PHYSICAL MODELS OF LIVING SYSTEMS

LECTURE NOTES

Amir Erez

Contents

List of Figures

0 Why model living systems, and what we will learn

Living systems are complicated, but not every biological question requires a model of every molecule, cell, or organism. A physical model identifies a small set of relevant variables and processes, states their relations mathematically, and asks which conclusions follow from those assumptions. The value of such a model is not only that it predicts. It also makes a proposed mechanism precise enough to test, revise, or reject.

Lopez and Erez (2026) describe modeling as an iterative process of formulation, solution, and validation, and distinguish four overlapping uses: enforcing logical consistency, making quantitative predictions, inferring hidden quantities from data, and building intuition. Although their discussion is framed around microbiology, these roles apply across living systems and provide a useful guide for this course.

This course is organized around the modeling philosophy of Nelson (2022). We will study deterministic, nonlinear, and stochastic models of living systems and repeatedly connect them to quantitative biological observations. The examples range from viral turnover and single-molecule motion to gene regulation, microbial competition, epidemics, ecological communities, and evolution.

0.1 Why learn physical models of living systems?

Biological data are shaped by dynamics and randomness. A steady concentration may hide rapid production and removal; two mechanisms with the same mean may produce completely different fluctuations; feedback may stabilize one system and create switches or oscillations in another. Modeling provides a disciplined way to distinguish these possibilities. A model can contribute in several distinct ways (Lopez and Erez 2026):

- **Consistency:** translating a verbal hypothesis into defined variables and mathematical relations exposes hidden assumptions and makes the consequences of "if–then" reasoning explicit.
- **Prediction:** a quantitative model can forecast observations under new conditions. Genuine prediction must be tested on outcomes that were not used to build or fit the model.
- **Inference:** a model can combine indirect measurements to estimate rates, fluxes, timescales, or other quantities that cannot be observed directly.
- **Intuition:** exploring a model's hypothetical world can reveal the dominant mechanism and produce a compact mental rule that transfers to other systems.

These goals are related but not identical. A highly detailed or data-driven model may predict accurately while offering little explanation, whereas a minimal mechanistic model may sacrifice numerical precision to expose a general principle. Machine learning and mechanistic modeling can therefore complement one another: patterns found in data can suggest mechanisms, while mechanistic constraints can guide and test predictions. The modeling goal should be stated before deciding what kind of model to build.

0.2 Choosing the level of description

There is no universally correct amount of detail. Depending on the question, the same living system may be represented at the level of molecules, cells, populations, or communities. Important choices include whether individuals must be distinguished or can be represented by a density; whether fluctuations matter or a deterministic mean is sufficient; whether molecular detail is essential or can

be coarse-grained; whether space must be represented explicitly; and whether state variables should be discrete or continuous (Lopez and Erez 2026).

A productive default is to begin with the simplest assumptions that could answer the question and relax them when the data, the scale, or the desired prediction demands it. Simplicity is not an end in itself: a model that omits the mechanism responsible for the phenomenon is too simple. Conversely, added parameters and processes are useful only when they change an inference, prediction, or explanation. Studying the same question with several models that preserve the biological mechanism but make different simplifications can reveal which conclusions are robust rather than artifacts of one formulation.

0.3 What we will learn

The course develops four connected strands.

1. **Dynamics and feedback.** We will formulate and analyze population models, turnover, fixed points, stability, negative and positive feedback, genetic switches, cellular oscillators, chemostats, epidemic dynamics, and ecological interactions.

2. **Randomness and inference.** We will use discrete and continuous distributions, entropy and diversity measures, Poisson and birth–death processes, Bayesian and likelihood-based inference, model selection, and stochastic simulation. Probability is introduced when it is needed for a biological problem rather than as a separate formal survey.

3. **Noise at the molecular and cellular scales.** We will connect random walks, first-passage processes, facilitated diffusion, master equations, and Gillespie dynamics to single-molecule experiments and stochastic gene expression, including transcriptional bursting and intrinsic and extrinsic noise.

4. **Evolutionary dynamics.** The Luria–Delbrück experiment will show how fluctuations distinguish mutation from directed adaptation. We will then study genetic drift and selection with the Moran process, fixation, molecular evolution and the molecular clock, and long-term evolution experiments.

By the end of the course, students should be able to turn a biological mechanism into a tractable mathematical model; derive and interpret its central predictions; compare those predictions with data; identify parameters that can or cannot be inferred; and reason about the roles of stochasticity, feedback, and finite populations in living systems.

0.4 How we will work

Most topics begin with an experiment or biological puzzle. We will decide what must be represented, construct a minimal model, analyze it, and return to the data. This follows the formulation–solution–validation cycle emphasized by Lopez and Erez (2026):

1. **Formulation:** define the biological question, variables, processes, assumptions, and the observations the model must explain.

2. **Solution:** derive consequences analytically when possible and explore the remaining behavior with computation.

3. **Validation:** compare those consequences with data, limiting cases, and known invariants, then revise the formulation when the comparison fails.

Mathematical reasoning and interpretation are the central skills. Computer simulation and data analysis will support that reasoning, but learning programming syntax is not a course objective. When code is useful, a large language model may generate an implementation; the student's responsibility is to specify the model unambiguously, test the output against theoretical limits and invariants, and recognize scientifically plausible-looking answers that are nevertheless wrong.

1 HIV dynamics: learning from a perturbation

This first case study puts the course strategy into practice. A nearly steady viral load appears uneventful, yet a targeted perturbation reveals rapid underlying turnover. A minimal dynamical model will let us turn that observation into a quantitative biological conclusion.

By the end of this week, students should be able to answer:

- How can a steady measured population conceal rapid biological turnover?
- How does a perturbation turn a dynamical model into an inference tool?
- Which combinations of HIV production and clearance parameters can viral-load data identify, and what additional information is needed to distinguish them?

1.1 The puzzle of the apparently latent stage

In June 1981 the U.S. Centers for Disease Control reported clusters of Pneumocystis pneumonia and Kaposi's sarcoma in young men, marking the recognition of AIDS. In 1983–1984 the causative retrovirus was isolated and was later given the unified name human immunodeficiency virus (HIV).

By 1994, Alan Perelson and collaborators were studying a basic puzzle. After the acute stage of infection, the concentration of virus in the blood can remain approximately steady for years before rising again as AIDS develops (Figure 1.1). Does this apparently latent stage mean that HIV is replicating only very slowly?

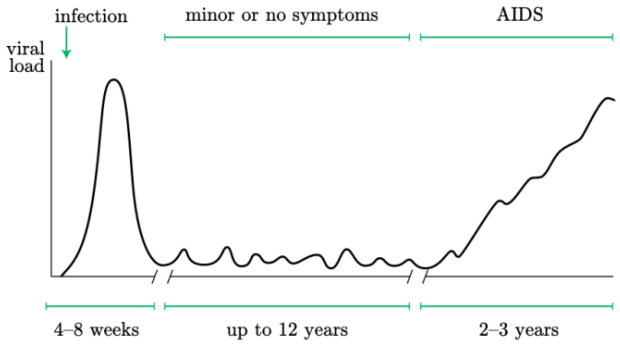

Figure 1.1: Schematic time course of HIV infection. After a brief acute peak, the viral load falls to a low, nearly steady level that can persist for years. Viral load eventually rises again as the symptoms of AIDS appear (Weiss 1993).

1.2 A physical analogy: the leaky container

Imagine pouring water into a leaky container (Figure 1.2). If the water level is constant, the system is in a steady state and

$$Q_{\text{in}} = Q_{\text{out}}. \tag{1.1}$$

The equality of the two rates does not tell us whether both rates are small or both are large. A constant amount of water can therefore conceal rapid turnover.

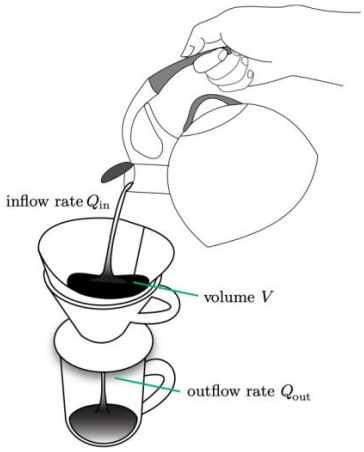

Figure 1.2: The leaky-container analogy. At steady state, the inflow and outflow rates match even though the water is continually being replaced (Nelson 2022).

The outflow grows with the amount of water in the container. For a fluid column of height h, mechanical equilibrium gives, schematically,

$$P_{\text{lower}}A = P_{\text{higher}}A + mg, \qquad P_{\text{lower}} - P_{\text{higher}} = \rho g h. \tag{1.2}$$

Thus a larger volume produces a larger pressure difference and a faster outflow. More generally, a stable steady state requires removal to become stronger as the population increases. A number-independent production rate together with removal proportional to population size is the simplest example. The same turnover logic helps explain how a roughly constant blood-cell population can coexist with the production of hundreds of billions of new cells per day.

This suggests a different hypothesis for HIV: the virus may multiply rapidly during the apparently latent stage while the body clears it at an equally rapid rate. Different patients can have different quasi-steady viral loads, but a nearly constant load alone cannot reveal the hidden flux. A perturbation is needed.

1.3 Ritonavir reveals the hidden turnover

Ritonavir is an antiviral drug that acts primarily as a protease inhibitor. It binds to the active site of an HIV protease and blocks cleavage of viral polyproteins. Viral particles may still be released by cells that were already infected, but those particles do not mature into functional, infectious virions. Protease inhibition therefore decouples the *quantity* of released particles from their *quality*, or infectivity.

In a 1994 clinical study led by David Ho, the viral load fell dramatically after ritonavir treatment (Figure 1.3). The drug later lost its effect as resistant virus became dominant, but the early decrease exposed the otherwise hidden rate of viral turnover.

The concentration fell by roughly a factor of 100 in about ten days, or by a factor of 10 in five to six days. Such rapid clearance means that the pre-treatment input flux must also have been large–of order 10^9 virions per day in a patient. The apparently quiet stage is therefore dynamically active.

Rapid replication also has evolutionary consequences. Reverse transcription is error-prone, so more infection events mean more opportunities for mutations that confer drug resistance or immune escape. If a particular resistance mutation occurs with probability $p \ll 1$, then simultaneous

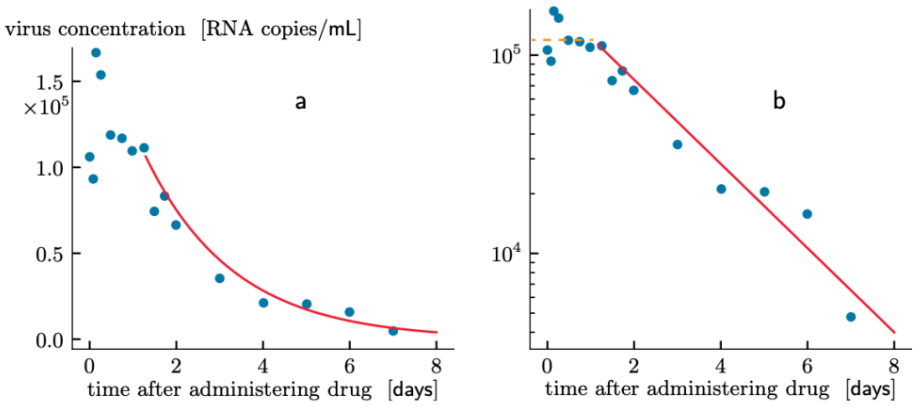

Figure 1.3: Viral load in one patient after treatment with a protease inhibitor. Panel (a) uses a linear vertical scale; panel (b) shows the same data on a semilogarithmic scale. The early plateau is a systematic deviation from a single exponential decay (Perelson et al. 1996).

resistance to two independent drugs is roughly of order p^2, and resistance to three is of order p^3. This is the basic logic behind combination therapy, or a drug cocktail.

1.4 What made the inference possible?

The HIV example combined several ingredients:

- an interdisciplinary biological and physical question;
- a simple physical picture–the leaky container;
- a mathematical description using differential equations;
- a controlled perturbation produced by a new drug; and
- quantitative data with which to support or reject hypotheses.

Fitting adjusts the model parameters so that the predicted trajectory agrees as closely as possible with the data. This is a parametric inverse problem: measured outcomes are used to infer otherwise hidden rates. A good fit does not prove that every component of the model is literally correct, but it can make the model promising for answering a specific question. Conversely, systematic structure in the residuals is evidence that the model is missing something.

The semilog plot in Figure 1.3 already shows two important features: a short plateau before the decline and a later region that is approximately, but not perfectly, exponential. We now ask whether a minimal dynamical model can explain both.

1.5 A minimal post-treatment model

We idealize the treatment as follows:

1. Before treatment, $t < 0$, the patient is in a quasi-steady state: infection of new T cells balances infected-cell clearance, and virion production balances free-virion clearance.
2. At $t = 0$, an ideal antiviral drug completely prevents new infections of target T cells.

3. Each infected T cell has a fixed probability per unit time of being cleared.

Once new infection is blocked, the population of uninfected target cells is decoupled from the two variables measured by the minimal model. Infected cells continue to release virions, infected cells are cleared, and free virions are cleared (Figure 1.4).

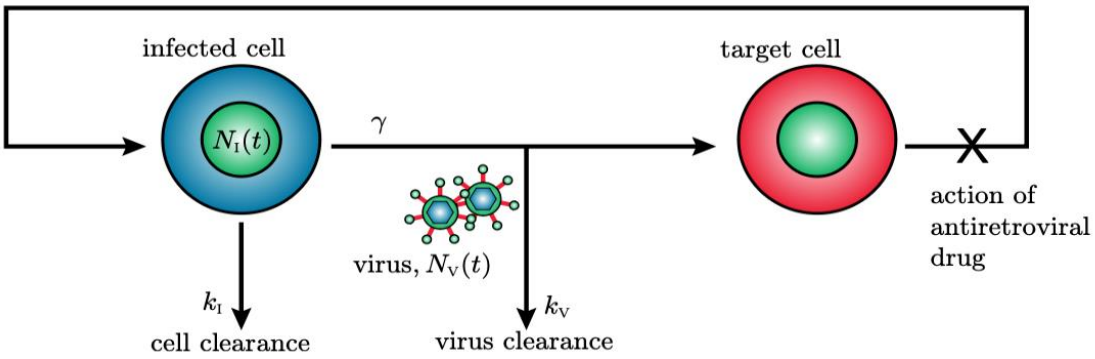

Figure 1.4: Simplified HIV life cycle after ideal antiretroviral treatment. Infected T cells release virions at rate γ; infected cells and free virions are cleared at rates k_I and k_V, respectively. The drug blocks infection of new target cells (Nelson 2022).

The quantities in the model are

t	time since treatment begins,
$N_I(t)$	population of infected T cells, with $N_I(0) = N_{I0}$,
$N_V(t)$	population of free virions, with $N_V(0) = N_{V0}$,
k_I	clearance rate constant for infected T cells,
k_V	clearance rate constant for free virions,
γ	rate of virion production per infected T cell, and
β	the abbreviation $\beta = \gamma N_{I0}$.

The symbol N_V denotes a number of virions in the modeled compartment, whereas a viral-load assay usually reports a concentration. For example, a measurement of 10^5 RNA copies per milliliter corresponds to $N_V = 10^5$ copies if the modeled compartment is 1 mL, or $N_V = 10^9$ copies if an effective compartment of 10 L is used. Because the equations are linear in N_I and N_V, one may instead formulate the same model entirely in concentrations, provided every source term is expressed per unit volume and the convention is used consistently.

The pre-treatment viral steady state supplies an important initial condition:

$$\gamma N_{I0} = k_V N_{V0}, \qquad \text{or equivalently} \qquad \beta = k_V N_{V0}. \tag{1.3}$$

Consequently, Equation (1.5) has zero initial slope immediately after infection is blocked: production by cells infected before treatment initially balances virion clearance. As those infected cells disappear, production falls and the viral load begins to decline.

For a short interval Δt, the probability that one infected cell is cleared is $k_I \Delta t + o(\Delta t)$. The expected infected-cell population therefore obeys

$$\frac{\mathrm{d}N_I}{\mathrm{d}t} = -k_I N_I, \qquad N_I(t) = N_{I0} e^{-k_I t}. \tag{1.4}$$

Virions are produced by infected cells and cleared from the blood:

$$\frac{\mathrm{d}N_V}{\mathrm{d}t} = -k_V N_V + \gamma N_I = -k_V N_V + \beta e^{-k_I t}. \tag{1.5}$$

To solve the forced linear equation, multiply by the integrating factor $e^{k_V t}$:

$$e^{k_V t}\left(\frac{\mathrm{d}N_V}{\mathrm{d}t} + k_V N_V\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(e^{k_V t} N_V\right) = \beta e^{(k_V - k_I)t}. \tag{1.6}$$

Integrating from 0 to t and using $N_V(0) = N_{V0}$ gives

$$e^{k_V t} N_V(t) - N_{V0} = \beta \int_0^t e^{(k_V - k_I)s}\,\mathrm{d}s = \frac{\beta}{k_V - k_I}\left[e^{(k_V - k_I)t} - 1\right]. \tag{1.7}$$

Multiplying by $e^{-k_V t}$ and collecting the two exponential terms yields

$$N_V(t) = X e^{-k_I t} + (N_{V0} - X)\, e^{-k_V t}, \qquad X = \frac{\beta}{k_V - k_I}, \tag{1.8}$$

for $k_V \neq k_I$. The observed viral load is a sum of two decaying exponentials, but one coefficient is negative when the quasi-steady condition above is imposed. The two exponentials then cancel in the initial derivative, producing the observed short plateau even though both modes eventually decay.

The apparently singular case $k_V = k_I \equiv k$ is obtained directly from the integral in Equation (1.7), or by taking the limit $k_V \to k_I$ in Equation (1.8):

$$N_V(t) = \big(N_{V0} + \beta t\big) e^{-kt}. \tag{1.9}$$

With the quasi-steady initial condition $\beta = kN_{V0}$, differentiation again gives $dN_V/dt = 0$ at $t = 0$. Thus the initial plateau is not lost when the two clearance rates happen to coincide.

The two limiting cases clarify the meaning of the rates:

- If $k_I \gg k_V$, virion production shuts off before many free virions have been cleared. After the short transient, the observed decline is controlled by k_V.
- If $k_V \gg k_I$, free virions equilibrate rapidly with their slowly changing input. Equation (1.5) is approximately in instantaneous balance, so $N_V \simeq (\beta/k_V)e^{-k_I t}$, and the observed decline is controlled by k_I.

1.6 Parameter inference and falsification

The initial values N_{I0} and N_{V0} act as integration constants. In practice, N_{V0} is measured while the infected-cell population can remain hidden. If the pre-treatment quasi-steady condition in Equation (1.3) is imposed exactly, then $\beta = k_V N_{V0}$; after measuring N_{V0}, the viral-load curve has only two free dynamical parameters, k_I and k_V. If the initial state is not assumed to be in exact balance, β remains a third fitted parameter. More independent data features than free parameters make the fit over-constrained. Fewer independent constraints than parameters make the model underidentified; overfitting is the distinct problem of using excess flexibility to fit noise rather than the underlying signal.

Even an over-constrained fit need not identify every biological label. After imposing $\beta = k_V N_{V0}$, Equation (1.8) becomes

$$\frac{N_V(t)}{N_{V0}} = \frac{k_V e^{-k_I t} - k_I e^{-k_V t}}{k_V - k_I}. \tag{1.10}$$

This expression is unchanged when k_I and k_V are exchanged. Viral load alone can therefore identify the unordered pair of decay rates, but cannot decide which rate belongs to infected-cell loss and which belongs to free-virion clearance. An independent cell measurement, a biological ordering assumption, or a richer model is needed to assign the two rates. This is a concrete example of *structural identifiability*: abundant, noise-free data cannot resolve a symmetry built into the observation model.

Figure 1.5 shows why matching one or two points is not enough. A physically motivated solution with a poor parameter choice rapidly departs from the data. A flexible but unphysical curve can be forced to match the beginning and end yet still miss the intermediate trajectory.

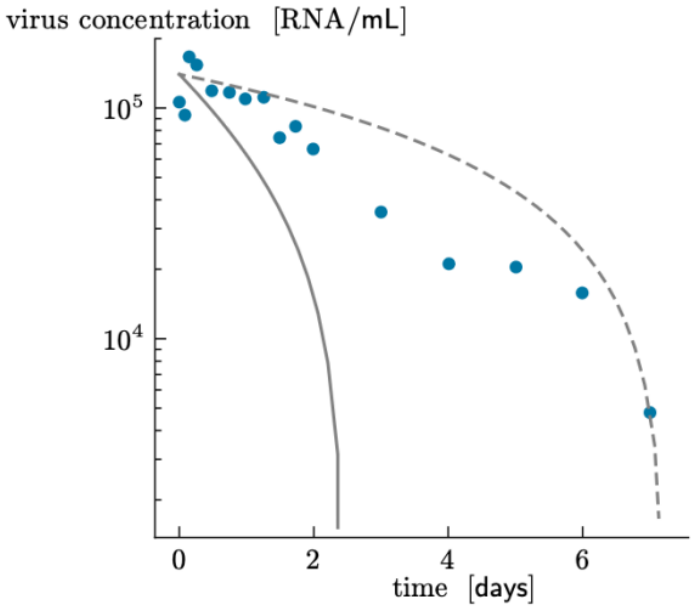

Figure 1.5: Examples of bad fits to the viral-load data. The solid curve is a solution of the physical model with an unsuitable parameter set. The dashed curve belongs to a different, unmechanistic family and cannot reproduce all of the observed features (Perelson et al. 1996; Nelson 2022).

Some useful terminology is:

Over-constrained. The data impose more independent constraints than there are free parameters.

Overfit. The model has enough flexibility to memorize the observations without reliably explaining or extrapolating them.

Blind fit. A curve summarizes a trend without being tied to a physical mechanism.

Successful mechanistic fit. The data are consistent with the theory and the inferred parameters answer the motivating question, even when some biological actors are not measured directly.

An informal falsification test is to count independent features rather than raw data points. The viral-load curve contains an initial value, the slope near the plateau, the sharpness of the transition, the later slope, and the intercept of the later exponential regime. After using one feature to set the initial condition, roughly four features remain to constrain three parameters. There is no guarantee that any parameter combination will fit them all. In particular, the data are inconsistent with an infected-cell lifetime $1/k_I$ of ten years: viral turnover is not slow.

1.7 Beyond the minimal model

The minimal model explains the short-time response to treatment, but it does not describe the full infection. As HIV exits the long quasi-steady stage, immune control deteriorates and evolutionary escape becomes increasingly important. A more realistic treatment model distinguishes uninfected cells, infected cells, infectious virions, and noninfectious virions. Let N_U, N_I, N_V, and N_X denote these four populations. One extension is

$$\frac{\mathrm{d}N_U}{\mathrm{d}t} = \lambda - k_U N_U - \epsilon N_V N_U, \tag{1.11}$$

$$\frac{\mathrm{d}N_I}{\mathrm{d}t} = \epsilon N_V N_U - k_I N_I, \tag{1.12}$$

$$\frac{\mathrm{d}N_V}{\mathrm{d}t} = \epsilon' \gamma N_I - k_V N_V, \tag{1.13}$$

$$\frac{\mathrm{d}N_X}{\mathrm{d}t} = (1-\epsilon')\gamma N_I - k_V N_X. \tag{1.14}$$

Here λ is the source rate of uninfected cells, k_U is their loss rate, and ϵ is the mass-action infection-rate coefficient. Because it multiplies $N_V N_U$, ϵ is not itself a fraction and carries the units needed to make $\epsilon N_V N_U$ a population per unit time. In contrast, ϵ' is the dimensionless fraction of newly produced virions that are competent to infect another cell. Protease inhibition reduces ϵ', redirecting production from N_V toward N_X.

Long-lived, latently infected cells make complete clearance much more difficult. HIV can persist in dormant immune cells that are largely invisible to standard treatment. Immune-checkpoint molecules such as PD-1 and CTLA-4 may help maintain this hidden state. Experimental checkpoint inhibition and "shock and kill" strategies attempt to expose or reactivate the latent reservoir so that infected cells can be removed. Reducing that reservoir is a possible path toward a cure, but the approach remains experimental and requires careful evaluation of efficacy and safety.

1.8 Aside: pulse-chase proliferation and differentiation

The same linear-dynamics ideas are useful far beyond viral infection. In a pulse-chase experiment, labeled cells appear in an upstream source population and later enter a downstream destination population. A minimal source-destination model is

$$\frac{\mathrm{d}D(t)}{\mathrm{d}t} = kS(t) - dD(t), \tag{1.15}$$

where $S(t)$ is the measured source signal, $D(t)$ is the destination signal, k is the transfer or differentiation rate, and d is the loss rate (Figure 1.6).

Estimating derivatives from noisy, discretely sampled data is unstable because differencing amplifies measurement noise. Integrating the dynamics gives a more useful form for data analysis:

$$D(t) = k\int_{t_0}^{t} e^{-d(t-t')} S(t')\,\mathrm{d}t' + e^{-d(t-t_0)} D(t_0). \tag{1.16}$$

The exponential kernel weights recent source measurements more strongly than old ones, while the second term propagates the initial destination population.

The HIV and pulse-chase examples share the same lesson: a perturbation reveals hidden fluxes, and a small dynamical model converts the resulting time series into estimates of biologically meaningful rates.

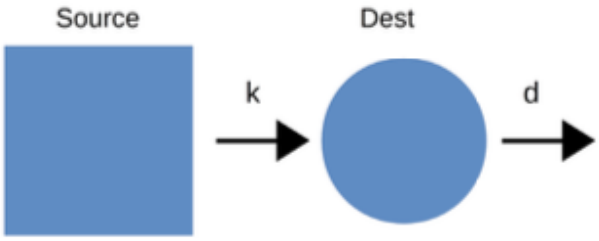

Figure 1.6: Minimal source-destination model for a pulse-chase experiment. Labeled cells enter the destination at rate $kS(t)$ and leave it at rate $dD(t)$ (course-generated).

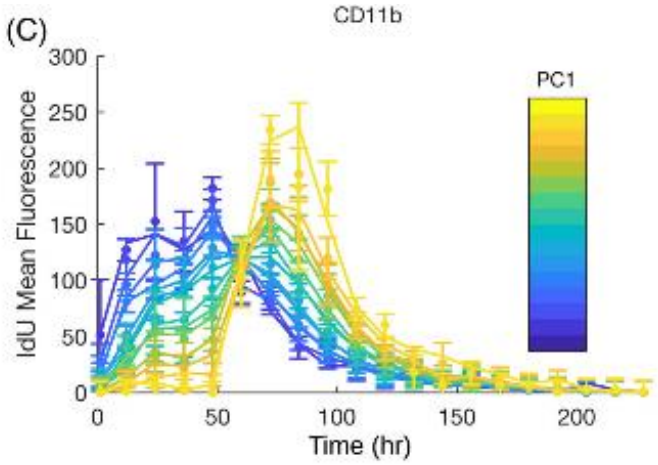

Figure 1.7: Example pulse-chase trajectories for CD11b-expressing cells. Each colored curve shows the measured label signal over time for a different population state (Erez, Mukherjee, et al. 2019).

2 Randomness: probability, information, and sampling

Randomness enters biological systems at many scales: inheritance selects one of two parental copies of a gene, molecular reactions occur after random waiting times, and measurements fluctuate from one sample to the next. This week develops a compact language for such variability.

By the end of this week, students should be able to answer:

- What outcomes are possible, and how should probability be assigned to them?
- How long must we wait for the next successful event?
- How does new evidence change the probability of a hypothesis?
- How much information do two measurements share?
- How accurately can a finite sample estimate a population property?

2.1 Bernoulli trials and random binary fractions

A Bernoulli trial has two outcomes. Write $s = 1$ for success and $s = 0$ for failure, with

$$\Pr(s = 1) = \xi, \qquad \Pr(s = 0) = 1 - \xi. \tag{2.1}$$

For a fair trial, $\xi = 1/2$. Two independent fair bits can be combined into the binary fraction

$$x = \frac{s_1}{2} + \frac{s_2}{4}, \tag{2.2}$$

which takes the values $0, 1/4, 1/2, 3/4$, each with probability $1/4$. More generally,

$$x_m = \sum_{i=1}^{m} \frac{s_i}{2^i} \tag{2.3}$$

is uniform on the 2^m binary fractions between 0 and $1-2^{-m}$. As m increases, the discrete distribution approaches a continuous uniform distribution on $[0, 1]$. The bars in Figure 2.1 become shorter because the same total probability is divided among more bins.

The same construction works in other bases. Independent decimal digits $d_i \in \{0, \ldots, 9\}$ generate

$$x = \frac{d_1}{10} + \frac{d_2}{100} + \cdots .$$

Mendelian inheritance supplies a biological Bernoulli example: ignoring recombination, mutation, duplication, and related complications, an offspring receives one of a parent's two gene copies with probability $1/2$.

2.2 Probability mass functions

For a reproducible experiment, let N_ℓ be the number of times outcome ℓ appears in N_{tot} repetitions. The frequentist interpretation gives the following operational characterization of its probability:

$$P(\ell) = \lim_{N_{\text{tot}} \to \infty} \frac{N_\ell}{N_{\text{tot}}}. \tag{2.4}$$

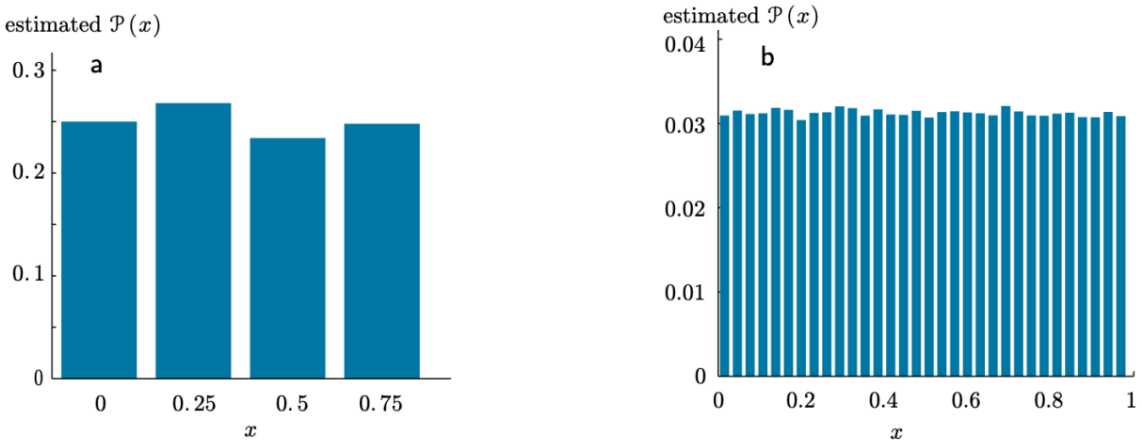

Figure 2.1: Empirical distributions of uniformly distributed binary fractions. Panel (a) shows 500 two-bit fractions, and panel (b) shows 250,000 five-bit fractions. The finer distribution has more bins, each containing a smaller share of the unit total probability (Nelson 2022).

A probability mass function for a discrete random variable satisfies

$$P(\ell) \geq 0, \qquad \sum_{\ell} P(\ell) = 1. \tag{2.5}$$

Probabilities are dimensionless. If E_1 and E_2 are events, then

$$P(E_1 \cup E_2) = P(E_1) + P(E_2), \qquad E_1 \cap E_2 = \varnothing, \tag{2.6}$$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2), \tag{2.7}$$

$$P(E^{\mathrm{c}}) = 1 - P(E). \tag{2.8}$$

The subtraction in Equation (2.7) prevents outcomes in the overlap from being counted twice.

2.3 Conditional probability and independence

Suppose a fair die has been rolled. Before receiving any other information, $P(5) = 1/6$. If we learn that the result is odd, the possible outcomes have been restricted to $\{1, 3, 5\}$, and

$$P(5 \mid \text{odd}) = \frac{1}{3}.$$

Information changes the relevant ensemble. In general, the probability of E conditioned on E' is

$$P(E \mid E') = \frac{P(E \cap E')}{P(E')}, \qquad P(E') > 0. \tag{2.9}$$

Equivalently, the general product rule is

$$P(E \cap E') = P(E \mid E')P(E'). \tag{2.10}$$

The events are statistically independent when conditioning on one does not alter the probability of the other:

$$P(E \mid E') = P(E) \quad \Longleftrightarrow \quad P(E \cap E') = P(E)P(E'). \tag{2.11}$$

Unrelated information, such as the result of an independent coin toss, does not change the die probability.

2.4 Waiting for success: the geometric distribution

Let independent Bernoulli trials be repeated until the first success. If J is the number of the successful trial, then $J = j$ requires $j - 1$ failures followed by one success:

$$P(J = j) = \xi(1 - \xi)^{j-1}, \qquad j = 1, 2, \ldots. \tag{2.12}$$

This is the geometric distribution. It is normalized because

$$\sum_{j=1}^{\infty} \xi(1 - \xi)^{j-1} = \xi \frac{1}{1 - (1 - \xi)} = 1. \tag{2.13}$$

The mean waiting time answers the question posed at the beginning of the lecture:

$$\mathbb{E}[J] = \frac{1}{\xi}. \tag{2.14}$$

A rare event with success probability ξ therefore requires $1/\xi$ attempts on average.

2.5 Joint and marginal distributions

For two discrete random variables X and Y, the joint mass function is

$$P_{XY}(x, y) = P(X = x, Y = y).$$

It is normalized, and summing over one variable produces the marginal distribution of the other:

$$\sum_{x} \sum_{y} P_{XY}(x, y) = 1, \tag{2.15}$$

$$P_X(x) = \sum_{y} P_{XY}(x, y), \qquad P_Y(y) = \sum_{x} P_{XY}(x, y). \tag{2.16}$$

The variables are independent precisely when the joint distribution factorizes:

$$P_{XY}(x, y) = P_X(x) P_Y(y). \tag{2.17}$$

For example, suppose $X \in \{0, 1\}$ records whether a regulator is active and $Y \in \{0, 1\}$ records whether a reporter is high. A possible joint distribution is

	$Y = 0$	$Y = 1$	$P_X(x)$
$X = 0$	0.50	0.10	0.60
$X = 1$	0.10	0.30	0.40
$P_Y(y)$	0.60	0.40	1

The marginal entries are obtained by summing rows or columns. Here $P(X = 1, Y = 1) = 0.30 \neq (0.40)(0.40)$, so the variables are not independent, and $P(Y = 1 \mid X = 1) = 0.30/0.40 = 0.75$. The table keeps the joint, marginal, and conditional probabilities visibly distinct.

2.6 Entropy and mutual information

The Shannon entropy of a discrete variable X is

$$H(X) = -\sum_{x} P_X(x) \log_2 P_X(x). \tag{2.18}$$

The base-2 logarithm measures entropy in bits. Entropy is large when many outcomes have similar probability and small when one outcome is nearly certain. The joint entropy is

$$H(X,Y) = -\sum_{x,y} P_{XY}(x,y) \log_2 P_{XY}(x,y). \tag{2.19}$$

The conditional entropy

$$H(Y \mid X) = -\sum_{x,y} P_{XY}(x,y) \log_2 P(Y = y \mid X = x) \tag{2.20}$$

is the uncertainty remaining in Y after X is known.

Mutual information measures how much knowing either variable reduces uncertainty about the other:

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{2.21}$$

$$= H(Y) - H(Y \mid X) = H(X) - H(X \mid Y). \tag{2.22}$$

Figure 2.2 summarizes these relations. If X determines Y completely, then $H(Y \mid X) = 0$ and $I(X;Y) = H(Y)$. If X and Y are independent, then $I(X;Y) = 0$.

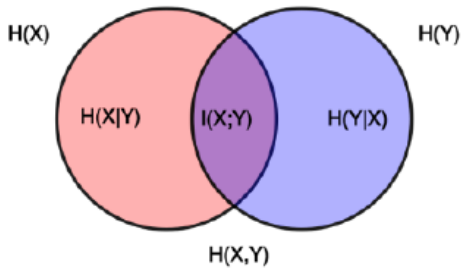

Figure 2.2: Information decomposition for two random variables. The overlap is the mutual information $I(X;Y)$; the nonoverlapping regions are the conditional entropies $H(X \mid Y)$ and $H(Y \mid X)$ (Shannon 1948). Diagram from Wikipedia/Wikimedia Commons.

An equivalent expression makes the comparison with independence explicit:

$$I(X;Y) = \sum_{x,y} P_{XY}(x,y) \log_2 \frac{P_{XY}(x,y)}{P_X(x) P_Y(y)}. \tag{2.23}$$

2.7 Kullback-Leibler divergence

The Kullback-Leibler divergence compares a distribution P with a reference distribution Q:

$$D_{\mathrm{KL}}(P\|Q) = \sum_{x\in\mathcal{X}} P(x)\log_2\frac{P(x)}{Q(x)}. \tag{2.24}$$

Terms with $P(x) > 0$ require $Q(x) > 0$. Throughout this week we use base-2 logarithms, so entropy, mutual information, and KL divergence are all measured in bits. The divergence is zero when $P = Q$, and it quantifies the additional surprise incurred by describing data generated by P using the model Q. It is not a distance: in general, $D_{\mathrm{KL}}(P\|Q) \neq D_{\mathrm{KL}}(Q\|P)$.

Nonnegativity from Jensen's inequality. The function $\phi(u) = -\log_2 u$ is convex for $u > 0$, since $\phi''(u) = 1/(u^2 \ln 2) > 0$. Jensen's inequality states

$$\phi(\mathbb{E}[Z]) \le \mathbb{E}[\phi(Z)] \tag{2.25}$$

for a random variable Z in the domain of a convex function. Draw i according to P and set $Z = Q(i)/P(i)$. Then

$$\mathbb{E}_P[Z] = \sum_i P(i)\frac{Q(i)}{P(i)} = \sum_i Q(i) = 1.$$

Applying Jensen's inequality gives Gibbs' inequality:

$$\begin{aligned} 0 &= -\log_2 \mathbb{E}_P\left[\frac{Q(i)}{P(i)}\right] \\ &\le \mathbb{E}_P\left[-\log_2\frac{Q(i)}{P(i)}\right] = \sum_i P(i)\log_2\frac{P(i)}{Q(i)} = D_{\mathrm{KL}}(P\|Q). \end{aligned} \tag{2.26}$$

Mutual information is a KL divergence between the joint distribution and the product of its marginals:

$$I(X;Y) = D_{\mathrm{KL}}(P_{XY}\|P_X P_Y)\,. \tag{2.27}$$

It therefore measures how far the joint distribution is from independence and is always nonnegative.

2.8 Bayes' theorem and a medical test

Bayes' theorem reverses a conditional probability:

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)}. \tag{2.28}$$

The denominator follows from the law of total probability. If mutually exclusive hypotheses $H_1, \dots, H_m$ cover all possibilities, then

$$P(E) = \sum_{j=1}^{m} P(E \mid H_j)P(H_j). \tag{2.29}$$

For the medical test below, the two hypotheses are disease and no disease, so the positive tests arising from these two groups must be added. Here H is a hypothesis and E is newly observed evidence:

Prior $P(H)$**.** The probability assigned to the hypothesis before observing the current evidence.

Likelihood $P(E \mid H)$**.** The probability of seeing the evidence if the hypothesis is true.

Evidence $P(E)$**.** The total or marginal probability of observing the evidence under all possible hypotheses.

Posterior $P(H \mid E)$**.** The updated probability after incorporating the evidence.

Bayesian updating is used in medical diagnosis, spam filtering, parameter inference, decision making, forensic reasoning, and many other settings. Ignoring the prior is especially dangerous for rare events.

Consider a screening test for a disease with prevalence

$$P(S) = 0.009.$$

Suppose the sensitivity and specificity are both 97%:

$$P(+ \mid S) = 0.97, \qquad P(- \mid S^{\mathrm{c}}) = 0.97. \tag{2.30}$$

The false-positive probability is therefore $P(+ \mid S^{\mathrm{c}}) = 0.03$. The total probability of a positive test is

$$\begin{aligned} P(+) &= P(+ \mid S)P(S) + P(+ \mid S^{\mathrm{c}})P(S^{\mathrm{c}}) \\ &= (0.97)(0.009) + (0.03)(0.991) = 0.03846. \end{aligned} \tag{2.31}$$

Bayes' theorem then gives

$$P(S \mid +) = \frac{(0.97)(0.009)}{0.03846} \approx 0.227. \tag{2.32}$$

Thus a positive result from a test described as 97% sensitive and specific implies only about a 23% chance of disease in this low-prevalence screening population. Nevertheless, the test has increased the probability from 0.9% to 22.7%, a roughly 25-fold update.

2.9 Expectation and moments

For a function $f(s)$ of a discrete random outcome s, its nth raw moment is

$$\langle f^n \rangle = \mathbb{E}[f^n] = \sum_s f^n(s) P(s). \tag{2.33}$$

The variance measures the mean squared fluctuation around the expectation:

$$\mathrm{Var}(f) = \Big\langle (f - \langle f \rangle)^2 \Big\rangle = \langle f^2 \rangle - \langle f \rangle^2. \tag{2.34}$$

For the Bernoulli variable in Equation (2.1),

$$\langle s \rangle = \xi, \qquad \langle s^2 \rangle = \xi, \qquad \mathrm{Var}(s) = \xi(1 - \xi).$$

The Bernoulli variance is maximal at $\xi = 1/2$.

If f and g are independent, their joint distribution factorizes and

$$\langle fg \rangle = \sum_{i,j} f(i) g(j) P_i P_j = \langle f \rangle \langle g \rangle. \tag{2.35}$$

Linearity of expectation does not require independence:

$$\langle f + g \rangle = \langle f \rangle + \langle g \rangle.$$

For independent variables, however, the variances add:

$$\mathrm{Var}(f + g) = \mathrm{Var}(f - g) = \mathrm{Var}(f) + \mathrm{Var}(g). \tag{2.36}$$

2.10 Relative noise and the coefficient of variation

The coefficient of variation, also called the relative standard deviation, is

$$\mathrm{CV}(f) = \frac{\sqrt{\mathrm{Var}(f)}}{|\langle f \rangle|}. \tag{2.37}$$

It is dimensionless and measures fluctuations relative to the mean scale. For two independent variables with nonnegative means, provided the two coefficients of variation below are defined,

$$\frac{\mathrm{CV}(f+g)}{\mathrm{CV}(f-g)} = \frac{|\langle f \rangle - \langle g \rangle|}{|\langle f \rangle + \langle g \rangle|} \le 1. \tag{2.38}$$

The difference of two noisy positive quantities can therefore have much larger relative noise than their sum. If their means are equal, $\langle f - g \rangle = 0$ and $\mathrm{CV}(f-g)$ is undefined; as the means approach cancellation, the denominator of that coefficient of variation approaches zero and the relative noise becomes arbitrarily large.

2.11 Sample means and standard errors

Given M independent measurements $f_1, \dots, f_M$, define the sample mean

$$\overline{f} = \frac{1}{M} \sum_{i=1}^{M} f_i. \tag{2.39}$$

The sample mean is itself a random variable: a new batch of M measurements generally produces a different value. If every measurement has mean μ and variance σ^2, then

$$\mathbb{E}[\overline{f}] = \mu, \tag{2.40}$$

$$\mathrm{Var}(\overline{f}) = \frac{1}{M^2} \sum_{i=1}^{M} \mathrm{Var}(f_i) = \frac{\sigma^2}{M}. \tag{2.41}$$

The theoretical standard deviation of the sampling distribution of the mean, called its standard error, is therefore

$$\mathrm{SEM}(\overline{f}) = \frac{\sigma}{\sqrt{M}}. \tag{2.42}$$

Averaging improves precision only as the square root of sample size: reducing the standard error by a factor of two requires four times as many independent observations.

When μ is estimated from the same data, an unbiased estimator of the population variance uses Bessel's correction:

$$s^2 = \frac{1}{M-1} \sum_{i=1}^{M} (f_i - \overline{f})^2. \tag{2.43}$$

Only $M-1$ deviations are independent because their sum is constrained to be zero. When the population standard deviation σ is unknown, the reported standard error is estimated by $s/\sqrt{M}$. This estimated standard error should not be confused with s, which describes the spread of individual observations. Both formulas also presume independent sampling; correlation reduces the effective amount of independent information.

2.12 Covariance and correlation

The covariance of two random variables ℓ and m is

$$\operatorname{Cov}(\ell, m) = \langle (\ell - \langle \ell \rangle)(m - \langle m \rangle) \rangle \,. \tag{2.44}$$

Normalizing by the standard deviations gives the Pearson correlation coefficient:

$$\operatorname{corr}(\ell, m) = \frac{\operatorname{Cov}(\ell, m)}{\sqrt{\operatorname{Var}(\ell)\,\operatorname{Var}(m)}}. \tag{2.45}$$

It lies between -1 and 1 and measures the direction and strength of a *linear* relationship. Figure 2.3 illustrates both its use and its limitations. Strong nonlinear dependence can have correlation zero.

