Published in final edited form as:

Sci Signal. 2011 September 27; 4(192): tr13. doi:10.1126/scisignal.2001973.

Simulations of Stochastic Biological Phenomena

Fernand Hayot*

Department of Neurology, Mount Sinai School of Medicine, New York, NY 10029, USA

Abstract

This Teaching Resource provides lecture notes, slides, and a student assignment for a two-part lecture that introduces stochastic modeling of biological systems. The first lecture uses biological examples to present the concept of cell-to-cell variability and makes the connection between the variability of single-cell measurements and concepts from statistical mechanics and probability theory. This section makes the point that for low copy number of a species, the usual differential equation formalism is no longer applicable and needs to be replaced by a probabilistic approach based on the so-called Master Equation. As an example, a simple model of gene transcription is discussed in detail, the different contributions to the relevant Master Equation are highlighted, and the equation itself is derived. The second lecture describes how, for more complex and biologically interesting applications, direct solution of the Master Equation becomes difficult. Gillespie's algorithm, which is used in most cases of biological interest, is then introduced as a practical alternative. The lecture delves into the crux of Gillespie's algorithm, which entails the drawing of two random numbers at each time step. It establishes the corresponding formalism, details the connection between chemical rate constants and Gillespie rates, and culminates in a description and explanation of a core MATLAB program for the transcriptional model considered in the first lecture. This core program, written for a single cell, is expanded by the students in the homework assignment to consider both transcription and translation.

Lecture Notes

We discuss the role of single-cell measurements (1), how cell variability can lead to different phenotypic outcomes (2), and how average behavior can mask what actually happens in single cells, such as a graded response masking an all-or-none individual response (3). For example, when a population of immature *Xenopus* oocytes is treated with progesterone, the average activity of mitogen-activated protein kinase (MAPK) varies smoothly with progesterone concentration. In contrast, at the single-cell level, the response to progesterone is either high or low (Slides 1 to 3) (4). The lecture then takes up the issue of measuring some species in individual, identically prepared cells, with definitions of average (population) behavior and standard deviation, and argues that, as the number of cells becomes large, the population can be considered a statistical ensemble. In this case, the probability density distribution of the measured species is well approximated by the data histogram (Slides 4 to 7).

Copyright 2008 by the American Association for the Advancement of Science; all rights reserved.

*Corresponding author. fernand.hayot@mssm.edu.

Supplementary Materials

<http://stke.sciencemag.org/cgi/content/full/sigtrans;4/192/tr13/DC1>

Slides. Simulations of Stochastic Biological Phenomena.

MATLAB code. A MATLAB file for stochastic modeling of biological phenomena.

Answer key. MATLAB code and PDF of the answers is available upon request.

Some distributions and their characteristics are discussed, as illustrated in a study of cell-to-cell variability in protein levels that determined the relation between the mean and the variance of protein abundance (Slide 8) (5). Thereafter, it is shown that fluctuations for small copy numbers can be important and could preclude the use of ordinary differential equations (ODEs) to model reactions. Instead, a new description has to be used, in which the major object is the probability for the system to be in a given state at a point in time. In this case, because the state of the system is the number of representatives of each species at that time, an equation for these probabilities replaces the usual ODEs for concentrations (Slides 9 and 10).

The equation satisfied by the probabilities is called the Master Equation. In a very simple example of transcription and mRNA decay (a birth-and-death process), the equation is derived in detail with the contributions from each reaction isolated (“birth,” “death,” or no reaction) (Slides 11 to 13). Average values, calculated across many individual cells, satisfy the same equations that one would write for the chemical concentrations. The first part of the lecture concludes with a presentation of the Thattai–Van Oudenaarden model (6), which is used in the problem set (Slide 14). This minimal model describes mRNA transcription, protein translation from mRNA, and degradation of mRNA and protein.

