourjik. Design principles of a bacterial signalling network. *Nature*, 438:504–507, 2005.

[16] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox. Real-time kinetics of gene activity in individual bacteria. *Cell*, 123:1025–1036, 2005.