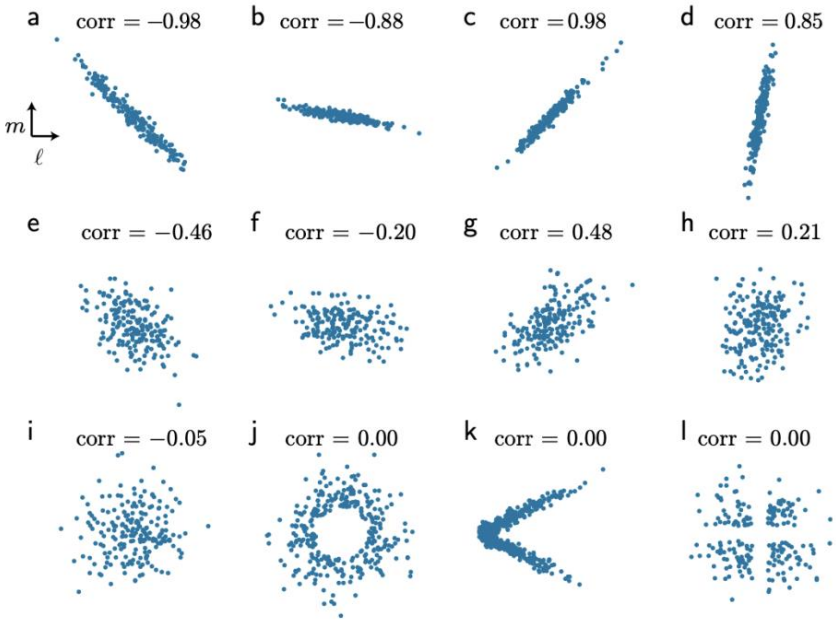

Figure 2.3: Pearson correlations for simulated joint distributions. Panels (a)–(h) show noisy linear relations of different signs and strengths. Panel (i) is independent, while panels (j)–(l) are dependent but have nearly zero linear correlation. Each coefficient was estimated from 5000 simulated points; the first 200 are shown (Nelson 2022).

Spearman correlation first replaces observations by their ranks and then computes Pearson correlation. It is sensitive to a general monotonic relation rather than only a linear one.

For a stationary time series with constant mean μ, the theoretical autocovariance is

$$C(j) = \mathbb{E}[(f_i - \mu)(f_{i+j} - \mu)], \tag{2.46}$$

which depends on the lag j, not on the starting index i. A common finite-sample estimate is

$$\widehat{C}(j) = \frac{1}{M - j} \sum_{i=1}^{M-j} (f_i - \overline{f})(f_{i+j} - \overline{f}). \tag{2.47}$$

A positive $\widehat{C}(j)$ means observations separated by j time steps tend to fluctuate in the same direction. Correlation alone does not establish causation; methods such as Granger analysis ask instead

whether the past of one time series improves prediction of another. If the mean or variance drifts with time, this single-lag summary mixes the drift with fluctuations. One should then model or remove the trend, analyze locally stationary segments, or estimate a two-time covariance rather than silently assuming stationarity.

2.13 Anscombe's quartet: always look at the data

Anscombe's quartet, introduced by Anscombe (1973), consists of four datasets with eleven (x, y) pairs each. To standard numerical precision, all four datasets share the same basic summaries:

Quantity	Common value
Number of observations	$n = 11$
Mean of x	$\overline{x} = 9.0$
Mean of y	$\overline{y} \simeq 7.5$
Sample variance of x	$s_x^2 = 11.0$
Sample variance of y	$s_y^2 \simeq 4.13$
Pearson correlation	$r \simeq 0.816$
Least-squares line	$\widehat{y} = 3.0 + 0.5x$

If we examined only these quantities, we might conclude that the datasets were interchangeable. Plotting them reveals four very different structures (Figure 2.4):

- Dataset I is approximately linear with ordinary scatter around the fitted line.
- Dataset II has a pronounced curved relation that a straight line fails to describe.
- Dataset III is nearly linear except for one influential vertical outlier.
- Dataset IV has almost no variation in x except for one high-leverage point; that single observation largely determines the fitted line and correlation.

The lesson is not that summary statistics are unhelpful. Rather, numerical summaries, model fits, residual analysis, and visualization answer different questions and should be used together. Waiting-time distributions describe events, Bayes' theorem combines data with prior knowledge, information measures compare variables and models, and moments and standard errors connect a stochastic population to finite experimental samples. Anscombe's quartet adds the final warning: before trusting a compact summary, look at the data.

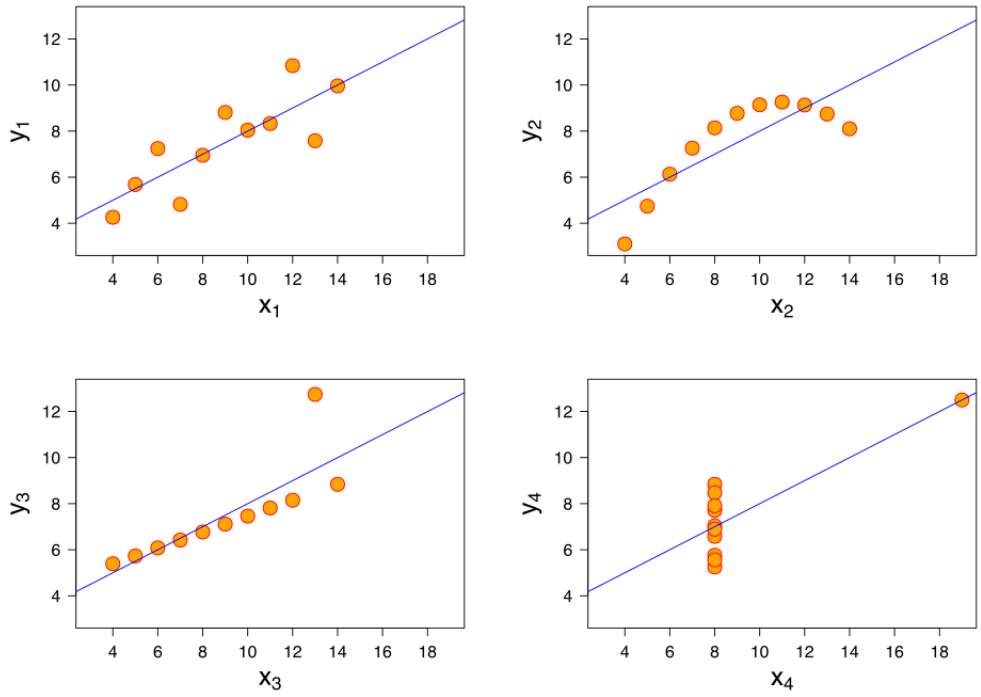

Figure 2.4: Anscombe's quartet. The four datasets have nearly identical means, variances, Pearson correlations, and least-squares regression lines, yet their geometries are qualitatively different. Figure from Wikipedia/Wikimedia Commons; the datasets were introduced by Anscombe (1973).

3 Using discrete distributions

Discrete distributions connect microscopic random events to quantities that can be measured in cells. We begin with Bernoulli trials and the binomial distribution, use partition noise to count fluorescent molecules, and take the rare-event limit that leads to the Poisson distribution. These tools then become the null model for the Luria–Delbrück fluctuation experiment, in which the *shape* of a distribution distinguishes induced adaptation from spontaneous mutation.

By the end of this week, students should be able to answer:

- How does a sum of independent yes-or-no events produce a binomial distribution?
- How can fluctuations at cell division calibrate fluorescence in units of molecule number?
- Why does the binomial distribution approach the Poisson distribution for many rare trials?
- How do adaptation after exposure and mutation before exposure generate different distributions of resistant bacteria?
- Why do mutation timing and exponential growth produce heavy-tailed jackpot events?

3.1 Bernoulli trials and the binomial distribution

Recall from Week 2 that a Bernoulli trial succeeds with probability ξ. The number L of successes in M independent trials has the binomial distribution

$$P_{\text{binom}}(L = \ell; M, \xi) = \binom{M}{\ell} \xi^{\ell} (1-\xi)^{M-\ell}, \qquad \ell = 0, 1, \ldots, M. \tag{3.1}$$

Independence makes the moments additive:

$$\mathbb{E}[L] = M\xi, \qquad \mathrm{Var}(L) = M\xi(1-\xi). \tag{3.2}$$

This short recall is enough for the partition-noise application below and for the rare-event limit later in the week.

3.2 Counting fluorescent molecules from partition noise

Suppose a parent bacterium contains M_0 identical fluorescent molecules and has measured intensity

$$y_0 = \alpha M_0, \tag{3.3}$$

where α is the detected intensity per molecule. The intensity y_0 is easy to measure, but M_0 cannot be inferred until α is calibrated.

At nearly symmetric cell division, each molecule independently enters daughter 1 with probability 1/2. Therefore

$$M_1 \sim \text{Binomial}\left(M_0, \frac{1}{2}\right), \qquad M_2 = M_0 - M_1, \tag{3.4}$$

and $\mathrm{Var}(M_1) = M_0/4$. Define the partitioning difference

$$\Delta M = M_1 - M_2 = 2M_1 - M_0. \tag{3.5}$$

Its mean is zero and its variance is

$$\mathrm{Var}(\Delta M) = 4\,\mathrm{Var}(M_1) = M_0. \tag{3.6}$$

Because $\Delta y = \alpha \Delta M$, Equation (3.3) gives

$$\mathrm{Var}(\Delta y) = \alpha^2 M_0 = \alpha y_0, \qquad \mathrm{SD}(\Delta y) = \sqrt{\alpha y_0}. \tag{3.7}$$

Thus a plot of partitioning variance against parental fluorescence has slope α. Once that single-molecule calibration is known, every measured intensity can be converted to molecule number using $M = y/\alpha$.

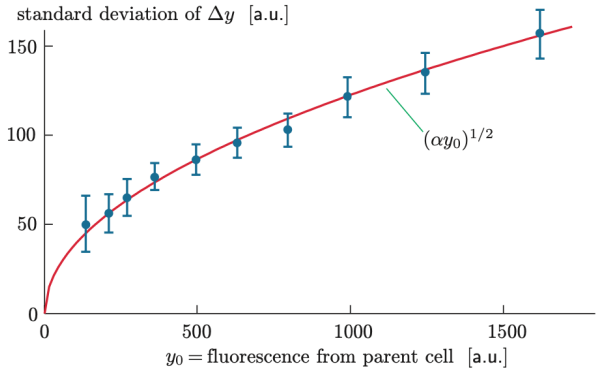

Figure 3.1: Calibration of a single-molecule fluorescence measurement using partition noise. Points show the measured standard deviation of the daughter-cell fluorescence difference at a given parent-cell intensity. The fitted curve $\mathrm{SD}(\Delta y) = \sqrt{\alpha y_0}$ corresponds to approximately 15 fluorescence units per tagged molecule (Rosenfeld, Young, et al. 2005).

The argument assumes independent partitioning, equal molecular brightness, and nearly equal daughter volumes. For asymmetric division, a molecule enters daughter 1 with probability $\xi \neq 1/2$, giving

$$\mathrm{Var}(\Delta M) = 4 M_0 \xi (1 - \xi). \tag{3.8}$$

Additional camera and background noise would add a separate contribution to the measured variance. These deviations do not invalidate the approach, but they must be included in the measurement model.

3.3 Sampling an arbitrary discrete distribution

The same probabilistic construction gives a general way to sample a discrete variable without making programming syntax part of the learning objective. For probabilities $p_0, p_1, \ldots$, define the cumulative distribution

$$F(k) = \sum_{j=0}^{k} p_j. \tag{3.9}$$

Draw U uniformly on $(0, 1)$ and choose

$$L = \min\{k : F(k) \geq U\}. \tag{3.10}$$

Geometrically, the unit interval is divided into adjacent bins of widths p_k. For $L \sim \mathrm{Binomial}(3, \xi)$, those widths are

$$(1 - \xi)^3, \quad 3\xi(1 - \xi)^2, \quad 3\xi^2(1 - \xi), \quad \xi^3. \tag{3.11}$$

An LLM can translate Equation (3.10) into any chosen language; the essential scientific checks are that the bins cover the unit interval without gaps or overlap and that their widths equal the target probabilities.

3.4 The Poisson rare-event limit

The binomial distribution contains two parameters, M and ξ. When the number of trials is large and each success is rare, let

$$M \to \infty, \qquad \xi = \frac{\mu}{M} \to 0, \qquad M\xi = \mu \tag{3.12}$$

with μ fixed. Substituting into Equation (3.1) and regrouping gives

$$P(L = \ell) = \frac{\mu^\ell}{\ell!} \left[\frac{M(M-1)\cdots(M-\ell+1)}{M^\ell} \right] \left(1 - \frac{\mu}{M}\right)^M \left(1 - \frac{\mu}{M}\right)^{-\ell}. \tag{3.13}$$

For fixed ℓ, the first bracket and the final factor tend to one, while

$$\lim_{M\to\infty} \left(1 - \frac{\mu}{M}\right)^M = e^{-\mu}. \tag{3.14}$$

Therefore

$$P_{\text{Pois}}(L = \ell; \mu) = e^{-\mu} \frac{\mu^\ell}{\ell!}, \qquad \ell = 0, 1, 2, \ldots. \tag{3.15}$$

The Poisson distribution is normalized by the exponential series and satisfies

$$\mathbb{E}[L] = \mu, \qquad \text{Var}(L) = \mu. \tag{3.16}$$

It describes the total number of successes generated by many independent opportunities, each with a small probability. This mean-equals-variance prediction will now serve as the benchmark against which the bacterial fluctuation data are compared.

3.5 The fluctuation experiment

Luria and Delbrück began each culture with a small number N_0 of *Escherichia coli* cells and allowed the population to grow for g generations. For synchronous binary division with doubling time τ_d,

$$N(t) = N_0 2^{t/\tau_d}, \qquad N = N_0 2^g. \tag{3.17}$$

The final cultures contained roughly 10^9 cells. Each culture was then exposed to a bacteriophage that killed sensitive bacteria. Resistant cells survived, formed visible colonies when plated, and passed their resistance to their descendants.

The number of resistant cells varied enormously from culture to culture. Many cultures had few or no survivors, whereas a small number contained jackpots with many survivors (Figure 3.2). In particular, the standard deviation was much larger than the mean.

The experiment compared two generative explanations:

1. **Induced adaptation.** Resistance is produced by contact with the virus. Every cell independently has a small probability of adapting after exposure.

2. **Spontaneous mutation.** Resistance is produced by a heritable mutation during growth, before the virus is present. A mutant's descendants inherit resistance.

Both hypotheses can produce a similar *average* number of survivors. Their fluctuations, however, are qualitatively different.

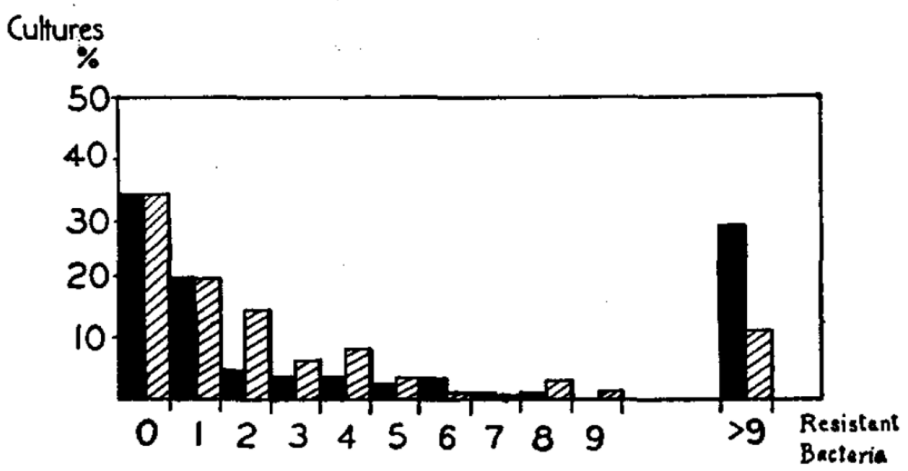

Figure 3.2: Distribution of resistant bacteria in the original fluctuation experiment. The black bars show the observed percentage of cultures, and the hatched bars show the prediction of the random-mutation model. The conspicuous group of cultures with more than nine resistant bacteria reflects rare jackpot events (Luria and Delbrück 1943).

3.6 The adaptation model: binomial and Poisson fluctuations

Suppose that each of the N cells present when the virus is added adapts independently with probability $p \ll 1$. Let $x_i = 1$ if cell i survives and $x_i = 0$ otherwise. The total number of resistant cells is

$$X = \sum_{i=1}^{N} x_i, \qquad X \sim \text{Binomial}(N, p), \tag{3.18}$$

with probability mass function

$$P_{\text{adapt}}(X = x) = \binom{N}{x} p^x (1-p)^{N-x}. \tag{3.19}$$

In the rare-event limit $N \to \infty$, $p \to 0$, with $\lambda = Np$ fixed, this becomes

$$P_{\text{adapt}}(X = x) \simeq e^{-\lambda} \frac{\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots, \tag{3.20}$$

the Poisson distribution. Its first two moments are

$$\mathbb{E}[X] = \lambda, \qquad \text{Var}(X) = \lambda. \tag{3.21}$$

Thus the Fano factor, the variance divided by the mean, is

$$F \equiv \frac{\text{Var}(X)}{\mathbb{E}[X]} = 1. \tag{3.22}$$

Independent last-minute adaptation therefore predicts ordinary fluctuations around the mean. It does not predict a mixture of many nearly empty cultures and occasional enormous jackpots.

3.7 The mutation model: time creates jackpots

Now suppose that resistance mutations occur during population growth. Write $\mu \ll 1$ for the mutation probability per daughter-cell lineage. The number of daughter cells born into generation n is $N_0 2^n$, so the number of new mutations in that generation is approximately

$$M_n \sim \text{Poisson}(\lambda_n), \qquad \lambda_n = \mu N_0 2^n. \tag{3.23}$$

We initially neglect a second mutation arising within an already resistant lineage. This is an excellent approximation when μ is small.

A mutation arising in generation n has $g - n$ generations in which to expand. Under the synchronous-growth approximation, its clone therefore contributes

$$A_n = 2^{g-n} \tag{3.24}$$

resistant cells at the end. The final resistant population is the weighted random sum

$$S = \sum_{n=1}^{g} A_n M_n = \sum_{n=1}^{g} 2^{g-n} M_n. \tag{3.25}$$

This expression contains the central mechanism. Mutations are far less likely in early generations because few cells are present, but each early mutation produces exponentially more descendants. A late mutation is common but produces only a small clone; a rare early mutation creates a jackpot.

There is an exact cancellation in the expected contribution from each generation:

$$\lambda_n A_n = (\mu N_0 2^n)\, 2^{g-n} = \mu N, \tag{3.26}$$

so every generation contributes the same amount to the mean. The variance does not share this cancellation because clone size enters it quadratically.

3.8 A generating function for the full distribution

The mutation model is more than a verbal explanation. Under the independent-generation approximation, its entire distribution is summarized by a probability-generating function. For any nonnegative integer-valued random variable X, define

$$G_X(z) = \mathbb{E}[z^X] = \sum_{x=0}^{\infty} P(X = x) z^x. \tag{3.27}$$

The coefficient of z^x is the probability $P(X = x)$. Evaluation and differentiation at $z = 1$ give

$$G_X(1) = 1, \qquad G'_X(1) = \mathbb{E}[X], \qquad G''_X(1) = \mathbb{E}[X(X-1)]. \tag{3.28}$$

If X and Y are independent, then $G_{X+Y}(z) = G_X(z) G_Y(z)$; this product rule is why generating functions are useful for the sum of mutant clones. For a Poisson variable M_n,

$$\mathbb{E}[u^{M_n}] = \exp[\lambda_n (u - 1)]\,. \tag{3.29}$$

Using $u = z^{A_n}$ and multiplying the independent contributions gives

$$G_S(z) \equiv \mathbb{E}[z^S] = \exp\left[\sum_{n=1}^{g} \lambda_n \left(z^{A_n} - 1\right)\right]. \tag{3.30}$$

The coefficient of z^s in $G_S(z)$ is $P(S = s)$. More importantly, several experimentally useful results follow without computing every coefficient:

- setting $z = 0$ gives the probability of no resistant cells;
- differentiating once at $z = 1$ gives the mean; and
- differentiating twice, or using Poisson cumulants, gives the variance.

The large gaps among the possible clone sizes $A_n = 1, 2, 4, \ldots$ are an artifact of perfectly synchronous division. Random division times smooth these discrete peaks but do not remove the broad tail.

3.9 The zero class and mutation-rate estimation

The number of mutation opportunities in the complete lineage tree is

$$K = N_0 \sum_{n=1}^{g} 2^n = 2N - 2N_0 \simeq 2N. \tag{3.31}$$

The total number of mutation events is therefore approximately Poisson with mean

$$\Lambda = \mu K \simeq 2\mu N. \tag{3.32}$$

There are no resistant cells if and only if there were no resistance mutations, hence

$$P_0 \equiv P(S = 0) = e^{-\Lambda}. \tag{3.33}$$

This is also obtained immediately by setting $z = 0$ in Equation (3.30).

Suppose C independent cultures are grown and C_0 contain no resistant cells. Treating zero versus nonzero as a Bernoulli observation gives the likelihood

$$L(\Lambda) = \binom{C}{C_0} \left(e^{-\Lambda}\right)^{C_0} \left(1 - e^{-\Lambda}\right)^{C - C_0}. \tag{3.34}$$

It is maximized when the model probability of a zero equals the observed zero fraction:

$$e^{-\widehat{\Lambda}} = \frac{C_0}{C}, \qquad \widehat{\Lambda} = -\log\!\left(\frac{C_0}{C}\right), \qquad \widehat{\mu} = \frac{\widehat{\Lambda}}{K} \simeq -\frac{1}{2N}\log\!\left(\frac{C_0}{C}\right). \tag{3.35}$$

The numerical factor relating μ to N depends on whether the mutation rate is defined per cell division or per daughter lineage. Equation (3.35) uses the latter convention.

The estimator also exposes two finite-sample edge cases. If every culture is nonzero ($C_0 = 0$), the point estimate is formally infinite; the experiment supplies only a lower bound on Λ. If every culture is zero ($C_0 = C$), the estimate is zero, but the data still allow a positive upper bound. In either case, reporting a confidence or credible interval and changing the number of cells plated or the number of replicate cultures is more informative than quoting the boundary estimate alone.

This *zero-class estimator* is robust because it records whether a mutation happened, not how large its descendant clone became. A single jackpot can dominate the sample mean and variance, but it changes the zero count by only one culture. This is a useful general principle in parameter estimation: choose a statistic that is directly tied to the hidden event of interest and is insensitive to the unstable part of the data.

3.10 Mean, variance, and the jackpot Fano factor

Because a Poisson variable has equal mean and variance, the mean mutant count is

$$\mathbb{E}[S] = \sum_{n=1}^{g} A_n \lambda_n = g\mu N, \tag{3.36}$$

$$\mathrm{Var}(S) = \sum_{n=1}^{g} A_n^2 \lambda_n = \mu N \sum_{n=1}^{g} 2^{g-n} = \mu N(2^g - 1). \tag{3.37}$$

The corresponding Fano factor is

$$F = \frac{\mathrm{Var}(S)}{\mathbb{E}[S]} = \frac{2^g - 1}{g} \simeq \frac{N}{N_0 g}. \tag{3.38}$$

For twenty generations this factor is of order 10^4–10^5, rather than the Poisson value $F = 1$. The mutation model therefore predicts exactly the qualitative feature that defeated the adaptation model.

Equation (3.36) also reveals a subtlety. Since every generation contributes equally to the theoretical mean, observing the mean accurately requires sampling rare early mutations as well as common late ones. Equation (3.37) is even more demanding: early mutations receive an additional factor of clone size and dominate the variance.

3.11 Finite experiments do not sample the ensemble evenly

The expectation values above describe an arbitrarily large ensemble of cultures. A real experiment may contain only tens or hundreds. In C cultures, generation n is typically represented only when the expected total number of mutations in that generation is at least of order one:

$$C\lambda_n = C\mu N_0 2^n \sim 1. \tag{3.39}$$

The earliest generation normally sampled is therefore approximately

$$n_{\min} \simeq \lceil -\log_2(C\mu N_0) \rceil, \tag{3.40}$$

with $n_{\min}$ restricted to the interval from 1 to g. For $N_0 = 100$, $\mu = 5 \times 10^{-9}$, and $C = 200$, this estimate gives $n_{\min} = 14$. Mutations from the first thirteen generations will therefore usually be absent even though they strongly affect the infinite-ensemble moments.

In a finite dataset, the sums in Equations (3.36) and (3.37) are effectively truncated below $n_{\min}$. The sample variance is then usually much smaller than its infinite-ensemble expectation, but it remains far larger than the mean. If another batch happens to include a still earlier mutation, its sample mean and variance can jump discontinuously. This explains why heavy-tailed data often converge very slowly as more replicates are added.

3.12 The heavy tail

Choose one mutation event at random from all mutation opportunities. Its probability of occurring in generation n is proportional to the number of cells in that generation, so $q_n \propto 2^n$. If it occurs in generation n, its clone size is $x = 2^{g-n}$. Treating n and x as continuous for a scaling argument,

$$2^n = \frac{2^g}{x}, \qquad \left|\frac{\mathrm{d}n}{\mathrm{d}x}\right| = \frac{1}{(\ln 2)x}. \tag{3.41}$$

Changing variables therefore gives

$$f_X(x) \propto q[n(x)] \left| \frac{\mathrm{d}n}{\mathrm{d}x} \right| \propto \frac{1}{x^2}, \tag{3.42}$$

or $P(X > x) \propto x^{-1}$. This power-law tail is the mathematical origin of the jackpots. For any finite g, clone size has an upper cutoff and all moments are finite. As g grows, however, the cutoff moves outward and the moments become increasingly dominated by rare events.

The long tail is not merely an artifact of the synchronous model used here. Mandelbrot (1974) derived explicit long-tailed mutation-count distributions and connected their asymptotic behavior to stable laws. A later fully stochastic analysis by Kessler and Levine (2013) showed why the central part of the distribution can be well behaved while rare outer-tail events control its formal moments. These results reinforce the practical lesson: the mean and variance can be poor summaries of a finite fluctuation experiment even when they exist for every finite culture.

The important conclusion for this course is qualitative: the usual Gaussian intuition for sums is unreliable when the tail is this broad. Rare early mutations remain influential, sample moments converge slowly, and collecting more cultures does not immediately produce a narrow, well-behaved distribution. We will use those consequences without developing the technical limit theorems.

Real fluctuation assays also depart from the ideal model through incomplete plating, variation in final culture size, different growth or death rates of resistant cells, phenotypic delay between mutation and detectable resistance, and pre-existing mutants in the inoculum. These effects can alter the zero class, clone sizes, or both, so a serious mutation-rate estimate should state which of them are negligible and incorporate the others in the observation model.

3.13 Reading a simulated fluctuation distribution

Figure 3.3 shows a binned simulated mutation distribution. The large zero class is accompanied by progressively smaller probabilities at ordinary counts and appreciable probability in the pooled rightmost tail bin. The irregular peaks reflect the synchronous-division approximation: mutations one generation apart produce clone sizes that differ by a factor of two.

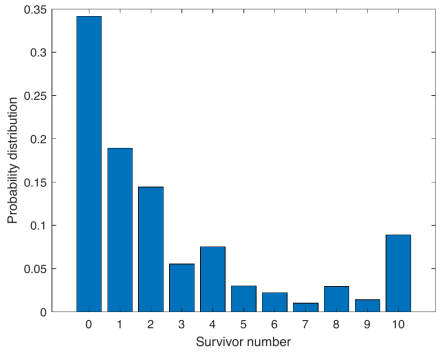

Figure 3.3: Simulated distribution of resistant-cell counts for $N_0 = 100$, $g = 20$, $\mu = 5 \times 10^{-9}$, and 10,000 cultures. Counts beyond the displayed interior range are pooled into the rightmost bar. Synchronous binary division produces visible structure at clone sizes related by powers of two. Course-generated from the model of Luria and Delbrück (1943).

The computational learning goal is to specify and test the generative model, not to memorize programming syntax. An LLM can translate the model into a chosen language, but the scientific

checks must come from the theory: a correct implementation should reproduce the zero probability in Equation (3.33), the mean scaling in Equation (3.36), the large Fano factor, and the jackpot tail. These invariants are also the best way to detect a plausible-looking but incorrect simulation.

3.14 What the experiment teaches

The fluctuation test illustrates several themes that recur throughout the course:

- **Model distributions, not only averages.** Two mechanisms can predict the same mean and radically different distributions.
- **Mechanism determines noise.** Independent adaptation gives Poisson-like fluctuations; heritable growth after a random mutation produces jackpots.
- **Timing can be amplified.** Exponential population growth converts a modest difference in mutation time into an enormous difference in clone size.
- **Choose stable estimators.** The fraction of zero cultures estimates the mutation rate more reliably than a mean dominated by rare events.
- **Finite samples can be unrepresentative.** Heavy tails make convergence slow and cause sample moments to vary strongly between experiments.

The striking variability that first looked like experimental noise was therefore the most informative part of the data. Quantitative modeling turned that variability into evidence that bacterial resistance existed before the selective challenge.

4 Entropy measures and biodiversity

Species richness—the number of species present—is only one aspect of biodiversity. A community dominated by one species is not equivalent to an evenly populated community with the same species list, and a collection of nearly identical strains is not equivalent to the same number of distantly related taxa. This unit develops measures that incorporate abundance, evenness, and biological similarity.

The information-theoretic quantities introduced in Week 2 now acquire a biological interpretation. Entropy measures the uncertainty in the identity of a randomly sampled individual, while its exponential converts that uncertainty into an effective number of equally abundant species. Similarity-sensitive extensions then account for functional, genetic, or phylogenetic overlap.

Week 2 used base-2 logarithms, so its entropy was measured in bits. This week uses natural logarithms and therefore measures entropy in nats. The numerical entropy changes by the constant factor $H_{\text{nats}} = (\ln 2) H_{\text{bits}}$, but the effective diversity does not when the matching exponential base is used:

$$\exp(H_{\text{nats}}) = 2^{H_{\text{bits}}}. \tag{4.1}$$

By the end of this week, students should be able to answer:

- How do Shannon entropy, Rényi entropy, and Hill numbers weight rare and common species differently?
- How do pairwise similarity and phylogenetic distance change a diversity comparison?
- How do finite sampling and sequencing depth affect estimated diversity?

4.1 Shannon entropy and the effective number of species

Consider a community with N_{sp} species. Let p_i be the relative abundance of species i, with

$$p_i \geq 0, \qquad \sum_{i=1}^{N_{\text{sp}}} p_i = 1. \tag{4.2}$$

The Shannon entropy, using natural logarithms, is

$$H = -\sum_{i=1}^{N_{\text{sp}}} p_i \ln p_i, \tag{4.3}$$

where $0 \ln 0$ is interpreted as zero. The surprise associated with observing species i is $-\ln p_i$, so H is the expected surprise when an individual is sampled at random.

The exponential

$$D_1 = \exp(H) \tag{4.4}$$

is the *effective number of species*, or Hill number of order one (Hill 1973). It is the number of equally abundant species that would produce the observed entropy. Expressing diversity as an effective number makes values directly interpretable and comparable across indices, rather than treating raw entropy as though it were itself a species count (Jost 2006). For example:

- If $p_1 = p_2 = 1/2$, then $H = \ln 2$ and $D_1 = 2$.
- If one of those species disappears, then $H = 0$ and $D_1 = 1$.

- If three species are equally abundant, then $H = \ln 3$ and $D_1 = 3$.

In general, a perfectly even community has $p_i = 1/N_{\mathrm{sp}}$, giving the maximum entropy $H_{\max} = \ln N_{\mathrm{sp}}$. A community concentrated in a single species has $H = 0$.

In data, the probabilities p_i are replaced by observed frequencies. The resulting plug-in entropy is typically biased downward because rare, unobserved species contribute nothing and observed frequencies fluctuate. Richness D_0 is especially sensitive to sampling depth. When communities have different numbers of reads or individuals, one should compare them at a common sampling effort (for example by rarefaction), report uncertainty from resampling or a fitted sampling model, and avoid interpreting a difference in detection depth as a biological difference in diversity.

4.2 Rényi entropies and Hill numbers

The Rényi entropies form a one-parameter family. For $q \geq 0$ and $q \neq 1$,

$$H_q = \frac{1}{1-q} \ln \left(\sum_{i=1}^{N_{\mathrm{sp}}} p_i^q \right), \tag{4.5}$$

and the corresponding Hill number is

$$D_q = \exp(H_q) = \left(\sum_{i=1}^{N_{\mathrm{sp}}} p_i^q \right)^{1/(1-q)}. \tag{4.6}$$

The order q controls sensitivity to abundance. Small q gives relatively more weight to rare species, whereas large q emphasizes common species.

The limit $q \to 1$. Let $S(q) = \sum_i p_i^q$. Because $S(1) = 1$, both the numerator and denominator of Equation (4.5) vanish as $q \to 1$. L'Hôpital's rule gives

$$\lim_{q \to 1} H_q = \lim_{q \to 1} \frac{S'(q)/S(q)}{-1} = -\sum_i p_i \ln p_i = H. \tag{4.7}$$

Hence $D_q \to D_1 = \exp(H)$.

The limit $q \to 0$. For every species with $p_i > 0$, $p_i^0 = 1$. Therefore

$$H_0 = \ln N_+, \qquad D_0 = N_+, \tag{4.8}$$

where N_+ is the number of species actually present. Thus order zero recovers ordinary species richness.

Order $q = 2$: Simpson concentration. At order two,

$$D_2 = \left(\sum_i p_i^2 \right)^{-1}. \tag{4.9}$$

The quantity

$$\lambda = \sum_i p_i^2 \tag{4.10}$$

is the Simpson concentration: the probability that two independent individuals sampled from the community belong to the same species. Its complement $1 - \lambda$ is the Gini–Simpson index, the probability that they belong to different species. Consequently, $D_2 = 1/\lambda$ is small when one or a few species dominate.

The limit $q \to \infty$: dominance and min-entropy. Let $p_{\max} = \max_i p_i$. For large q, the sum in Equation (4.6) is dominated by $p_{\max}^q$, so

$$H_\infty = -\ln p_{\max}, \qquad D_\infty = \frac{1}{p_{\max}}. \tag{4.11}$$

This effective species number depends only on the most abundant species and is closely related to the Berger–Parker index. All orders give $D_q = N_{\mathrm{sp}}$ for a perfectly even community, but they distinguish nonuniform communities in different ways. Noninteger orders, such as $q = 1/2$, interpolate smoothly between richness and dominance.

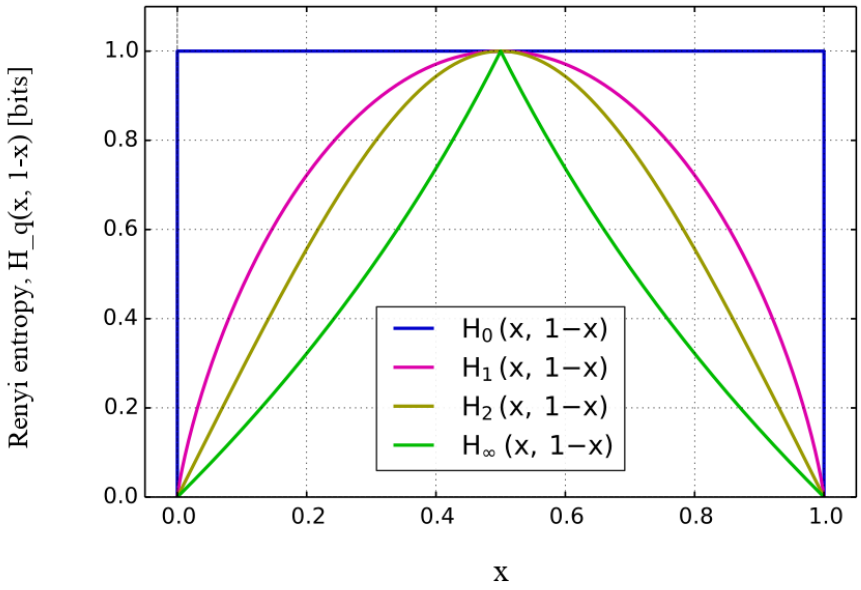

Figure 4.1: Rényi entropies for a binary distribution $P = (x, 1 - x)$. The curves show orders $q = 0, 1, 2$, and ∞. The plot uses base-2 entropy units, for which the maximum binary entropy is one bit (Rényi 1961). Figure from Wikipedia/Wikimedia Commons.

4.3 Rao's quadratic diversity

Abundance-based indices treat every pair of distinct species as equally different. Rao (1982) introduced a quadratic diversity that assigns a dissimilarity d_{ij} to species i and j, with $d_{ii} = 0$:

$$Q = \sum_{i=1}^{N_{\mathrm{sp}}} \sum_{j=1}^{N_{\mathrm{sp}}} p_i p_j d_{ij}. \tag{4.12}$$

This is the expected dissimilarity between two independently sampled individuals. The dissimilarity can describe functional traits, genetic distance, or phylogenetic separation.

If every pair of distinct species is assigned $d_{ij} = 1$, then

$$Q = \sum_{i \neq j} p_i p_j = 1 - \sum_i p_i^2, \tag{4.13}$$

which is exactly the Gini–Simpson index. Rao's index therefore extends that familiar measure by allowing some species pairs to be more distinct than others.

4.4 Similarity-sensitive diversity

An equivalent approach begins with a similarity matrix Z, where $0 \leq Z_{ij} \leq 1$, $Z_{ii} = 1$, and larger values mean that species i and j are more alike. Following Leinster and Cobbold (2012), define the *ordinariness* of species i as

$$\widetilde{p}_i = (Zp)_i = \sum_{j=1}^{N_{\mathrm{sp}}} Z_{ij} p_j. \tag{4.14}$$

A rare species can still have high ordinariness if it resembles one common species or many individually rare species whose combined abundance is large.

For $q \neq 1$, the similarity-sensitive Hill diversity is

$$D_q^Z = \left[\sum_{i=1}^{N_{\mathrm{sp}}} p_i \left(\sum_{j=1}^{N_{\mathrm{sp}}} Z_{ij} p_j \right)^{q-1} \right]^{1/(1-q)}. \tag{4.15}$$

The continuous order-one limit is

$$D_1^Z = \exp\left[-\sum_i p_i \ln \widetilde{p}_i \right]. \tag{4.16}$$

When Z is the identity matrix, $\widetilde{p}_i = p_i$ and Equation (4.15) reduces to the ordinary Hill number. If all species are regarded as identical, $Z_{ij} = 1$ for every pair, then $D_q^Z = 1$.

A two-species example makes the interpolation concrete. Let $p = (1/2, 1/2)$ and

$$Z = \begin{pmatrix} 1 & z \\ z & 1 \end{pmatrix}, \qquad 0 \leq z \leq 1. \tag{4.17}$$

Both species then have ordinariness $(1 + z)/2$, and every order gives

$$D_q^Z = \frac{2}{1 + z}. \tag{4.18}$$

Thus two unrelated, equally abundant species give diversity 2 when $z = 0$, two half-similar species give $4/3$ when $z = 1/2$, and two biologically indistinguishable labels give diversity 1 when $z = 1$.

Similarity can be constructed in several ways:

- **Genetic similarity:** Z_{ij} can measure shared sequence identity.
- **Functional similarity:** for k binary traits, Z_{ij} can be the fraction of traits for which the two species agree.
- **Phylogenetic similarity:** Z_{ij} can measure shared evolutionary history before two lineages diverged.

- **Distance conversion:** a nonnegative distance can be mapped to similarity, for example by $Z_{ij} = e^{-d_{ij}}$.

When only a taxonomic classification is available, a deliberately simple approximation is

$$Z_{ij} = \begin{cases} 1, & \text{the same species,} \\ 0.8, & \text{different species in the same genus,} \\ 0.5, & \text{different genera in the same family,} \\ 0, & \text{otherwise.} \end{cases} \tag{4.19}$$

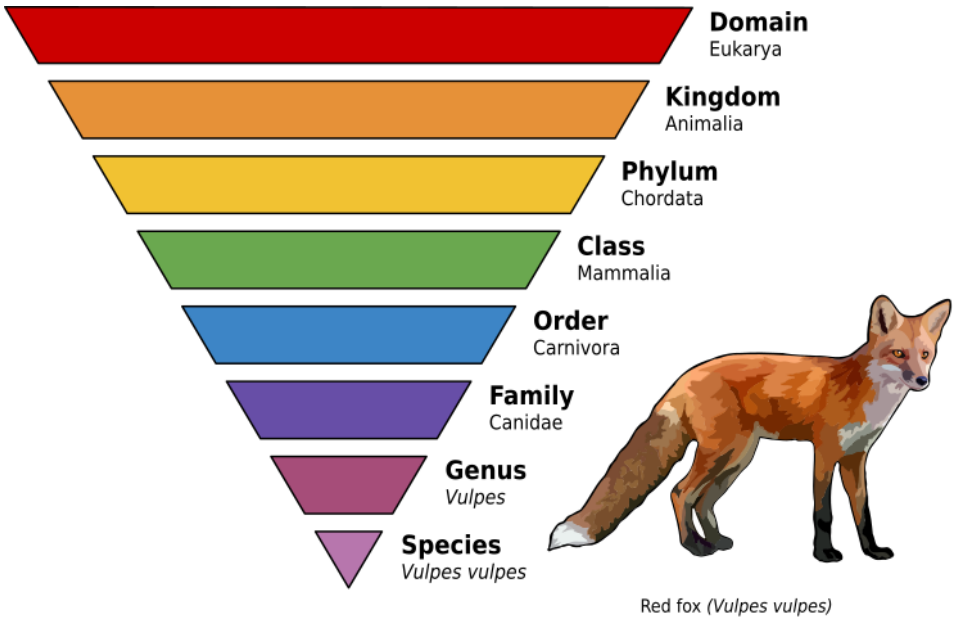

Figure 4.2: The major taxonomic ranks—domain, kingdom, phylum, class, order, family, genus, and species—illustrated for the red fox, *Vulpes vulpes*. Taxonomy supplies a coarse similarity hierarchy when direct functional, genetic, or phylogenetic distances are unavailable. Source: Wikipedia.

Two limiting orders have direct ordinariness interpretations:

$$D_0^Z = \sum_{i:p_i>0} \frac{p_i}{\widetilde{p}_i}, \qquad D_\infty^Z = \frac{1}{\max\limits_{i:p_i>0} \widetilde{p}_i}. \tag{4.20}$$

The ratio $p_i/\widetilde{p}_i$ is large when species i is both present and unusual. The maximum in the infinite-order expression is taken only over species that are present. This diversity becomes small not only when one species is highly abundant, but also when one highly similar cluster dominates the community.

Similarity-sensitive order-two diversity also connects directly to Rao's index. Define $d_{ij} = 1 - Z_{ij}$. Then

$$Q = \sum_{ij} p_i p_j (1 - Z_{ij}) = 1 - \sum_i p_i \widetilde{p}_i, \qquad D_2^Z = \left(\sum_i p_i \widetilde{p}_i \right)^{-1} = \frac{1}{1-Q}. \tag{4.21}$$

Rao's expected dissimilarity and the similarity-sensitive Hill number are therefore two representations of the same order-two information.

4.5 Why the similarity matrix matters in practice

Diversity is a profile D_q, not necessarily a single ranking. Figure 4.3 shows two applications from Leinster and Cobbold (2012): butterfly communities in the canopy and understory of an Ecuadorian rainforest, and gut microbial communities from a lean child (TS1) and an overweight mother (TS3).

For the butterfly data, the naive calculation treats all species as unrelated and ranks the canopy above the understory over most of the displayed range. Once species in the same genus receive nonzero similarity, the profiles cross and the understory can be judged more diverse. The canopy is concentrated in one genus, whereas understory abundance is spread more evenly among genera.

The gut microbiome profiles likewise change after DNA sequence similarity is included. With the naive matrix, the profiles cross near $q \approx 1$, indicating a tradeoff between richness and evenness. With genetic similarity included, the lean-child community has greater effective diversity throughout the plotted range. Absolute diversity values also fall when overlap among taxa is recognized.

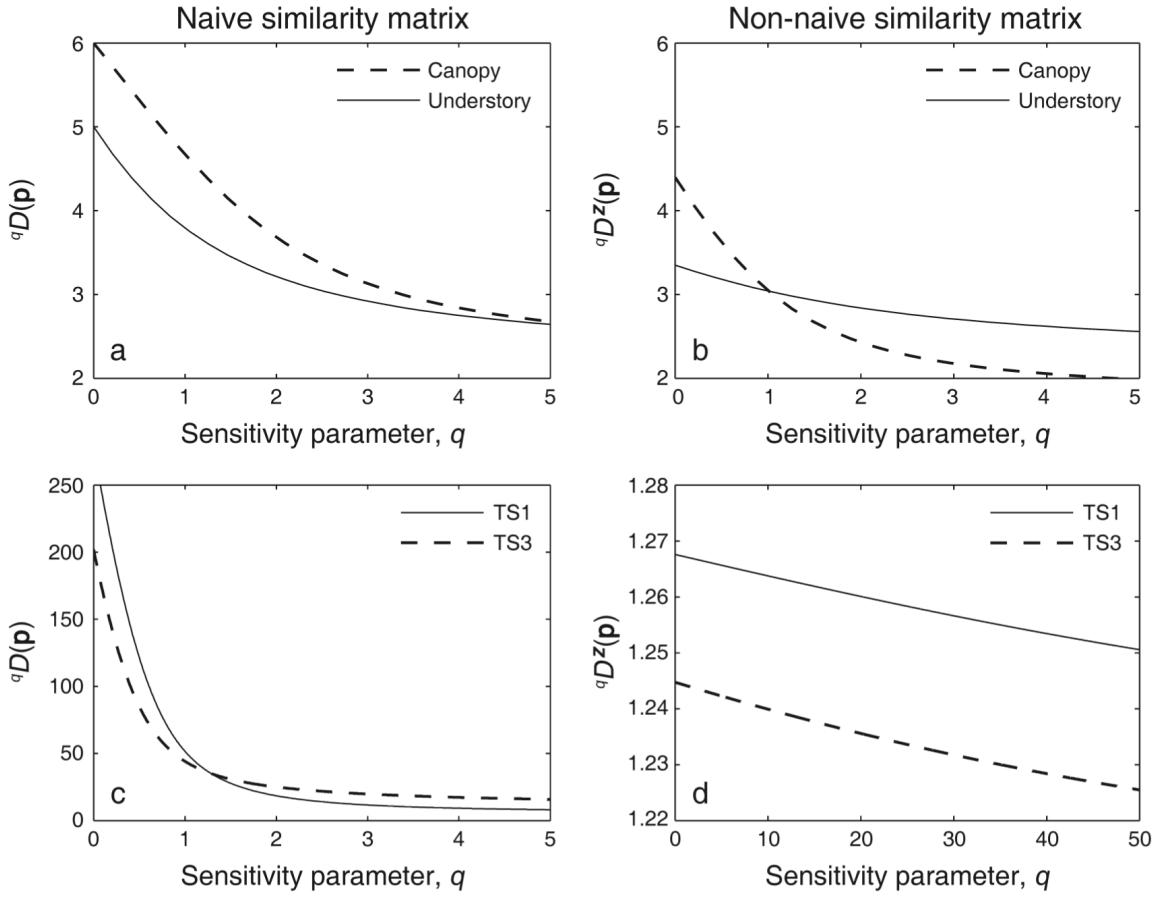

Figure 4.3: Similarity-sensitive diversity profiles. Top: six butterfly species in canopy and understory communities using (a) the naive identity similarity matrix and (b) a taxonomic similarity matrix. Bottom: gut microbiomes from a lean child (TS1) and an overweight mother (TS3) using (c) naive and (d) genetic similarity matrices. Note the different horizontal and vertical scales (Leinster and Cobbold 2012).