The second part of the lecture explains the main aspects of Gillespie’s algorithm (7–11), a Monte-Carlo–type algorithm that, if sufficient simulation runs are performed, provides an accurate estimate of the solution obtained through directly solving the Master Equation. Because the Master Equation is too complicated to solve for most situations of biological interest, the Gillespie algorithm, which is numerically straightforward to apply, has become the algorithm of choice for the simulation of many stochastic phenomena (Slides 15 to 17). The crux of the algorithm consists of choosing two random numbers at each time step (Slide 18). One random number, drawn from an exponential distribution, determines the delay until the next reaction will take place (Slide 19). The second, drawn from a uniform distribution, determines which reaction will take place. Once the reaction is chosen, the copy numbers of those species participating in the reaction change, and consequently the state of the system is updated. Next, a core MATLAB program implementing Gillespie’s algorithm for gene transcription is presented, and its organization is dissected to clarify the steps involved in implementing the algorithm (Slide 20). The lecture then briefly discusses recent improvements to the original algorithm (Slide 21) (7, 10, 12).

This core implementation of the Gillespie algorithm serves as the basis for the derivation and numerical implementation of the full Thattai–Van Oudenaarden model (6), which also includes translation of mRNA into protein and the latter’s degradation (Slide 14). The assignment requires the students to perform simulations with this model, to generate simulated histograms, and to interpret the results in terms of the underlying biochemical rates. Cell-to-cell variability in this case is due to small copy-number noise, which is an example of so-called intrinsic noise. The lecture concludes by discussing another potential source of intrinsic noise, transcriptional bursting noise, by showing results from studies that have examined this phenomenon (Slides 22 and 23) (13–15). This source of noise results from the fact that transcription does not occur continuously, but instead stochastically, in bursts.

Problem Set

Introductory Details

The homework assignment requires the students to apply the concepts of stochastic simulations that are described in the lectures. The students must derive important analytical results based on equations described in the lectures, and they must expand existing

MATLAB programs so that these models simulate additional biological processes. The Supplementary Material includes a MATLAB program called “transcription.m,” which uses Gillespie’s algorithm to implement a simple model of mRNA transcription and degradation. This program is based on the program “core_gillespie.m” presented in the lecture notes (Slide 20).

Student Assignment: Part 1

1. For the transcription-degradation model (described in Slides 9 to 13), derive from the Master Equation the equations satisfied by the average value of the number of mRNA molecules and the variance of this distribution. At steady state, what is the relation between variance and average value?

Note: Keep in mind the following equality:

$$\Sigma n \cdot P(n) = \Sigma (n-1) \cdot P(n-1) = \Sigma (n+1) \cdot P(n+1).$$

This can help to simplify the derivation.

2. Consider the steady state of the same Master Equation. Notice that $P(n+1)$ is determined once $P(n)$ and $P(n-1)$ are known. Therefore, find the steady-state solution recursively starting with $P(1)$ expressed as a function of $P(0)$ and $P(-1) = 0$. Express $P(n)$ as a function of $P(0)$, determine $P(0)$ from normalization, and show that $P(n)$ is a Poisson distribution.

Student Assignment: Part 2

The MATLAB program “transcription.m” simulates the transcription-degradation model described in Slides 9 to 13. For a more biologically complete description, a simple model of protein translation (6) can be added to this:

where D = DNA; M = mRNA; P = protein, and Φ = degradation products (see also Slide 14).

1. Expand “transcription.m” to also simulate protein translation and degradation. Run this model with the following rate constants: $k_1 = 0.01 \text{ s}^{-1}$, $k_2 = 0.006 \text{ s}^{-1}$, $k_3 = 0.0058 \text{ s}^{-1}$, and $k_4 = 0.000192 \text{ s}^{-1}$.
2. Show the time courses of protein and mRNA up to steady state, both for an average of more than 200 cells and for two particular cells. Compute and display histograms at steady state for both protein and mRNA. Why do the distributions have these shapes?

Student Assignment: Part 3

1. Test 10 different values of the rate of mRNA production, k_1 , from 0.01 to 0.1 s^{-1} . For these simulations, it is sufficient to compute a smaller population of 50 to 100 cells. At the end of each simulation, calculate the following quantities: (i) the mean mRNA abundance across the population of cells, (ii) the variance of mRNA abundance, (iii) the mean protein abundance across the population, and (iv) the variance of protein abundance.