4.6 Average nucleotide identity and its limitations

Average nucleotide identity (ANI) estimates genetic similarity between microbial genomes by averaging sequence identity over aligned homologous regions. It is widely used to delineate microbial species, often with thresholds near 95–96%. Its usefulness does not make it a complete diversity metric:

- very dissimilar genomes may not contain enough alignable sequence for a reliable comparison;
- nucleotide identity need not capture functional or ecological differences;
- a sharp species threshold is partly conventional and need not correspond to a sharp phenotypic or ecological boundary; and
- a pairwise average does not by itself represent the full evolutionary history of the organisms.

ANI can supply one component of a similarity matrix, but the biological question should determine whether sequence, function, ecology, or phylogeny is the relevant notion of similarity.

4.7 UniFrac: phylogenetic community dissimilarity

Introduced by Lozupone and Knight (2005), UniFrac compares microbial communities on a phylogenetic tree. Branch lengths represent evolutionary divergence, so the measure can distinguish communities whose species counts are similar but whose evolutionary histories differ.

Unweighted UniFrac. For two communities, mark every branch whose descendant taxa occur in only one community. Then

$$d_{\mathrm{UniFrac}} = \frac{\text{branch length unique to either community}}{\text{total branch length in the combined tree}}. \tag{4.22}$$

This presence–absence version asks what fraction of the represented evolutionary history is exclusive to one community.

Weighted UniFrac. The weighted version additionally multiplies each branch length by the difference in the relative abundances of its descendant taxa between the two communities. It therefore responds both to phylogenetic separation and to how much of each lineage is present. More explicitly, let L_b be the length of branch b, and let A_b and B_b be the relative abundances in communities A and B descending from that branch. One common normalized convention is

$$d_{\mathrm{wUF}}(A, B) = \frac{\sum_b L_b |A_b - B_b|}{\sum_b L_b (A_b + B_b)}. \tag{4.23}$$

The literature also uses an unnormalized numerator and related normalizations, so the chosen convention should be stated when values are compared across analyses.

As a simple unweighted example, suppose a combined tree has three branches: a branch of length 0.3 unique to community 1, a shared branch of length 0.5, and a branch of length 0.2 unique to community 2. The total length is 1.0, the unique length is $0.3 + 0.2 = 0.5$, and hence

$$d_{\mathrm{UniFrac}} = \frac{0.5}{1.0} = 0.5. \tag{4.24}$$

Half of the evolutionary history in the combined tree is unique to one of the communities. UniFrac distances are often analyzed using principal-coordinate analysis or clustering to visualize relationships among many microbial communities.

5 Using continuous distributions

Continuous distributions describe measurements such as fluorescence intensity, protein abundance, cell size, and phase in a biological cycle. This week develops the probability density function, introduces several important continuous distributions, and then uses flow cytometry to motivate covariance, multivariate Gaussians, and principal component analysis. The final part of the week studies how probability densities change under transformations and how the same idea can be used to draw samples from a specified distribution.

By the end of this week, students should be able to answer:

- How does a probability density differ from a probability?
- Why is differential entropy coordinate dependent, and which information quantities remain invariant?
- How do covariance and correlated Gaussian fluctuations arise in a measurement?
- What do the eigenvectors and eigenvalues of a covariance matrix tell us?
- How does a nonlinear change of variables reshape a density?
- How can a uniform random number be transformed into a draw from a chosen distribution?

5.1 Probability density functions

Let N_{tot} independent measurements of a continuous variable X be collected, and let ΔN of them fall in the interval $[x_0 - \Delta x/2, x_0 + \Delta x/2]$. The probability density is

$$p_X(x_0) = \lim_{\Delta x \to 0} \left(\lim_{N_{\mathrm{tot}} \to \infty} \frac{\Delta N}{N_{\mathrm{tot}} \Delta x} \right). \tag{5.1}$$

For a sufficiently small interval,

$$\Pr\left(x_0 - \frac{\Delta x}{2} \le X \le x_0 + \frac{\Delta x}{2} \right) \simeq p_X(x_0) \Delta x. \tag{5.2}$$

The density is normalized according to

$$p_X(x) \ge 0, \qquad \int_{-\infty}^{\infty} p_X(x)\, dx = 1. \tag{5.3}$$

A density is not itself a probability. If x has physical units, then p_X has the reciprocal units, and its numerical value can exceed one. Only an integral such as $\int_a^b p_X(x)\, dx$ is a dimensionless probability bounded by one.

Estimating a density from data. A histogram estimates p_X by dividing a count by both the number of observations and the bin width. Wide bins hide structure, while narrow bins reveal sampling noise. Logarithmically spaced bins require particular care because their widths are unequal. Any conclusion that depends on a histogram should therefore be checked across several reasonable bin choices.

The cumulative distribution function

$$F_X(x) = \Pr(X \le x) = \int_{-\infty}^{x} p_X(u)\, du \tag{5.4}$$

is often easier to estimate: it requires no arbitrary bin edges and every observation contributes directly. Where F_X is differentiable, $p_X(x) = dF_X/dx$.

5.2 The uniform distribution and differential entropy

The uniform density on the interval $[x_{\min}, x_{\max}]$ is

$$p_{\text{unif}}(x) = \begin{cases} \dfrac{1}{x_{\max} - x_{\min}}, & x_{\min} \le x \le x_{\max}, \\ 0, & \text{otherwise.} \end{cases} \tag{5.5}$$

All equal-width subintervals of the same size have equal probability. Among densities supported on a fixed bounded interval, the uniform density has maximum differential entropy. This is a preview of the general constrained maximum-entropy calculation in Week 7; here we use the result only to motivate differential entropy.

For a continuous variable, differential entropy is

$$h(X) = -\int p_X(x) \ln p_X(x)\, dx. \tag{5.6}$$

If X has physical units, the logarithm of the dimensionful density is shorthand for a choice of reference measure. Introducing a reference scale x_0 with the same units as X gives the explicitly dimensionless form

$$h_{x_0}(X) = -\int p_X(x) \ln[p_X(x) x_0]\, dx. \tag{5.7}$$

For a uniform interval of length L, this is $h_{x_0} = \ln(L/x_0)$. The usual formula $h = \ln L$ therefore presumes that numerical values are being expressed in a fixed unit; changing that unit changes the entropy. For a uniform variable on an interval of length L,

$$h(X) = \ln L. \tag{5.8}$$

Thus $h = 0$ for a uniform density on $[0, 1]$, whereas $h = -\ln 2$ on $[0, 1/2]$. Unlike discrete Shannon entropy, differential entropy can be negative.

More importantly, it depends on the coordinate. If $Y = aX$ with $a > 0$, then

$$p_Y(y) = \frac{1}{a} p_X\left(\frac{y}{a}\right) \quad \Longrightarrow \quad h(Y) = h(X) + \ln a. \tag{5.9}$$

Merely changing units from meters to centimeters shifts the entropy by a constant. Kullback–Leibler divergence and mutual information are invariant under smooth one-to-one changes of variables. Differential entropy is not: even differences of differential entropies are not generally invariant under a nonlinear change of variables. Such comparisons therefore require a fixed coordinate system, or more generally a fixed reference measure. Under a common linear rescaling the same additive shift appears in both entropies and cancels, but that cancellation does not extend to arbitrary nonlinear transformations.

5.3 Gaussian and von Mises distributions

5.3.1 The Gaussian distribution

A Gaussian variable is described by a mean μ and a variance σ^2:

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \qquad \mathbb{E}[X] = \mu, \quad \operatorname{Var}(X) = \sigma^2. \tag{5.10}$$

Its differential entropy is

$$h(X) = \frac{1}{2}\ln\bigl(2\pi e\sigma^2\bigr) = \ln\sigma + \frac{1}{2}\ln(2\pi e). \tag{5.11}$$

Every reasonable measure of the width of a fixed-shape Gaussian peak is therefore a constant times σ. For example, the full width at half maximum is $2\sqrt{2\ln 2}\,\sigma$.

5.3.2 A Gaussian-like distribution on a circle

Angles and phases require a periodic distribution. The von Mises density is

$$p(\theta \mid \mu, \kappa) = \frac{\exp[\kappa\cos(\theta-\mu)]}{2\pi I_0(\kappa)}, \qquad -\pi \le \theta < \pi, \tag{5.12}$$

where I_0 is the modified Bessel function of the first kind. The parameter μ sets the preferred angle and κ is a concentration parameter. When $\kappa = 0$, the density is uniform on the circle. For large κ, expanding $\cos(\theta-\mu)$ near its maximum gives a Gaussian with approximate variance $1/\kappa$.

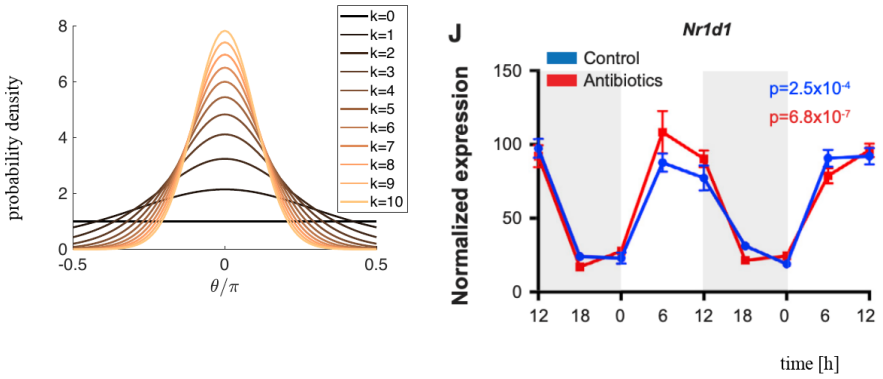

Figure 5.1: The von Mises distribution describes circular data, while rhythmic gene expression provides a biological example of phase-dependent observations. Left: the von Mises family changes continuously from a uniform circular density at $\kappa = 0$ to a sharply localized, Gaussian-like density at large κ (course-generated). Right: periodic expression of the clock gene *Nr1d1* provides a biological example of phase-dependent data (Thaiss et al. 2016).

5.4 Flow cytometry and fluorescence compensation

Flow cytometry sends cells in a fluid stream past one or more lasers. Scattered light reports physical properties, and fluorescent antibodies or reporters mark molecular features of each cell. Each cell becomes a point in a high-dimensional measurement space.

Different fluorochromes have overlapping emission spectra (Figure 5.4). Consequently, a molecule intended to be read in one detector can contribute signal to another detector. Fluorescence compensation estimates this spillover and subtracts it from the measured channels.

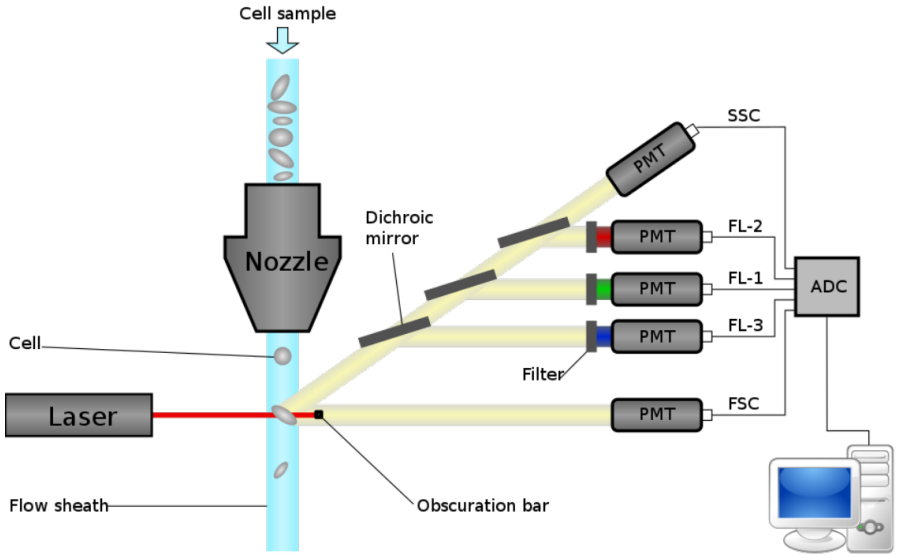

Figure 5.2: Schematic of a flow cytometer. Cells pass the laser one at a time; filters, dichroic mirrors, and photomultiplier tubes route scattered and fluorescent light to separate detectors. Figure from Wikipedia/Wikimedia Commons.

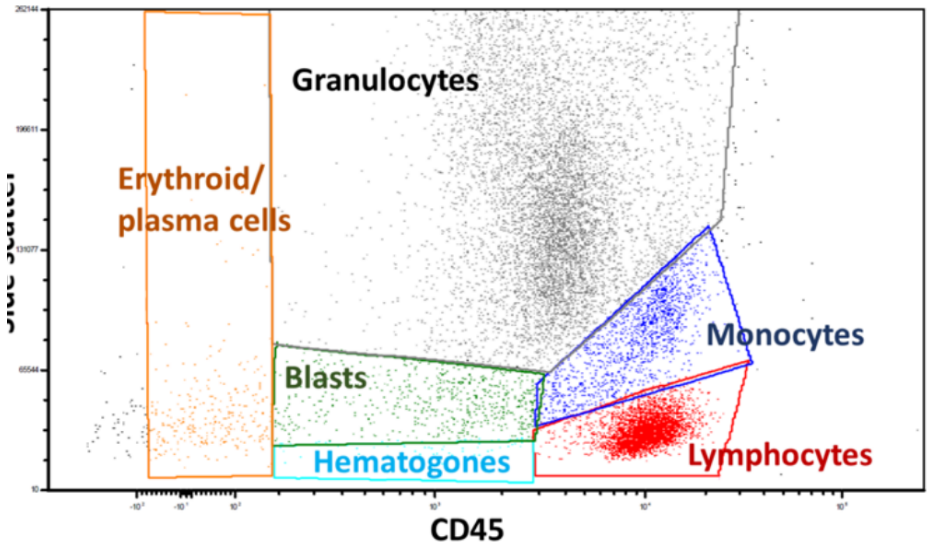

Figure 5.3: A two-channel flow-cytometry measurement. Cell types form partially separated clouds in side-scatter and CD45 coordinates, so the joint distribution contains more structure than either marginal distribution alone. Figure from Wikipedia/Wikimedia Commons.

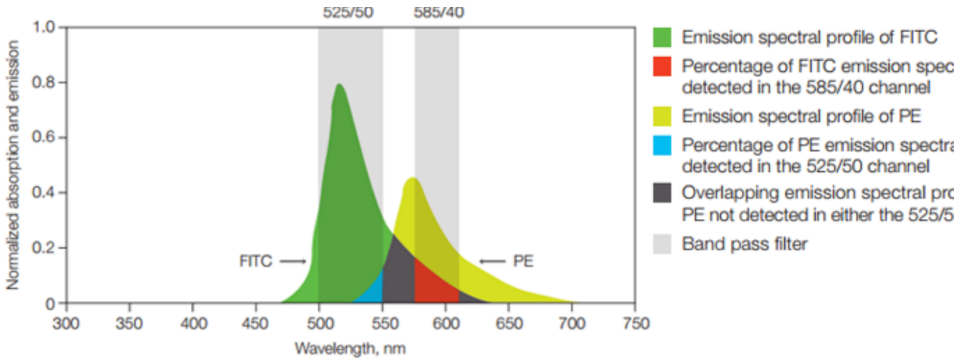

Figure 5.4: Overlapping FITC and PE emission spectra. Each detector integrates light through a finite band-pass filter, so both fluorochromes can contribute to a measured channel (Drescher et al. 2021).

5.4.1 A two-channel mixing model

Let X_0 and Y_0 be the signals that would be measured with perfectly separated channels. A simple symmetric spillover model is

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix} \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix}, \tag{5.13}$$

or

$$X = (1-\alpha)X_0 + \alpha Y_0, \qquad Y = (1-\alpha)Y_0 + \alpha X_0. \tag{5.14}$$

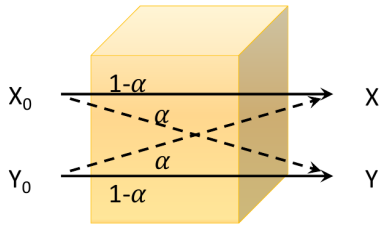

Figure 5.5: Symmetric two-channel spillover model. A fraction $1-\alpha$ remains in the intended detector and a fraction α contributes to the other detector (course-generated).

If X_0 and Y_0 are independent, then

$$\langle X_0 Y_0 \rangle = \langle X_0 \rangle \langle Y_0 \rangle, \tag{5.15}$$

and the mixing alone creates covariance:

$$\mathrm{Cov}(X,Y) = \alpha(1-\alpha)\left[\mathrm{Var}(X_0) + \mathrm{Var}(Y_0)\right], \tag{5.16}$$

$$\mathrm{Var}(X) = (1-\alpha)^2 \mathrm{Var}(X_0) + \alpha^2 \mathrm{Var}(Y_0). \tag{5.17}$$

An analogous expression holds for $\mathrm{Var}(Y)$. Thus observed correlation need not imply biological interaction; it can be produced by the instrument.

Compensation applies the inverse of the mixing matrix. When $\alpha \neq 1/2$,

$$\begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} = \frac{1}{1-2\alpha} \begin{pmatrix} 1-\alpha & -\alpha \\ -\alpha & 1-\alpha \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}. \tag{5.18}$$

The subtraction can produce negative compensated values, especially when a true signal is small relative to spillover and measurement noise. Fluorescence intensity is proportional to antibody number only after channel-specific gains and molecular brightnesses are calibrated; these proportionality constants can differ widely.

5.5 The multivariate Gaussian

For a two-dimensional measurement, define the mean vector and covariance matrix

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}, \tag{5.19}$$

where

$$\rho = \frac{\operatorname{Cov}(X,Y)}{\sigma_X \sigma_Y} \tag{5.20}$$

is the correlation coefficient. In k dimensions, the covariance matrix must be positive definite (and hence nonsingular) for the multivariate Gaussian to have an ordinary full-dimensional density,

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{k/2} \det(\boldsymbol{\Sigma})^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu}) \right]. \tag{5.21}$$

The inverse covariance matrix sets the quadratic form and therefore the orientation and widths of equal-density ellipsoids. A positive-semidefinite but singular covariance matrix instead defines a degenerate Gaussian supported on a lower-dimensional subspace; it has no density with respect to full k-dimensional volume.

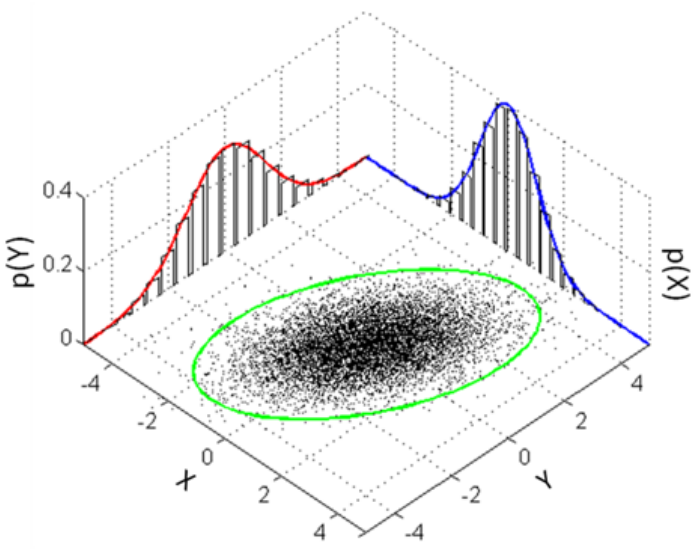

Figure 5.6: A correlated bivariate Gaussian. The green contour encloses an elliptical joint distribution; the red and blue curves are its one-dimensional marginals (course-generated).

The differential entropy of a k-dimensional Gaussian is

$$h(\mathbf{X}) = \frac{k}{2}\ln(2\pi e) + \frac{1}{2}\ln\det(\mathbf{\Sigma}). \tag{5.22}$$

For two Gaussian variables, the density may be written explicitly as

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{z}{2(1-\rho^2)}\right], \tag{5.23}$$

$$z = \frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}. \tag{5.24}$$

Their mutual information depends only on the magnitude of the correlation:

$$I(X;Y) = -\frac{1}{2}\ln(1-\rho^2). \tag{5.25}$$

This expression diverges as $|\rho| \to 1$, when one ideal continuous Gaussian variable becomes an exact linear function of the other.

5.6 Diagonalization, whitening, and principal components

Because a covariance matrix is symmetric and positive semidefinite, it has an orthonormal eigendecomposition

$$\mathbf{\Sigma} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{\mathsf{T}}, \qquad \mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_k), \tag{5.26}$$

where the columns of $\mathbf{P}$ are eigenvectors. Rotating centered data into principal-component coordinates,

$$\mathbf{y} = \mathbf{P}^{\mathsf{T}}(\mathbf{x} - \boldsymbol{\mu}), \tag{5.27}$$

gives $\operatorname{Cov}(\mathbf{Y}) = \mathbf{\Lambda}$. The Gaussian exponent factorizes:

$$\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] = \prod_{i=1}^{k}\exp\left(-\frac{y_i^2}{2\lambda_i}\right). \tag{5.28}$$

Full whitening includes an additional rescaling:

$$\mathbf{z} = \mathbf{\Lambda}^{-1/2}\mathbf{P}^{\mathsf{T}}(\mathbf{x} - \boldsymbol{\mu}), \qquad \operatorname{Cov}(\mathbf{Z}) = \mathbf{I}. \tag{5.29}$$

Equation (5.29) assumes that every eigenvalue is positive. If the covariance is singular, one must restrict the transformation to principal components with nonzero eigenvalues or replace $\mathbf{\Lambda}^{-1/2}$ by the corresponding pseudoinverse. The eigenvectors rotate the cloud; the positive eigenvalues specify the rescaling. In the simple Gaussian spillover example, this removes observed correlations (Figure 5.7). Recovering the true biological channels, however, additionally requires a calibrated mixing model such as Equation (5.13); covariance alone does not identify the physical source of every correlation.

Principal component analysis. PCA ranks the eigenvectors by decreasing eigenvalue. Keeping the first $m < k$ coordinates produces the rank-m approximation

$$\mathbf{x} \approx \boldsymbol{\mu} + \sum_{i=1}^{m} y_i \mathbf{p}_i. \tag{5.30}$$

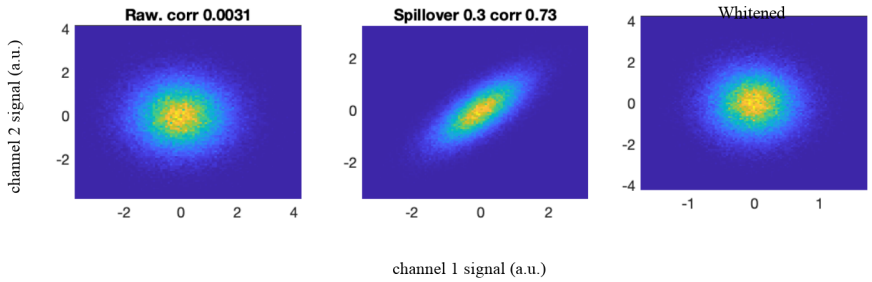

Figure 5.7: A synthetic compensation and whitening example. Independent raw signals (left) become strongly correlated after spillover (middle); a covariance-based linear transformation returns a round, uncorrelated cloud (right; course-generated).

The retained variance is $\sum_{i=1}^{m} \lambda_i$, and the fraction of total variance explained is

$$R_m = \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{k} \lambda_i}. \tag{5.31}$$

PCA therefore provides a principled linear dimension reduction when most variation lies along a small number of directions. It describes variance, not necessarily biological importance or causality.

Because variance carries squared units, covariance-based PCA depends on how each variable is scaled. A channel measured in large numerical units can dominate the first component even when it is not the most biologically informative. Standardizing every variable to unit variance makes PCA equivalent to diagonalizing the correlation matrix, but it also assigns the same initial weight to a precise variable and a noisy one. The choice between raw, calibrated, log-transformed, or standardized coordinates is therefore part of the model and should be reported.

5.7 Transformations of a probability density

Biological measurements often span orders of magnitude. Protein abundance, gene expression, tissue size, and particle size are frequently closer to Gaussian after a logarithmic transformation. A positive variable X is log-normally distributed when

$$\ln X \sim \mathcal{N}(\mu, \sigma^2), \qquad X = e^{\mu+\sigma Z}, \quad Z \sim \mathcal{N}(0,1). \tag{5.32}$$

Its density in the original coordinate is

$$p_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], \qquad x > 0. \tag{5.33}$$

The parameters μ and σ^2 are the mean and variance of $\ln X$, not of X. The corresponding summaries on the original scale are

$$\mathbb{E}[X] = e^{\mu+\sigma^2/2}, \qquad \mathrm{Var}(X) = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}, \tag{5.34}$$

$$\mathrm{median}(X) = e^{\mu}, \qquad \mathrm{mode}(X) = e^{\mu-\sigma^2}. \tag{5.35}$$

The separation of mean, median, and mode is a direct measure of the right skew. One reason log-normal structure is common is that a product of many positive factors becomes a sum after taking logarithms. Under the usual central-limit conditions, that sum can approach a Gaussian. This is a mechanism, not a guarantee: dependence, heavy-tailed factors, and selection can lead to other distributions.

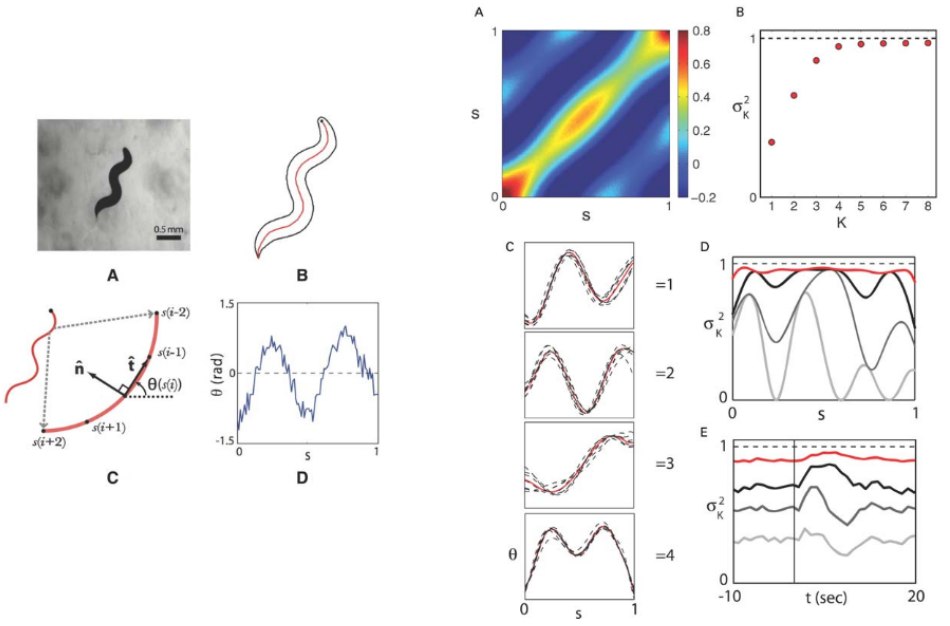

Figure 5.8: An application of PCA to posture dynamics. A small set of shape modes captures most of the variance in a moving worm, and their time-dependent coefficients compactly describe behavior (Stephens et al. 2008).

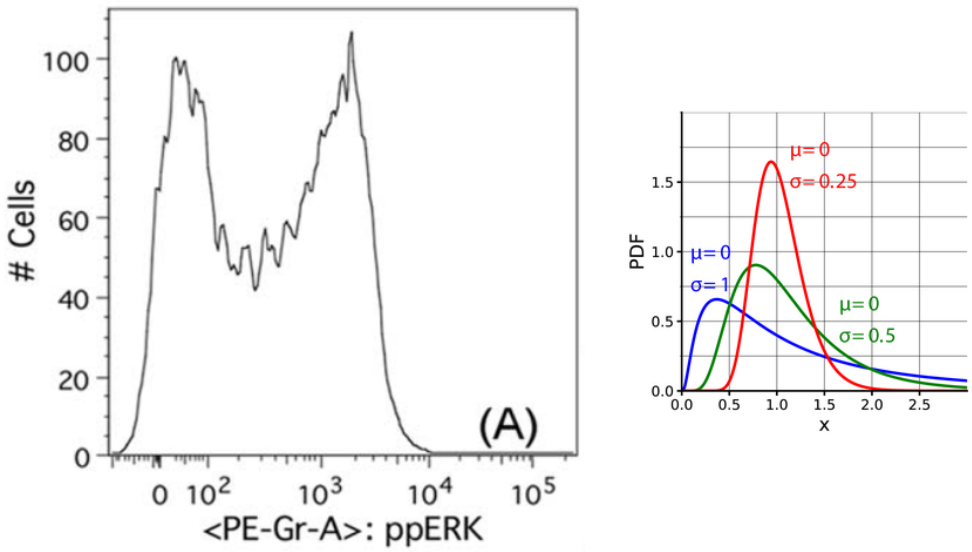

Figure 5.9: Logarithmic coordinates reveal structure in biological measurements that span multiple orders of magnitude. Left: flow-cytometry measurements of ppERK (Erez, Vogel, et al. 2018). Right: log-normal densities with different μ and σ (course-generated). The logarithmic coordinate separates scale changes from changes in shape.

5.7.1 The Jacobian rule

Let $Y = G(X)$, with G differentiable and monotonic. Conservation of probability requires

$$p_Y(y)\,|dy| = p_X(x)\,|dx|. \tag{5.36}$$

Therefore

$$p_Y(y) = p_X\big(G^{-1}(y)\big)\left|\frac{dG^{-1}(y)}{dy}\right| = \left.\frac{p_X(x)}{|G'(x)|}\right|_{x=G^{-1}(y)}. \tag{5.37}$$

The absolute value is essential for decreasing transformations.

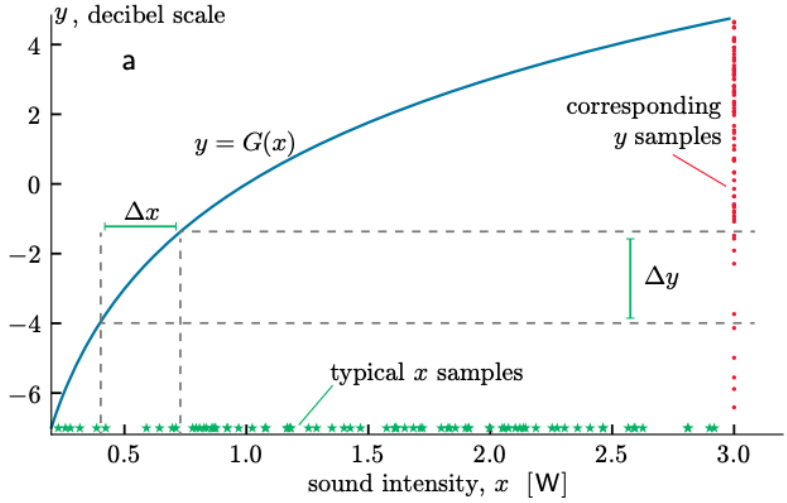

Figure 5.10: A nonlinear transformation $y = G(x)$ changes local bin widths. Equal probability in the corresponding x- and y-intervals produces the Jacobian factor in Equation (5.37) (Nelson 2022).

For $Y = \ln X$, $dY/dX = 1/X$, so

$$p_{\ln X}(\ln x) = x\,p_X(x), \qquad p_X(x) = \frac{p_{\ln X}(\ln x)}{x}. \tag{5.38}$$

The factor $1/x$ changes the relative height of peaks. Consequently, a density can be bimodal in $\ln I$ but unimodal in the original intensity I, or vice versa (Figure 5.11). Peak counting is therefore not invariant under a change of coordinate.

5.8 Inverse-transform sampling

The change-of-variables rule also gives a general method for sampling a desired continuous density. Let

$$F_X(x) = \int_{x_{\min}}^{x} p_X(u)\,du \tag{5.39}$$

be the cumulative distribution on its support. If U is uniform on $(0,1)$, then

$$X = F_X^{-1}(U) \tag{5.40}$$

has density p_X. Indeed, $F_X'(x) = p_X(x)$, so the transformation maps each probability interval in x to an interval of the same length in the uniform coordinate.

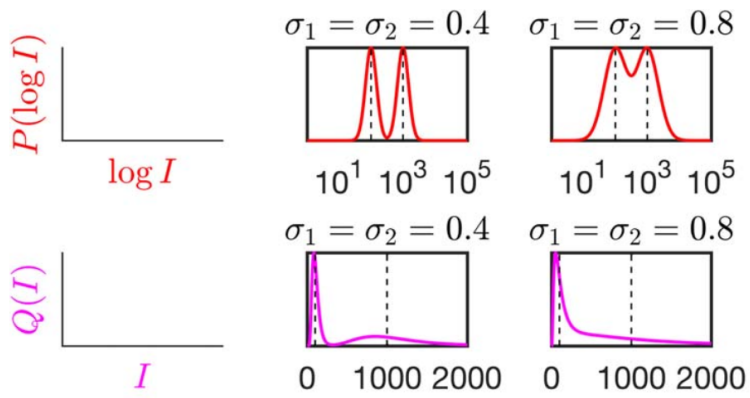

Figure 5.11: Two Gaussian peaks in log intensity (red) transform into densities in linear intensity (magenta). Increasing the log-space widths can merge the two linear-space peaks because the Jacobian weights low intensities more strongly (Erez, Vogel, et al. 2018).

For the exponential density

$$p_X(x) = e^{-x}, \qquad x \geq 0, \tag{5.41}$$

the cumulative distribution is $F_X(x) = 1 - e^{-x}$. Hence

$$X = -\ln(1 - U). \tag{5.42}$$

Because $1 - U$ is also uniform on $(0, 1)$, the equivalent expression $X = -\ln U$ is often used.

An LLM can translate Equation (5.40) into code in any chosen language; the scientific task is to verify the result. Useful checks are that all draws lie in the stated support, the empirical cumulative distribution approaches F_X, and summary quantities such as the exponential mean and variance approach their theoretical values.

6 Model selection and parameter estimation

A physical model predicts more than an average: together with its parameters, it specifies a probability distribution for the possible experimental outcomes. This week develops likelihood as a quantitative measure of agreement between a stochastic model and data. We compare frequentist and Bayesian interpretations, estimate the bias of a coin, use photon statistics to localize a molecular motor, derive least-squares fitting from a Gaussian likelihood, and connect the curvature of the log-likelihood to Fisher information and the Cramér–Rao bound.

By the end of this week, students should be able to answer:

- How should a family of stochastic models be compared with one dataset?
- What is held fixed when a likelihood is maximized?
- How do prior information and model complexity enter an inference?
- Why can one localize a point source much more accurately than the diffraction limit for resolving two sources?
- Under what assumptions is least-squares fitting a maximum-likelihood method?
- How does likelihood curvature set the best possible precision of an estimator?

6.1 Models, parameters, and likelihood

Let a model $\mathcal{M}$ with parameter vector $\boldsymbol{\theta}$ predict a density or mass function

$$p(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M})$$

for a dataset $\mathcal{D}$. After the data have been observed, the same expression, viewed as a function of the parameter while holding the data fixed, is the *likelihood*:

$$L(\boldsymbol{\theta}; \mathcal{D}, \mathcal{M}) \equiv p(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M}). \tag{6.1}$$

A likelihood is not generally a normalized probability density over $\boldsymbol{\theta}$. It ranks the parameter values by how well they predict the observed data.

The maximum-likelihood estimator is

$$\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \mathcal{D}, \mathcal{M}) = \arg\max_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}; \mathcal{D}, \mathcal{M}), \tag{6.2}$$

where

$$\log L(\boldsymbol{\theta}; \mathcal{D}, \mathcal{M}) \equiv \ln[L(\boldsymbol{\theta}; \mathcal{D}, \mathcal{M})] \tag{6.3}$$

is the log-likelihood. Logarithms turn products over independent observations into sums, improve numerical stability, and do not change the maximizing parameter.

Different parameter values can fit different portions of a dataset well (Figure 6.1). The likelihood combines the agreement over all observations into one objective, but a flexible model can also fit noise. Parameter estimation and model selection are therefore related but distinct problems.

6.2 Frequentist and Bayesian viewpoints

Both viewpoints begin with the same sampling model $p(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M})$, but they attach probability to different objects.

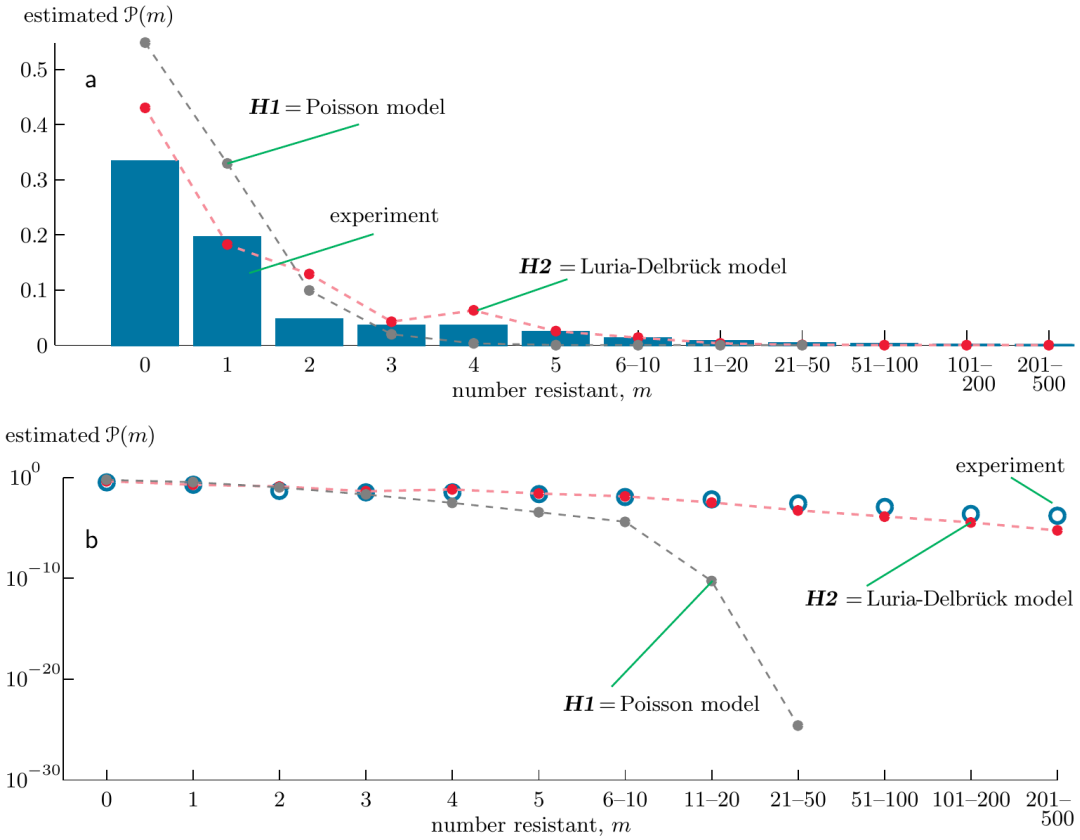

Figure 6.1: Two models compared with the Luria–Delbrück fluctuation data. Agreement over the central range can conceal large discrepancies in the tail. Experimental data are from Luria and Delbrück (1943); fits and plotting follow Nelson (2022).

6.2.1 Frequentist inference

In the frequentist viewpoint, $\boldsymbol{\theta}$ is fixed but unknown and the data are random under repeated realizations of the experiment. Common procedures include:

- maximum-likelihood point estimates;
- confidence intervals constructed to have a stated long-run coverage;
- hypothesis tests that quantify how surprising a statistic is under a null model; and
- model-selection criteria that penalize unnecessary flexibility.

A 95% confidence interval is produced by a procedure that covers the true parameter in 95% of repeated experiments under its assumptions. It is not, after the data have been observed, a statement that the fixed parameter has probability 0.95 of lying in that particular interval.

For a model with k fitted parameters and maximum likelihood $L_{\max}$, two commonly used criteria are

$$\mathrm{AIC} = 2k - 2\ln L_{\max}, \tag{6.4}$$

$$\mathrm{BIC} = k\ln N - 2\ln L_{\max}, \tag{6.5}$$

where N is the number of observations. Smaller values are preferred. These criteria answer specific predictive or asymptotic questions; they are not universal measures of biological truth.

For a numerical comparison, suppose the same $N = 30$ observations are fit by a two-parameter model with $\log L_{\max} = -42$ and a four-parameter model with $\log L_{\max} = -39.5$. Then

	two parameters	four parameters
AIC	$2(2) - 2(-42) = 88$	$2(4) - 2(-39.5) = 87$
BIC	$2\ln 30 - 2(-42) \simeq 90.8$	$4\ln 30 - 2(-39.5) \simeq 92.6$

The extra parameters improve the fit, but only enough for AIC to prefer the larger model slightly; BIC prefers the smaller model. The criteria need not agree because their penalties answer different questions.

AIC or BIC differences are meaningful only for models fit to the same response data with comparable, fully normalized likelihoods. A residual sum of squares cannot be compared directly with a count likelihood, and changing variables requires the appropriate Jacobian in the likelihood. The usual BIC expression also treats N as the number of effectively independent observations; strong dependence or hierarchical sampling makes that choice less automatic.

6.2.2 Bayesian inference

In the Bayesian viewpoint, uncertainty about the parameter is represented by a prior density $p(\boldsymbol{\theta} \mid \mathcal{M})$. Bayes' theorem gives the posterior

$$p(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}) = \frac{p(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M}) p(\boldsymbol{\theta} \mid \mathcal{M})}{p(\mathcal{D} \mid \mathcal{M})}. \tag{6.6}$$

The denominator

$$p(\mathcal{D} \mid \mathcal{M}) = \int p(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M}) p(\boldsymbol{\theta} \mid \mathcal{M})\, d\boldsymbol{\theta} \tag{6.7}$$

is the *marginal likelihood* or *evidence*. It normalizes the posterior and averages the fit over the parameter values allowed by the prior. A model that fits only inside a very narrow portion of a

large parameter space can therefore have less evidence than a model that predicts the data well over a broader region. This is the Bayesian form of an Occam penalty.

For a discrete set of competing models,

$$p(\mathcal{M}_j \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid \mathcal{M}_j)p(\mathcal{M}_j)}{\sum_i p(\mathcal{D} \mid \mathcal{M}_i)p(\mathcal{M}_i)}. \tag{6.8}$$

The prior can encode physical constraints, previous measurements, or symmetry. It must also be stated and tested, since the posterior can depend strongly on the prior when the dataset is small.

6.2.3 When the distinction matters

Likelihood and posterior answer different questions:

$$p(\mathcal{D} \mid \mathcal{M}) \neq p(\mathcal{M} \mid \mathcal{D}).$$

The distinction is especially important for rare events, where a base rate can dominate a diagnostic posterior; for models with different flexibility; for parameters with physical boundaries such as positive rate constants; and when evidence is combined sequentially. In sequential Bayesian inference, today's posterior becomes tomorrow's prior.

In practice, scientific analyses often combine tools: an MLE for a point estimate, a frequentist goodness-of-fit test, and a Bayesian sensitivity analysis can all be useful. The choice should follow the scientific question rather than a preference for one label.

6.3 Bernoulli parameter estimation

Suppose M independent Bernoulli trials produce ℓ successes. If the success probability is ξ, then

$$L(\xi; \ell, M) = \Pr(\ell \mid \xi, M) = \binom{M}{\ell} \xi^{\ell}(1-\xi)^{M-\ell}. \tag{6.9}$$

The data ℓ and M are now fixed; ξ is the variable over which the likelihood is optimized. Terms that do not depend on ξ can be omitted:

$$\log L(\xi; \ell, M) = \text{constant} + \ell \ln \xi + (M-\ell)\ln(1-\xi). \tag{6.10}$$

For $0 < \ell < M$,

$$0 = \frac{d \log L}{d\xi} = \frac{\ell}{\xi} - \frac{M-\ell}{1-\xi} \quad \Longrightarrow \quad \widehat{\xi}_{\mathrm{MLE}} = \frac{\ell}{M}. \tag{6.11}$$

The same maximum is obtained from a uniform prior $p(\xi) = 1$ on $0 \le \xi \le 1$. The normalized posterior is then

$$p(\xi \mid \ell, M) = \frac{\xi^{\ell}(1-\xi)^{M-\ell}}{B(\ell+1, M-\ell+1)}, \tag{6.12}$$

where B is the beta function. Equivalently,

$$\xi \mid \ell, M \sim \mathrm{Beta}(\ell+1, M-\ell+1). \tag{6.13}$$

Its mode, posterior mean, and variance are

$$\xi_{\mathrm{mode}} = \frac{\ell}{M}, \tag{6.14}$$

$$\mathbb{E}[\xi \mid \ell, M] = \frac{\ell+1}{M+2}, \tag{6.15}$$

$$\mathrm{Var}(\xi \mid \ell, M) = \frac{(\ell+1)(M-\ell+1)}{(M+2)^2(M+3)}. \tag{6.16}$$

The mode formula assumes $0 < \ell < M$. The difference between the MLE and posterior mean is largest for small datasets and vanishes as M grows.

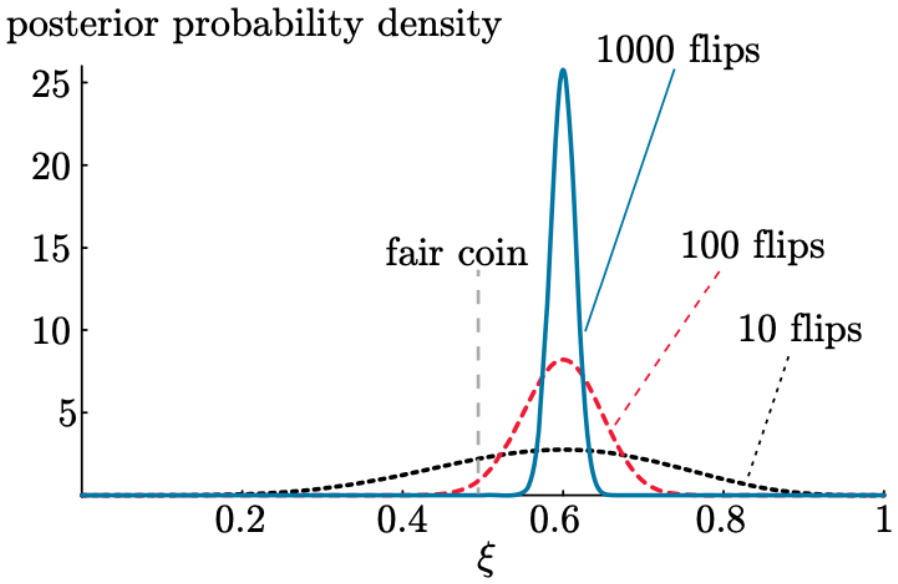

Figure 6.2: Posterior densities for observing a success fraction 0.6 after $M = 10$, 100, and 1000 Bernoulli trials, using a uniform prior. The location remains near 0.6, while the width decreases approximately as $M^{-1/2}$. The dashed line marks the fair-coin value $\xi = 1/2$ (Nelson 2022).

6.3.1 Credible intervals

A Bayesian credible interval $[a, b]$ contains a chosen posterior probability:

$$\int_a^b p(\xi \mid \ell, M)\, d\xi = 1 - \alpha. \tag{6.17}$$

The endpoints can define an equal-tail interval, a highest-posterior-density interval, or another stated convention. These intervals differ slightly for an asymmetric posterior.

Figure 6.3 illustrates a symmetric range $[\widehat{\xi} - \Delta, \widehat{\xi} + \Delta]$. Its contained probability is

$$C(\Delta) = \int_{\widehat{\xi}-\Delta}^{\widehat{\xi}+\Delta} p(\xi \mid \ell, M)\, d\xi. \tag{6.18}$$

Cancer example. Suppose a control population has disease probability $\xi_0 = 0.17$, and 6 of 25 animals exposed to a suspected carcinogen develop the disease. The MLE is $6/25 = 0.24$, but the one-sided exact binomial probability under the control model is

$$\Pr_{\xi_0}(L \geq 6) = \sum_{k=6}^{25} \binom{25}{k} \xi_0^k (1 - \xi_0)^{25-k} \approx 0.24. \tag{6.19}$$

The observed fraction is larger than 0.17, but this dataset alone is not unusual enough to establish a carcinogenic effect by a conventional one-sided test. A complete analysis would also specify

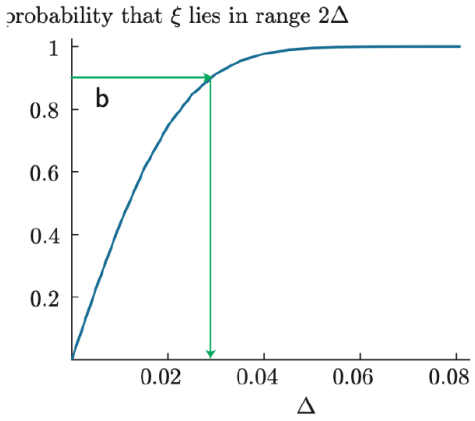

Figure 6.3: Posterior probability contained in a symmetric interval of half-width Δ. Reading horizontally from a desired probability, such as 0.9, gives the corresponding credible width (Nelson 2022).

the control uncertainty, experimental design, and whether the direction of the test was chosen in advance.

6.4 Localization microscopy

The diffraction-limited separation of two point sources is of order

$$d = \frac{\lambda}{2n \sin\theta} = \frac{\lambda}{2\,\mathrm{NA}}, \tag{6.20}$$

where n is the refractive index, θ is the collection half-angle, and $\mathrm{NA} = n \sin\theta$ is the numerical aperture. For visible light and a high-numerical-aperture objective, d is a few hundred nanometers. This prevents two simultaneous nearby sources from being resolved as distinct spots.

Localization asks a different question: if there is one isolated point source, how precisely can the center of its blurred point-spread function be inferred? A wide distribution can have a very precisely estimated mean when many photons are observed.

6.4.1 A Gaussian point-spread model

In one dimension, approximate the detected photon positions as

$$x_i \mid \mu, \sigma \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2), \qquad i = 1, \dots, M, \tag{6.21}$$

where μ is the unknown fluorophore position and σ describes the point-spread width. The likelihood and log-likelihood are

$$L(\mu, \sigma; \mathbf{x}) = \prod_{i=1}^{M} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i - \mu)^2}{2\sigma^2}\right], \tag{6.22}$$

$$\log L(\mu, \sigma; \mathbf{x}) = -\frac{M}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{M} (x_i - \mu)^2. \tag{6.23}$$

Differentiating with respect to μ gives

$$\widehat{\mu}_{\mathrm{MLE}} = \overline{x} = \frac{1}{M}\sum_{i=1}^{M} x_i. \tag{6.24}$$

If σ^2 is also estimated by maximum likelihood,

$$\widehat{\sigma^2}_{\mathrm{MLE}} = \frac{1}{M}\sum_{i=1}^{M}(x_i - \overline{x})^2. \tag{6.25}$$

The denominator is M for the MLE; the unbiased sample-variance estimator instead uses $M - 1$.

The identity

$$\sum_{i=1}^{M}(x_i - \mu)^2 = \sum_{i=1}^{M}(x_i - \overline{x})^2 + M(\mu - \overline{x})^2 \tag{6.26}$$

shows that, for known σ and a locally flat prior,

$$p(\mu \mid \mathbf{x}, \sigma) \propto \exp\left[-\frac{M}{2\sigma^2}(\mu - \overline{x})^2\right]. \tag{6.27}$$

Thus

$$\mathrm{SD}(\mu \mid \mathbf{x}, \sigma) = \frac{\sigma}{\sqrt{M}}. \tag{6.28}$$

Collecting more photons improves localization as $M^{-1/2}$. This does not make the spot itself narrower and does not, by itself, resolve two emitters that are active at the same time.

Real localization models also account for pixelation, background light, detector noise, and the fact that photon counts per pixel are often Poisson distributed. The Gaussian calculation isolates the central statistical principle.

6.4.2 FIONA and molecular-motor steps

FIONA—fluorescence imaging with one-nanometer accuracy—fits the point-spread function of a single fluorophore in each video frame. Yildiz and coauthors used this method to track myosin-V. The inferred trajectory shows plateaus separated by steps near 74 nm, revealing motor motion far below the diffraction-limited spot width (Figure 6.4).

6.4.3 PALM, FPALM, and STORM

To reconstruct a complete super-resolved image, many fluorophores must be localized without their point-spread functions overlapping strongly. Localization-based super-resolution microscopy separates nearby emitters in time:

1. densely label the structure with photo-switchable fluorophores;
2. activate a sparse, random subset so that the bright spots are isolated;
3. fit and localize each active point-spread function;
4. switch or bleach those fluorophores off; and
5. repeat and accumulate the inferred coordinates.

Combining the sparse frames yields a map of molecular positions beyond the conventional diffraction limit. The experimental physics creates identifiable single-source likelihoods; the statistical inference converts each blurred spot into a localized coordinate.

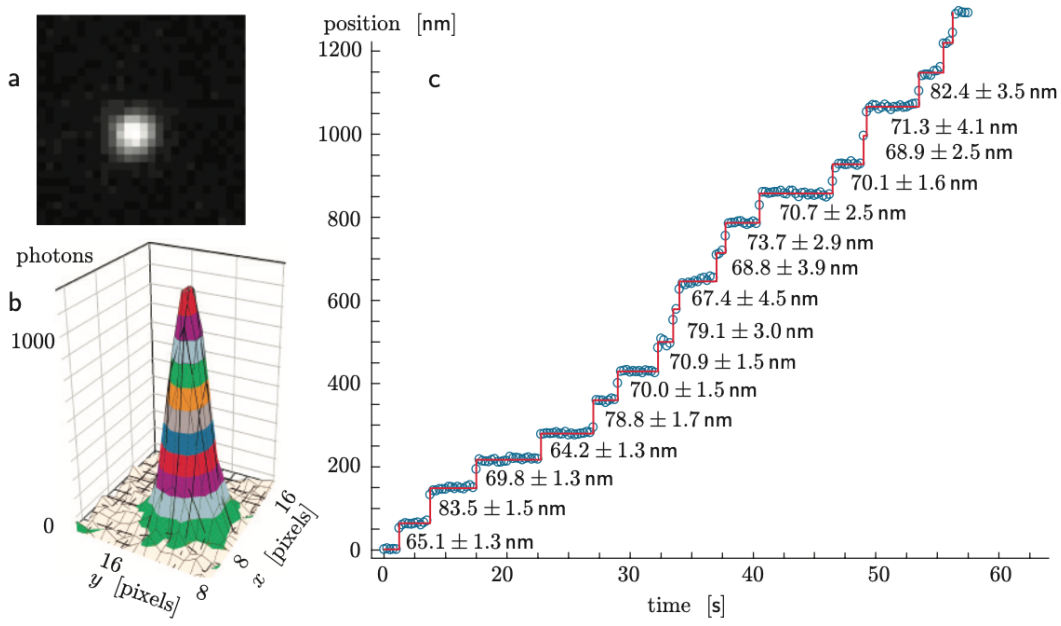

Figure 6.4: FIONA imaging of myosin-V. Panel (a) is one diffraction-limited fluorescent spot, panel (b) shows its point-spread function, and panel (c) shows the localized position over time. The staircase consists of molecular-motor steps near 74 nm (Yildiz et al. 2003).

6.5 Least squares as maximum likelihood

Suppose a physical model predicts

$$y = F(x;\boldsymbol{\theta})$$

and the observations have independent Gaussian errors:

$$y_i = F(x_i;\boldsymbol{\theta}) + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0,\sigma_i^2). \tag{6.29}$$

The likelihood is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(y_i - F(x_i;\boldsymbol{\theta}))^2}{2\sigma_i^2}\right]. \tag{6.30}$$

If the error scales σ_i are known and do not depend on the fitted parameters, maximizing this likelihood is equivalent to minimizing

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{[y_i - F(x_i;\boldsymbol{\theta})]^2}{\sigma_i^2}. \tag{6.31}$$

Ordinary unweighted least squares is the special case in which all σ_i are equal. Weighted least squares follows directly when the variances differ.

For the straight-line model $F(x;A,B) = Ax + B$,

$$\chi^2(A,B) = \sum_{i=1}^{N} \frac{(y_i - Ax_i - B)^2}{\sigma_i^2}. \tag{6.32}$$

The same likelihood logic applies to nonlinear physical models; only the optimization changes.

6.5.1 Goodness of fit

If the model is correct, the σ_i are correct, and the standardized residuals are independent Gaussians, then the minimized statistic is approximately chi-square distributed with

$$\nu = N - k \tag{6.33}$$

degrees of freedom, where k fitted parameters have been estimated from the same data. The reduced chi-square is

$$\chi^2_{\text{red}} \equiv \frac{\chi^2_{\min}}{\nu}. \tag{6.34}$$

It has expectation one and fluctuations of approximate standard deviation $\sqrt{2/\nu}$. To define the upper-tail goodness-of-fit probability, let $X \sim \chi^2_\nu$ be a chi-square random variable with ν degrees of freedom. Then

$$p_{\text{gof}} = \Pr\left(X \geq \chi^2_{\min}\right). \tag{6.35}$$

The p-value compares the unscaled minimized statistic with the unscaled chi-square reference distribution. Writing the reduced statistic as χ^2_{red}, rather than reusing χ^2_ν, prevents these two quantities from being confused. A *small* p_{gof} means that residuals this large would be unusual under the model. Qualitatively:

- $\chi^2_{\text{red}} \gg 1$: the noise may be underestimated, observations may be correlated, or the model may be too simple;
- $\chi^2_{\text{red}} \ll 1$: the noise may be overestimated, correlations may reduce the effective degrees of freedom, or the model may be overfit.

Chi-square calibration is not justified when the residual distribution, independence, or stated variances are wrong. For example, with $N = 22$ observations and $k = 2$ fitted parameters, $\nu = 20$. A minimized value $\chi^2_{\min} \simeq 31.4$ has upper-tail probability about 0.05. This does not mean that the model has a 5% probability of being true. It means that, if the model and noise assumptions were correct, only about 5% of repeated datasets would yield a minimized statistic at least this large.

6.5.2 Unequal measurement precision

Suppose independent measurements satisfy

$$x_i \sim \mathcal{N}(\mu, \sigma_i^2)$$

with known, unequal σ_i. Maximizing the likelihood gives the inverse-variance weighted mean. Up to terms independent of μ,

$$\log L(\mu) = \text{constant} - \frac{1}{2}\sum_i \frac{(x_i - \mu)^2}{\sigma_i^2}, \qquad \frac{d\log L}{d\mu} = \sum_i \frac{x_i - \mu}{\sigma_i^2}. \tag{6.36}$$

Setting the derivative to zero gives

$$\widehat{\mu} = \frac{\sum_i x_i/\sigma_i^2}{\sum_i 1/\sigma_i^2}. \tag{6.37}$$

To obtain its variance explicitly, write $\widehat{\mu} = \sum_i w_i x_i$, where $w_i = (1/\sigma_i^2)/(\sum_j 1/\sigma_j^2)$. Independence then gives

$$\text{Var}(\widehat{\mu}) = \sum_i w_i^2 \sigma_i^2 = \frac{1}{\sum_i 1/\sigma_i^2}. \tag{6.38}$$

Precise measurements carry larger weight, and independent information adds.

6.6 Likelihood curvature and Fisher information

Near a well-behaved maximum, the log-likelihood is approximately quadratic:

$$\log L(\theta) \approx \log L(\widehat{\theta}) - \frac{1}{2} J(\widehat{\theta})(\theta - \widehat{\theta})^2, \tag{6.39}$$

where

$$J(\theta) \equiv -\frac{\partial^2 \log L(\theta)}{\partial \theta^2} \tag{6.40}$$

is the observed information. Large curvature means a narrow peak and therefore a precise estimate; small curvature means that many parameter values predict the data similarly well.

For one observation X drawn from $p(x \mid \theta)$, the score is

$$s(X;\theta) = \frac{\partial}{\partial \theta} \ln p(X \mid \theta). \tag{6.41}$$

Under regularity conditions that allow differentiation through the normalization integral,

$$\begin{aligned} \mathbb{E}_\theta[s(X;\theta)] &= \int p(x \mid \theta) \frac{\partial}{\partial \theta} \ln p(x \mid \theta)\, dx \\ &= \int \frac{\partial p(x \mid \theta)}{\partial \theta}\, dx = \frac{\partial}{\partial \theta} \int p(x \mid \theta)\, dx = 0. \end{aligned} \tag{6.42}$$

Differentiating this zero expectation requires differentiating both the score and the sampling density:

$$0 = \frac{\partial}{\partial \theta} \mathbb{E}_\theta[s] = \mathbb{E}_\theta\left[\frac{\partial s}{\partial \theta}\right] + \mathbb{E}_\theta[s^2]. \tag{6.43}$$

The Fisher information can therefore be written in either equivalent form:

$$I_1(\theta) = \mathbb{E}_\theta\big[s(X;\theta)^2\big] = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial \theta^2} \ln p(X \mid \theta)\right]. \tag{6.44}$$

For n independent observations,

$$I_n(\theta) = n I_1(\theta). \tag{6.45}$$

Thus independent data add information linearly, while typical standard errors decrease as $n^{-1/2}$.

6.6.1 The Cramér–Rao bound

For n observations, let $S_n = \partial \log L / \partial \theta$ be the total score. If an estimator $\widehat{\theta}$ is unbiased, differentiation of $\mathbb{E}_\theta[\widehat{\theta}] = \theta$ gives, under the same regularity conditions,

$$\begin{aligned} 1 &= \frac{\partial}{\partial \theta} \int \widehat{\theta}(\mathbf{x}) p(\mathbf{x} \mid \theta)\, d\mathbf{x} \\ &= \int \widehat{\theta}(\mathbf{x}) p(\mathbf{x} \mid \theta) \frac{\partial \log p(\mathbf{x} \mid \theta)}{\partial \theta}\, d\mathbf{x} = \mathbb{E}_\theta[\widehat{\theta} S_n]. \end{aligned} \tag{6.46}$$

Because $\mathbb{E}_\theta[S_n] = 0$, subtracting the constant θ inside the expectation gives

$$1 = \mathbb{E}_\theta\Big[(\widehat{\theta} - \theta) S_n\Big] = \mathrm{Cov}_\theta(\widehat{\theta}, S_n). \tag{6.47}$$

Cauchy–Schwarz then implies $1 \le \mathrm{Var}(\widehat{\theta})\,\mathrm{Var}(S_n)$, and $\mathrm{Var}(S_n) = I_n(\theta)$. Hence, for any unbiased estimator satisfying the same regularity conditions,

$$\mathrm{Var}_\theta(\widehat{\theta}) \ge \frac{1}{I_n(\theta)}. \tag{6.48}$$

An estimator that attains this lower bound is called *efficient*. The bound is a statement about a specified stochastic model and a specified parameterization. It provides an experiment-design limit even when the best estimator is difficult to derive.

For the Gaussian location model with known σ,

$$I_M(\mu) = \frac{M}{\sigma^2}, \qquad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{M} = \frac{1}{I_M(\mu)}. \tag{6.49}$$

The sample mean is therefore efficient, reproducing the localization precision in Equation (6.28).

6.7 Poisson example

Let

$$X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathrm{Poisson}(\lambda), \qquad p(x_i \mid \lambda) = e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}. \tag{6.50}$$

The joint log-likelihood is

$$\log L(\lambda; \mathbf{x}) = -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\ln\lambda - \sum_{i=1}^{n}\ln(x_i!). \tag{6.51}$$

Its first two derivatives are

$$\frac{\partial \log L}{\partial \lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i, \tag{6.52}$$

$$\frac{\partial^2 \log L}{\partial \lambda^2} = -\frac{1}{\lambda^2}\sum_{i=1}^{n} x_i. \tag{6.53}$$

Here the lowercase x_i are the observed, fixed counts. When the curvature is averaged to calculate Fisher information, they are replaced conceptually by the random variables X_i; since $\mathbb{E}[X_i] = \lambda$,

$$I_n(\lambda) = -\mathbb{E}_\lambda\left[\frac{\partial^2 \log L}{\partial \lambda^2}\right] = \frac{n}{\lambda}. \tag{6.54}$$

The MLE is the sample mean,

$$\widehat{\lambda}_{\mathrm{MLE}} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \tag{6.55}$$

and

$$\mathbb{E}[\overline{X}] = \lambda, \qquad \mathrm{Var}(\overline{X}) = \frac{\lambda}{n} = \frac{1}{I_n(\lambda)}. \tag{6.56}$$

The Poisson sample mean is therefore an efficient unbiased estimator of λ.

Interpreting the information. For the absolute rate parameter λ, the information in one count is $I_1(\lambda) = 1/\lambda$, so the best absolute variance is λ. Relative precision tells a complementary story:

$$\frac{\mathrm{Var}(\widehat{\lambda})}{\lambda^2} \geq \frac{1}{n\lambda}. \tag{6.57}$$

Large expected counts improve fractional precision even though their absolute fluctuations are larger. This difference reflects parameterization: for $\eta = \ln \lambda$, the chain rule gives the general transformation law

$$I_\eta(\eta) = I_\theta(\theta) \left(\frac{d\theta}{d\eta} \right)^2. \tag{6.58}$$

Here $d\lambda/d\eta = \lambda$, so the information per observation is

$$I_1(\eta) = \lambda. \tag{6.59}$$

Fisher information must therefore always be interpreted together with the parameter being estimated.

LLM-assisted implementation. An LLM can implement the log-likelihood, optimization, Hessian, and synthetic-data checks without making programming syntax a course objective. Validation should use log-likelihoods, enforce the parameter domains, compare numerical derivatives with the analytic results, recover known parameters from simulated data, and reproduce the predicted $n^{-1/2}$ uncertainty scaling. The model and its checks—not the generated code—are the transferable scientific products.

6.8 Returning to Luria–Delbrück: choose a stable statistic

Week 3 used the fluctuation experiment to distinguish induced adaptation from spontaneous mutation and derived the zero-class estimator in Section 3; see Equation (3.35). Here the experiment illustrates a general inference lesson. Under the adaptation model, resistant-cell counts are approximately Poisson, so the sample mean efficiently estimates the expected count. Under the mutation model, rare early mutations produce a heavy right tail: one jackpot can dominate the sample mean and variance.

The zero class avoids that unstable tail by recording only the fraction of cultures with no resistant cells. Its estimate is derived in Week 3 and need not be repeated here. The key point is that a statistic well suited to one generative model need not be stable under another.

This example separates two tasks that are sometimes conflated. The Fano factor and the shape of the survivor distribution help select between adaptation and mutation models; once the mutation model is accepted, the zero fraction estimates its mutation-rate parameter. Good inference therefore requires both a plausible mechanism and a statistic whose sampling behavior is stable under that mechanism.

7 Poisson processes and their simulation

Many biological events occur at irregular times even when their long-time average rate is stable. Molecular motors step, receptors bind ligands, photons arrive at a detector, and enzymes release products one event at a time. A Poisson process is the simplest model of such event sequences. It is not merely a distribution of counts: it is a probability model for an entire random set of event times.

This week uses myosin-V as a physical example and develops four complementary descriptions of the same process: Bernoulli trials in short time slots, exponentially distributed waiting times, Poisson-distributed counts, and a random counting trajectory. We then use thinning, superposition, and convolution to analyze more realistic kinetic schemes, derive a diffusion-limited first-passage result, and develop maximum entropy as a complementary route to several familiar distributions.

By the end of this week, students should be able to answer:

- How does a continuous-time Poisson process emerge from repeated Bernoulli trials?
- Why are exponential waiting times and Poisson counts two views of the same process?
- What physical assumptions are encoded by a constant event rate?
- Why do thinning and superposition preserve the Poisson form?
- How can hidden kinetic substeps be inferred from a waiting-time distribution?
- How do absorbing boundaries turn diffusion into a target-finding problem?
- How does constrained entropy maximization generate exponential, geometric, Gaussian, Gamma, and log-normal families?
- How can an LLM help implement and test an event-driven simulation without making programming syntax a course objective?

7.1 A single-molecule machine

Myosin-V is a dimeric motor protein with two motor heads connected by flexible necks and a common tail. It walks processively toward the barbed end of an actin filament while carrying intracellular cargo. The actin track is polar, and neighboring binding sites along the helical repeat are separated by approximately 36 nm. The geometry of the two heads allows one head to remain attached while the other explores possible binding sites (Figure 7.1).

At nanometer scales, the detached head is continually driven by thermal fluctuations. ATP hydrolysis does not eliminate this Brownian motion. Instead, an energy-consuming chemical cycle biases it: favorable forward configurations are captured, backward motion is suppressed, and detailed balance is broken. This is the basic idea of a *Brownian ratchet*. A nonzero mean velocity can therefore emerge from strongly fluctuating microscopic motion because the motor is maintained out of equilibrium.

The mechanical work in one step has the scale

$$W \sim Fd \sim (2\,\mathrm{pN})(36\,\mathrm{nm}) \simeq 7 \times 10^{-20}\,\mathrm{J}, \tag{7.1}$$

which is comparable to the cellular free energy released by hydrolyzing one ATP molecule. Myosin is thus an efficient chemomechanical converter: molecular binding and conformational change are coupled to directed mechanical work.

Single-molecule localization measurements show an irregular staircase (Figure 7.2). The rises have nearly reproducible heights, whereas the horizontal waiting intervals fluctuate widely. Because

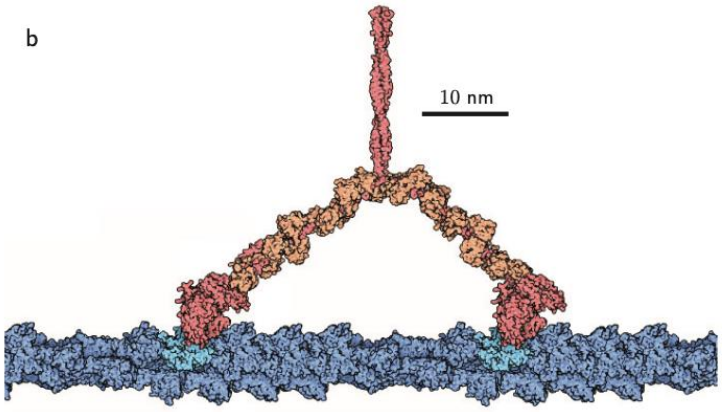

Figure 7.1: Molecular structure of myosin-V spanning two actin-binding sites. The two motor heads act as feet, their lever arms form the legs, and their common junction is connected to the cargo-binding tail. Art by David S. Goodsell (Nelson 2022).

the fluorescent label follows one head, successive visible displacements can be close to 74 nm: the labeled head moves past its partner to the next equivalent binding position. Despite the variable waiting times, the staircase has a well-defined average slope.

A single realization is therefore not summarized by one random number. It is described by an ordered list of event times

$$0 < t_1 < t_2 < \cdots < t_n < \cdots , \tag{7.2}$$

or equivalently by the counting trajectory

$$N(t) = \sum_{\alpha \geq 1} \mathbf{1}_{\{t_\alpha \leq t\}}, \tag{7.3}$$

where $N(t)$ is the number of events observed by time t. A probability law over such time-dependent outcomes is a *stochastic process*.

7.2 From Bernoulli trials to continuous time

7.2.1 Discrete waiting times

First divide time into slots of duration Δt. In each slot, suppose that a step occurs with probability ξ, independently of all other slots. As reviewed in Section 2.4, the number J of slots through the next step is geometric: $\Pr(J = j) = \xi(1 - \xi)^{j-1}$, with mean $1/\xi$. Its memoryless property means that after any run of failed trials, the probability law for the remaining wait is unchanged. We now use this discrete result only as the starting point for a continuous-time limit.

7.2.2 The continuous-time limit

If Δt is made smaller while ξ is held fixed, the expected number of events in a fixed interval diverges. A finite limit requires the success probability to shrink with the slot width:

$$\xi = \beta \Delta t + o(\Delta t), \tag{7.4}$$

where the rate β has dimensions of inverse time. In this limit the artificial time slot disappears; only the physical rate remains.

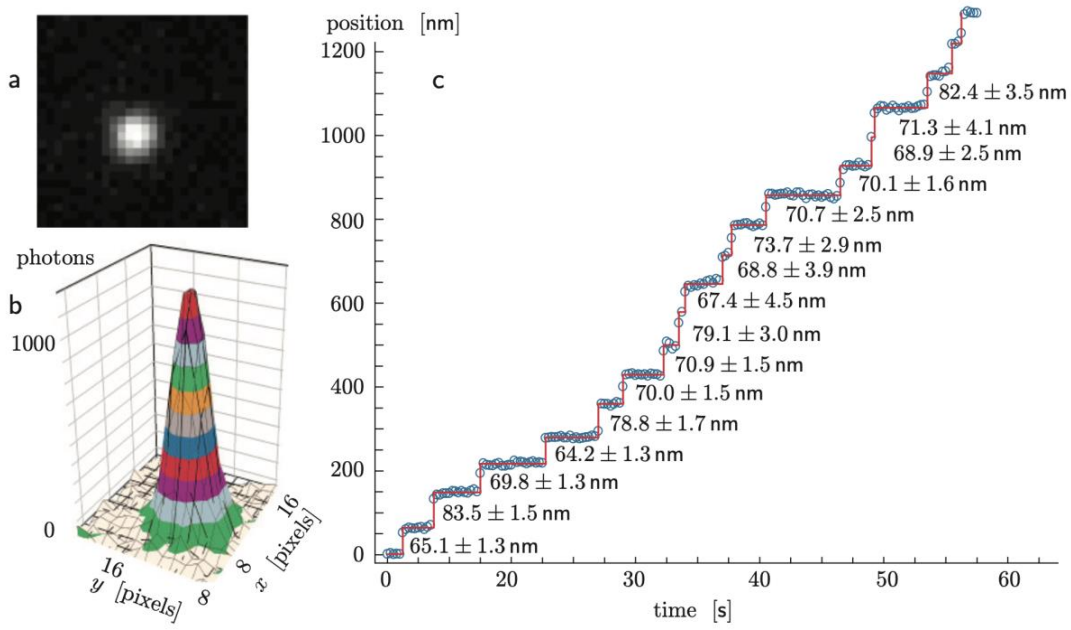

Figure 7.2: Localization of a fluorescently labeled myosin-V head. The position trace is a sequence of abrupt displacements separated by random waiting intervals (Yildiz et al. 2003).

Let $T_w = J\Delta t$ be the waiting time. Its survival probability is

$$\Pr(T_w > t) = (1 - \beta\Delta t)^{t/\Delta t} \longrightarrow e^{-\beta t}. \tag{7.5}$$

Differentiating the survival function gives the exponential density

$$f_{T_w}(t) = \beta e^{-\beta t}, \qquad t \geq 0. \tag{7.6}$$

Its first two moments are

$$\mathbb{E}[T_w] = \frac{1}{\beta}, \qquad \mathbb{E}[T_w^2] = \frac{2}{\beta^2}, \qquad \mathrm{Var}(T_w) = \frac{1}{\beta^2}. \tag{7.7}$$

The exponential distribution is the only continuous distribution with the memoryless property

$$\Pr(T_w > s + t \mid T_w > s) = \Pr(T_w > t). \tag{7.8}$$

Equivalently, its hazard rate is constant:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T_w < t + \Delta t \mid T_w \geq t)}{\Delta t} = \beta. \tag{7.9}$$

The longer the motor has already waited, the more surprising the total elapsed wait becomes, but the instantaneous probability of stepping in the next short interval does not increase.

7.3 Poisson-distributed counts

Consider a time interval of length T, divided into $M = T/\Delta t$ slots. The number of events is binomial with M trials and success probability $\beta\Delta t$. The binomial-to-Poisson limit was derived in Section 3.4. Applying it here as $M \to \infty$ and $\Delta t \to 0$, with $M\Delta t = T$, gives

$$\Pr(N(T) = n) = e^{-\beta T} \frac{(\beta T)^n}{n!}, \qquad n = 0, 1, 2, \ldots. \tag{7.10}$$

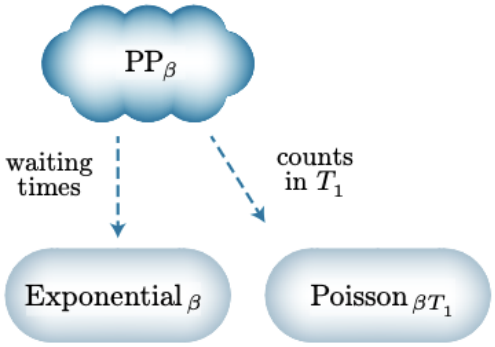

Figure 7.3: Two summaries of one Poisson process. Waiting times are exponentially distributed, while the number of events in a fixed interval is Poisson distributed (Nelson 2022).

Therefore

$$N(T) \sim \text{Poisson}(\beta T), \qquad \mathbb{E}[N(T)] = \text{Var}(N(T)) = \beta T. \tag{7.11}$$

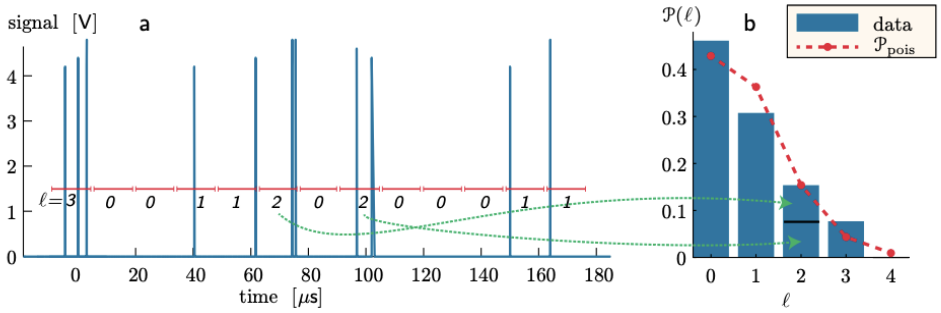

Figure 7.4: Counts of experimental events in equal time bins. The left panel shows the event sequence and binning; the right panel compares the empirical count frequencies with a Poisson distribution (Nelson 2022); data courtesy John F. Beausang.

The *Poisson distribution* in Equation (7.10) describes one random integer. A *Poisson process* describes the entire random event sequence. A homogeneous Poisson counting process is characterized by

1. $N(0) = 0$;
2. independent increments in disjoint time intervals;
3. stationary increments, so the law of $N(t+s) - N(s)$ depends only on t; and
4. the short-time probabilities

$$\begin{aligned} \Pr(\Delta N = 0) &= 1 - \beta \Delta t + o(\Delta t), \\ \Pr(\Delta N = 1) &= \beta \Delta t + o(\Delta t), \\ \Pr(\Delta N \geq 2) &= o(\Delta t). \end{aligned} \tag{7.12}$$

These properties formalize a constant-rate, memoryless stream of isolated events.

If every myosin event advances the observed coordinate by a distance d, then $X(t) = dN(t)$, and

$$\mathbb{E}[X(t)] = d\beta t, \qquad \mathrm{Var}(X(t)) = d^2\beta t. \tag{7.13}$$

The mean trajectory is a straight line with velocity $v = d\beta$, even though every individual trajectory is an irregular staircase.

7.3.1 Estimating the event rate

Suppose n events are observed during a fixed duration T. Apart from factors that do not depend on the rate, the likelihood is

$$L(\beta) \propto \beta^n e^{-\beta T}. \tag{7.14}$$

The maximum-likelihood rate is therefore

$$\widehat{\beta}_{\mathrm{MLE}} = \frac{n}{T}. \tag{7.15}$$

For a genuine homogeneous Poisson process,

$$\mathbb{E}[\widehat{\beta}] = \beta, \qquad \mathrm{Var}(\widehat{\beta}) = \frac{\beta}{T}. \tag{7.16}$$

Longer observation improves the absolute precision as $T^{-1/2}$. More importantly, model validation should compare both the count variance with the count mean and the observed waiting-time distribution with Equation (7.6). Agreement of the mean rate alone is not enough.

7.4 Thinning and superposition

7.4.1 Thinning

Begin with a Poisson process of rate β. Independently retain each event with probability p. In a small interval,

$$\Pr(\text{retained event in } \Delta t) = (\beta\Delta t)p + o(\Delta t), \tag{7.17}$$

so the retained process is Poisson with rate

$$\beta_{\mathrm{thin}} = p\beta. \tag{7.18}$$

The rejected events independently form another Poisson process with rate $(1-p)\beta$. This result explains why random detection loss changes a count rate without changing the Poisson form. Photon arrivals can be scattered, absorbed, or missed by a detector; independent loss simply thins the original stream.

7.4.2 Superposition and event labels

If independent Poisson processes have rates $\beta_1, \ldots, \beta_K$, the probability of any event in a small interval is

$$\sum_{k=1}^{K} \beta_k \Delta t + o(\Delta t). \tag{7.19}$$

Their merged process is therefore Poisson with rate

$$\beta_{\mathrm{tot}} = \sum_{k=1}^{K} \beta_k. \tag{7.20}$$

Conditional on an event having occurred, its type is distributed as

$$\Pr(\text{type} = k \mid \text{an event occurred}) = \frac{\beta_k}{\beta_{\text{tot}}}. \tag{7.21}$$

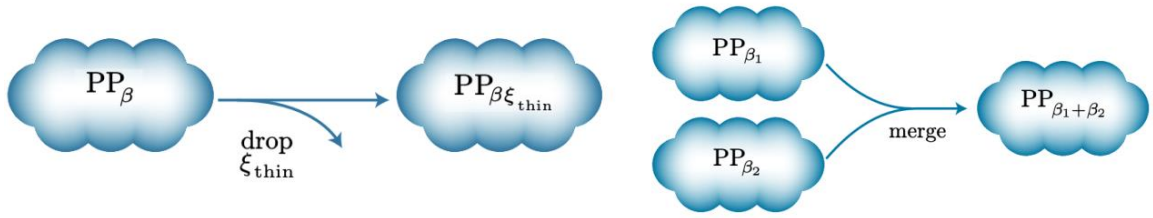

Figure 7.5: Closure properties of Poisson processes. Independent thinning multiplies the rate by the retention probability (left), while merging independent streams adds their rates (right) (Nelson 2022).

These properties have direct biological interpretations. ATP encounters may be approximately Poisson, while only a fraction pass the activation barrier and trigger a motor step; successful encounters form a thinned Poisson process. Several identical enzymes may each release product in independent Poisson streams; the total production process is their superposition.

7.5 Multistep processes and convolution

A single visible event can require several hidden molecular transitions. If two independent steps have waiting-time densities f_1 and f_2, the density of their total duration is the convolution

$$f_{1+2}(t) = \int_0^t f_1(s) f_2(t-s)\, ds. \tag{7.22}$$

For two exponential steps with the same rate β,

$$\begin{aligned} f_2(t) &= \int_0^t \beta e^{-\beta s}\, \beta e^{-\beta(t-s)}\, ds \\ &= \beta^2 t e^{-\beta t}. \end{aligned} \tag{7.23}$$

Unlike an exponential density, this density vanishes at $t = 0$: two transitions cannot be completed instantaneously with appreciable probability.

More generally, the sum of m independent exponential waiting times with common rate β has the Gamma, or Erlang for integer m, density

$$f_m(t) = \frac{\beta^m}{\Gamma(m)} t^{m-1} e^{-\beta t}, \qquad t \geq 0. \tag{7.24}$$

Its mean, variance, and coefficient of variation are

$$\mathbb{E}[T_m] = \frac{m}{\beta}, \qquad \text{Var}(T_m) = \frac{m}{\beta^2}, \qquad \text{CV} = \frac{\sqrt{\text{Var}(T_m)}}{\mathbb{E}[T_m]} = \frac{1}{\sqrt{m}}. \tag{7.25}$$

Thus the relative width of a waiting-time distribution can reveal hidden substeps without first knowing β. An exponential one-step process has CV $= 1$; m identical rate-limiting steps have CV $= m^{-1/2}$.

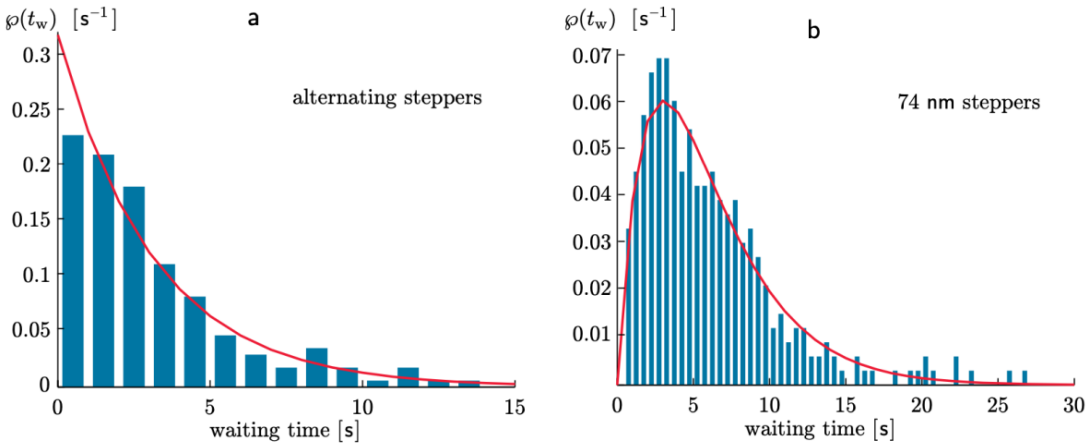

Figure 7.6: Waiting-time densities for two observed myosin-V subpopulations. Alternating short and long visible steps follow an exponential density (left). When only every second physical step is visible, the apparent wait is the sum of two exponential waits and follows a two-step Gamma density (right) (Yildiz et al. 2003).

This result resolved an apparent conflict in the myosin-V experiment. In one subpopulation, the fluorescent marker made alternating short and long displacements visible, and each waiting interval corresponded to one molecular step. In the 74 nm subpopulation, alternate physical steps were invisible because of the marker position. Each observed wait was consequently the sum of two underlying waits. The same motor cycle generated both datasets; only the observation map differed.

The coefficient-of-variation test must be interpreted with care. If two sequential steps have unequal rates $\beta_1 \neq \beta_2$, their total waiting-time density is

$$f(t) = \frac{\beta_1 \beta_2}{\beta_2 - \beta_1} \left(e^{-\beta_1 t} - e^{-\beta_2 t} \right). \tag{7.26}$$

A very slow transition can dominate the total variability and make a multistep process look nearly exponential. Therefore CV^{-2} is an effective number of comparable rate-limiting steps, not automatically the literal number of biochemical reactions.

7.6 First-passage processes: when does diffusion find a target?

Waiting-time models need not describe a molecule changing internal state. They can also describe a molecule diffusing through space until it reaches a target for the first time. The target is an *absorbing boundary*: the trajectory stops when it arrives. This is a different question from asking for the mean-squared displacement at a fixed time. Two walks can spread at the same rate while having very different probabilities of finding a small target.

Consider a diffusing particle in three dimensions, starting a distance $r > a$ from an absorbing sphere of radius a. Let $H(r)$ be the probability that the particle ever hits the sphere. Away from the target, the backward equation is Laplace's equation,

$$\nabla^2 H = 0, \qquad H(a) = 1, \qquad H(r) \longrightarrow 0 \quad \text{as } r \longrightarrow \infty. \tag{7.27}$$

Spherical symmetry gives

$$\nabla^2 H = \frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dH}{dr} \right) = 0. \tag{7.28}$$

Integrating once gives $r^2 H'(r) = C_1$, and integrating again gives $H(r) = C_2 - C_1/r$. The condition at infinity sets $C_2 = 0$; the absorbing-boundary condition $H(a) = 1$ then sets $-C_1 = a$. Therefore

$$H(r) = \frac{a}{r}. \tag{7.29}$$

The probability is less than one because a three-dimensional random walk can escape to infinity. This contrasts with one-dimensional diffusion, where a point target is eventually reached with probability one.

The same calculation gives a diffusion-limited reaction rate. If the concentration far from an absorbing sphere is c_∞, the steady concentration profile is

$$c(r) = c_\infty \left(1 - \frac{a}{r}\right). \tag{7.30}$$

Here

$$\left.\frac{dc}{dr}\right|_{r=a} = \frac{c_\infty}{a}, \qquad j_{\text{in}} = D \left.\frac{dc}{dr}\right|_{r=a} = \frac{D c_\infty}{a}, \tag{7.31}$$

where j_{in} is the positive inward flux magnitude; the vector Fick flux is $-D\boldsymbol{\nabla} c$. Multiplying by the surface area gives

$$J = 4\pi a^2 j_{\text{in}} = 4\pi D a\, c_\infty, \qquad k_{\text{Smol}} = \frac{J}{c_\infty} = 4\pi D a. \tag{7.32}$$

The rate constant is proportional to the target radius, not its cross-sectional area. A diffusing molecule can approach the target from every direction and can make repeated encounters.

Repeated encounters also explain why a cell need not cover its entire surface with receptors. For a spherical cell of radius a carrying N_R receptors of radius s, the sparse-receptor calculation of Berg and Purcell (1977) gives the capture efficiency

$$\eta = \frac{N_R s}{\pi a + N_R s}, \qquad k_{\text{cell}} = 4\pi D a\, \eta. \tag{7.33}$$

With $a = 1\,\mu\text{m}$, $s = 10\,\text{nm}$, and $N_R = 10^4$, the formula gives $\eta \simeq 0.970$, while the geometric area fraction $N_R s^2/(4a^2) \simeq 0.25$. A genuinely sparse illustration is $a = 1\,\mu\text{m}$, $s = 1\,\text{nm}$, and $N_R = 8 \times 10^3$: the receptors then cover about 2×10^{-3} of the surface yet $\eta \simeq 0.718$. A heuristic repeated-encounter argument in the same paper replaces πa by $4a$; its close numerical agreement with the sparse-receptor result is useful for intuition but is not the same calculation. Geometry and return probability make sparse receptors surprisingly effective.