2. If your script keeps track of simulation time, you will notice that the last simulation (highest k_1) takes much longer to run than the first simulation (smallest k_1). Based on your understanding of Gillespie's algorithm, how do you interpret this?
3. How does the mean abundance of mRNA vary with changes in k_1 ? How does the mean abundance of protein vary with changes in k_1 ? Does an increase in k_1 cause, on average, a larger increase in the number of mRNA molecules or the number of protein molecules?
4. Plot the 10 mean mRNA abundances versus the 10 mRNA variances. What type of relationship is observed? Why?

Educational Details

Learning Resource Type: Lecture, assignment, digital presentation

Context: Graduate

Intended Users: Teacher, learner

Intended Educational Use: Learn, plan, teach

Discipline: Biochemistry, cell biology, developmental biology, education, enzymology, molecular biology, theoretical biology

Keywords: Dynamical systems, mathematical modeling, differential equations, MATLAB, simulation

Technical Details

Format: PowerPoint (.ppt)

Size: 700 KB

Requirements: Microsoft PowerPoint

Format: MATLAB (.m)

Requirements: Mathworks MATLAB

Format: PDF (.pdf)

Requirements: Adobe Acrobat Reader

Limits for Use

Cost: Free

Rights: This material may be downloaded for noncommercial, course-teaching purposes only, provided credit to *Science Signaling* is included by listing the citation for the Teaching Resource.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The author thanks K. Zhang for help in developing the problem set and Y.-S. Lee for testing the problem set.

Funding: The development of this course was supported by a Systems Biology Center grant (P50 GM071558) and a training grant in Pharmacological Sciences (T32GM062754). The author acknowledges the support of a contract from the National Institute of Allergy and Infectious Diseases (HHSN272201000054C).

References and Notes

1. Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. *Nature*. 2006; 440:358–362. [PubMed: 16541077]
2. Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. *Nature*. 2010; 467:167–173. [PubMed: 20829787]
3. Rao CV, Wolf DM, Arkin AP. Control, exploitation and tolerance of intracellular noise. *Nature*. 2002; 420:231–237. [PubMed: 12432408]
4. Ferrell JE Jr, Machleder EM. The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes. *Science*. 1998; 280:895–898. [PubMed: 9572732]
5. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, Barkai N. Noise in protein expression scales with natural protein abundance. *Nat. Genet.* 2006; 38:636–643. [PubMed: 16715097]
6. Thattai M, van Oudenaarden A. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.* 2001; 98:8614–8619. [PubMed: 11438714]
7. Gibson MA, Bruck J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A*. 2000; 104:1876–1889.
8. Gillespie DT. Exact stochastic simulation of coupled chemical-reactions. *J. Phys. Chem.* 1977; 81:2340–2361.
9. Gillespie DT. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.* 2001; 115:1716–1733.
10. Gillespie DT, Petzold LR. Improved leap-size selection for accelerated stochastic simulation. *J. Chem. Phys.* 2003; 119:8229–8234.
11. Gillespie DT. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* 2007; 58:35–55. [PubMed: 17037977]
12. Ramsey S, Orrell D, Bolouri H. Dizzy: Stochastic simulation of large-scale genetic regulatory networks. *J. Bioinform. Comput. Biol.* 2005; 3:415–436. [PubMed: 15852513]
13. Hu J, Sealfon SC, Hayot F, Jayaprakash C, Kumar M, Pendleton AC, Ganee A, Fernandez-Sesma A, Moran TM, Wetmur JG. Chromosome-specific and noisy IFNB1 transcription in individual virus-infected human primary dendritic cells. *Nucleic Acids Res.* 2007; 35:5232–5241. [PubMed: 17675303]
14. Iyer-Biswas S, Hayot F, Jayaprakash C. Stochasticity of gene products from transcriptional pulsing. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 2009; 79:031911. [PubMed: 19391975]
15. Hu J, Iyer-Biswas S, Sealfon SC, Wetmur J, Jayaprakash C, Hayot F. Power-laws in interferon-B mRNA distribution in virus-infected dendritic cells. *Biophys. J.* 2009; 97:1984–1989. [PubMed: 19804729]