7.7 Event-driven simulation

The time-slot construction is useful for derivations but inefficient for simulation: most small slots contain no event. A direct simulation draws only the waiting times. If $U \sim \text{Uniform}(0, 1)$, inverse-transform sampling gives

$$T_w = -\frac{1}{\beta} \ln U \sim \text{Exponential}(\beta). \tag{7.34}$$

Successive event times are cumulative sums,

$$t_n = \sum_{i=1}^{n} T_{w,i}. \tag{7.35}$$

For K competing event types, one can use the superposition property:

1. calculate $\beta_{\mathrm{tot}} = \sum_k \beta_k$;
2. draw the next waiting time from Exponential(β_{tot});
3. choose event type k with probability $\beta_k/\beta_{\mathrm{tot}}$; and
4. update the physical state and repeat.

When the rates depend on the current molecular populations, they are recalculated after every event. This is the conceptual core of the Gillespie stochastic simulation algorithm used for chemical reaction networks.

An LLM can translate these equations into a short simulation and plotting workflow, while the student remains responsible for the scientific specification and the checks. A valid implementation should recover $\mathbb{E}[T_w] = 1/\beta$, $\mathrm{Var}(T_w) = 1/\beta^2$, equal mean and variance for counts in fixed windows, the rate $p\beta$ after thinning, the summed rate after merging, and $\mathrm{CV} = m^{-1/2}$ for an m-step Erlang process. These invariant predictions are more transferable than any particular generated code.

7.8 Maximum entropy and waiting-time models

The maximum-entropy principle offers a complementary interpretation of familiar distributions. For a continuous density $p(x)$, the differential entropy is

$$h[p] = -\int p(x) \ln p(x)\, dx. \tag{7.36}$$

Suppose that the support $\mathcal{S}$ is known and that the available information consists of expectation constraints

$$\mathbb{E}[f_k(X)] = \int_{\mathcal{S}} p(x) f_k(x)\, dx = c_k. \tag{7.37}$$

Introduce a multiplier for normalization and one for every constraint:

$$\mathcal{L}[p] = -\int_{\mathcal{S}} p(x) \ln p(x)\, dx - \lambda_0 \left(\int_{\mathcal{S}} p(x)\, dx - 1 \right) - \sum_k \lambda_k \left(\int_{\mathcal{S}} p(x) f_k(x)\, dx - c_k \right). \tag{7.38}$$

To see how the functional is optimized, write $p_\varepsilon(x) = p(x) + \varepsilon \eta(x)$, where $\eta(x)$ is an arbitrary small test function. The derivative of the entropy integrand is

$$\frac{d}{d\varepsilon} \left[-p_\varepsilon(x) \ln p_\varepsilon(x) \right] \Big|_{\varepsilon=0} = -\eta(x) \left[\ln p(x) + 1 \right]. \tag{7.39}$$

The constraints are linear in p, so their first variations simply replace $p(x)$ by $\eta(x)$. Therefore

$$\frac{d\mathcal{L}[p_\varepsilon]}{d\varepsilon} \Big|_{\varepsilon=0} = -\int_{\mathcal{S}} \eta(x) \left[\ln p(x) + 1 + \lambda_0 + \sum_k \lambda_k f_k(x) \right] dx. \tag{7.40}$$

Because the multipliers already enforce normalization and the moment constraints, $\eta(x)$ is otherwise arbitrary. The integral can vanish for every η only if the expression in square brackets vanishes pointwise. Thus

$$\ln p(x) = -1 - \lambda_0 - \sum_k \lambda_k f_k(x), \tag{7.41}$$

and exponentiating gives the exponential-family form

$$p(x) = \frac{1}{Z} \exp\left[-\sum_k \lambda_k f_k(x)\right], \tag{7.42}$$

where normalization determines

$$Z(\boldsymbol{\lambda}) = \int_{\mathcal{S}} \exp\left[-\sum_k \lambda_k f_k(x)\right] dx, \qquad e^{-1-\lambda_0} = \frac{1}{Z}. \tag{7.43}$$

The remaining multipliers are fixed by the measured constraints. Differentiating the partition function provides a compact set of equations for them:

$$c_k = -\frac{\partial \ln Z}{\partial \lambda_k}. \tag{7.44}$$

Indeed, differentiating Z brings down $-f_k(x)$, and division by Z turns the integral into an expectation under p.

The stationary density is a maximum rather than a minimum. For a nonzero perturbation that respects the constraints, the second variation of the entropy is

$$\left.\frac{d^2 h[p_\varepsilon]}{d\varepsilon^2}\right|_{\varepsilon=0} = -\int_{\mathcal{S}} \frac{\eta(x)^2}{p(x)}\, dx < 0. \tag{7.45}$$

Thus entropy is strictly concave on positive densities, and the feasible stationary solution is the unique maximum whenever it exists.

7.8.1 Uniform distribution: finite support only

If $X \in [a, b]$ and no information is imposed beyond normalization, there are no nonconstant functions f_k in Equation (7.42). Equation (7.41) then says $p(x) = C$, while normalization gives $1 = C(b-a)$. Hence

$$p(x) = \frac{1}{b-a}, \qquad a \le x \le b. \tag{7.46}$$

Its differential entropy is $h = \ln(b-a)$. Week 5 introduced this uniform density and its entropy; the variational argument here shows why it is selected when finite support is the only information. The finite support matters: a constant density cannot be normalized on the whole real line.

7.8.2 Exponential distribution: nonnegative support and a mean

Let $T \ge 0$ and fix $\mathbb{E}[T] = \mu$. The only nonconstant constraint function is $f(T) = T$, so Equation (7.42) first gives

$$p(t) = Ae^{-\lambda t}, \qquad t \ge 0, \tag{7.47}$$

with $\lambda > 0$ required for normalization. The normalization and mean constraints are

$$1 = A\int_0^\infty e^{-\lambda t}\, dt = \frac{A}{\lambda}, \tag{7.48}$$

$$\mu = A\int_0^\infty t e^{-\lambda t}\, dt = \frac{A}{\lambda^2}. \tag{7.49}$$

The first equation gives $A = \lambda$; substituting this into the second gives $\lambda = 1/\mu$. Therefore

$$p(t) = \frac{1}{\mu} e^{-t/\mu} = \beta e^{-\beta t}, \qquad \beta = \frac{1}{\mu}. \tag{7.50}$$

The exponential density is therefore the least structured nonnegative waiting-time model with a specified mean, relative to the usual time coordinate. This maximum-entropy statement is consistent with, but logically distinct from, the memoryless Poisson-process derivation.

7.8.3 Discrete analogue: the geometric distribution

Week 2 described the trial number $J \in \{1, 2, \ldots\}$ of the first success. Here it is convenient to count the number of failures $N = J - 1 \in \{0, 1, 2, \ldots\}$; the two conventions describe the same geometric family. The maximum-entropy calculation on the nonnegative integers does not produce a Poisson law. For probabilities p_n, maximize $-\sum_{n=0}^{\infty} p_n \ln p_n$ subject to $\sum_n p_n = 1$ and $\sum_n n p_n = \mu$. Varying each p_n gives

$$p_n = A e^{-\lambda n} = A q^n, \qquad q \equiv e^{-\lambda} \in (0, 1). \tag{7.51}$$

Using the geometric-series identities

$$\sum_{n=0}^{\infty} q^n = \frac{1}{1-q}, \qquad \sum_{n=0}^{\infty} n q^n = \frac{q}{(1-q)^2}, \tag{7.52}$$

normalization gives $A = 1 - q$, and the mean constraint gives $\mu = q/(1-q)$, or $q = \mu/(1+\mu)$. Hence

$$p_n = \frac{1}{1+\mu} \left(\frac{\mu}{1+\mu} \right)^n, \qquad n = 0, 1, 2, \ldots. \tag{7.53}$$

This is the geometric distribution for the number of failures before a success. It is the maximum-entropy distribution on $\{0, 1, 2, \ldots\}$ with a fixed mean. The Poisson law, by contrast, enters this course through the independent rare-event mechanism developed earlier in the week.

7.8.4 Gaussian, Gamma, and log-normal distributions

Week 5 introduced Gaussian and log-normal densities as continuous models and derived their change-of-variable relation. We now recover those same forms from specified information, which is a different justification from either a central-limit mechanism or a coordinate transformation.

On the full real line, fixing the mean μ and variance σ^2 is equivalent to fixing $\mathbb{E}[X] = \mu$ and $\mathbb{E}[X^2] = \mu^2 + \sigma^2$. The general solution therefore begins as

$$p(x) = A \exp(-\lambda_1 x - \lambda_2 x^2), \qquad \lambda_2 > 0. \tag{7.54}$$

Completing the square,

$$-\lambda_2 x^2 - \lambda_1 x = -\lambda_2 \left(x + \frac{\lambda_1}{2\lambda_2} \right)^2 + \frac{\lambda_1^2}{4\lambda_2}. \tag{7.55}$$

The last term is independent of x and can be absorbed into the normalization constant. Comparison with a Gaussian centered at μ with variance σ^2 gives

$$-\frac{\lambda_1}{2\lambda_2} = \mu, \qquad \frac{1}{2\lambda_2} = \sigma^2, \tag{7.56}$$

so $\lambda_2 = 1/(2\sigma^2)$ and $\lambda_1 = -\mu/\sigma^2$. Normalizing then gives

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \qquad -\infty < x < \infty. \tag{7.57}$$

The quadratic exponent is therefore a direct consequence of constraining the first and second moments; it is not an independent assumption about the shape.

On positive support, fixing both $\mathbb{E}[T]$ and $\mathbb{E}[\ln T]$ gives a Gamma family. In this case the two constraint functions are t and $\ln t$, and the general solution is

$$p(t) = A \exp[-\lambda_1 t - \lambda_2 \ln t] = A t^{-\lambda_2} e^{-\lambda_1 t}. \tag{7.58}$$

Set $\beta = \lambda_1 > 0$ and $\alpha = 1 - \lambda_2 > 0$. The Gamma-integral identity

$$\int_0^\infty t^{\alpha-1} e^{-\beta t}\, dt = \frac{\Gamma(\alpha)}{\beta^\alpha} \tag{7.59}$$

then fixes $A = \beta^\alpha/\Gamma(\alpha)$, yielding

$$p(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}, \qquad t > 0. \tag{7.60}$$

The two measured constraints determine α and β through

$$\mathbb{E}[T] = \frac{\alpha}{\beta}, \qquad \mathbb{E}[\ln T] = \psi(\alpha) - \ln\beta, \tag{7.61}$$

The logarithmic-moment identity follows by differentiating the normalizing integral

$$I(\alpha,\beta) \equiv \int_0^\infty t^{\alpha-1} e^{-\beta t}\, dt = \frac{\Gamma(\alpha)}{\beta^\alpha}. \tag{7.62}$$

On one hand,

$$\frac{\partial \ln I}{\partial \alpha} = \psi(\alpha) - \ln\beta. \tag{7.63}$$

On the other,

$$\begin{aligned} \frac{\partial \ln I}{\partial \alpha} &= \frac{1}{I}\int_0^\infty t^{\alpha-1} e^{-\beta t} \ln t\, dt \\ &= \int_0^\infty p(t) \ln t\, dt = \mathbb{E}[\ln T]. \end{aligned} \tag{7.64}$$

Here $\psi(\alpha) = d\ln\Gamma(\alpha)/d\alpha$ is the digamma function. These equations usually determine the multipliers numerically from the specified arithmetic and logarithmic means.

In the multistep model of Section 7.5, however, the Gamma law has a stronger mechanistic origin: it is generated by the sum of sequential exponential steps. Maximum entropy explains why a distribution is a minimally committed choice under stated information; convolution explains why it arises from a particular kinetic scheme.

Finally, if $Y = \ln X$ has specified mean m and variance s^2, the Gaussian result above gives

$$p_Y(y) = \frac{1}{s\sqrt{2\pi}} \exp\left[-\frac{(y-m)^2}{2s^2}\right]. \tag{7.65}$$

Since $y = \ln x$, the change-of-variables factor is $|dy/dx| = 1/x$. Therefore $p_X(x) = p_Y(\ln x)/x$, which gives the log-normal density

$$p_X(x) = \frac{1}{xs\sqrt{2\pi}} \exp\left[-\frac{(\ln x - m)^2}{2s^2}\right], \qquad x > 0. \tag{7.66}$$

This model is natural when fluctuations are multiplicative and relative changes matter more than absolute changes. Notice precisely what was maximized: it was $h(Y)$ in the logarithmic coordinate, not $h(X)$ in the original coordinate.

Differential entropy depends on the coordinate and reference measure. A maximum-entropy statement is meaningful only after specifying the variable, its support, and the constraints. Physical modeling supplies those choices; entropy maximization does not choose them for us.

7.9 Relations among common distributions

Probability distributions are connected by sums, transformations, mixtures, special cases, and limiting procedures. These relations are not interchangeable: a sum describes a generative mechanism, a transformation changes the observed variable, and a limit describes an approximation regime. The relations most useful in this course are collected below.

Starting variables	**Operation or limit**	**Resulting distribution**
Bernoulli(p)	Sum of n independent trials	Binomial(n, p)
Binomial(n, p)	$n \to \infty$, $p \to 0$, with $np = \lambda$ fixed	Poisson(λ)
Trial index $J \sim$ Geometric(p)	$T = J\Delta t$, $p = \beta\Delta t$, and $\Delta t \to 0$	Exponential(β)
Independent Exponential(β) waiting times	Sum of m waiting times	Gamma with shape m and rate β (Erlang)
Poisson(λ_i)	Sum of independent counts	Poisson($\sum_i \lambda_i$)
Standard normal Z_i	$\sum_{i=1}^{k} Z_i^2$	Chi-squared with k degrees of freedom
Independent $Z \sim \mathcal{N}(0,1)$, $V \sim \chi^2_\nu$	$Z/\sqrt{V/\nu}$	Student t_ν
Independent $X \sim$ Gamma(α, β) and $Y \sim$ Gamma(γ, β) with a common rate	$X/(X+Y)$	Beta(α, γ)
Normal Y	Transformation $X = e^Y$	Log-normal distribution

Figure 7.7 gives a much broader map. The portion most directly relevant to this course connects Bernoulli, Binomial, Poisson, Geometric, Exponential, Gamma, Normal, Chi-squared, Student t, and Beta families. The arrows encode special cases, transformations, convolution, mixture constructions, and limiting relations; the legend is therefore essential when reading the diagram.

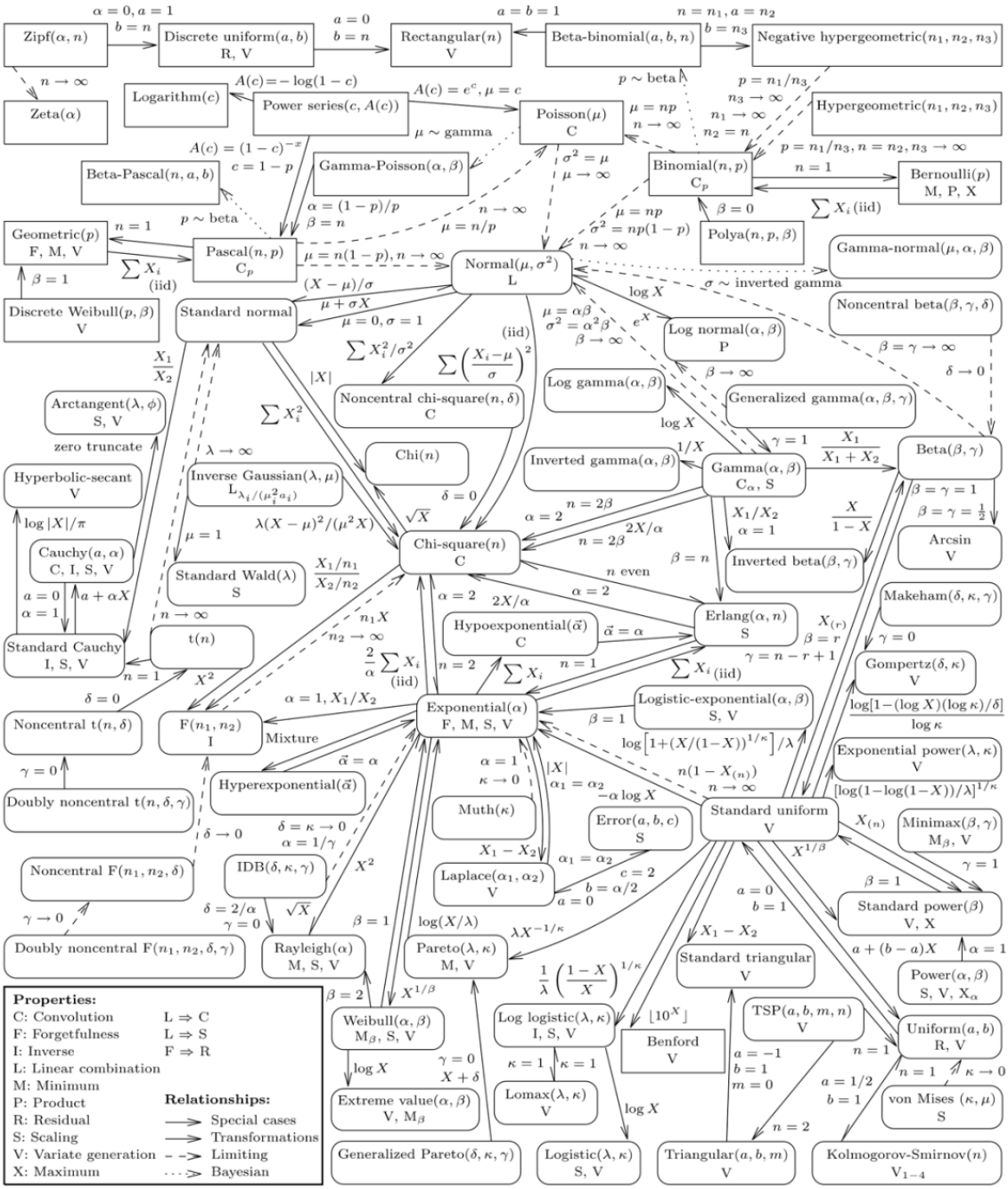

Figure 7.7: Relationships among common univariate probability distributions. Arrows represent special cases, transformations, limits, convolutions, mixtures, and other stochastic constructions, as specified by the legend (Leemis and McQueston 2008).

7.10 Summary

- A homogeneous Poisson process of rate β has independent exponential waiting times with mean $1/\beta$ and Poisson counts with mean and variance βT in a duration T.
- Independent thinning multiplies the rate by the retention probability.
- Independent superposition adds rates.
- Sequential hidden steps produce convolutions and, for equal rates, Gamma waiting times.
- Event-driven simulation draws the next waiting time directly instead of advancing through empty time slots.
- First-passage calculations use absorbing boundaries; in three dimensions a particle starting at radius r hits a spherical target of radius a with probability a/r, and the corresponding diffusion-limited rate constant is $4\pi Da$.
- Maximum entropy produces an exponential family once the variable, support, reference measure, and expectation constraints are specified.

The Poisson process is a useful baseline precisely because its assumptions are strong. When data violate its constant hazard, independent increments, or mean-equals-variance prediction, the pattern of failure points toward additional biological structure.

8 Randomness in cellular processes

Cellular populations are small enough that individual reaction events can be visible. A deterministic concentration may describe the average behavior of many cells while missing the fluctuations, waiting times, correlations, and rare states that distinguish competing mechanisms. This week develops the birth-death process as the simplest state-dependent continuous-time Markov process, derives its master equation, and uses it as a falsifiable model of mRNA dynamics.

By the end of this week, students should be able to answer:

- How does a continuous-time random walk differ from a walk with regularly timed steps?
- How can alternating one-dimensional and three-dimensional diffusion accelerate a protein's search for a particular DNA sequence?
- Why does a reaction network remain Markovian when its rates depend on the current molecular population?
- How does Gillespie's direct algorithm generate exact reaction trajectories?
- How does the master equation determine the mean and the full probability distribution?
- Why is the steady population of a linear birth-death process Poisson distributed?
- Which single-cell measurements falsify the simple birth-death model of transcription?
- How do promoter switching and transcriptional bursts explain the excess noise?
- Why do protein bursts produce gamma-like or negative-binomial distributions?
- How can total variance be separated into intrinsic molecular noise and extrinsic cell-to-cell variation?

8.1 Two random walks with the same long-time variance

Consider a walker with step length L and no directional bias. In a clocked model, one step occurs every $1/\beta$ units of time and each direction is chosen independently with probability $1/2$. After $n = \beta t$ steps,

$$X(t) = L\sum_{i=1}^{n} S_i, \qquad \Pr(S_i = +1) = \Pr(S_i = -1) = \frac{1}{2}. \tag{8.1}$$

The mean and variance are

$$\mathbb{E}[X(t)] = 0, \qquad \mathrm{Var}(X(t)) = L^2\beta t. \tag{8.2}$$

For large n, the central limit theorem gives an approximately Gaussian position.

A more physical continuous-time model assigns rate $\beta/2$ to rightward jumps and rate $\beta/2$ to leftward jumps. Their merged event stream has total rate β. The event time is random, and the event direction is a Bernoulli choice. Equivalently,

$$N_+(t) \sim \mathrm{Poisson}\left(\frac{\beta t}{2}\right), \qquad N_-(t) \sim \mathrm{Poisson}\left(\frac{\beta t}{2}\right), \tag{8.3}$$

independently, with

$$X(t) = L\big[N_+(t) - N_-(t)\big]. \tag{8.4}$$

The exact distribution of X/L is a symmetric Skellam distribution,

$$\Pr(X(t) = mL) = e^{-\beta t} I_{|m|}(\beta t), \qquad m \in \mathbb{Z}, \tag{8.5}$$

where I_m is a modified Bessel function. Its mean and variance are again

$$\mathbb{E}[X(t)] = 0, \qquad \mathrm{Var}(X(t)) = L^2\beta t. \tag{8.6}$$

Thus two processes can have the same Gaussian long-time limit and the same variance growth while differing in their short-time waiting statistics and higher-order temporal structure. The entire process, not only its endpoint variance, is part of the model.

8.2 Facilitated diffusion: how a protein finds a DNA target

The first-passage calculation of Week 7 treated a target as an absorbing boundary. A DNA-binding protein faces a particularly difficult search problem: it must find one short target sequence among roughly M possible binding sites. Three-dimensional diffusion brings the protein to DNA quickly, but usually to the wrong sequence. One-dimensional sliding along DNA examines neighboring base pairs accurately, but sliding across the entire genome would be very slow. Cells combine the two modes.

Model one search round as a period of nonspecific sliding with mean duration $\tau_{1\mathrm{D}}$, followed by a three-dimensional excursion with mean duration $\tau_{3\mathrm{D}}$. Suppose a sliding round examines an average of $\bar{n} \ll M$ distinct sites and each reassociation begins at an approximately independent position. The success probability per round is

$$p = \frac{\bar{n}}{M}. \tag{8.7}$$

The number K of rounds required to find the target is geometric, so

$$\mathbb{E}[K] = \frac{1}{p} = \frac{M}{\bar{n}}. \tag{8.8}$$

The mean search time is therefore

$$\overline{T}_{\mathrm{search}} = \frac{M}{\bar{n}}\left(\tau_{1\mathrm{D}} + \tau_{3\mathrm{D}}\right). \tag{8.9}$$

For ordinary one-dimensional diffusion with sliding coefficient $D_{1\mathrm{D}}$, the typical scanned length is

$$\bar{n} \simeq \sqrt{2D_{1\mathrm{D}}\tau_{1\mathrm{D}}}. \tag{8.10}$$

Here M and $\bar{n}$ are measured in binding sites (or base pairs), so $D_{1\mathrm{D}}$ has units of sites squared per unit time. If distance along DNA is instead measured in meters, both M and $\bar{n}$ must be replaced by physical contour lengths. The numerical factor in the scanned range depends on how revisits and the two directions along DNA are counted; the square-root scaling and the time-allocation tradeoff are the robust conclusions of this minimal model (Nelson 2022). Substituting this relation into Equation (8.9) reveals a tradeoff. Very short binding events scan too little DNA, while very long binding events trap the protein in one genomic neighborhood. Explicitly,

$$\overline{T}_{\mathrm{search}} = \frac{M}{\sqrt{2D_{1\mathrm{D}}}}\left(\sqrt{\tau_{1\mathrm{D}}} + \frac{\tau_{3\mathrm{D}}}{\sqrt{\tau_{1\mathrm{D}}}}\right). \tag{8.11}$$

Differentiating with respect to $\tau_{1\mathrm{D}}$ gives

$$\frac{d\overline{T}_{\mathrm{search}}}{d\tau_{1\mathrm{D}}} = \frac{M}{2\sqrt{2D_{1\mathrm{D}}}}\left(\tau_{1\mathrm{D}}^{-1/2} - \tau_{3\mathrm{D}}\tau_{1\mathrm{D}}^{-3/2}\right). \tag{8.12}$$

The derivative vanishes only when the two mean durations are equal. Because the search time diverges as $\tau_{1D} \to 0$ and as $\tau_{1D} \to \infty$, this stationary point is the minimum:

$$\tau_{1D} = \tau_{3D}, \tag{8.13}$$

so the fastest idealized search divides its time equally between sliding and spatial relocation. At this optimum,

$$\overline{T}_{\text{opt}} = \frac{2M\tau_{3D}}{\bar{n}_{\text{opt}}}. \tag{8.14}$$

Compared with a three-dimensional-only search time $M\tau_{3D}$, sliding gives an ideal speedup of approximately $\bar{n}_{\text{opt}}/2$. This is the *antenna effect*: binding near the target is sufficient because one-dimensional motion can finish the search. Nonspecific binding is nevertheless beneficial only up to a point; excessive affinity reduces spatial mobility and slows the search. The optimum emerges from balancing those two effects, not from maximizing either one alone.

8.3 Molecular populations as Markov processes

A process $Y(t)$ is Markovian if, conditional on its present state, the future is independent of the earlier history:

$$\Pr(Y(t+\tau) \mid Y(t), Y(t_1), \ldots, Y(t_n)) = \Pr(Y(t+\tau) \mid Y(t)), \qquad t_i < t. \tag{8.15}$$

Chemical reactions in a well-mixed volume satisfy this idealization when the chosen state contains all variables that determine the reaction rates. The waiting time is exponential while the state is fixed, but its rate can change after a reaction changes the state.

8.3.1 The immigration-death process

Let $\ell(t)$ be the number of molecules of species X. The simplest regulated population has two reactions:

$$\varnothing \xrightarrow{\beta_s} X, \qquad X \xrightarrow{k_\phi} \varnothing. \tag{8.16}$$

Synthesis has a constant zeroth-order propensity β_s. Each molecule is cleared independently with rate k_ϕ, so when the population is ℓ, the total clearance propensity is $k_\phi \ell$. This addition of independent molecular hazards is the superposition property of Poisson processes.

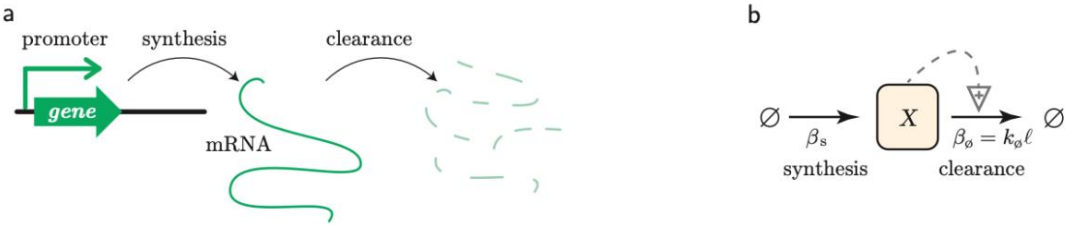

Figure 8.1: A cellular birth-death process and its reaction-network representation. Synthesis increases the molecular inventory by one at constant rate β_s; clearance decreases it by one at total rate $k_\phi \ell$ (Nelson 2022).

Over a short interval Δt, the transition law is

$$\Pr(\ell(t+\Delta t)=m \mid \ell(t)=\ell)=\begin{cases}\beta_s\Delta t+o(\Delta t), & m=\ell+1,\\ k_\phi\ell\Delta t+o(\Delta t), & m=\ell-1,\\ 1-(\beta_s+k_\phi\ell)\Delta t+o(\Delta t), & m=\ell,\\ o(\Delta t), & \text{otherwise.}\end{cases} \tag{8.17}$$

The right-hand side depends on the present population ℓ, but not on earlier populations. The process is therefore Markovian even though the total event rate is state-dependent.

8.4 Deterministic trajectory and exact mean

Replacing the discrete population by a continuous variable gives the deterministic rate equation

$$\frac{d\ell}{dt}=\beta_s-k_\phi\ell. \tag{8.18}$$

Its fixed point and relaxation time are

$$\ell_*=\frac{\beta_s}{k_\phi},\qquad \tau_\phi=\frac{1}{k_\phi}. \tag{8.19}$$

For initial value $\ell(0)=\ell_0$,

$$\ell(t)=\ell_*+\big(\ell_0-\ell_*\big)e^{-k_\phi t}. \tag{8.20}$$

In particular, starting from zero,

$$\ell(t)=\frac{\beta_s}{k_\phi}\left(1-e^{-k_\phi t}\right). \tag{8.21}$$

For a single stochastic trajectory, Equation (8.18) is an approximation: the population moves in integer jumps. For the ensemble mean, however, the same equation will turn out to be exact because both propensities are linear:

$$\frac{d}{dt}\mathbb{E}[\ell(t)]=\beta_s-k_\phi\mathbb{E}[\ell(t)]. \tag{8.22}$$

Equality of the deterministic solution and the stochastic mean does not imply that the deterministic model captures the distribution.

8.5 Gillespie's direct algorithm

Gillespie's direct method generates exact trajectories of a well-mixed Markov reaction network (Gillespie 1976). At population ℓ, the synthesis and clearance clocks are independent exponential variables with rates β_s and $k_\phi\ell$. Their minimum is exponential with total rate

$$a_0(\ell)=\beta_s+k_\phi\ell. \tag{8.23}$$

If U_1, U_2 are independent uniform variables on $(0,1)$, the next waiting time is

$$\tau=-\frac{\ln U_1}{a_0(\ell)}. \tag{8.24}$$

Conditional on an event occurring, it is a synthesis event with probability

$$p_{\text{synth}} = \frac{\beta_s}{\beta_s + k_\phi \ell}, \tag{8.25}$$

and a clearance event otherwise. At $\ell = 0$, the clearance propensity vanishes, so the hard boundary is enforced automatically.

The direct algorithm is:

1. evaluate all propensities in the current state;
2. draw the next waiting time from an exponential distribution with rate equal to their sum;
3. select reaction r with probability a_r/a_0;
4. update time and the molecular state according to that reaction; and
5. recompute the propensities and repeat.

For a general reaction network, the state is a vector and each reaction has a stoichiometric change vector. This is an exact simulation of the continuous-time Markov chain under the well-mixed assumptions; it does not discretize time.

An LLM can implement the reaction list and generate ensembles without making programming syntax a course objective. The scientific checks remain essential: no trajectory may cross the boundary $\ell = 0$; the empirical waiting time at fixed ℓ must have mean $1/a_0(\ell)$; reaction frequencies must approach a_r/a_0; the ensemble mean must follow Equation (8.22); and the steady histogram must converge to the Poisson law derived below.

8.6 The master equation

Let

$$P_\ell(t) = \Pr(\ell(t) = \ell), \qquad \ell = 0, 1, 2, \ldots. \tag{8.26}$$

Probability enters state ℓ through synthesis from $\ell - 1$ and clearance from $\ell + 1$, and leaves through either reaction from ℓ. Taking the $\Delta t \to 0$ limit of Equation (8.17) gives

$$\frac{dP_\ell}{dt} = \beta_s P_{\ell-1} + k_\phi(\ell + 1)P_{\ell+1} - \big(\beta_s + k_\phi \ell\big)P_\ell, \tag{8.27}$$

with $P_{-1} \equiv 0$. Every term is a probability flux into or out of state ℓ. Summing Equation (8.27) over all ℓ gives $d\sum_\ell P_\ell/dt = 0$, so normalization is conserved.

8.6.1 Evolution of the mean and variance

Multiply the master equation by ℓ and sum:

$$\begin{aligned} \frac{d}{dt}\langle \ell \rangle &= \sum_{\ell=0}^{\infty} \ell \frac{dP_\ell}{dt} \\ &= \beta_s - k_\phi \langle \ell \rangle. \end{aligned} \tag{8.28}$$

The index shifts behind the last line are worth displaying. With $j = \ell - 1$ in the birth inflow and $j = \ell + 1$ in the death inflow,

$$\sum_{\ell=0}^{\infty} \ell P_{\ell-1} = \sum_{j=0}^{\infty} (j+1) P_j = \langle \ell \rangle + 1,$$
$$\sum_{\ell=0}^{\infty} \ell(\ell+1) P_{\ell+1} = \sum_{j=1}^{\infty} (j-1) j P_j = \langle \ell^2 \rangle - \langle \ell \rangle. \tag{8.29}$$

Substitution into the four terms of Equation (8.27) cancels the remaining $\beta_s \langle \ell \rangle$ and $k_\phi \langle \ell^2 \rangle$ contributions, leaving Equation (8.28). This proves Equation (8.22). Applying the same method to ℓ^2 yields

$$\frac{d}{dt}\langle \ell^2 \rangle = \beta_s \big(2\langle \ell \rangle + 1\big) + k_\phi \big(-2\langle \ell^2 \rangle + \langle \ell \rangle\big). \tag{8.30}$$

Writing $m(t) = \langle \ell(t) \rangle$ and $V(t) = \langle \ell^2(t) \rangle - m(t)^2$, the chain rule gives

$$\begin{aligned} \frac{dV}{dt} &= \frac{d}{dt}\langle \ell^2 \rangle - 2m \frac{dm}{dt} \\ &= \beta_s(2m+1) + k_\phi\big(-2\langle \ell^2 \rangle + m\big) - 2m(\beta_s - k_\phi m) \\ &= \beta_s + k_\phi m - 2k_\phi V. \end{aligned} \tag{8.31}$$

The terms proportional to $2\beta_s m$ cancel, and $\langle \ell^2 \rangle - m^2 = V$ closes the equation without requiring a third moment. At steady state,

$$\langle \ell \rangle_* = \frac{\beta_s}{k_\phi}, \qquad V_* = \frac{\beta_s}{k_\phi}. \tag{8.32}$$

Thus the variance equals the mean, a prediction that can be tested independently of the relaxation curve.

8.6.2 The stationary distribution

At equilibrium, adjacent probability fluxes balance:

$$\beta_s P^*_{\ell-1} = k_\phi \ell P^*_\ell. \tag{8.33}$$

Let

$$\mu = \frac{\beta_s}{k_\phi}. \tag{8.34}$$

The recurrence relation is

$$P^*_\ell = \frac{\mu}{\ell} P^*_{\ell-1} = \frac{\mu^\ell}{\ell!} P^*_0. \tag{8.35}$$

Normalization gives $P^*_0 = e^{-\mu}$, hence

$$P^*_\ell = e^{-\mu} \frac{\mu^\ell}{\ell!}. \tag{8.36}$$

The stationary population is Poisson distributed. Its coefficient of variation is $\mu^{-1/2}$, so the deterministic approximation becomes relatively accurate when the mean population is large.

If the process starts from $\ell(0) = 0$, the complete time-dependent distribution is also Poisson:

$$P_\ell(t) = e^{-\mu(t)} \frac{\mu(t)^\ell}{\ell!}, \qquad \mu(t) = \frac{\beta_s}{k_\phi}\left(1 - e^{-k_\phi t}\right). \tag{8.37}$$

One way to see this result is to regard births as points of a Poisson process. A molecule born at time s survives until time t with probability $e^{-k_\phi(t-s)}$. Independent thinning therefore makes the number of surviving births Poisson, with mean

$$\int_0^t \beta_s e^{-k_\phi(t-s)}\, ds = \frac{\beta_s}{k_\phi}\left(1 - e^{-k_\phi t}\right) = \mu(t). \tag{8.38}$$

This establishes both the Poisson form and the time-dependent parameter in Equation (8.37).

For a deterministic nonzero initial count $\ell(0) = \ell_0$, the surviving initial molecules and surviving new molecules contribute independently. Writing $q(t) = e^{-k_\phi t}$,

$$S_0(t) \sim \text{Binomial}(\ell_0, q(t)), \qquad S_{\text{new}}(t) \sim \text{Poisson}[\mu(1 - q(t))]\,, \tag{8.39}$$

and $\ell(t) = S_0(t) + S_{\text{new}}(t)$. Equivalently, its generating function is

$$G_\ell(z,t) = [1 - q(t) + q(t)z]^{\ell_0} \exp\{\mu[1 - q(t)](z-1)\}\,. \tag{8.40}$$

The initial binomial memory disappears as $q(t) \to 0$, leaving the stationary Poisson distribution. Equation (8.37) is the special case $\ell_0 = 0$.

At steady state the temporal covariance is

$$\text{Cov}(\ell(t), \ell(t+\tau)) = \mu e^{-k_\phi|\tau|}, \tag{8.41}$$

To derive it for $\tau \geq 0$, condition on $\ell(t)$. Each molecule present at time t survives the interval τ with probability $e^{-k_\phi\tau}$, while molecules born during the interval contribute independently. Hence

$$\mathbb{E}[\ell(t+\tau) \mid \ell(t)] = \ell(t)e^{-k_\phi\tau} + \mu\left(1 - e^{-k_\phi\tau}\right). \tag{8.42}$$

The conditional-covariance identity then gives

$$\begin{aligned}\text{Cov}(\ell(t), \ell(t+\tau)) &= \text{Cov}(\ell(t), \mathbb{E}[\ell(t+\tau) \mid \ell(t)]) \\ &= e^{-k_\phi\tau}\,\text{Var}(\ell(t)) = \mu e^{-k_\phi\tau}.\end{aligned} \tag{8.43}$$

Symmetry of covariance replaces τ by $|\tau|$ for either time ordering, so the same clearance rate controls both deterministic relaxation and loss of stochastic memory.

8.7 Gene expression one molecule at a time

The central dogma connects DNA, RNA, and protein, but each arrow represents individual biochemical events and is therefore noisy (Figure 8.2). A simple first hypothesis is that mRNA is synthesized at constant rate and cleared independently, exactly as in the birth-death model.

Golding and coauthors measured mRNA in living *Escherichia coli*. The transcript contained 96 repeated binding sequences in a noncoding region. Each sequence folded into a site for an MS2 protein fused to green fluorescent protein. A newly transcribed molecule therefore recruited many fluorescent proteins and appeared as a bright spot.

The total spot fluorescence in each cell formed a sequence of nearly equally spaced peaks (Figure 8.3). The lowest nonzero peak calibrated one transcript; successive peaks represented integer multiples. The fluorescence measurement could therefore be converted to an absolute single-cell mRNA count.

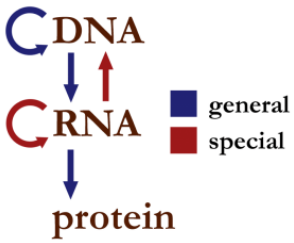

Figure 8.2: Information flow among DNA, RNA, and protein. Each biochemical step may be stochastic and regulated. Art by David S. Goodsell (Nelson 2022).

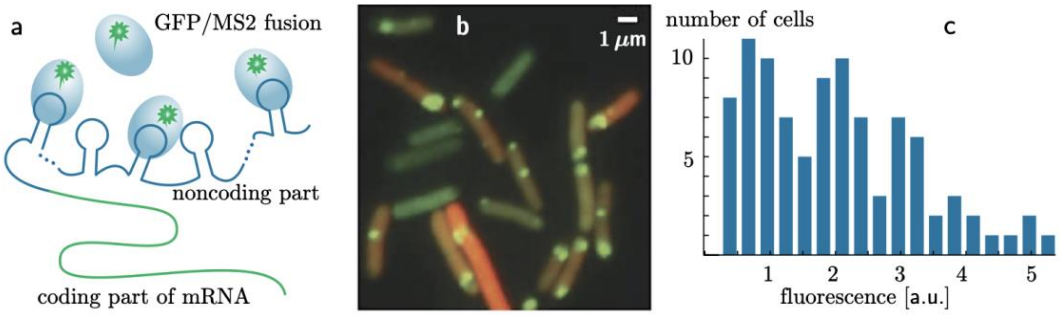

Figure 8.3: Single-molecule quantification of mRNA. Repeated MS2 binding sites recruit GFP fusion proteins to each transcript (left); labeled transcripts appear as bright spots in living cells (middle); and separated fluorescence peaks calibrate integer mRNA copy numbers (right) (Golding et al. 2005).

8.8 A model can fit the mean and still be wrong

After transcription was induced, the observed mean mRNA count was fit by Equation (8.21), giving apparent parameters

$$\beta_s \simeq 0.15\,\mathrm{min}^{-1}, \qquad k_\phi \simeq 0.014\,\mathrm{min}^{-1}. \tag{8.44}$$

Panel (a) of Figure 8.4 shows that the simple curve describes the mean reasonably well. Two predictions for the distribution nevertheless fail.

First, define the Fano factor

$$F = \frac{\mathrm{Var}(\ell)}{\mathbb{E}[\ell]}. \tag{8.45}$$

The birth-death model predicts $F = 1$, but the observed value was approximately 5 across a range of expression levels. The population was strongly overdispersed relative to a Poisson distribution.

Second, Equation (8.37) predicts

$$P_0(t) = \Pr(\ell(t) = 0) = e^{-\mu(t)}. \tag{8.46}$$

At early times, $\mu(t) \simeq \beta_s t$, so

$$\ln P_0(t) \simeq -\beta_s t. \tag{8.47}$$

The observed zero-cell fraction decayed much more slowly than the apparent synthesis rate inferred from the mean. The three panels therefore overconstrain and falsify the constant-synthesis birth-death hypothesis.

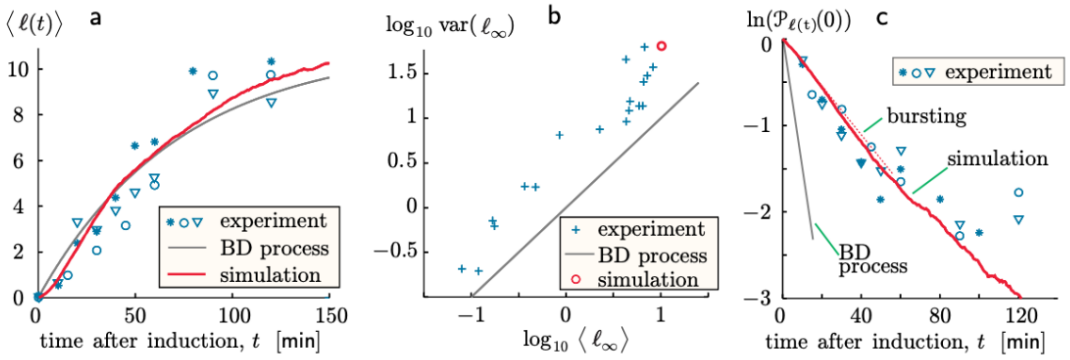

Figure 8.4: Indirect evidence for transcriptional bursting. A simple birth-death model can fit the induction curve for the mean (a), but it predicts Fano factor one rather than the observed value near five (b), and it incorrectly predicts the loss of cells with zero transcripts (c) (Golding et al. 2005).

8.9 Promoter switching and transcriptional bursts

Introduce a promoter state $G(t) \in \{0, 1\}$, with $G = 0$ inactive and $G = 1$ active:

$$0 \underset{\beta_{\mathrm{stop}}}{\overset{\beta_{\mathrm{start}}}{\rightleftharpoons}} 1. \tag{8.48}$$

Transcription occurs at rate rG, while each transcript is cleared at rate k_ϕ. The complete reaction set is

$$\begin{aligned} G = 0 &\longrightarrow G = 1 \quad \text{at rate } \beta_{\text{start}}, \\ G = 1 &\longrightarrow G = 0 \quad \text{at rate } \beta_{\text{stop}}, \\ \ell &\longrightarrow \ell + 1 \quad \text{at rate } rG, \\ \ell &\longrightarrow \ell - 1 \quad \text{at rate } k_\phi \ell. \end{aligned} \tag{8.49}$$

The on and off episode durations are exponentially distributed:

$$\mathbb{E}[T_{\text{off}}] = \frac{1}{\beta_{\text{start}}}, \qquad \mathbb{E}[T_{\text{on}}] = \frac{1}{\beta_{\text{stop}}}. \tag{8.50}$$

At stationarity, the active fraction is

$$p_{\text{on}} = \frac{\beta_{\text{start}}}{\beta_{\text{start}} + \beta_{\text{stop}}}, \tag{8.51}$$

and the mean mRNA population is

$$\langle \ell \rangle_* = \frac{r\, p_{\text{on}}}{k_\phi}. \tag{8.52}$$

Unlike the simple birth-death process, the telegraph model is overdispersed. Its stationary Fano factor is

$$F = 1 + \frac{r \beta_{\text{stop}}}{(\beta_{\text{start}} + \beta_{\text{stop}})(k_\phi + \beta_{\text{start}} + \beta_{\text{stop}})}. \tag{8.53}$$

Slow promoter switching creates cell-to-cell variation in the time-averaged synthesis rate. In the short-on-state limit, an active episode produces a burst with mean size $b \simeq r/\beta_{\text{stop}}$, and F approaches $1 + b$ when promoter switching is fast compared with mRNA clearance. More generally, the exact conversion from a measured Fano factor to burst size depends on the burst-size distribution and the switching regime. The original analysis inferred an effective burst scale of order five.

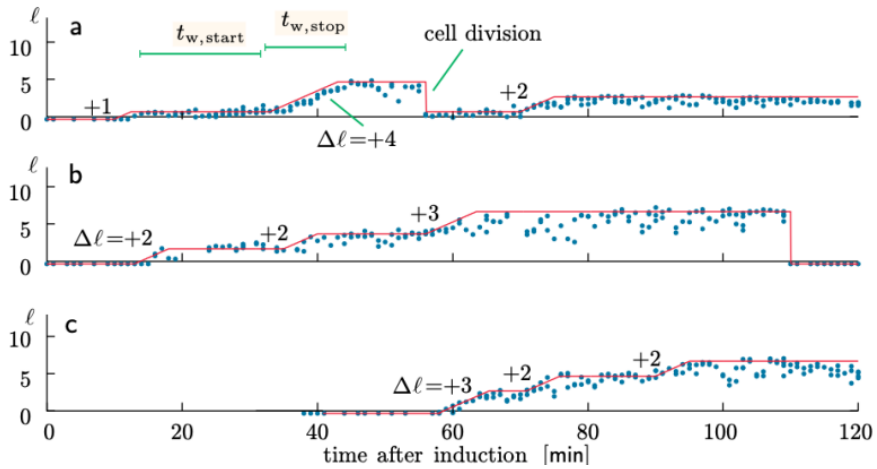

Figure 8.5: Direct single-cell evidence for bursts. Individual cells alternate between quiet intervals and periods of approximately steady transcript production. Downward jumps at division reflect partitioning between daughter cells (Golding et al. 2005).

The individual traces show the structure hidden by population averages (Figure 8.5). Quiet intervals alternate with sloped production episodes, and both durations vary between events. The

semilog waiting-time plots are approximately linear, consistent with exponential on and off times (Figure 8.6). In that figure, the synthesis label β_s corresponds to the active-state transcription rate denoted by r in Equation (8.49).

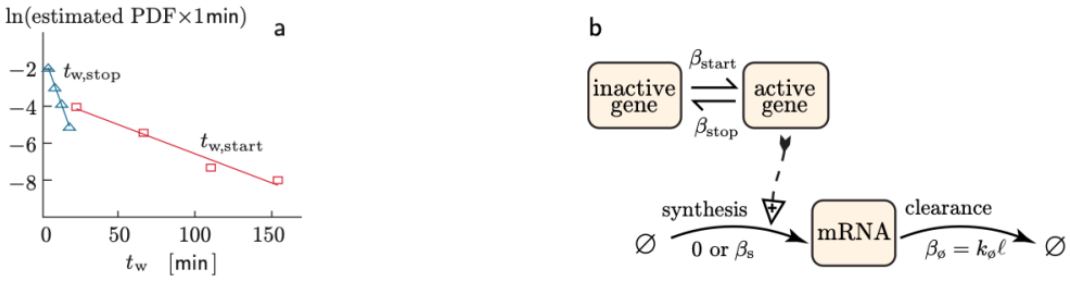

Figure 8.6: Promoter-switching model and measured episode durations. The semilog waiting-time plots support exponential off and on intervals (left); the reaction network couples a two-state promoter to mRNA synthesis and clearance (right) (Golding et al. 2005).

8.9.1 Quantitative cross-checks

Fits to the episode durations gave approximately

$$\beta_{\text{start}} \simeq \frac{1}{37\,\text{min}}, \qquad \beta_{\text{stop}} \simeq \frac{1}{6\,\text{min}}. \tag{8.54}$$

The initial slope of $\ln P_0(t)$ was about $-0.028\,\text{min}^{-1}$, consistent with $-\beta_{\text{start}}$. If each activation produces an effective $m \simeq 5$ transcripts, the coarse burst arrival rate predicts

$$m\beta_{\text{start}} \simeq \frac{5}{37\,\text{min}} \simeq 0.135\,\text{min}^{-1}, \tag{8.55}$$

close to the apparent $0.15\,\text{min}^{-1}$ obtained by fitting the mean. The same parameters therefore explain the induction curve, excess variance, zero-cell fraction, and direct waiting-time measurements.

8.10 From mRNA bursts to protein distributions

Transcriptional bursts describe how mRNAs appear. Translation creates a second layer of bursting because one mRNA can produce many proteins before it disappears. Let k_p be the translation rate per mRNA, γ_m the mRNA degradation rate, and γ_p the protein degradation or dilution rate. The minimal reactions are

$$M \xrightarrow{k_p} M + P, \qquad M \xrightarrow{\gamma_m} \varnothing, \qquad P \xrightarrow{\gamma_p} \varnothing. \tag{8.56}$$

The lifetime T of one mRNA is exponentially distributed with rate γ_m. Given that lifetime, translation events form a Poisson process:

$$T \sim \text{Exponential}(\gamma_m), \qquad B \mid T \sim \text{Poisson}(k_p T), \tag{8.57}$$

where B is the number of proteins made from that transcript. Mixing the Poisson count over the random lifetime gives

$$\begin{aligned} P(B=b) &= \int_0^\infty \gamma_m e^{-\gamma_m t} e^{-k_p t} \frac{(k_p t)^b}{b!}\, dt \\ &= \frac{\gamma_m k_p^b}{b!} \int_0^\infty t^b e^{-(\gamma_m + k_p)t}\, dt \\ &= \frac{\gamma_m k_p^b}{b!} \frac{\Gamma(b+1)}{(\gamma_m + k_p)^{b+1}} \\ &= \frac{\gamma_m}{\gamma_m + k_p} \left(\frac{k_p}{\gamma_m + k_p}\right)^b, \qquad b = 0, 1, 2, \ldots . \end{aligned} \tag{8.58}$$

Thus the protein burst size is geometric, with

$$\overline{b} = \frac{k_p}{\gamma_m}, \qquad \mathrm{Var}(B) = \overline{b}(1+\overline{b}). \tag{8.59}$$

If mRNAs are produced at rate k_m, are short lived relative to proteins, and can be coarse-grained as instantaneous protein bursts, the stationary protein distribution is negative binomial. The connection can be derived with probability-generating functions. This burst limit assumes $\gamma_m \gg \gamma_p$, Poisson creation of mRNAs at constant rate k_m, independent geometric translation bursts, and first-order protein loss. Promoter switching or slowly varying cell states would add further layers of overdispersion. Outside this separation of timescales, the full joint mRNA–protein process must be retained and the stationary protein marginal is not generally the simple negative-binomial law below. The geometric burst law in Equation (8.58) has generating function

$$G_B(z) \equiv \mathbb{E}[z^B] = \frac{1}{1 + \overline{b}(1-z)}. \tag{8.60}$$

Let $G_P(z,t) = \sum_{n=0}^\infty \Pr(P(t) = n) z^n$. A burst-arrival event multiplies the current generating function by $G_B(z)$, whereas independent protein loss contributes the usual linear-death term. The master equation therefore becomes

$$\frac{\partial G_P}{\partial t} = k_m [G_B(z) - 1] G_P + \gamma_p (1-z) \frac{\partial G_P}{\partial z}. \tag{8.61}$$

At stationarity, canceling the common factor $1 - z$ gives

$$\frac{G_P'(z)}{G_P(z)} = \frac{k_m}{\gamma_p} \frac{\overline{b}}{1 + \overline{b}(1-z)}. \tag{8.62}$$

Writing

$$a = \frac{k_m}{\gamma_p}, \tag{8.63}$$

and integrating Equation (8.62) from $z = 1$, where $G_P(1) = 1$, yields

$$G_P(z) = [1 + \overline{b}(1-z)]^{-a}. \tag{8.64}$$

To read off the probabilities, set $q = \overline{b}/(1+\overline{b})$ and use

$$(1 - qz)^{-a} = \sum_{n=0}^\infty \frac{\Gamma(n+a)}{\Gamma(a) n!} q^n z^n. \tag{8.65}$$

Since $1+\bar{b}(1-z)=(1+\bar{b})(1-qz)$, the coefficient of z^n is

$$P(P=n)=\frac{\Gamma(n+a)}{\Gamma(a)n!}\left(\frac{1}{1+\bar{b}}\right)^a\left(\frac{\bar{b}}{1+\bar{b}}\right)^n. \tag{8.66}$$

Finally, a probability-generating function obeys $\mathbb{E}[P]=G'_P(1)$ and $\mathrm{Var}(P)=G''_P(1)+G'_P(1)-G'_P(1)^2$. Differentiating Equation (8.64) therefore gives

$$\mathbb{E}[P]=a\bar{b}, \qquad \mathrm{Var}(P)=a\bar{b}(1+\bar{b}), \qquad F=1+\bar{b}. \tag{8.67}$$

For abundant proteins, and specifically in the large-burst regime $\bar{b}\gg 1$, the negative-binomial mass function has a useful continuous approximation. For relevant counts $n\gg 1$,

$$\frac{\Gamma(n+a)}{\Gamma(a)n!}\simeq\frac{n^{a-1}}{\Gamma(a)}, \qquad \left(\frac{\bar{b}}{1+\bar{b}}\right)^n\simeq e^{-n/\bar{b}}, \qquad \left(\frac{1}{1+\bar{b}}\right)^a\simeq\bar{b}^{-a}. \tag{8.68}$$

Replacing the integer count by a continuous variable x therefore gives the Gamma density

$$p(x)=\frac{x^{a-1}e^{-x/\bar{b}}}{\bar{b}^a\Gamma(a)}, \qquad x\geq 0. \tag{8.69}$$

This Gamma distribution has

$$\mathbb{E}[X]=a\bar{b}, \qquad \mathrm{Var}(X)=a\bar{b}^2. \tag{8.70}$$

Its mean matches the exact negative-binomial mean, while its variance neglects the additional $a\bar{b}$ term. The relative variance error is therefore of order $1/\bar{b}$. A large mean $a\bar{b}$ alone is not sufficient for this particular Gamma approximation if it arises from many small bursts rather than from $\bar{b}\gg 1$. The shape parameter a counts the effective number of bursts per protein lifetime, while $\bar{b}$ sets their scale. The same mathematical architecture appears one level upstream when promoter switching creates bursts of mRNA: a randomly switching upstream state drives production of a downstream molecular species.

8.11 Cell division as dilution and partition noise

The labeled transcripts were rarely degraded; cell division was the main process reducing their concentration. For division time $T_d\simeq 50\,\mathrm{min}$, halving the mean once per generation is equivalent to continuous dilution rate

$$k_{\mathrm{dil}}=\frac{\ln 2}{T_d}\simeq 0.014\,\mathrm{min}^{-1}, \tag{8.71}$$

in agreement with the fitted k_ϕ.

Continuous dilution captures the mean but not the full division event. If a followed daughter receives each of the ℓ transcripts independently with probability $1/2$, then

$$\ell_{\mathrm{after}}\mid\ell_{\mathrm{before}}\sim\mathrm{Binomial}\left(\ell_{\mathrm{before}},\frac{1}{2}\right). \tag{8.72}$$

Hence

$$\mathbb{E}[\ell_{\mathrm{after}}\mid\ell_{\mathrm{before}}]=\frac{\ell_{\mathrm{before}}}{2}, \qquad \mathrm{Var}(\ell_{\mathrm{after}}\mid\ell_{\mathrm{before}})=\frac{\ell_{\mathrm{before}}}{4}. \tag{8.73}$$

The discrete partitioning event adds a source of noise that a continuous loss term omits.

8.12 Intrinsic and extrinsic contributions to gene-expression noise

The distinction between intrinsic and extrinsic gene-expression noise can be measured with paired reporters (Elowitz, Levine, et al. 2002). The reaction models above assume fixed kinetic rates. Real cells differ in age, volume, gene copy number, polymerase abundance, and other slowly varying properties. Let Y be an mRNA or protein copy number and let X collect these cell-level variables. The law of total variance gives

$$\operatorname{Var}(Y) = \mathbb{E}[\operatorname{Var}(Y \mid X)] + \operatorname{Var}(\mathbb{E}[Y \mid X]) . \tag{8.74}$$

The first term is *intrinsic noise*: even cells with the same effective parameters experience different molecular reaction histories. The second is *extrinsic noise*: cells with different values of X have different conditional means.

For two identically regulated reporters C and Y with equal mean $\langle C \rangle = \langle Y \rangle = \bar{y}$, reporter-specific fluctuations tend to separate the two signals, whereas shared cell state moves them together. The standard two-reporter estimators are

$$\eta^2_{\text{int}} = \frac{\langle (C - Y)^2 \rangle}{2\bar{y}^2}, \tag{8.75}$$

$$\eta^2_{\text{ext}} = \frac{\langle CY \rangle - \langle C \rangle \langle Y \rangle}{\bar{y}^2} = \frac{\operatorname{Cov}(C, Y)}{\bar{y}^2}. \tag{8.76}$$

Under the symmetric reporter model, with independent reporter-specific noise conditional on the shared state, the normalized variance of either reporter satisfies $\eta^2_{\text{tot}} = \eta^2_{\text{int}} + \eta^2_{\text{ext}}$. Unequal reporter means, spectral cross-talk, or reporter-specific maturation require a modified measurement model rather than direct use of these formulas.

Genome replication provides a concrete example. After a gene is copied, its dosage can double, changing the transcription rate. Cells sampled at different stages of the cell cycle therefore need not share the same conditional mean expression. Division adds both a reset of gene dosage and the binomial partition noise in Equation (8.72). If the mRNA lifetime is not negligible compared with the cell-cycle time, expression does not instantaneously follow the dosage change; the birth-death dynamics must be solved across replication and division events.

It is also essential to distinguish copy number from concentration. Copy number can grow with cell volume even when concentration is nearly constant. Conditioning on cell volume and gene dosage can therefore remove predictable population heterogeneity before the remaining fluctuations are interpreted as evidence for a molecular mechanism.

8.13 Summary

- A continuous-time random walk is specified by both its jump directions and its random event times.
- State-dependent propensities define a Markov process when the present state contains all rate-determining information.
- Gillespie's algorithm draws the next event time and event type directly from the current propensities.
- The master equation gives the full probability dynamics; moments are derived from it rather than assumed.

- Constant synthesis plus independent clearance produces a Poisson population with equal mean and variance.
- Single-cell mRNA data fit the predicted mean but violate the Poisson Fano factor and zero-state dynamics.
- A switching promoter explains quiet intervals, transcriptional bursts, exponential episode durations, and excess population noise.
- Alternating three-dimensional relocation and one-dimensional sliding can accelerate target search; optimal facilitated diffusion balances the time spent in the two modes.
- A random mRNA lifetime converts Poisson translation into geometric protein bursts, which produce negative-binomial or Gamma-like protein distributions.
- The law of total variance separates intrinsic reaction noise from extrinsic variation in cell state, gene dosage, and volume.

The lesson is methodological: fitting a mean trajectory estimates parameters, while testing distributional predictions distinguishes mechanisms.

9 Negative feedback control

Cells must maintain molecular inventories despite fluctuating environments, noisy reaction events, growth, and division. Negative feedback provides a general physical solution: a deviation from a desired state changes the dynamics in a direction that opposes the deviation. The result is a stable fixed point. This week connects that dynamical-systems picture to transcription-factor binding, Hill functions, autoregulated gene expression, and a second continuous-time Markov-chain application: the evolution of DNA sequences.

By the end of this week, students should be able to answer:

- How does negative feedback create a stable setpoint and determine a relaxation time?
- How do microscopic binding and unbinding rates produce a gene-regulation function?
- Why does cooperative binding create a sharp, sigmoidal response?
- How do clearance and growth dilution enter the dynamics of a protein concentration?
- Why can an autoregulated gene recover faster and fluctuate less than an unregulated gene?
- How can the same continuous-time Markov-chain reasoning describe neutral DNA substitutions?

9.1 Negative feedback and stable setpoints

A feedback controller compares the present output with a target and feeds a corrective signal back into the dynamics. The simplest model for a state variable $v(t)$ is

$$\frac{dv}{dt} = -k\big(v - v_*\big), \qquad k > 0, \tag{9.1}$$

where v_* is the setpoint. Writing $x = v - v_*$ gives

$$\frac{dx}{dt} = -kx, \qquad v(t) = v_* + \big(v(0) - v_*\big)e^{-kt}. \tag{9.2}$$

The sign is the essential feature. If $v > v_*$, then $dv/dt < 0$; if $v < v_*$, then $dv/dt > 0$. Perturbations decay with relaxation time $\tau = 1/k$, so v_* is a stable fixed point.

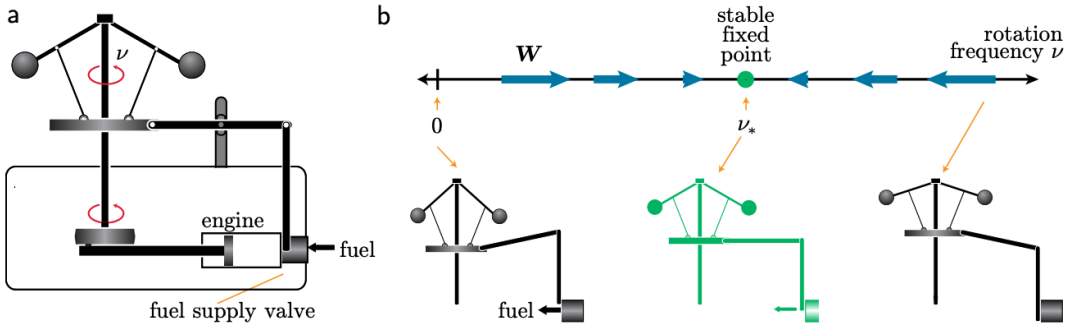

Figure 9.1: Negative feedback in a centrifugal governor. Increasing rotation speed moves the masses outward and reduces the fuel supply; decreasing speed has the opposite effect. The phase-line representation shows flow toward a stable fixed point (Nelson 2022).

Homeostasis in a cell or organism uses the same principle, although the controller is a reaction network rather than a mechanical linkage. Blood-cell production, ion balance, and protein abundance

all require an output-dependent correction. Even the linear birth-death process from Week 8 contains a weak form of feedback: the total clearance rate increases with the molecular population. Evolved regulatory circuits can add a stronger response through transcription factors and allosteric proteins.

9.2 Molecular control of gene activity

Cells use molecular counts and concentrations as state variables. DNA supplies the rules for how those variables influence production, but biochemical signals share the same crowded medium. Specific molecular recognition is therefore essential, while cross-talk, nonspecific binding, and signal interference remain possible.

In allosteric modulation, an effector binds away from an enzyme's active site and changes the enzyme's conformation. A small change in effector concentration can thereby alter catalytic activity or the ability of a transcription factor to bind DNA.

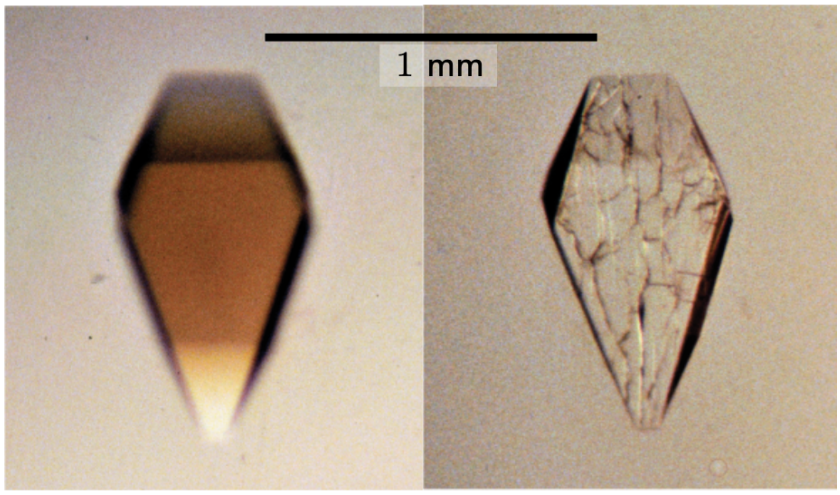

Figure 9.2: Allosteric conformational change in a crystal of lac repressor. Exposure to the effector IPTG changes the conformation enough to destabilize the crystal. Courtesy of Pace et al. (1990).

The portability of such control is striking. In the transgenic-mouse experiment of Figure 9.3, the bacterial LacI repressor controls a mammalian tyrosinase transgene. Tyrosinase is required to synthesize melanin. LacI represses the gene, while IPTG binds LacI and relieves repression; changing a molecule in the drinking water therefore changes coat pigmentation.

9.3 From binding kinetics to a regulation function

Let R be a repressor and O its operator sequence on the DNA. Binding is stochastic: thermal motion brings R near O, binding occurs with a concentration-dependent rate, and a thermal fluctuation eventually breaks the complex apart. We first derive the average operator occupancy and then assume that binding equilibrates faster than protein production and clearance.

9.3.1 One binding site

Suppose the operator is bound at time $t = 0$. Over a short interval Δt,

$$\Pr(\text{bound at } \Delta t \mid \text{bound at } 0) = 1 - \beta_{\text{off}} \Delta t + o(\Delta t), \tag{9.3}$$

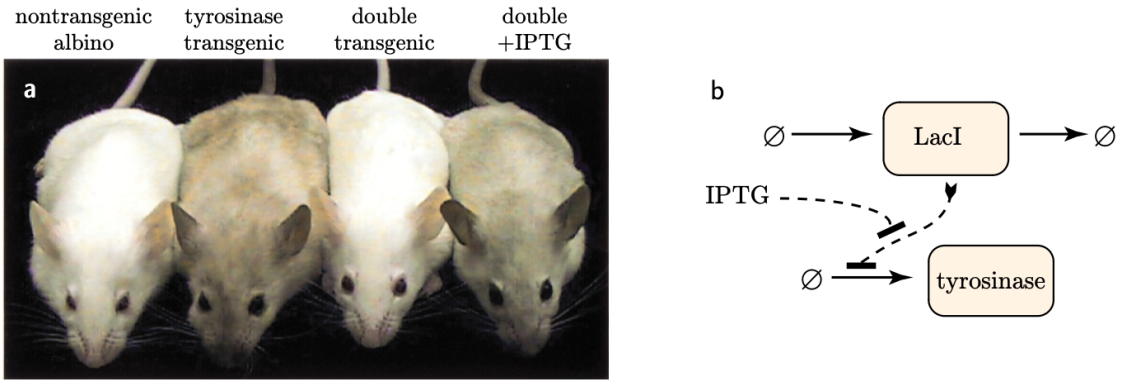

Figure 9.3: Control of a mammalian gene with a bacterial regulatory module. The first mouse is a nontransgenic albino, the second carries a functional tyrosinase transgene, and the two on the right also carry LacI. IPTG prevents LacI repression in the rightmost mouse. The network diagram summarizes the control logic (Cronin et al. 2001).

where β_{off} is the dissociation rate. If the operator is unbound and the repressor concentration is c, mass-action kinetics gives

$$\Pr(\text{bound at } \Delta t \mid \text{unbound at } 0) = k_{\text{on}} c\, \Delta t + o(\Delta t). \tag{9.4}$$

The factor c represents availability; k_{on} represents the geometry and stickiness of the encounter. The reaction scheme is

$$O \underset{\beta_{\text{off}}}{\overset{k_{\text{on}} c}{\rightleftharpoons}} OR. \tag{9.5}$$

Here the repressor is treated as a reservoir held at concentration c, so binding is pseudo-first-order and the effective forward rate is $k_{\text{on}}c$. An equally valid explicit mass-action notation is $O + R \underset{\beta_{\text{off}}}{\overset{k_{\text{on}}}{\rightleftharpoons}} OR$, with the factor c appearing in the rate law rather than above the arrow. These are two descriptions of the same kinetics; writing both R and $k_{\text{on}}c$ in one scheme would mix the conventions.

Let $p(t) = \Pr(\text{operator bound at } t)$. The two-state master equation is

$$\frac{dp}{dt} = k_{\text{on}} c(1 - p) - \beta_{\text{off}} p. \tag{9.6}$$

At equilibrium,

$$p_* = \frac{k_{\text{on}} c}{k_{\text{on}} c + \beta_{\text{off}}} = \frac{c}{c + K_d}, \qquad K_d = \frac{\beta_{\text{off}}}{k_{\text{on}}}. \tag{9.7}$$

The dissociation constant K_d has units of concentration. A smaller K_d means stronger binding. Transcription is allowed while the operator is unbound, so

$$P_{\text{unbound}}(c) = 1 - p_* = \frac{K_d}{c + K_d} = \frac{1}{1 + c/K_d}. \tag{9.8}$$

The relaxation of the binding state is also explicit. For fixed c,

$$p(t) - p_* = \big(p(0) - p_*\big) e^{-(k_{\text{on}} c + \beta_{\text{off}})t}. \tag{9.9}$$

The rapid-equilibrium approximation used below requires this timescale to be short compared with protein production, clearance, and dilution.

9.3.2 Cooperative binding and Hill functions

If n repressors bind cooperatively, a phenomenological extension is

$$P_{\text{unbound}}(c) = \frac{K_{1/2}^n}{c^n + K_{1/2}^n} = \frac{1}{1 + (c/K_{1/2})^n}. \tag{9.10}$$

For a microscopic simultaneous-binding model, n is an integer. In experimental fits, the Hill coefficient may be noninteger because it summarizes more complicated molecular interactions. Here $K_{1/2}$ is the concentration at half response. It equals the microscopic dissociation constant $K_d = \beta_{\text{off}}/k_{\text{on}}$ for the one-site model above, but a fitted cooperative Hill curve does not generally identify a single microscopic binding constant. With $\bar{c} = c/K_{1/2}$,

$$P_{\text{unbound}} = \frac{1}{1 + \bar{c}^n}. \tag{9.11}$$

The limits are informative:

$$P_{\text{unbound}} \simeq \begin{cases} 1, & \bar{c} \ll 1, \\ \bar{c}^{-n}, & \bar{c} \gg 1. \end{cases} \tag{9.12}$$

Thus the high-concentration branch has slope $-n$ on a log-log plot. Larger n gives a sharper transition near $c = K_{1/2}$.

For $n > 1$, the decreasing Hill curve has an inflection point. Setting its second derivative to zero gives

$$\bar{c}_{\text{inf}} = \left(\frac{n-1}{n+1}\right)^{1/n}. \tag{9.13}$$

For $n = 1$, this value lies at the boundary $\bar{c} = 0$, so there is no interior inflection point.

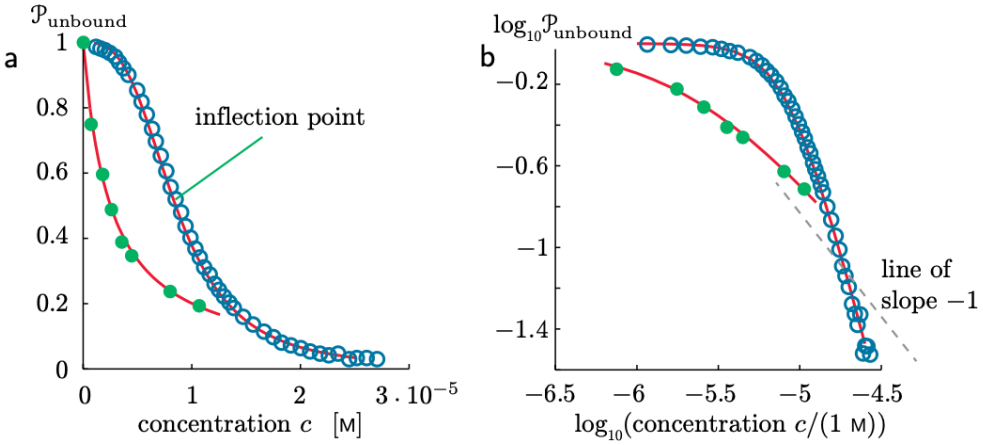

Figure 9.4: Binding curves for oxygen-binding proteins. The cooperative curve has an interior inflection point, while the noncooperative curve does not. At high concentration, the log-log representation approaches the predicted power-law slope (Mills et al. 1976; Rossi-Fanelli and Antonini 1958).

9.4 Gene regulation, clearance, and dilution

Let $g_{\max}$ be the effective maximum rate of increase of protein concentration. In a balanced-growth concentration model, average gene dosage and protein-synthesis capacity scale with cell volume, so $g_{\max}$ is treated as constant over the cell cycle. Under rapid binding equilibration, the gene-regulation function is

$$f(c) = \frac{g_{\max}}{1 + (c/K_{1/2})^n}. \tag{9.14}$$

An explicit cell-cycle model could instead retain time-dependent volume and gene dosage; that added detail is outside the present coarse-grained description. Production alone cannot yield a steady concentration. Protein removal contributes

$$\left.\frac{dc}{dt}\right|_{\text{clearance}} = -k_\phi c. \tag{9.15}$$

Growth dilutes concentration even if protein molecules are not degraded. Week 8 derived the equivalent continuous dilution rate $k_{\text{dil}} = \ln 2/\tau_d$ for a division time τ_d; see Equation (8.71). Writing $\tau_e = \tau_d/\ln 2$, dilution therefore contributes $-c/\tau_e$. Clearance and dilution combine into

$$\frac{1}{\tau_{\text{tot}}} = k_\phi + \frac{1}{\tau_e}, \qquad \left.\frac{dc}{dt}\right|_{\text{loss}} = -\frac{c}{\tau_{\text{tot}}}. \tag{9.16}$$

The self-repressed gene obeys

$$\frac{dc}{dt} = \frac{g_{\max}}{1 + (c/K_{1/2})^n} - \frac{c}{\tau_{\text{tot}}}. \tag{9.17}$$

9.5 Graphical stability and noise suppression

Figure 9.5: Unregulated and self-repressed TetR–GFP gene circuits. In the feedback circuit, the protein product represses its own expression (Becskei and Serrano 2000).

A fixed point c_* occurs where production equals loss. In the unregulated model, production is constant while loss grows linearly with c. In the regulated model, production also decreases with c. Both curves therefore push the system back toward their intersection:

- if $c > c_*$, loss exceeds production and c decreases;
- if $c < c_*$, production exceeds loss and c increases.

For a general one-dimensional model $dc/dt = F(c)$, write $c = c_* + \delta c$ and linearize:

$$\frac{d\,\delta c}{dt} \simeq F'(c_*)\delta c. \tag{9.18}$$

The fixed point is stable when $F'(c_*) < 0$, with local relaxation rate

$$\lambda = -F'(c_*). \tag{9.19}$$

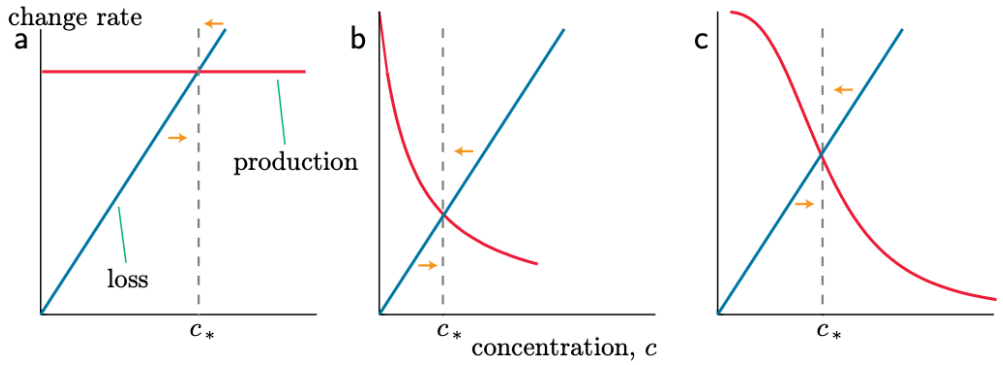

Figure 9.6: Production-loss diagrams for an unregulated gene, noncooperative negative feedback, and cooperative negative feedback. In each case the arrows point toward the unique intersection c_*, demonstrating stable fixed-point behavior (Nelson 2022).

Negative autoregulation makes the production derivative negative, increasing λ above the loss-only value $1/\tau_{\text{tot}}$. This establishes faster local relaxation, but noise suppression is not automatic: feedback can also change the event rates that generate fluctuations. In a local diffusion approximation with noise intensity $D(c_*)$, one expects $\text{Var}(c) \simeq D(c_*)/(2\lambda)$. Increasing λ narrows the distribution only when the corresponding change in $D(c_*)$ does not offset the stronger restoring drift. The comparison must therefore match the mean and relevant noise sources, or be tested directly in a stochastic model.

Becskei and Serrano tested this prediction with a synthetic *E. coli* circuit expressing a TetR–GFP fusion protein. TetR repressed the same promoter that produced it, while GFP reported the protein abundance in individual cells. Control strains lacked effective feedback but retained comparable maximum production and dilution rates.

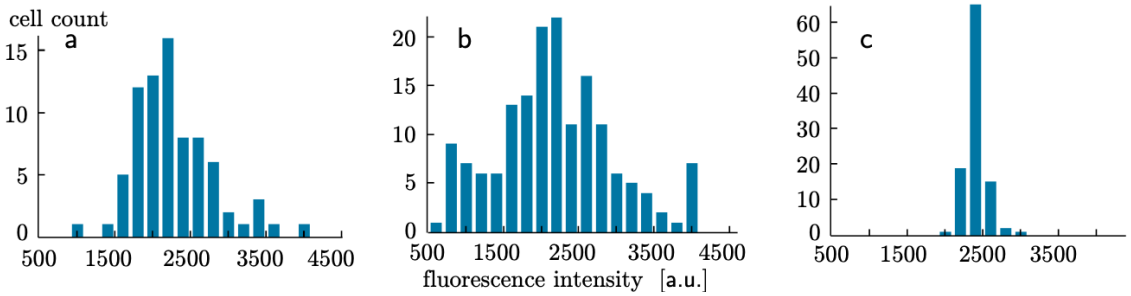

Figure 9.7: Negative autoregulation narrows the distribution of protein abundance across individual cells. The feedback-disabled controls are shown in panels (a) and (b), and the negatively autoregulated circuit is shown in panel (c). Data from Becskei and Serrano (2000); figure reproduced from Nelson (2022).

9.6 Regulated and unregulated recovery dynamics

For a noncooperative repressor, $n = 1$, so $K_{1/2} = K_d$. In the strong-repression regime $c \gg K_d$, Equation (9.17) becomes

$$\frac{dc}{dt} \simeq \frac{g_{\max} K_d}{c} - \frac{c}{\tau_{\rm tot}}. \tag{9.20}$$

Letting $y = c^2$ gives a linear equation,

$$\frac{dy}{dt} = 2\left(g_{\max} K_d - \frac{y}{\tau_{\rm tot}}\right). \tag{9.21}$$

Its fixed point is

$$y_* = \tau_{\rm tot} g_{\max} K_d, \qquad c_* = \sqrt{\tau_{\rm tot} g_{\max} K_d}. \tag{9.22}$$

Let t_0 be a time at which the full dynamics has reached $c_0 = c(t_0) \gg K_d$, so that the strong-repression approximation has become valid. Its solution from that point onward is

$$\frac{c(t)}{c_*} = \left[1 + \left(\frac{c_0^2}{c_*^2} - 1\right) e^{-2(t-t_0)/\tau_{\rm tot}}\right]^{1/2}. \tag{9.23}$$

The approximate production term $g_{\max} K_d / c$ is singular near $c = 0$, so it must not be initialized at zero. Early induction from zero is governed by the full Hill function in Equation (9.17); Equation (9.23) describes the later strong-repression regime. Near the fixed point,

$$\frac{c(t)}{c_*} \simeq 1 + \frac{1}{2}\left(\frac{c_0^2}{c_*^2} - 1\right) e^{-2(t-t_0)/\tau_{\rm tot}}. \tag{9.24}$$

By contrast, an unregulated gene with the same linear loss obeys

$$\frac{c(t)}{c_*^{(0)}} = 1 - e^{-t/\tau_{\rm tot}}. \tag{9.25}$$

Thus the self-repressed system has twice the late-time relaxation rate in this regime. More generally, if the high-concentration production law scales as $f(c) \propto c^{-n}$, then write $f(c) = Ac^{-n}$ and $F(c) = f(c) - c/\tau_{\rm tot}$. At the fixed point, $f(c_*) = c_*/\tau_{\rm tot}$, while

$$f'(c_*) = -n\frac{f(c_*)}{c_*} = -\frac{n}{\tau_{\rm tot}}. \tag{9.26}$$

Therefore $F'(c_*) = -(n+1)/\tau_{\rm tot}$, and the local relaxation rate is

$$\lambda \simeq \frac{n+1}{\tau_{\rm tot}}. \tag{9.27}$$

This local statement is robust even when the early-time strong-repression approximation is not valid.

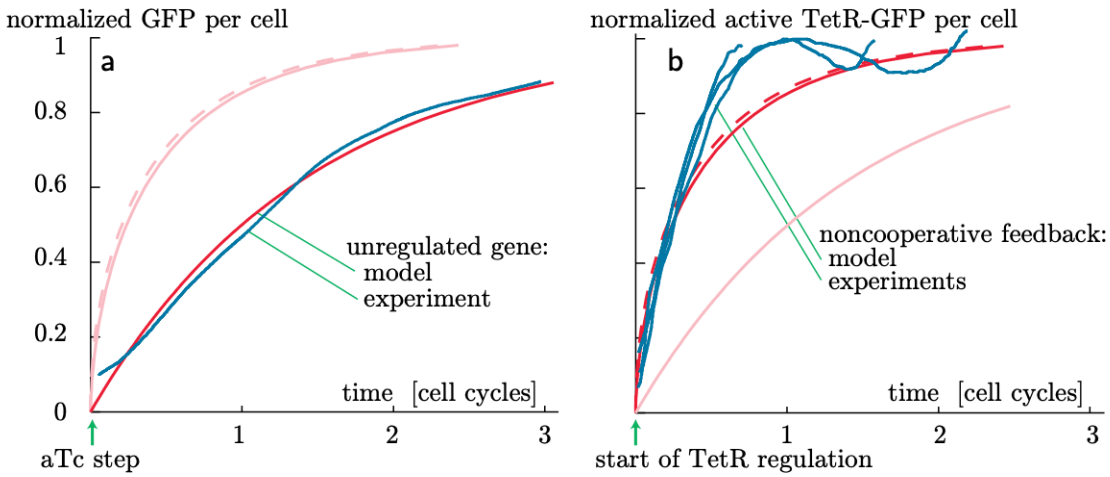

Figure 9.8: Theory and experiment for the kinetics of gene autoregulation. The unregulated gene approaches its setpoint slowly, while TetR negative feedback produces a much faster initial response. Overshoot in the regulated traces indicates dynamics beyond the one-variable approximation (Rosenfeld, Elowitz, et al. 2002).

The comparison illustrates a general modeling lesson. Agreement in the initial relaxation supports the feedback mechanism, while systematic overshoot identifies missing state variables or delays. A useful model should expose both what it explains and where it fails.

9.7 The same mathematics in DNA evolution

Feedback regulation and sequence evolution address different biological questions, but both can be formulated with continuous-time Markov chains. The Jukes–Cantor model (JC69) (Jukes and Cantor 1969) is a minimal neutral model for nucleotide substitutions at one site. Its state space is $\{A, C, G, T\}$, and it assumes:

- equal equilibrium base frequencies, $\pi_A = \pi_C = \pi_G = \pi_T = 1/4$;
- equal rates from a base to each of the other three bases;
- exponential waiting times and the Markov property; and
- independent evolution of sites.

The parameter μ below is a *substitution rate*: it counts changes that become fixed in the population and therefore appear along a lineage. It is not automatically the raw mutation probability in one reproductive event. Week 13 will derive the population- genetic relation $R = Nu\rho$ between mutation supply, fixation probability, and substitution rate. In the neutral Moran model, $\rho = 1/N$, so the population size cancels and the substitution rate equals the individual mutation rate, $R = u$. That result supplies a microscopic basis for the molecular clock used here.

Let μ be the total substitution rate out of the current base. With the row-vector convention, the generator is

$$Q = \begin{pmatrix} -\mu & \mu/3 & \mu/3 & \mu/3 \\ \mu/3 & -\mu & \mu/3 & \mu/3 \\ \mu/3 & \mu/3 & -\mu & \mu/3 \\ \mu/3 & \mu/3 & \mu/3 & -\mu \end{pmatrix}. \tag{9.28}$$

The transition matrix after time t is $P(t) = e^{Qt}$. Symmetry lets us obtain its entries from a single scalar differential equation.

9.7.1 Probability of retaining the ancestral base

Fix the ancestral nucleotide i, and let

$$p(t) = \Pr(X_t = i \mid X_0 = i). \tag{9.29}$$

In a short interval Δt, a site already in state i remains there with probability $1 - \mu\Delta t + o(\Delta t)$. From any one of the other three states, it changes specifically to i with probability $(\mu/3)\Delta t + o(\Delta t)$. Therefore

$$p(t + \Delta t) = (1 - \mu\Delta t)p(t) + \frac{\mu}{3}\Delta t\,[1 - p(t)] + o(\Delta t), \tag{9.30}$$

$$\frac{dp}{dt} = \frac{\mu}{3} - \frac{4\mu}{3}p = -\frac{4\mu}{3}\left(p - \frac{1}{4}\right). \tag{9.31}$$

With $p(0) = 1$,

$$p(t) = \frac{1}{4} + \frac{3}{4}e^{-4\mu t/3}. \tag{9.32}$$

The probability of a mismatch from the ancestral base is

$$P_{\text{diff}}(t) = 1 - p(t) = \frac{3}{4}\left(1 - e^{-4\mu t/3}\right). \tag{9.33}$$

As $t \to \infty$, the ancestral base is retained with probability $1/4$, and the mismatch probability saturates at $3/4$. This saturation reflects repeated and reversing substitutions, not a cessation of evolution.

9.7.2 The Jukes–Cantor distance correction

For an alignment of two homologous sequences, estimate the observed mismatch proportion as

$$\widehat{P}_{\text{diff}} = \frac{\text{number of mismatches}}{\text{number of aligned sites}}. \tag{9.34}$$

Equation (9.33) describes one lineage whose branch length is $K = \mu t$. For two sequences, the same transition law applies along the complete path joining them: if their branch lengths from a common ancestor are K_1 and K_2, the pairwise distance is $d = K_1 + K_2$. Inverting the mismatch probability therefore gives

$$\widehat{d}_{\text{JC}} = -\frac{3}{4}\ln\left(1 - \frac{4}{3}\widehat{P}_{\text{diff}}\right). \tag{9.35}$$

Thus d_{JC} is the expected number of substitutions per site along the total path between the sampled sequences. For an ancestor–descendant comparison it equals μt. For two contemporaneous descendants with equal elapsed time t, it equals $2\mu t$; division by two is justified only under that equal-clock assumption.

For example, 100 mismatches among 1000 aligned sites give

$$\widehat{P}_{\text{diff}} = 0.1, \qquad \widehat{d}_{\text{JC}} = -\frac{3}{4}\ln\left(1 - \frac{4}{3}(0.1)\right) \simeq 0.107. \tag{9.36}$$

The corrected distance exceeds the raw mismatch fraction because some sites have changed more than once. For small divergence, $-\ln(1-x) \simeq x$, so $\widehat{d}_{\mathrm{JC}} \simeq \widehat{P}_{\mathrm{diff}}$. The correction requires $\widehat{P}_{\mathrm{diff}} < 3/4$; as the observed fraction approaches the saturation value, the inferred distance becomes arbitrarily uncertain. For an alignment of L independent sites, a binomial approximation gives $\mathrm{SE}(\widehat{P}_{\mathrm{diff}}) \simeq \sqrt{\widehat{P}_{\mathrm{diff}}(1-\widehat{P}_{\mathrm{diff}})/L}$. Error propagation through Equation (9.35) then gives

$$\mathrm{SE}(\widehat{d}_{\mathrm{JC}}) \simeq \frac{\sqrt{\widehat{P}_{\mathrm{diff}}(1-\widehat{P}_{\mathrm{diff}})/L}}{1-\frac{4}{3}\widehat{P}_{\mathrm{diff}}}. \tag{9.37}$$

The denominator makes the loss of precision near saturation explicit. Correlated sites, alignment uncertainty, and rate variation make this binomial standard error optimistic.

9.7.3 What JC69 leaves out

JC69 is analytically transparent and useful as a neutral baseline, but real sequence evolution often violates its assumptions:

- equilibrium base frequencies can be unequal;
- transitions and transversions need not occur at the same rate;
- mutation rates can depend on neighboring bases and genomic context;
- different sites can have different rates or selective constraints; and
- sites may not evolve independently.

The Kimura two-parameter model (Kimura 1980) separates transitions ($A \leftrightarrow G,\ C \leftrightarrow T$) from transversions. The general time-reversible (GTR) model allows unequal equilibrium frequencies and a symmetric exchangeability parameter $r_{ij} = r_{ji}$ for every nucleotide pair. Under the row convention,

$$q_{ij} = r_{ij}\pi_j \quad (i \neq j), \qquad q_{ii} = -\sum_{j \neq i} q_{ij}. \tag{9.38}$$

Then $\pi_i q_{ij} = \pi_j q_{ji}$, which is the detailed-balance condition. These extensions retain the Markov-chain framework while relaxing specific biological assumptions.

An LLM can help construct a rate matrix, calculate a matrix exponential, or compare predicted and observed mismatch fractions. The modeler must still check that every off-diagonal rate is nonnegative, every row of Q sums to zero, $P(t)$ is stochastic, and the chosen parameter convention matches the meaning assigned to branch length.

9.8 Summary

- Negative feedback creates a stable fixed point by making deviations generate a restoring drift.
- Microscopic binding and unbinding rates determine one-site occupancy and K_d, whereas a phenomenological Hill curve uses a half-response scale $K_{1/2}$ that need not be a single microscopic dissociation constant.
- Cooperative binding produces a Hill response; the Hill coefficient controls the sharpness and high-concentration power-law slope.

- Protein concentration changes through regulated production, molecular clearance, and dilution by cell growth.
- Negative autoregulation can accelerate relaxation and narrow cell-to-cell protein variation, while overshoot reveals missing dynamics.
- The Jukes–Cantor model applies the same continuous-time Markov reasoning to DNA substitution and yields a correction for hidden multiple changes.

10 Positive feedback, chemostats, and epidemic dynamics

Positive feedback occurs whenever a system's present activity promotes still more activity. Cells reproduce, infective individuals create new infective individuals, and autocatalytic reactions make more of their own catalyst. Left unconstrained, the simplest positive-feedback law gives exponential growth. Real biological systems are constrained by resources, clearance, dilution, or depletion of susceptible hosts. Those constraints create thresholds, stable states, and bifurcations.

This week develops two closely related examples. The chemostat couples a growing microbial population to a limiting nutrient and to continuous dilution. Epidemic models couple an infective population to susceptible hosts and recovery. In both cases, the active population reinforces its own production while simultaneously changing the environment that permits that production.

By the end of this week, students should be able to answer:

- How does positive feedback create exponential growth, and what arrests it?
- How do mass-balance arguments produce the chemostat equations?
- When does a microbial population persist, and when is it washed out?
- How do chemostats and serial dilution create different laboratory environments for microbial competition and evolution?
- How does competition for one limiting resource lead to the R^* rule?
- How do the SIS, SIR, and SIRS epidemic models encode transmission, recovery, and immunity?
- What does the basic reproduction number R_0 tell us about invasion and epidemic peaks?

10.1 Autocatalysis and positive feedback

Let $B(t)$ be the number or concentration of bacteria in an environment where every cell divides at a constant rate g. The mean-field law

$$\frac{dB}{dt} = gB \tag{10.1}$$

has the solution

$$B(t) = B(0)e^{gt}. \tag{10.2}$$

The production rate is proportional to the amount already present: bacteria make bacteria. This is the defining structure of positive feedback. Equation (10.1) cannot remain valid indefinitely, however, because every new bacterium requires matter and energy. A physically complete model must track at least one limiting resource and the routes by which cells and resource enter or leave the system.

10.2 The chemostat: a controlled microbial ecosystem

A chemostat is a well-mixed vessel of fixed volume V. Fresh medium with nutrient concentration c_{in} enters at volumetric flow rate Q, while an equal volume of culture leaves. Let $c(t)$ be the nutrient concentration and $\rho(t)$ the bacterial concentration in the chamber. The dilution rate and residence time are

$$D = \frac{Q}{V}, \qquad T = \frac{1}{D} = \frac{V}{Q}. \tag{10.3}$$

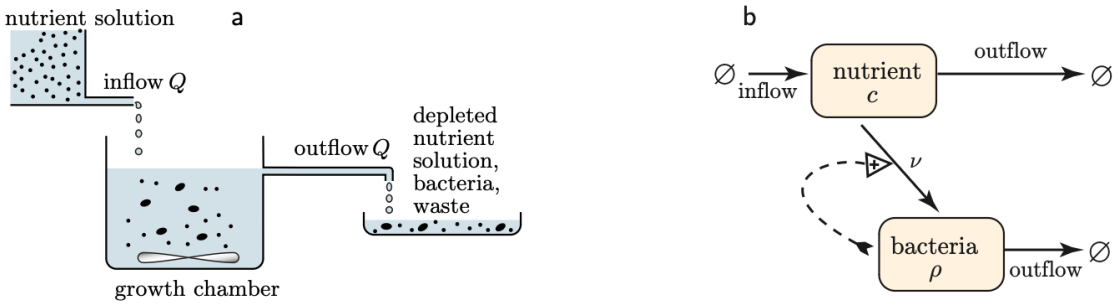

Figure 10.1: A chemostat couples microbial growth to continuous nutrient inflow and dilution. Nutrient solution enters a stirred growth chamber at rate Q, and an equal outflow maintains constant volume. The network view emphasizes that nutrient is converted into bacteria while both components are diluted (Nelson 2022).

10.2.1 Growth law and dimensional balances

Growth is limited by nutrient availability. A standard saturating law is the Monod function

$$g(c) = g_{\max}\frac{c}{K + c}, \tag{10.4}$$

where $g_{\max}$ is the maximal growth rate and K is the nutrient concentration at half-maximal growth. Suppose ν nutrient molecules are required to produce one bacterium. Equivalently, one may use a yield $Y = 1/\nu$ measured in bacteria produced per unit nutrient consumed.

The bacterial balance is production minus washout:

$$\frac{d\rho}{dt} = \rho g(c) - D\rho = \rho\big[g(c) - D\big]. \tag{10.5}$$

The nutrient balance contains inflow, outflow, and consumption:

$$\frac{dc}{dt} = D\big(c_{\text{in}} - c\big) - \nu\rho g(c). \tag{10.6}$$

Each term follows directly from a flux. In particular, the outflow removes bacteria at rate $D\rho$, even if the cells do not die inside the vessel.

10.2.2 Nondimensionalization

Use K as the concentration scale and D^{-1} as the time scale:

$$\tau = Dt, \qquad x = \frac{c}{K}, \qquad x_{\text{in}} = \frac{c_{\text{in}}}{K}, \qquad y = \frac{\nu\rho}{K}, \qquad \theta = \frac{g_{\max}}{D} = g_{\max}T. \tag{10.7}$$

Equations (10.5) and (10.6) become

$$\frac{dx}{d\tau} = x_{\text{in}} - x - y\,\theta\frac{x}{1 + x}, \tag{10.8}$$

$$\frac{dy}{d\tau} = y\left(\theta\frac{x}{1 + x} - 1\right). \tag{10.9}$$

The model now has only two control parameters: the scaled input nutrient x_{in} and the ratio θ of maximum growth rate to dilution rate.

Adding the two equations reveals a useful balance law:

$$\frac{d}{d\tau}(x+y) = x_{\mathrm{in}} - (x+y). \tag{10.10}$$

Thus the total scaled nutrient plus biomass relaxes to $x + y = x_{\mathrm{in}}$. The nonlinear growth terms merely transfer material between the two pools.

10.2.3 Nullclines and fixed points

The bacterial nullcline is

$$y = 0 \qquad \text{or} \qquad \theta\frac{x}{1+x} = 1, \tag{10.11}$$

while the nutrient nullcline is

$$y = \frac{(x_{\mathrm{in}} - x)(1+x)}{\theta x}. \tag{10.12}$$

Their intersections give two possible steady states. The washout state is

$$(x_{\mathrm{w}}, y_{\mathrm{w}}) = (x_{\mathrm{in}}, 0). \tag{10.13}$$

The coexistence state of nutrient and bacteria is

$$x_* = \frac{1}{\theta - 1}, \qquad y_* = x_{\mathrm{in}} - \frac{1}{\theta - 1}. \tag{10.14}$$

It is physically admissible only when $y_* > 0$, or

$$\theta > 1 + \frac{1}{x_{\mathrm{in}}}. \tag{10.15}$$

The stability calculation makes the threshold transparent. At washout, a rare bacterial population grows according to

$$\frac{1}{y}\frac{dy}{d\tau} = \theta\frac{x_{\mathrm{in}}}{1+x_{\mathrm{in}}} - 1. \tag{10.16}$$

Washout is stable when this quantity is negative. At equality, the washout and coexistence branches exchange stability in a transcritical bifurcation. Biologically, cells persist only when their growth rate in the inflowing environment exceeds the rate at which the vessel is replaced.

The positive-biomass branch is also stable whenever it is physically admissible. If the right-hand sides of Equations (10.8)–(10.9) are denoted by $F(x,y)$ and $G(x,y)$, their Jacobian is

$$J(x,y) = \begin{pmatrix} -1 - y\dfrac{\theta}{(1+x)^2} & -\dfrac{\theta x}{1+x} \\[2ex] y\dfrac{\theta}{(1+x)^2} & \dfrac{\theta x}{1+x} - 1 \end{pmatrix}. \tag{10.17}$$

For any real 2×2 Jacobian, the characteristic polynomial is $\lambda^2 - (\operatorname{tr} J)\lambda + \det J = 0$. Both eigenvalues have negative real parts precisely when

$$\operatorname{tr} J < 0, \qquad \det J > 0. \tag{10.18}$$

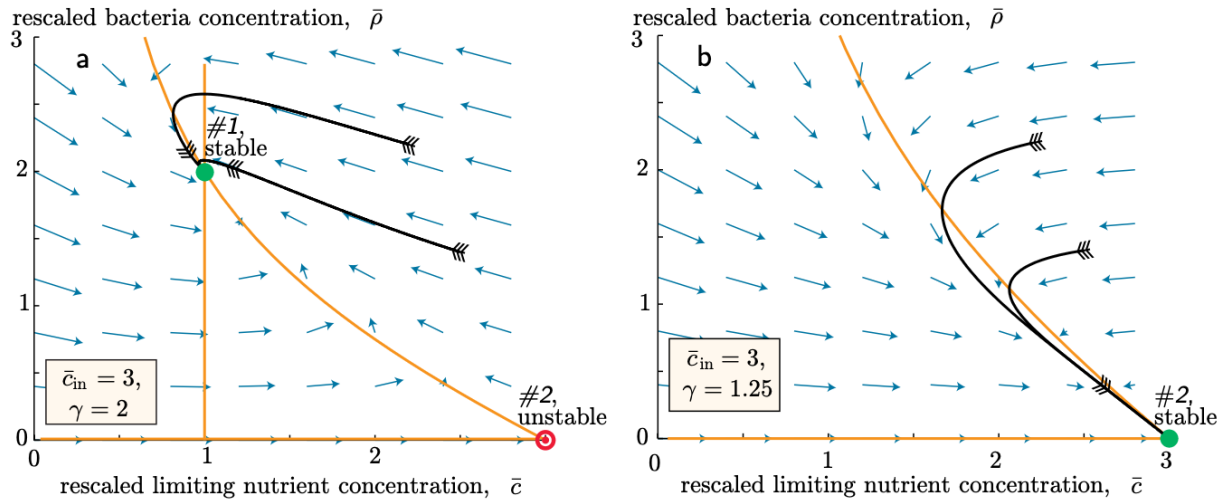

Figure 10.2: Chemostat phase portraits. At sufficiently large θ, trajectories approach the positive-biomass fixed point (left). When dilution is too fast relative to growth, only the washout state is stable (right). Orange curves are nullclines and black curves are representative trajectories (Nelson 2022).

If the eigenvalues are real, a positive determinant makes their signs agree and a negative trace makes both negative. If they are a complex-conjugate pair, each has real part $\operatorname{tr} J/2 < 0$. At coexistence, $\theta x_*/(1 + x_*) = 1$. Defining

$$\kappa_* = \frac{y_* \theta}{(1 + x_*)^2} > 0, \tag{10.19}$$

gives

$$J_* = \begin{pmatrix} -1 - \kappa_* & -1 \\ \kappa_* & 0 \end{pmatrix}, \qquad \operatorname{tr} J_* = -1 - \kappa_* < 0, \qquad \det J_* = \kappa_* > 0. \tag{10.20}$$

For a two-dimensional system, negative trace and positive determinant place both eigenvalues in the left half-plane, so the coexistence state is locally asymptotically stable. As the threshold is approached from above, $y_* \to 0$, hence $\kappa_* \to 0$, and one eigenvalue reaches zero at the transcritical exchange of stability.

At the coexistence state, $g(c_*) = D$: the population adjusts the residual nutrient until its per-capita birth rate exactly balances dilution. The mass balance then sets the biomass. This is a useful separation of roles: growth kinetics sets the resource level, whereas input resource sets the amount of biomass that can be supported.

10.3 Alternative uptake laws and experimental protocols

The Monod function is linear at low nutrient and saturates at high nutrient:

$$g(c) \simeq \begin{cases} (g_{\max}/K)c, & c \ll K, \\ g_{\max}, & c \gg K. \end{cases} \tag{10.21}$$

Other biological mechanisms produce different functional responses. A strictly linear law does not saturate, while a cooperative Hill law

$$g_h(c) = g_{\max} \frac{c^h}{K^h + c^h} \tag{10.22}$$

is more strongly suppressed at low concentration when $h > 1$.

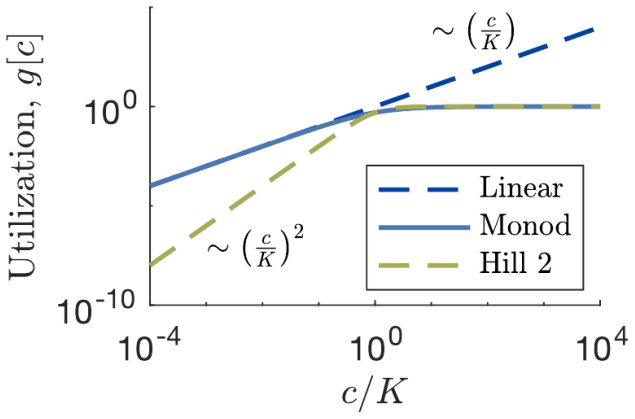

Figure 10.3: Linear, Monod, and Hill-2 nutrient-utilization laws. Monod growth is linear for $c \ll K$ and saturates for $c \gg K$; a cooperative Hill response has a steeper low-resource dependence (Lopez, Hein, et al. 2024).

The ideal chemostat assumes a well-mixed environment, one limiting resource, and constant parameters. Many laboratory evolution experiments instead use serial dilution. Cells grow in a closed batch after a nutrient bolus; a fraction of the culture then seeds the next batch. Steady state is therefore a fixed point of a discrete batch-to-batch map, even though the within-batch dynamics are continuous and nonstationary.

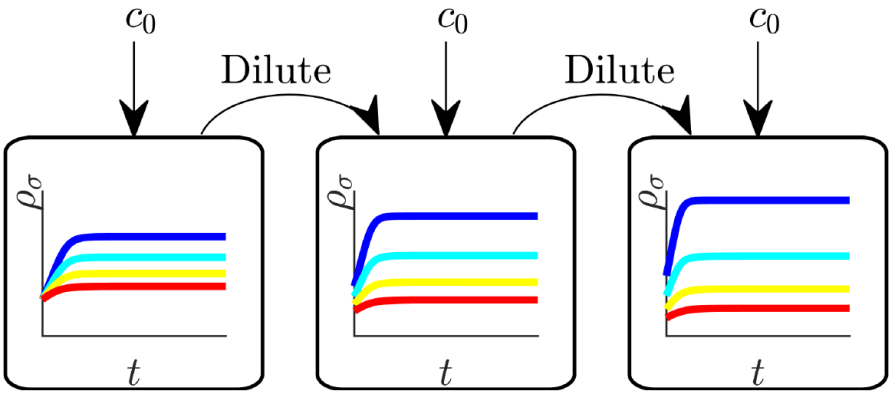

Figure 10.4: Serial-dilution protocol. Nutrient is added at the beginning of each batch and a fraction of the grown culture inoculates the next batch. The periodic environment can favor strategies different from those selected in a constant chemostat (Erez, Lopez, et al. 2020).

10.3.1 The Lenski long-term evolution experiment

The long-term evolution experiment with *E. coli* (Lenski et al. 1991) turns serial dilution into an experimental fossil record. Replicate populations grow each day from roughly 10^6 to 10^8 cells and

are then diluted about one hundredfold into fresh medium. Samples are periodically frozen, so ancestral and evolved cells can later be revived and competed in the same environment. Fitness is operationally defined by that competition: if two strains begin at abundances $N_1(0)$ and $N_2(0)$, their ratio after a common growth period T is approximately, during a phase in which both have constant exponential growth rates,

$$\frac{N_1(T)}{N_2(T)} = \frac{N_1(0)}{N_2(0)} e^{(\Lambda_1 - \Lambda_2)T}, \tag{10.23}$$

where Λ_i is the strain's exponential growth rate in that environment. A common dilution factor does not change the ratio, so repeated cycles amplify small fitness differences. Across a complete daily growth cycle, however, lag time, resource-dependent growth, yield, and stationary-phase survival can all contribute. A realized Malthusian selection rate can still be defined from the measured change in the abundance ratio,

$$s_{\text{cycle}} = \frac{1}{T} \ln\left[\frac{N_1(T)/N_2(T)}{N_1(0)/N_2(0)}\right], \tag{10.24}$$

but interpreting it as $\Lambda_1 - \Lambda_2$ requires the constant-rate approximation.

A related constant-population idealization makes the deterministic selection dynamics especially transparent. Let $x = N_1/N$ be the fraction of type 1 in a population whose total size is held fixed by replacing one cell whenever another divides. Mean-field dynamics give

$$\frac{dx}{dt} = (\Lambda_1 - \Lambda_2)x(1 - x). \tag{10.25}$$

If $\Lambda_1 > \Lambda_2$, every initial condition with $x > 0$ approaches fixation at $x = 1$. This deterministic statement is appropriate once both types are abundant. It fails at the moment a mutation first appears, when $x = 1/N$ represents one discrete cell. That lineage can disappear by chance even when it has a growth advantage. Week 13 replaces Equation (10.25) by the finite-population Moran process and calculates the probability of fixation.

10.4 Competition for one limiting resource

Now let two microbial types, A and B, consume the same nutrient. Their growth laws and yields may differ:

$$\frac{dc}{dt} = D(c_{\text{in}} - c) - \nu_A \rho_A g_A(c) - \nu_B \rho_B g_B(c), \tag{10.26}$$

$$\frac{d\rho_A}{dt} = \rho_A\big[g_A(c) - D\big], \tag{10.27}$$

$$\frac{d\rho_B}{dt} = \rho_B\big[g_B(c) - D\big]. \tag{10.28}$$

For Monod growth,

$$g_j(c) = g_{\max,j}\frac{c}{K_j + c}, \qquad j \in \{A, B\}. \tag{10.29}$$

The key quantity is the resource concentration R_j^* at which type j just balances dilution:

$$g_j(R_j^*) = D \qquad \Longrightarrow \qquad R_j^* = \frac{DK_j}{g_{\max,j} - D} = \frac{K_j}{g_{\max,j}T - 1}, \tag{10.30}$$

provided $g_{\max,j} > D$. Below R_j^*, that type declines; above R_j^*, it grows.

If $R_A^* < R_B^*$, type A can reduce the common resource to a level at which type B has negative net growth. Type A therefore excludes type B at steady state. This is the R^* rule: in a constant, well-mixed environment with one limiting resource, the competitor capable of persisting at the lowest resource concentration wins. Maximum growth rate alone does not determine the outcome.

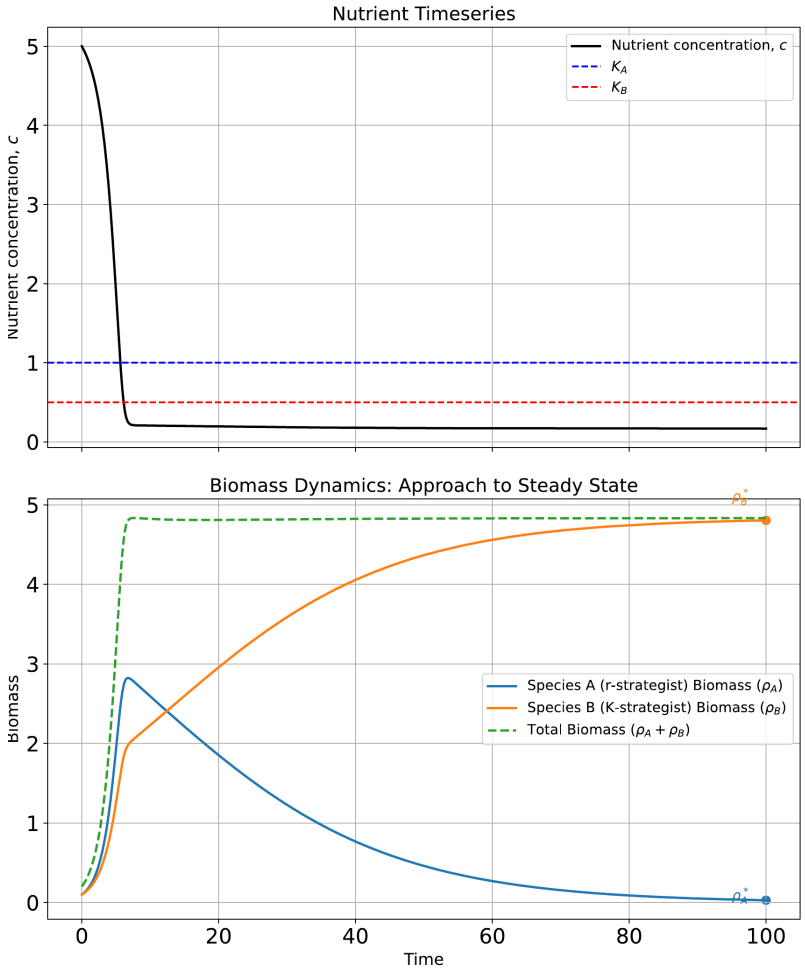

Figure 10.5: Competition between a fast-growing type and a more resource-efficient type. Nutrient depletion eventually places the fast-growing type below its break-even resource level, while the lower-R^* competitor persists (course-generated).

The simple result is also a puzzle. Natural communities often contain many species despite apparently sharing a small number of limiting resources—the paradox of the plankton. Coexistence can be maintained when the assumptions behind competitive exclusion fail:

- several resources limit growth, and species have different resource requirements;
- the environment varies in time, so no single strategy is always best;
- space is heterogeneous and dispersal is incomplete;

- predators, parasites, or phages preferentially suppress abundant competitors;
- physiological tradeoffs couple high maximal growth to poor low-resource performance.

The R^* rule is therefore a baseline prediction, not a universal claim of low diversity. Its value is that observed coexistence points directly to missing ecological structure.

10.5 Epidemic growth as positive feedback

An epidemic is another autocatalytic process: infective hosts generate new infective hosts. Susceptible depletion and recovery play the roles of resource depletion and dilution. We describe the population by fractions, so each model conserves total population.

10.5.1 The SIS model

In an SIS infection, recovery returns an individual immediately to the susceptible class. Let s and $i = 1 - s$ be the susceptible and infective fractions. Transmission occurs at rate βsi and recovery at rate γi:

$$\frac{ds}{dt} = -\beta s(1-s) + \gamma(1-s) = (1-s)(\gamma - \beta s). \tag{10.31}$$

$$\frac{di}{dt} = \beta si - \gamma i = i(\beta s - \gamma) = -\frac{ds}{dt}. \tag{10.32}$$

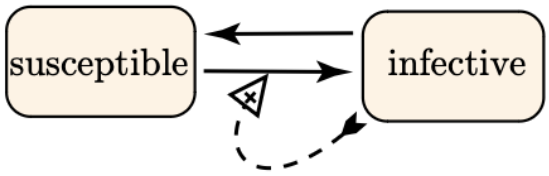

Figure 10.6: SIS network. Infective individuals transmit infection to susceptible individuals; recovery returns them to susceptibility. The dashed loop indicates positive feedback in the production of infectives. Adapted from Nelson (2022).

Define the basic reproduction number

$$R_0 = \frac{\beta}{\gamma}. \tag{10.33}$$

The fixed points are the disease-free state $s_* = 1$ and, when $R_0 > 1$, the positive endemic state

$$s_* = \frac{\gamma}{\beta} = \frac{1}{R_0}, \qquad i_* = 1 - \frac{1}{R_0}. \tag{10.34}$$

Linearization shows that the disease-free state is stable for $R_0 < 1$, while the endemic state is stable for $R_0 > 1$. The branches exchange stability at $R_0 = 1$, again through a transcritical bifurcation. The analogy with the chemostat is exact at the conceptual level: an active population can invade only when its per-capita production exceeds its removal rate.

10.5.2 The SIR model

For infections that confer lasting immunity, recovered individuals enter a third class:

$$\frac{ds}{dt} = -\beta si, \tag{10.35}$$

$$\frac{di}{dt} = \beta si - \gamma i, \tag{10.36}$$

$$\frac{dr}{dt} = \gamma i, \qquad s + i + r = 1. \tag{10.37}$$

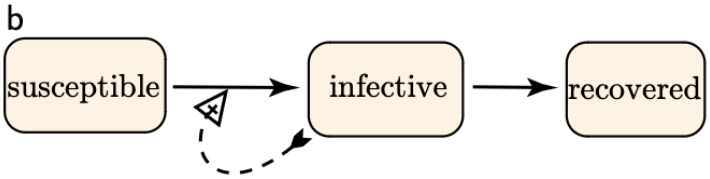

Figure 10.7: SIR network. Transmission moves susceptible individuals to the infective class; recovery then moves them to a permanently immune class (Nelson 2022).

Rescale time by the mean infectious period, $\tau = \gamma t$. Then

$$\frac{ds}{d\tau} = -R_0 si, \tag{10.38}$$

$$\frac{di}{d\tau} = (R_0 s - 1)i, \tag{10.39}$$

$$\frac{dr}{d\tau} = i. \tag{10.40}$$

At the beginning of an outbreak, $s \simeq s_0$ and

$$i(\tau) \simeq i_0 e^{(R_0 s_0 - 1)\tau}. \tag{10.41}$$

Thus invasion requires $R_0 s_0 > 1$. Vaccination or previous immunity acts by reducing s_0; the corresponding herd-immunity threshold in this homogeneous model is

$$1 - s_0 \geq 1 - \frac{1}{R_0}. \tag{10.42}$$

This threshold is an invasion condition, not a guarantee that no infections will occur. It assumes homogeneous mixing, a vaccine that removes the stated fraction completely from the susceptible class, a fixed R_0, and no demographic, spatial, age, or contact-network heterogeneity. Imperfect protection and heterogeneous contact patterns require a next-generation or structured model and generally change the coverage calculation.

Unlike SIS, SIR has no endemic fixed point in a closed population: i eventually vanishes. The infective fraction reaches its peak when

$$\frac{di}{d\tau} = 0, \qquad i > 0, \qquad \Longrightarrow \qquad s = \frac{1}{R_0}. \tag{10.43}$$

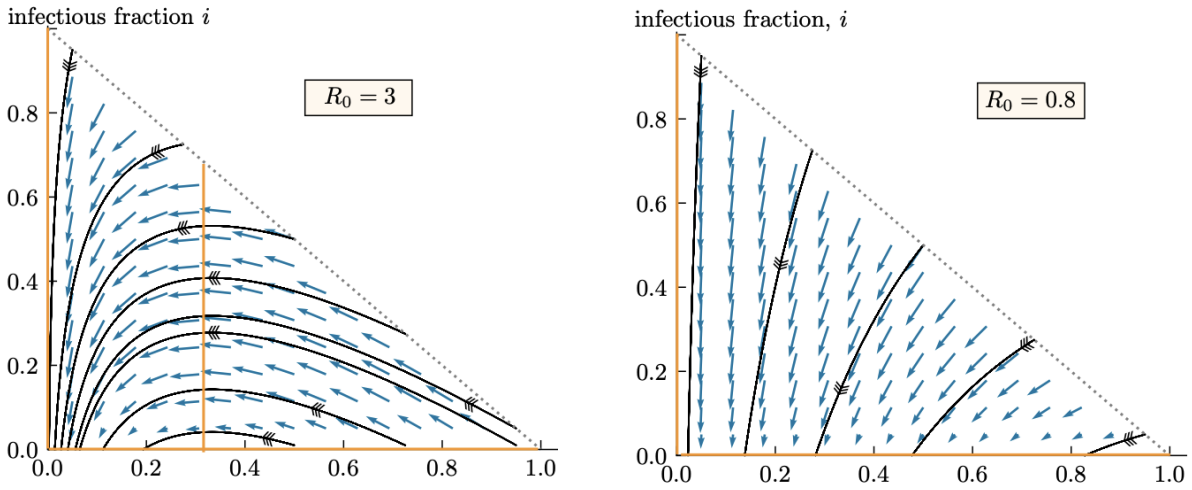

Figure 10.8: SIR phase portraits. For $R_0 = 3$, trajectories initially rise when $s > 1/R_0$ and fall after susceptible depletion carries them across the vertical nullcline (left). For $R_0 = 0.8$, the infective fraction decreases from every initial condition (right) (Nelson 2022).

Eliminating time from Equations (10.38) and (10.39) gives

$$\frac{di}{ds} = \frac{di/d\tau}{ds/d\tau} = \frac{R_0 si - i}{-R_0 si} = -1 + \frac{1}{R_0 s}, \tag{10.44}$$

where the factor i cancels wherever $i > 0$. Integrating from the initial point (s_0, i_0) to (s, i) gives

$$i - i_0 = -(s - s_0) + \frac{1}{R_0} \ln\left(\frac{s}{s_0}\right). \tag{10.45}$$

Rearranging yields the trajectory invariant

$$i + s - \frac{1}{R_0} \ln s = i_0 + s_0 - \frac{1}{R_0} \ln s_0. \tag{10.46}$$

At the end of the epidemic, $i_\infty = 0$. If $r_0 = 0$, the final susceptible fraction satisfies

$$\ln\left(\frac{s_\infty}{s_0}\right) = -R_0(1 - s_\infty), \qquad \text{or} \qquad s_\infty = s_0 e^{-R_0(1-s_\infty)}. \tag{10.47}$$

The threshold $s = 1/R_0$ marks the epidemic peak, not its endpoint: infections continue while $i > 0$, so the susceptible fraction overshoots below the threshold.

The measles example illustrates both the power and the limits of a minimal model. A small number of parameters captures the broad epidemic shape, but a homogeneous SIR model suppresses age structure, contact networks, reporting delays, spatial spread, and stochastic variation. Agreement with an aggregate curve should therefore be interpreted as support for the mechanism at that scale, not proof that every biological detail is correct.

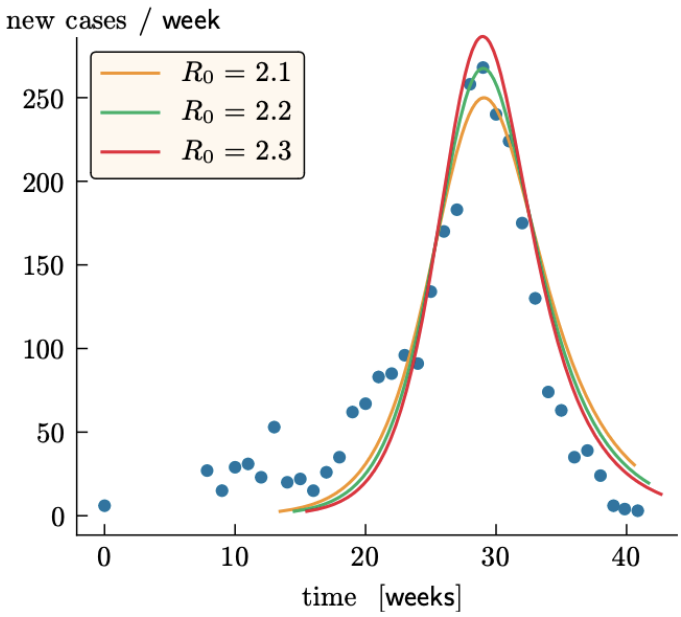

Figure 10.9: Measles outbreak in an undervaccinated subpopulation of 2816 people. Dots show weekly cases; curves show SIR predictions with mean infectious period $\gamma^{-1} = 20$ days and three values of R_0. The curves are aligned at their peaks because the outbreak start time was unknown. Data are from the CDC measles report; fits follow Nelson (2022).

10.5.3 Temporary immunity: the SIRS model

If immunity wanes at rate $\alpha\gamma$, recovered individuals return to susceptibility. With $r = 1 - s - i$ and $\tau = \gamma t$,

$$\frac{ds}{d\tau} = -R_0 si + \alpha(1 - s - i), \tag{10.48}$$

$$\frac{di}{d\tau} = (R_0 s - 1)i. \tag{10.49}$$

The disease-free state $(s, i) = (1, 0)$ is stable for $R_0 < 1$. For $R_0 > 1$, the endemic state is

$$s_* = \frac{1}{R_0}, \qquad i_* = \frac{\alpha(R_0 - 1)}{R_0(1 + \alpha)}, \qquad r_* = \frac{R_0 - 1}{R_0(1 + \alpha)}. \tag{10.50}$$

The SIRS model can approach the endemic state through damped oscillations: susceptible hosts accumulate while prevalence is low, transmission then rises, and the next wave depletes the susceptible pool.

An additional exposed but not yet infectious compartment gives the SEIR model. The latent period changes epidemic timing and transient growth, but the same modeling logic applies: identify states, enumerate transitions, write flux balances, and examine thresholds and fixed points.

10.6 Connecting chemostats and epidemics

The two applications share a common dynamical structure:

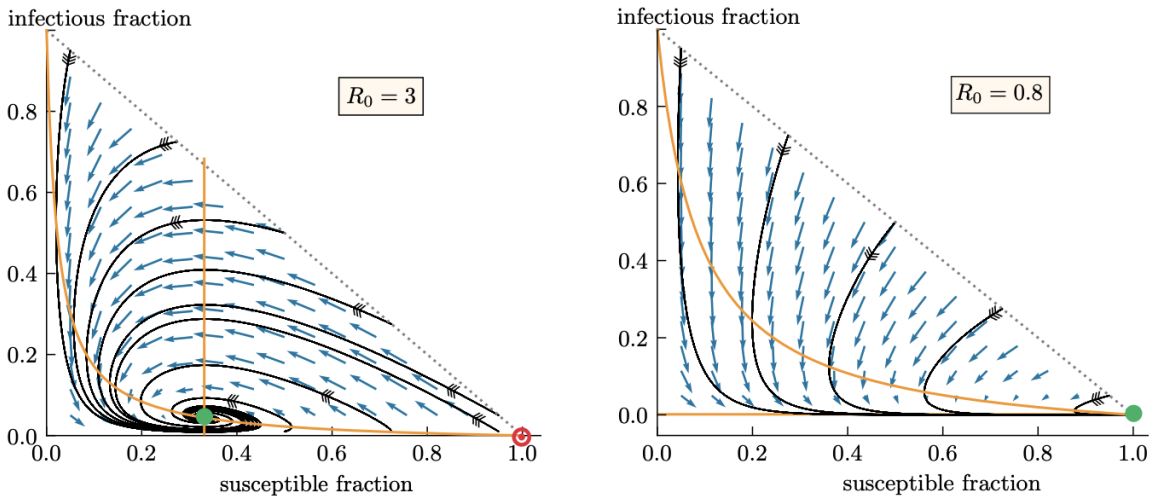

Figure 10.10: SIRS phase portraits with temporary immunity. For $R_0 = 3$, trajectories approach an endemic fixed point, often by damped oscillations (left). For $R_0 = 0.8$, infection is eradicated (right). Orange curves are nullclines (Nelson 2022).

Concept	Chemostat	Epidemic
Active population	bacteria	infective hosts
Enabling pool	nutrient	susceptible hosts
Positive feedback	bacteria produce bacteria	infectives produce infectives
Loss process	dilution	recovery
Invasion criterion	$g(c_{\text{in}}) > D$	$R_0 s_0 > 1$
Feedback constraint	nutrient depletion	susceptible depletion
Persistent state	positive biomass	SIS/SIRS endemic infection

In both systems, the threshold is a statement about a rare active population placed in an otherwise unperturbed environment. Above threshold, positive feedback amplifies that population. Growth then modifies the enabling pool, and the nonlinear feedback determines whether the system reaches a new steady state or produces a transient pulse. This pattern—an invasion calculation followed by resource or susceptible depletion—is a reusable modeling template far beyond the two examples considered here.

11 Switches, bistability, and gene expression

Cells do more than respond proportionally to their environment. A regulatory network can make a discrete choice, retain a memory of that choice, and switch only when a signal crosses a threshold. The physical basis of this behavior is positive feedback combined with nonlinearity. In a bistable regime, two stable states are separated by a watershed—the separatrix—so nearly identical initial conditions can lead to different fates.

This week develops that idea through the lac system in *E. coli*, the classic Novick-Weiner induction experiments, a general stochastic description of biochemical feedback, and a synthetic two-gene toggle.

By the end of this week, students should be able to answer:

- Why does bacterial growth on two sugars occur in two distinct phases?
- Can a partially induced population consist of partially active cells, or of a mixture of fully on and fully off cells?
- How do dilution and random switching determine the time course of a bulk enzyme measurement?
- What are bistability, hysteresis, and cellular memory?
- How can a stationary single-cell distribution reveal an effective feedback law?
- Why can two mutually repressing genes form a toggle switch?

11.1 Diauxic growth: bacteria change metabolic programs

When a single nutrient is abundant, bacterial abundance grows approximately exponentially, as in Week 10. Jacques Monod discovered that a mixture of two sugars can produce a more surprising trajectory. The culture first grows on its preferred carbon source, pauses after that source is exhausted, and then resumes growth using the second source. Monod called this two-stage behavior *diauxie*.

The lag is a regulatory delay, not merely a shortage of material. Each nutrient requires a specific set of transporters and metabolic enzymes. Carrying every possible pathway at high abundance would waste cellular resources and slow growth, so *E. coli* normally expresses only the machinery appropriate to its present environment. Lactose utilization requires, among other proteins, β-galactosidase (β-gal), which cleaves lactose into simpler sugars. A fully induced cell can contain thousands of β-gal molecules, whereas an uninduced cell contains fewer than ten.

Population growth is often measured by optical density. This method is fast, quantitative, and highly reproducible, but it measures light scattering rather than living-cell number directly. Dead cells and environment-dependent changes in cell shape or refractive index can affect the reading. Plating and flow cytometry provide useful complementary measurements.

Lactose complicates an induction experiment because it is both a signal and a food. The nonmetabolizable analogs TMG and IPTG solve this problem: they trigger the lac regulatory system without being consumed. They therefore allow the input signal to be controlled independently of metabolism.

11.2 The Novick-Weiner induction experiment

Novick and Weiner grew *E. coli* to steady state in a chemostat and then added TMG. At regular intervals they sampled the culture and measured its total β-gal content. The chemostat maintained

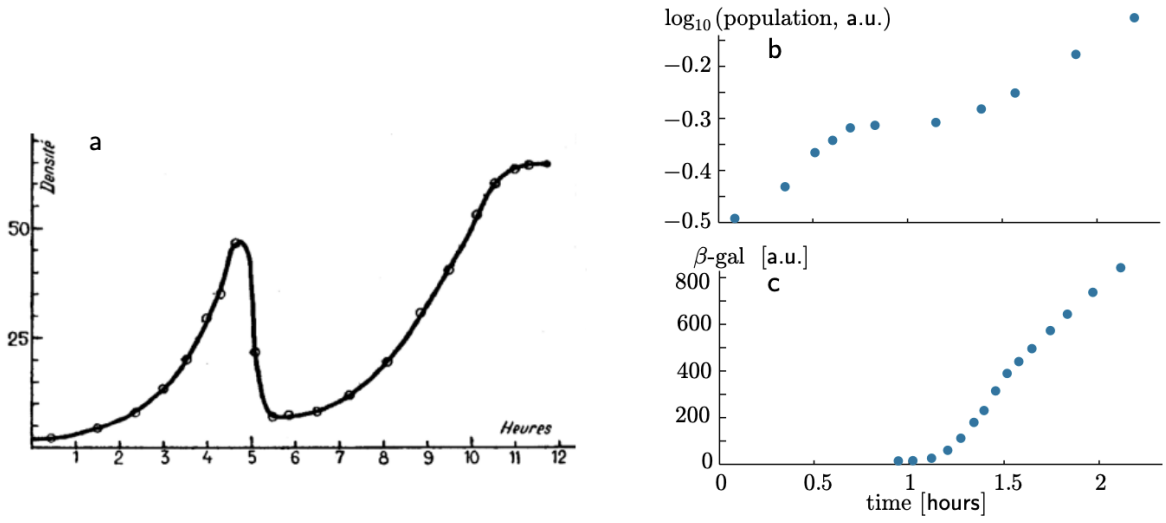

Figure 11.1: Diauxic growth. Monod's original population measurement shows two growth phases separated by a lag (left). In a later *E. coli* experiment, the two approximately linear segments on a semilog plot identify exponential-growth phases, while production of β-galactosidase begins only after glucose is depleted (right). The left and right panels are from Monod (1942) and Epstein et al. (1966), respectively.

a fixed bacterial density, so the bulk measurement could be reported as the average number of enzyme molecules per bacterium.

At high TMG, enzyme abundance initially rose almost linearly and then saturated. At moderate TMG, it began with pronounced upward curvature before approaching a nearly linear regime.

The population mean alone does not identify what individual cells are doing. Two hypotheses can give the same intermediate average:

All-or-none hypothesis. Each cell is either fully on or fully off. After inducer is added, cells switch randomly from off to on, at a rate that depends on inducer concentration.

Graded hypothesis. Every cell has the same intermediate expression level, which may change continuously with time.

The first hypothesis introduces heterogeneity among genetically identical cells even in a well-mixed, uniform environment. The second keeps the cells identical at every instant.

11.2.1 A chemostat balance for total enzyme

Let N_* be the steady number of bacteria in a chemostat of volume V and flow rate Q. Let $z(t)$ be the total number of β-gal molecules in the chamber and $S(t)$ the mean enzyme-production rate per cell. Production adds $N_*S(t)$ molecules per unit time, while outflow removes the fraction Q/V of the chamber contents per unit time:

$$\frac{dz}{dt} = N_*S(t) - \frac{Q}{V}z. \tag{11.1}$$

As in the chemostat model, introduce dimensionless time

$$\bar{t} = \frac{Q}{V}t. \tag{11.2}$$

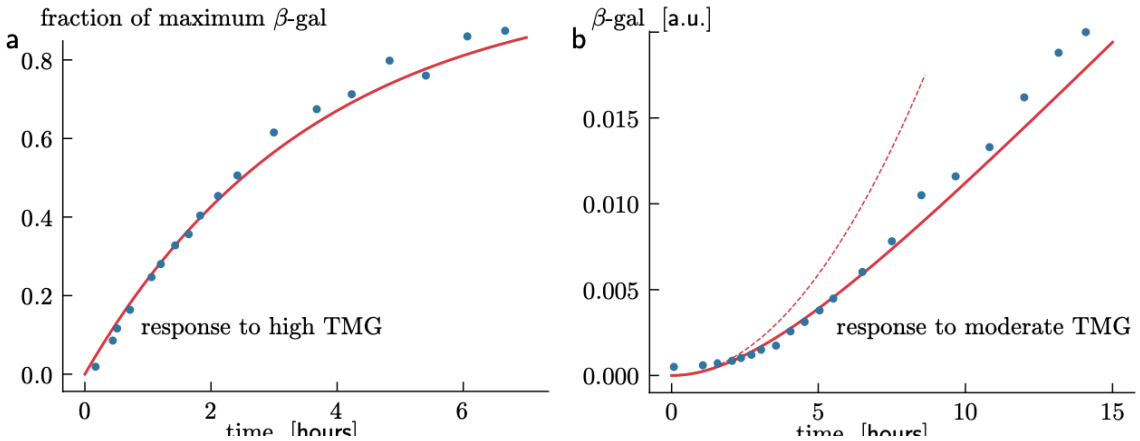

Figure 11.2: Novick and Weiner's induction data. High TMG rapidly places nearly every cell in the induced state, giving a saturating rise in mean β-gal (left). Moderate TMG gives a slow, curved onset followed by an approximately linear regime (right). Solid curves are the physical models derived below; the dashed curve shows the short-time quadratic approximation. Data from Novick and Weiner (1957); model curves from Nelson (2022).

Then

$$\frac{dz}{d\bar{t}} = \frac{V}{Q} N_* S(\bar{t}) - z. \tag{11.3}$$

High inducer. At sufficiently high TMG, switching is rapid compared with the residence time V/Q, so all cells are effectively on and $S = S_{\max}$ is constant. For $z(0) = 0$,

$$z(\bar{t}) = \frac{V N_* S_{\max}}{Q} \left(1 - e^{-\bar{t}}\right). \tag{11.4}$$

Equivalently,

$$\frac{z(\bar{t})}{z(\infty)} = 1 - e^{-\bar{t}}. \tag{11.5}$$

The exponential approach reflects continuous production opposed by washout.

Moderate inducer. Suppose initially off cells switch on independently with a constant hazard k, and ignore switching back off during the measurement. The exact on fraction is

$$f_{\text{on}}(t) = 1 - e^{-kt}, \qquad S(t) = S_{\max}(1 - e^{-kt}). \tag{11.6}$$

Let $\delta = Q/V$, $\bar{t} = \delta t$, and $\kappa = k/\delta$. Scaling enzyme abundance by its eventual value $z_\infty = N_* S_{\max}/\delta$, so $\bar{z} = z/z_\infty$, gives

$$\frac{d\bar{z}}{d\bar{t}} = 1 - e^{-\kappa \bar{t}} - \bar{z}. \tag{11.7}$$

For $\kappa \neq 1$, imposing $\bar{z}(0) = 0$ yields

$$\bar{z}(\bar{t}) = 1 - \frac{\kappa e^{-\bar{t}} - e^{-\kappa \bar{t}}}{\kappa - 1}. \tag{11.8}$$

When $\kappa = 1$, the continuous limit is $\bar{z} = 1 - (1 + \bar{t})e^{-\bar{t}}$. At short times,

$$\bar{z}(\bar{t}) = \frac{\kappa}{2}\bar{t}^2 + O(\bar{t}^3). \tag{11.9}$$

The quadratic onset is not an extra fitted mechanism. It follows because the number of on cells grows linearly and each on cell subsequently accumulates enzyme. Eventually the on fraction saturates automatically in the exact switching model. Allowing cells to switch off again would replace the one-way fraction in Equation (11.6) by the solution of a two-state Markov model.

11.3 Single-cell evidence: all-or-none induction

The time-course fit supports the all-or-none model but does not prove it. Novick and Weiner used a second property of the lac system to test the single-cell prediction directly. At high TMG, initially off cells turn on; at zero TMG, initially on cells eventually turn off. At an intermediate inducer concentration, however, cells preserve whichever state they had before entering the medium. This is a *maintenance medium*.

They diluted a partially induced culture until each new culture was founded by a single cell, then expanded every founder in maintenance medium. If expression were graded, each clone should have retained an intermediate activity. Instead, each clone was either fully induced or nearly uninduced. A culture whose bulk activity had been 30% of maximum produced about 30% on clones and 70% off clones.

Modern fluorescent reporters reveal the same behavior cell by cell. The induction state can remain bimodal over a range of TMG concentrations, and the switching threshold depends on the direction in which the inducer is changed.

11.3.1 Bistability, hysteresis, and memory

A bistable dynamical system has two stable fixed points for the same parameter values. An unstable state or separatrix divides their basins of attraction. A perturbation that remains inside one basin decays back to the same state; a perturbation that crosses the separatrix relaxes to the other state.

When an input is slowly increased and then decreased, a bistable system generally switches at different thresholds in the two directions. This path dependence is *hysteresis*. The maintenance regime is the interval between the thresholds: the present output depends on the cell's history. Cellular memory is therefore a dynamical property, not a change in DNA sequence.

The lac system contains a natural positive-feedback loop. Permease imports inducer. Internal inducer inactivates LacI repression, which increases expression of permease and β-galactosidase, which in turn promotes still more inducer entry. Cooperative binding and saturating regulation make this feedback nonlinear enough to support two stable states.

11.4 Why population means can be misleading

Bimodal responses are not unique to bacterial metabolism. In T cells, doubly phosphorylated ERK (ppERK) participates in the decision to mount a proliferative immune response. A population-averaged dose-response curve can vary smoothly even when individual cells occupy two distinct response states.

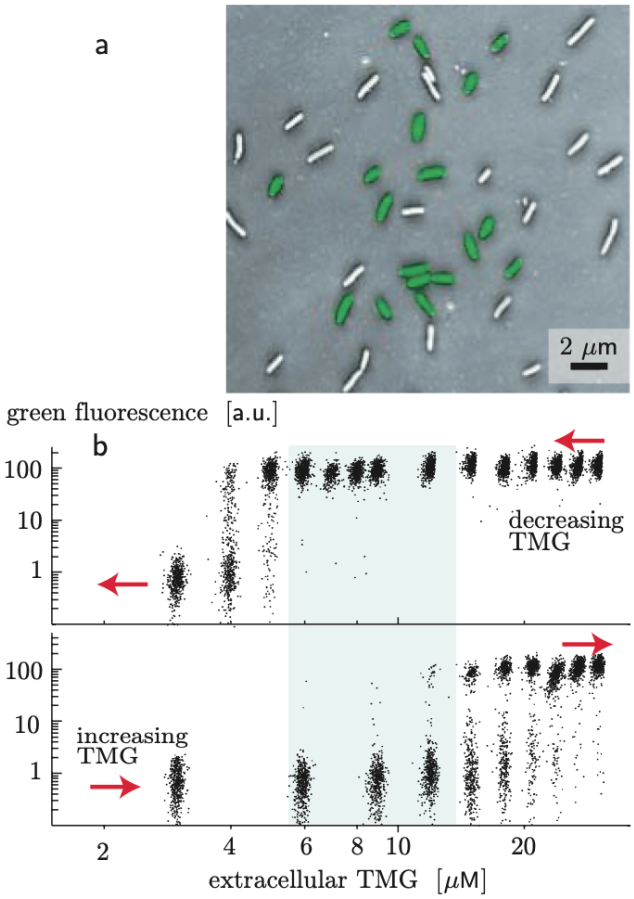

Figure 11.3: Single-cell evidence for all-or-none induction. Fluorescent *E. coli* cells are visibly on or off rather than uniformly intermediate (top). Populations initialized in the on state and populations initialized in the off state follow different branches over a range of TMG concentrations (bottom). The shaded range is a maintenance regime (Ozbudak et al. 2004).

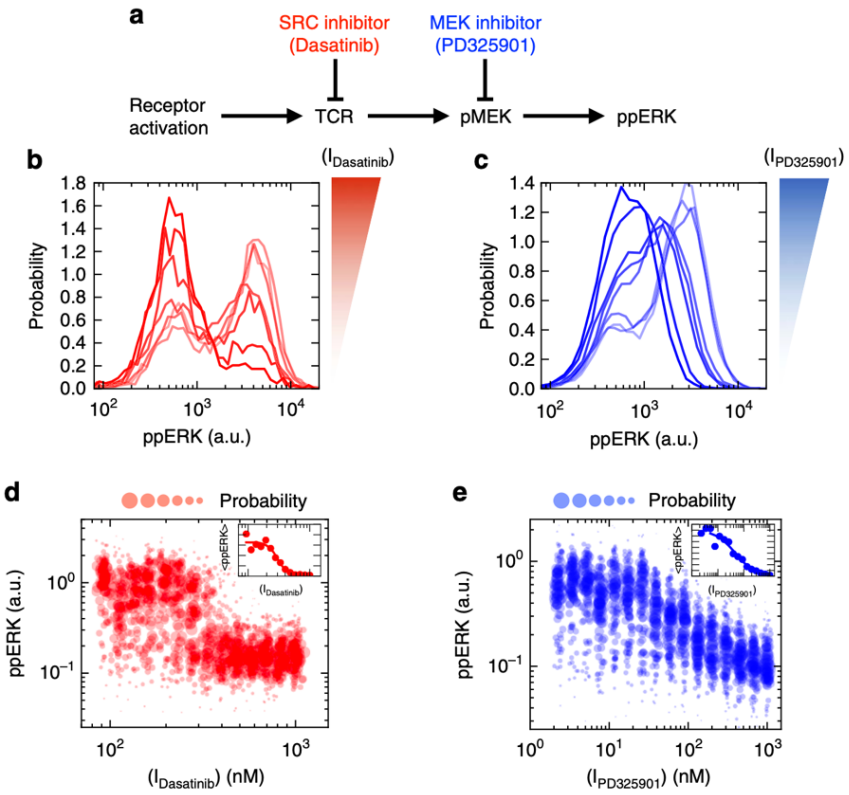

Figure 11.4: Single-cell ppERK responses under two modes of kinase inhibition. Histograms and single-cell scatter plots expose whether an inhibitor changes the fraction of responding cells or shifts the response of every cell. A population mean alone suppresses this distinction (Vogel et al. 2016).

The lesson is broader than this experiment. The same mean can arise from a narrow unimodal distribution, a broad distribution, or a mixture of low and high states. Single-cell distributions are therefore essential whenever feedback may create thresholds or discrete fates.

11.4.1 Bimodality is not the same as bistability

Week 8 introduced a promoter that switches stochastically between inactive and active states. Slow switching can produce a broad or even bimodal expression histogram without two stable fixed points in a deterministic feedback model. The two peaks then reflect cells caught in different promoter histories, not two self-maintaining attractors.

Conversely, a genuinely bistable circuit can look unimodal if transitions are fast compared with the measurement time or if one basin is only weakly populated. A histogram is therefore evidence about states, but not a complete dynamical diagnosis. Stronger evidence for bistability comes from hysteresis, perturbations that reveal two basins of attraction, long residence times, and an inferred drift field with two stable zeros separated by an unstable one. This distinction is essential when a two-state promoter model and a positive-feedback switch can both fit the same snapshot distribution.

Detailed mechanistic models can explain how a particular inhibitor alters a regulatory network, but they may contain many unmeasured states and parameters. The example in Figure 11.5 shows how the position of an inhibited reaction within a feedback network can change the number and stability of fixed points.

11.5 A general stochastic description of biochemical feedback

Let X be the molecular species of interest and let $A, B, C, \ldots$ denote the surrounding chemical bath. We assume the cell is well mixed and that each elementary reaction changes the copy number n of X by one. A general reaction channel can be written

$$j_r X + Y_r^+ \underset{k_r^-}{\overset{k_r^+}{\rightleftharpoons}} (j_r + 1)X + Y_r^-. \tag{11.10}$$

The integer j_r encodes the reaction order and hence the nonlinearity. The bath species are absorbed into effective forward and backward propensities.

The Schlögl model is a classic special case: X is produced either spontaneously from a bath species or autocatalytically through a reaction requiring two existing X molecules. Its nonlinear production can create two stable concentration states separated by an unstable one.

Let b_n be the total propensity for $n \to n+1$, and d_n the total propensity for $n \to n-1$. The one-step master equation is

$$\frac{dp_n}{dt} = b_{n-1}p_{n-1} + d_{n+1}p_{n+1} - (b_n + d_n)p_n. \tag{11.11}$$

At a stationary distribution with zero probability current between neighboring states,

$$b_{n-1}p_{n-1} = d_n p_n. \tag{11.12}$$

Therefore

$$p_n = p_0 \prod_{m=1}^{n} \frac{b_{m-1}}{d_m}, \qquad p_0^{-1} = 1 + \sum_{n=1}^{\infty} \prod_{m=1}^{n} \frac{b_{m-1}}{d_m}. \tag{11.13}$$

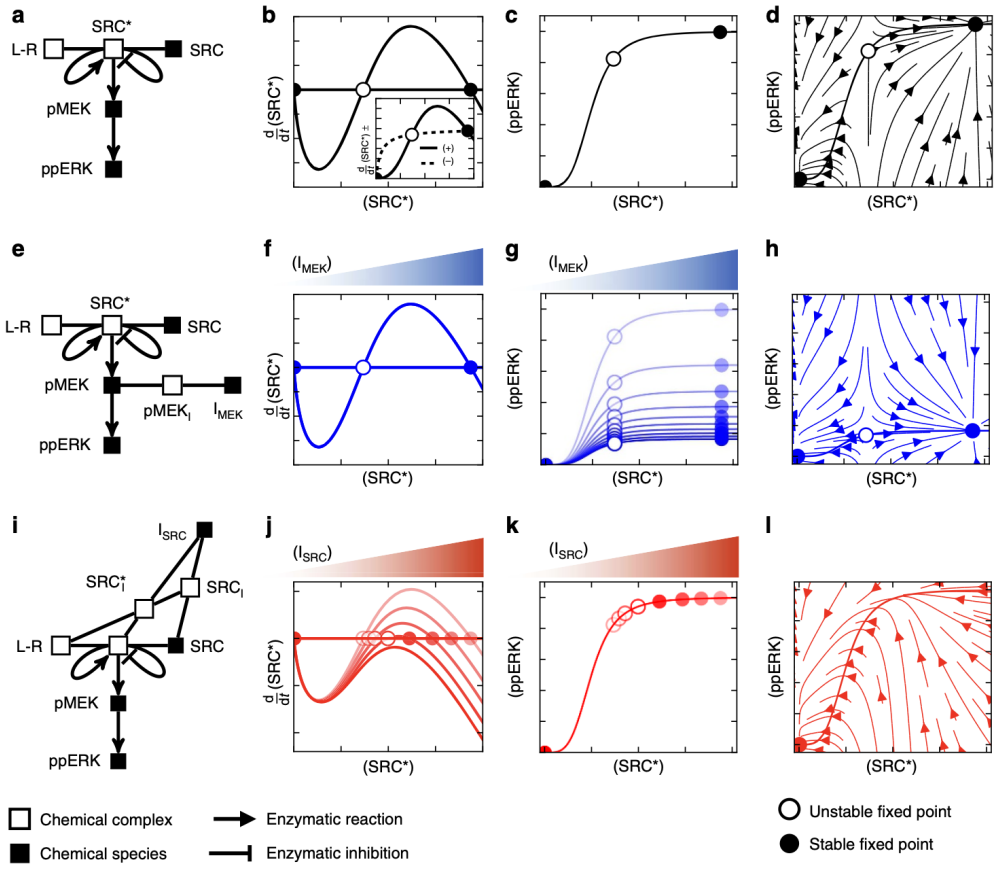

Figure 11.5: A mechanistic feedback model for ppERK signaling. Different inhibitor targets alter the effective feedback, fixed points, and phase flow in different ways. The detailed network is biologically informative but motivates a more general description that can be inferred directly from single-cell distributions (Vogel et al. 2016).

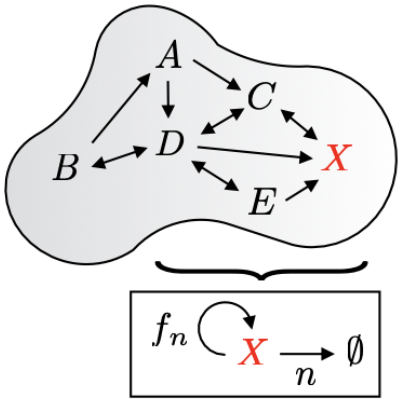

Figure 11.6: Coarse-graining a biochemical network. All bath species and reactions are replaced by an effective copy-number-dependent production propensity f_n for X, paired with linear loss (Erez, Byrd, et al. 2019).

This relation can be inverted by choosing linear loss as a reference representation and absorbing the complete nonlinear birth-to-death ratio into one effective production curve:

$$f_n \equiv (n+1)\frac{b_n}{d_{n+1}} = (n+1)\frac{p_{n+1}}{p_n}, \qquad n \geq 0. \tag{11.14}$$

The last expression contains only the observed stationary distribution. This construction does not assume that the microscopic loss propensity d_n was originally linear. It is a transformation that assigns the reference rates

$$\widetilde{d}_n = \frac{n}{\tau_0}, \qquad \widetilde{b}_n = \frac{f_n}{\tau_0}, \tag{11.15}$$

where the arbitrary reference time τ_0 cannot be inferred from a stationary histogram. All nonlinear state dependence in b_n/d_{n+1} is moved into f_n. Indeed, the transformed rates obey $\widetilde{b}_n p_n = \widetilde{d}_{n+1} p_{n+1}$, so they preserve the stationary neighboring probability ratios, and hence the stationary distribution, although they need not preserve transient kinetics or correlation times. Thus a single-cell histogram can reveal the shape of an effective feedback curve without specifying every biochemical reaction, but it cannot set the absolute kinetic clock.

Equation (11.14) is literal only when neighboring histogram bins represent adjacent molecular counts. If fluorescence is calibrated as $I = \alpha n$, bins of width α can be mapped to p_n, and the inferred curve may be displayed on the intensity axis by the same rescaling. Real measurements have detector noise, background, and bins wider than one molecule; neighboring intensity bins then need not represent the ratio p_{n+1}/p_n exactly. In practice one first calibrates the molecular scale when possible, estimates a smooth or regularized stationary distribution, avoids ratios involving empty bins, and propagates sampling uncertainty, for example by bootstrap resampling. When those steps are not possible, a curve labeled $f(I) - I$ should be interpreted as a phenomenological drift in the measured coordinate rather than as a uniquely identified microscopic propensity.

The corresponding deterministic proxy for this transformed representation is

$$\frac{dn}{d\bar{t}} = f(n) - n, \qquad \bar{t} = \frac{t}{\tau_0}. \tag{11.16}$$

Fixed points occur where the inferred production curve crosses the linear-loss line, $f(n) = n$. A crossing is stable when the local slope of $f(n) - n$ is negative. Three crossings—stable, unstable, stable—are the characteristic signature of bistability. For small copy numbers, there is a one-count distinction between this continuous proxy and the exact histogram criterion. A discrete mode lies where the neighboring ratio crosses $p_{n+1}/p_n = 1$, which by Equation (11.14) corresponds to $f_n = n + 1$, not $f_n = n$. The offset is negligible on a finely sampled, high-copy-number axis, but the exact neighboring probabilities should be used when only a few molecules are present.

11.6 Two mutually repressing genes form a toggle

Gardner and collaborators engineered a genetic switch from two transcriptional repressors. The product of gene 1 represses gene 2, and the product of gene 2 represses gene 1. Two negative interactions form an overall positive-feedback loop: an increase in repressor 1 reduces repressor 2, which releases gene 1 from repression and reinforces the original change.

Let c_1 and c_2 be the repressor concentrations. Repressor 2 inhibits production of repressor 1 and vice versa. A Hill-function model is

$$\frac{dc_1}{dt} = -\frac{c_1}{\tau_1} + \frac{\Gamma_1/V}{1 + (c_2/K_{d,2})^{n_1}}, \tag{11.17}$$

$$\frac{dc_2}{dt} = -\frac{c_2}{\tau_2} + \frac{\Gamma_2/V}{1 + (c_1/K_{d,1})^{n_2}}. \tag{11.18}$$

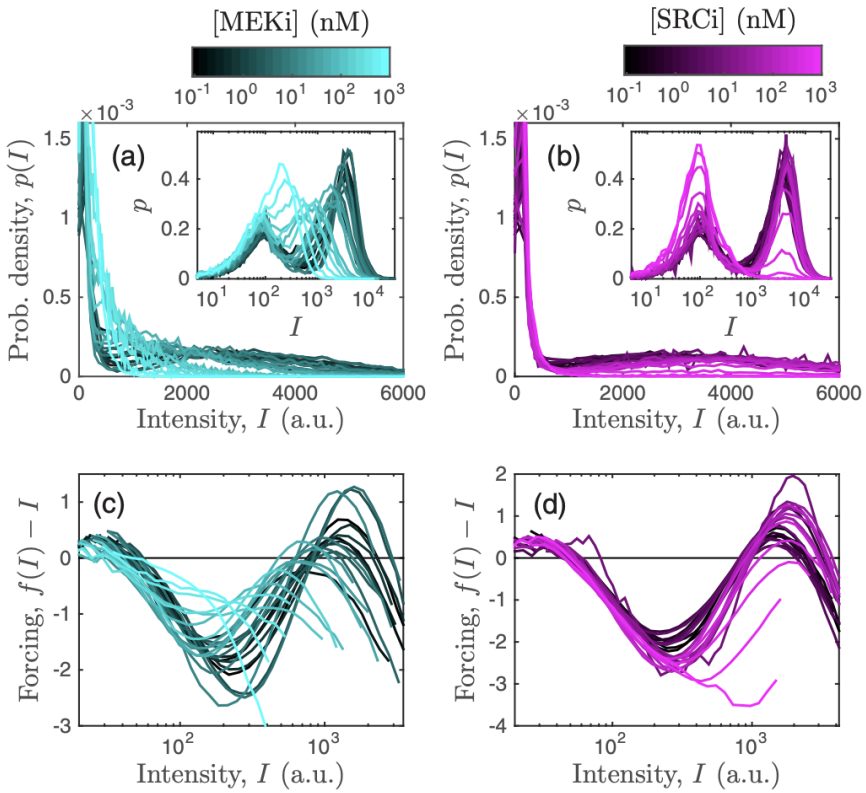

Figure 11.7: Inferring feedback from single-cell distributions. Measured intensity distributions change with inhibitor dose (top), while the ratio of neighboring probability bins yields the effective forcing $f(I) - I$ (bottom). Its zero crossings identify candidate fixed points and their stability (Erez, Byrd, et al. 2019).

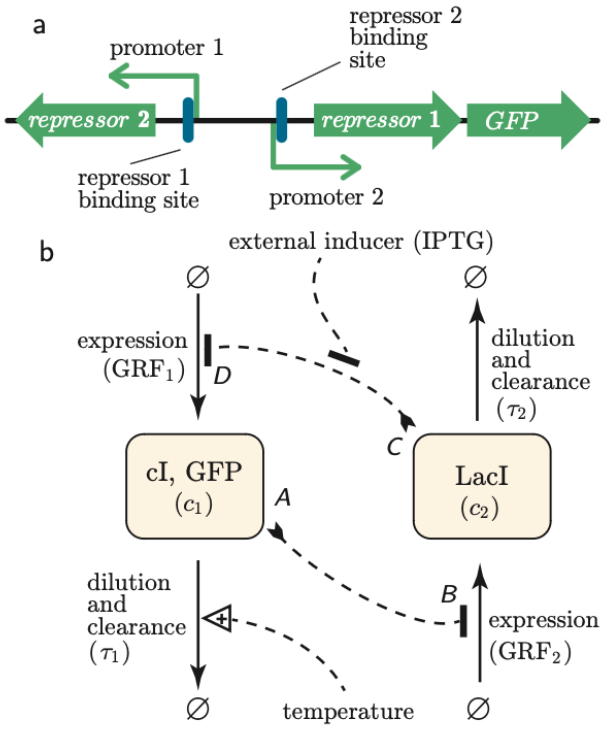

Figure 11.8: Synthetic two-gene toggle. Each repressor binds the other gene's promoter (top). The network representation (bottom) includes production, dilution and clearance, and an external inducer that can inactivate LacI to command a switch. Figure reproduced from Nelson (2022), after Gardner et al. (2000).

Here Γ_i/V is the maximal production rate per volume, $K_{d,i}$ is a regulatory concentration scale, n_i is a Hill coefficient, and τ_i is the combined clearance and dilution time.

Define

$$\bar{c}_i = \frac{c_i}{K_{d,i}}, \qquad \bar{\Gamma}_i = \frac{\Gamma_i \tau_i}{K_{d,i} V}. \tag{11.19}$$

For equal Hill coefficients $n_1 = n_2 = n$, the nullclines are

$$\bar{c}_1 = \frac{\bar{\Gamma}_1}{1 + \bar{c}_2^n}, \qquad \bar{c}_2 = \frac{\bar{\Gamma}_2}{1 + \bar{c}_1^n}. \tag{11.20}$$

Their intersections are the fixed points.

The symmetric case makes the stability threshold explicit. Let $\tau_1 = \tau_2 = \tau$, $K_{d,1} = K_{d,2}$, and $\bar{\Gamma}_1 = \bar{\Gamma}_2 = \bar{\Gamma}$. A symmetric fixed point has $\bar{c}_1 = \bar{c}_2 = u$, where

$$u = \frac{\bar{\Gamma}}{1 + u^n}. \tag{11.21}$$

In time units t/τ, define the magnitude of the off-diagonal repression slope

$$a = \frac{n \bar{\Gamma} u^{n-1}}{(1 + u^n)^2} = \frac{n u^n}{1 + u^n}, \tag{11.22}$$

where the fixed-point relation was used in the second equality. The Jacobian is

$$J_* = \begin{pmatrix} -1 & -a \\ -a & -1 \end{pmatrix}, \qquad \lambda_{\text{sym}} = -1 - a, \qquad \lambda_{\text{antisym}} = -1 + a. \tag{11.23}$$

The symmetric mode always decays. The antisymmetric mode—one repressor rising while the other falls—becomes unstable when

$$a > 1 \quad \Longleftrightarrow \quad (n-1)u^n > 1. \tag{11.24}$$

For $n = 1$, the inequality is impossible, so noncooperative regulation does not produce toggle behavior in this minimal symmetric model. For $n > 1$ and sufficiently strong production, the symmetric intersection becomes a saddle and two asymmetric, stable intersections appear: high c_1/low c_2 and low c_1/high c_2. The saddle's stable manifold is the separatrix dividing their basins of attraction.

An external inducer changes one repression arm and shifts the nullclines. A sufficiently strong pulse can eliminate the current attractor or push the state across the separatrix. When the pulse is removed, the system remains in the new basin. This is how a transient chemical signal can write a persistent biological memory.

11.7 The lac network performs a logic operation

The natural lac circuit combines its inner positive-feedback switch with glucose-dependent control. When glucose is absent, cyclic AMP accumulates and the cAMP-CRP complex activates the lac promoter. When glucose is present, cAMP-CRP activation is suppressed and inducer entry is reduced through inducer exclusion. These outer controls can override the inner lac switch.

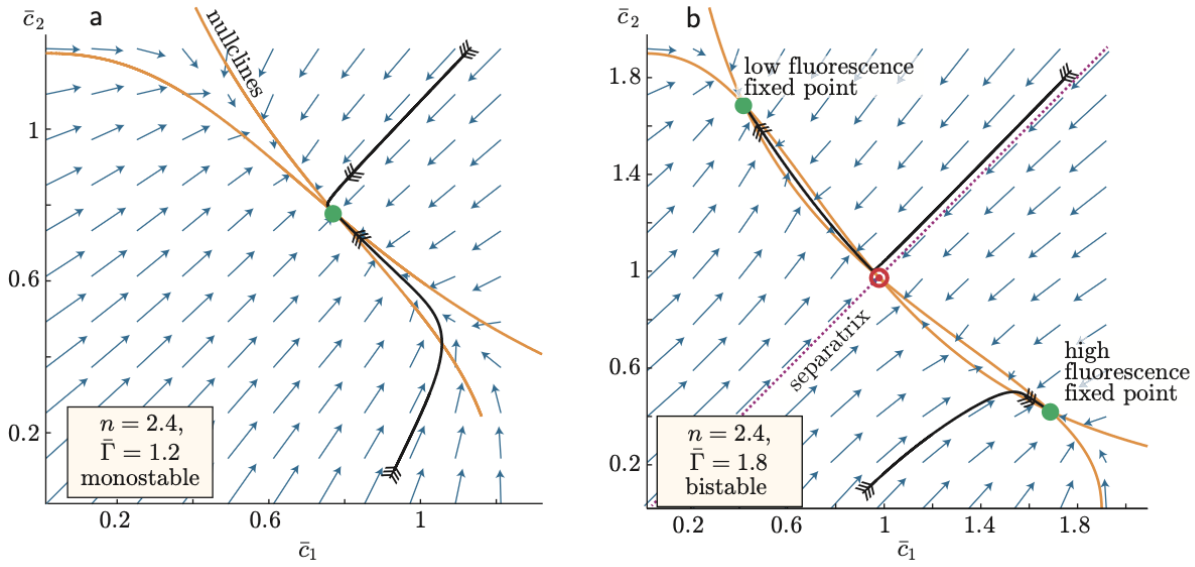

Figure 11.9: Phase portraits of the idealized toggle. At lower production strength the nullclines intersect once and the system is monostable (left). Stronger production with the same cooperative response creates two stable fixed points separated by an unstable saddle and its separatrix (right). Figure from Nelson (2022); model based on Gardner et al. (2000).

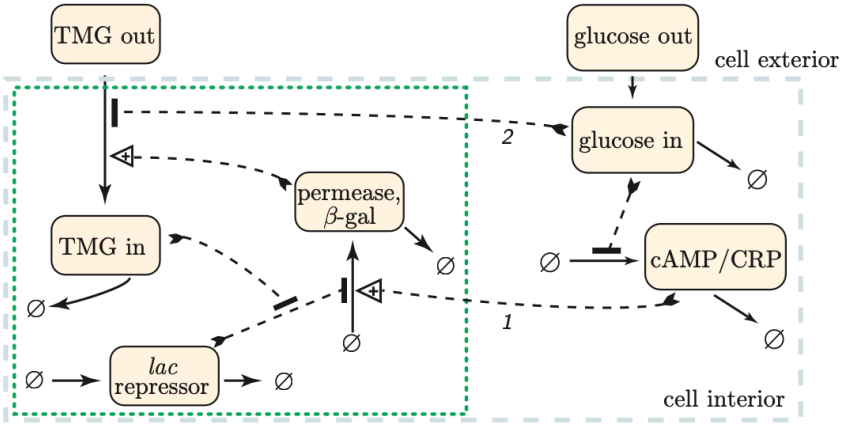

Figure 11.10: Extended decision-making circuit in the lac system. The inner box is the positive-feedback induction module. Glucose acts through cAMP-CRP regulation and inducer exclusion, preventing the inner loop from turning on when the preferred carbon source is available (Nelson 2022).

At the phenomenological level, the circuit implements

$$\text{lac operon on} = (\text{lactose present}) \ \textbf{AND} \ (\text{glucose absent}). \tag{11.25}$$

The molecular network therefore performs a logic operation while its bistable core can retain memory. Diauxic growth, all-or-none induction, hysteresis, and genetic toggles are different views of the same physical idea: nonlinear feedback reshapes the phase portrait so that cells can make and remember discrete decisions.

12 Cellular oscillators, network stability, and ecological dynamics

Living systems contain clocks on many time scales: the heartbeat, circadian rhythms, monthly and yearly cycles, and the much slower synchronized emergence of periodical insects. Some rhythms are driven by a periodic external signal, such as the daily light–dark cycle. Others are *autonomous*: the underlying network continues to oscillate even when the environment is held constant.

This week connects three versions of the same dynamical question. We begin with a synthetic gene circuit that generates oscillations through delayed negative feedback. We then ask how the eigenvalues of a large interaction network determine whether a steady state is stable. Finally, we apply these ideas to interacting populations, from the classical predator–prey model to generalized Lotka–Volterra communities.

By the end of this week, students should be able to answer:

- Why can negative feedback stabilize a system in one regime but generate oscillations in another?
- How does a ring of three repressors create a robust cellular clock?
- How can stability of a many-variable system be decided from its Jacobian?
- Why did Robert May argue that large, strongly connected random networks are difficult to stabilize?
- What produces neutral cycles, damped oscillations, extinction, or sustained cycles in interacting populations?

12.1 Oscillations from negative feedback and delay

Negative feedback normally opposes a perturbation. If the response is sufficiently fast, a variable displaced from its steady value returns smoothly. If sensing, transmission, and response take time, however, the correction is based on an outdated state. The controller can then overshoot the target, reverse its action too late, and overshoot in the opposite direction. A room controlled by a person who turns an air conditioner fully on or fully off only after becoming uncomfortable is a familiar example.

Biochemical feedback contains many natural sources of delay. Transcription must be followed by translation; eukaryotic mRNA must be exported from the nucleus; a protein must fold; and some regulators must dimerize or assemble into larger complexes before binding DNA. A simple schematic delay equation is

$$\frac{dx}{dt} = b\, f\big(x(t-\tau)\big) - \gamma x(t), \tag{12.1}$$

where f is a decreasing function for negative autoregulation and τ is the total response delay. Linearizing around a steady state gives a characteristic equation of the form

$$\lambda + \gamma + k e^{-\lambda \tau} = 0, \tag{12.2}$$

with $k > 0$ measuring feedback strength. Increasing either k or τ can move a complex-conjugate pair of roots toward the imaginary axis. The return to steady state first becomes oscillatory; beyond a threshold, the steady state can lose stability and a sustained oscillation can appear.

An important experimental warning follows. Oscillators in different cells need not remain in phase. If each cell oscillates with nearly the same period but a different phase, the population average may look constant even though every individual cell is rhythmic. A bulk measurement

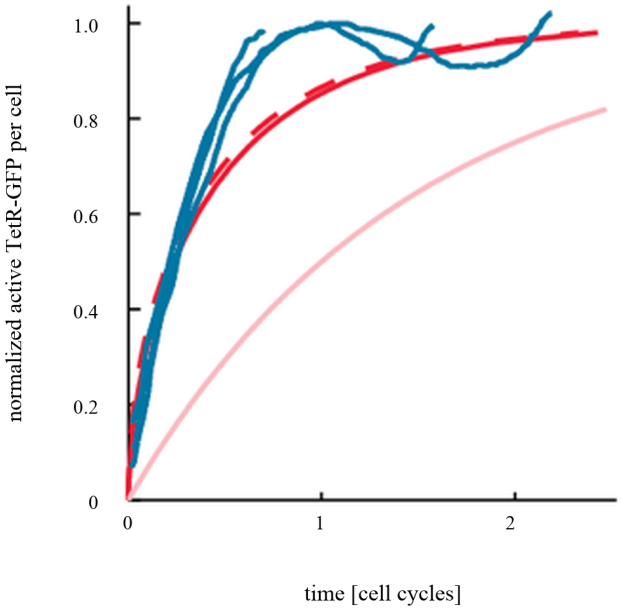

Figure 12.1: Overshoot after induction of a negatively autoregulated gene. Individual traces can rise above their long-time level before feedback restores the steady state. The slow pale curve illustrates the corresponding unregulated response (Rosenfeld, Elowitz, et al. 2002).

can therefore hide the very dynamics one is trying to detect. Fluorescent reporters and long-term single-cell microscopy are essential tools for separating loss of amplitude within cells from loss of synchrony between cells.

12.2 The repressilator: a three-gene oscillator

A single negatively autoregulated gene can show overshoot, but its oscillations are often weak and irregular. Elowitz and Leibler introduced a more tunable design: three repressors placed in a ring. Protein A represses gene B, protein B represses gene C, and protein C represses gene A. If A is abundant, B falls; the loss of B allows C to rise; the rise of C suppresses A; and the loss of A allows B to return. The state travels around the ring. Closely related architectures occur in natural biological clocks.

Let m_A, m_B, m_C be dimensionless mRNA concentrations and p_A, p_B, p_C the corresponding protein concentrations, scaled by the repression threshold. The standard Elowitz–Leibler model is

$$\frac{dm_A}{dt} = \alpha_0 + \frac{\alpha}{1 + p_C^n} - m_A, \qquad \frac{dp_A}{dt} = \beta(m_A - p_A), \tag{12.3}$$

$$\frac{dm_B}{dt} = \alpha_0 + \frac{\alpha}{1 + p_A^n} - m_B, \qquad \frac{dp_B}{dt} = \beta(m_B - p_B), \tag{12.4}$$

$$\frac{dm_C}{dt} = \alpha_0 + \frac{\alpha}{1 + p_B^n} - m_C, \qquad \frac{dp_C}{dt} = \beta(m_C - p_C). \tag{12.5}$$

Time is measured in mRNA lifetimes. The parameter α_0 is basal transcription, α is the additional regulated transcription rate, n is the Hill coefficient, and β is the ratio of protein to mRNA removal rates. The factor β multiplies both protein production and removal; it cannot be removed by

the same time scaling that sets the mRNA removal coefficient to one. The Hill functions provide nonlinearity, while the separate mRNA and protein variables provide an effective delay.

The model has a symmetric fixed point

$$m_A = m_B = m_C = p_A = p_B = p_C = x^*, \qquad x^* = \alpha_0 + \frac{\alpha}{1 + (x^*)^n}. \tag{12.6}$$

The right-hand side decreases continuously from $\alpha_0 + \alpha$ to α_0, while the left-hand side increases, so this positive fixed point is unique. Define the magnitude of the repression slope there by

$$\chi = -\left.\frac{d}{dp}\left(\alpha_0 + \frac{\alpha}{1+p^n}\right)\right|_{p=x^*} = \frac{\alpha n (x^*)^{n-1}}{[1+(x^*)^n]^2} > 0. \tag{12.7}$$

To test stability, consider the cyclic shift that maps a perturbation at one gene to the preceding gene. Applying the shift three times returns to the original component, so a normal-mode multiplier must obey $q^3 = 1$. The three roots of unity are

$$q_k = e^{2\pi i k/3}, \qquad k = 0, 1, 2. \tag{12.8}$$

The mode $q_0 = 1$ perturbs every gene in phase. The two complex modes $q_{1,2} = e^{\pm 2\pi i/3}$ describe phase-shifted waves traveling in opposite directions around the ring; real perturbations are obtained by combining this conjugate pair. For a selected mode, the protein perturbation entering the preceding gene is q times the local perturbation. Substitution of a time dependence $e^{\lambda t}$, with local amplitudes M and P, gives the two linearized mode equations

$$(\lambda + 1)M = -\chi q P, \tag{12.9}$$

$$(\lambda + \beta)P = \beta M. \tag{12.10}$$

The first equation contains the negative repression slope and the phase shift to the upstream protein; the second is the linear mRNA-to-protein response. Eliminating M and P gives

$$(\lambda + 1)(\lambda + \beta) + \beta\chi q = 0. \tag{12.11}$$

The uniform mode $q = 1$ is stable. The two traveling modes form a complex-conjugate pair and are the ones that can destabilize the fixed point. To locate the crossing explicitly, choose $q = e^{-2\pi i/3} = -1/2 - i\sqrt{3}/2$ and set $\lambda = i\omega$ in Equation (12.11). Separating real and imaginary parts gives

$$\beta - \omega^2 - \frac{\beta\chi}{2} = 0, \tag{12.12}$$

$$(1 + \beta)\omega - \frac{\sqrt{3}}{2}\beta\chi = 0. \tag{12.13}$$

The imaginary part first gives

$$\omega = \frac{\sqrt{3}\,\beta\chi}{2(1+\beta)}. \tag{12.14}$$

Substitution into the real part produces one quadratic equation for the critical slope,

$$3\beta\chi^2 + 2(1+\beta)^2\chi - 4(1+\beta)^2 = 0. \tag{12.15}$$

Only the positive root is physically relevant. The traveling modes therefore reach the imaginary axis when

$$\chi = \chi_c(\beta) = \frac{4(1+\beta)}{1 + \beta + \sqrt{(1+\beta)^2 + 12\beta}}, \qquad \omega_c = \frac{\sqrt{3}\,\beta\chi_c}{2(1+\beta)}. \tag{12.16}$$

For $\chi < \chi_c$, perturbations decay; at $\chi = \chi_c$, a traveling mode undergoes a Hopf instability; and for $\chi > \chi_c$, the symmetric fixed point is unstable and the nonlinear system can approach a limit cycle. Thus sufficiently steep repression, together with the mRNA–protein delay, produces oscillations. The six-dimensional model stores signal history in intermediate molecular states, something a one-variable instantaneous-feedback model cannot do (Elowitz and Leibler 2000).

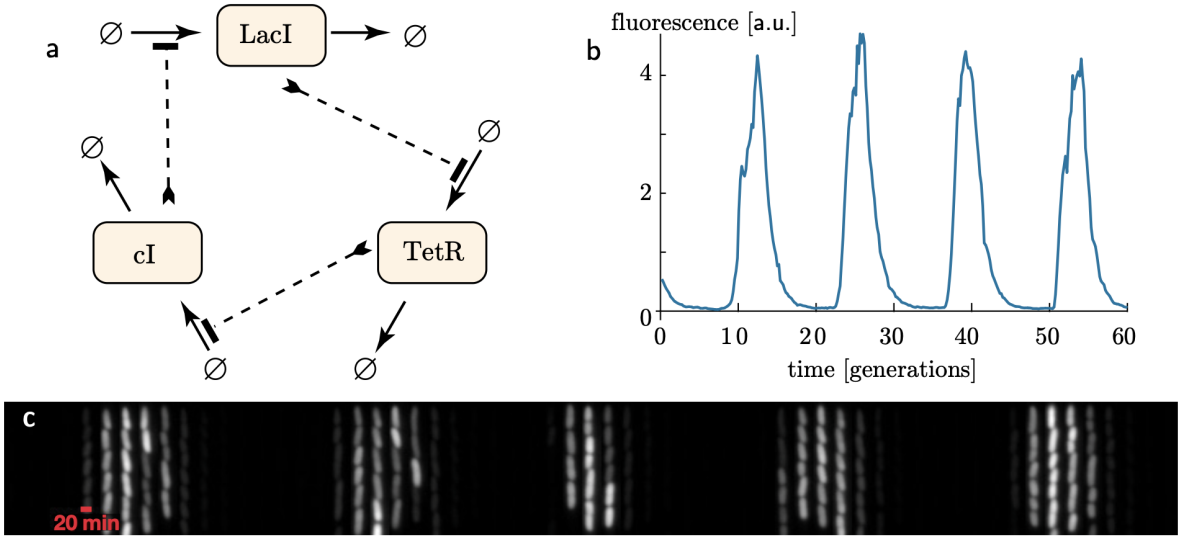

Figure 12.2: The repressilator. Three repressors form a directed ring (left). A fluorescent reporter records regular pulses over many generations (right), while successive microscopy frames show the oscillation through repeated cell divisions (bottom). The circuit diagram is from Elowitz and Leibler (2000); the single-cell traces and microscopy are from Potvin-Trottier et al. (2016).

12.2.1 The mother machine and single-cell time series

A microfluidic "mother machine" holds one end of a narrow channel closed. A mother cell remains trapped at that end while its descendants grow in a line and are pushed toward an open flow channel. Fresh medium can be supplied continuously and changed during the experiment. This design has three major advantages:

1. the same mother cell can be followed for many generations;
2. the one-dimensional arrangement makes segmentation and lineage tracking easier than in a disordered colony; and
3. nutrients, drugs, and stressors can be changed without losing the tracked lineage.

It thereby converts a static population snapshot into a controlled, long single-cell time series—exactly what is needed to distinguish a genuine oscillator from population-level averaging.

12.3 Local stability in many dimensions

The repressilator has six dynamical variables. Ecological, biochemical, and neural networks may have hundreds or thousands. We therefore need a general language for stability. Consider

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}), \tag{12.17}$$

and a fixed point $\mathbf{x}^*$ satisfying $\mathbf{F}(\mathbf{x}^*) = 0$. Write $\mathbf{x} = \mathbf{x}^* + \boldsymbol{\eta}$. To first order,

$$\frac{d\boldsymbol{\eta}}{dt} = J\boldsymbol{\eta}, \qquad J_{ij} = \left.\frac{\partial F_i}{\partial x_j}\right|_{\mathbf{x}^*}. \tag{12.18}$$

The normal modes have the form $\boldsymbol{\eta} = \mathbf{v}e^{\lambda t}$, where

$$J\mathbf{v} = \lambda\mathbf{v}. \tag{12.19}$$

Thus a fixed point is locally asymptotically stable precisely when every eigenvalue obeys

$$\operatorname{Re}\lambda < 0. \tag{12.20}$$

A real positive eigenvalue produces monotonic departure. A complex pair with positive real part produces an oscillatory departure. Eigenvalues on the imaginary axis require higher order analysis: linearization alone cannot decide their nonlinear fate.

12.4 May's random-network argument

Robert May asked a deliberately idealized question: if a large community matrix has many random interactions, when should its equilibrium be stable? Write

$$J = -dI + A, \tag{12.21}$$

where $-d < 0$ represents self-regulation. For $i \neq j$, let A_{ij} be nonzero with probability C, the *connectance*, and zero otherwise. Conditional on being nonzero, assume the interactions have mean zero and variance σ^2. Each off-diagonal entry therefore has total variance $C\sigma^2$.

This ensemble suppresses biological detail on purpose. It asks how the number of variables n, the fraction of possible interactions C, their typical strength σ, and the self-regulation d compete before any special network structure is introduced.

12.4.1 Two random-matrix laws

For a real symmetric random matrix whose entries have variance σ^2/n, the eigenvalues are real and approach Wigner's semicircle distribution,

$$\rho(\lambda) = \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - \lambda^2}, \qquad -2\sigma \le \lambda \le 2\sigma. \tag{12.22}$$

Symmetry is a strong assumption: it would require the effect of species j on species i to equal the reverse effect. Ecological interaction matrices are generally not symmetric, so their eigenvalues can be complex.

For a large nonsymmetric matrix with independent, centered entries of variance σ^2/n and no dominant heavy-tailed entries, the circular law places the bulk eigenvalues approximately uniformly inside a disk of radius σ in the complex plane. The statement is an asymptotic law: finite matrices have edge fluctuations, and correlations, nonzero means, very sparse structure, or heavy tails can create distorted clouds or isolated eigenvalue outliers. In May's sparse, unnormalized matrix, when the mean number nC of interactions per row is large enough for a bulk approximation, the corresponding radius is approximately

$$R \simeq \sigma\sqrt{nC}. \tag{12.23}$$

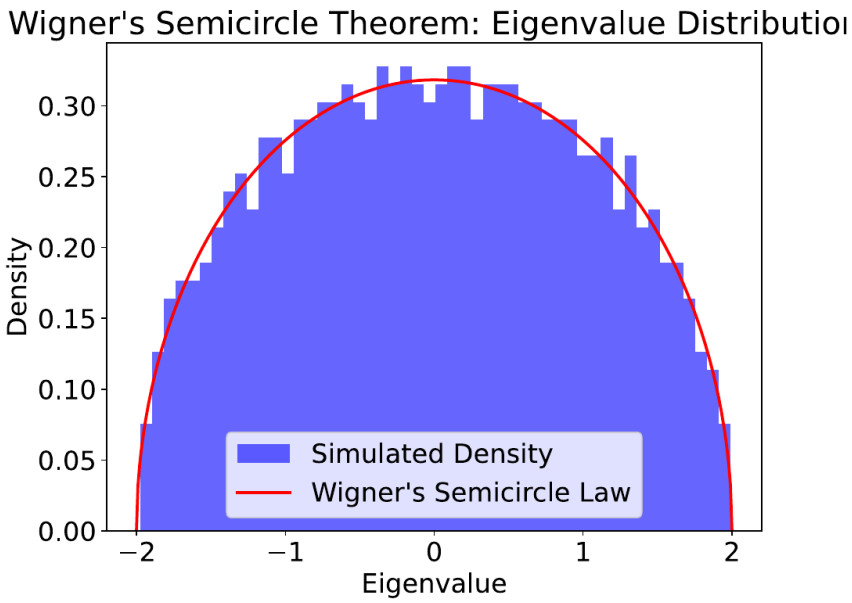

Figure 12.3: Wigner's semicircle law for a normalized random symmetric matrix. The histogram shows simulated eigenvalues and the red curve is the limiting density in Equation (12.22). The plot is course-generated from Wigner (1955).

The diagonal term $-dI$ shifts the center of the disk from the origin to $-d$. The fixed point is expected to be stable when the rightmost edge remains in the negative half-plane:

$$\boxed{\sigma\sqrt{nC} < d.} \tag{12.24}$$

With the common scaling $d = 1$, this becomes $\sigma\sqrt{nC} < 1$. Increasing species number, connectance, or typical interaction strength pushes the eigenvalue cloud toward the unstable half-plane; stronger self-regulation pulls it back.

The boundary is probabilistic for finite n, not perfectly sharp. Different random draws have different extreme eigenvalues. As the system grows, however, the transition from mostly stable to mostly unstable ensembles becomes increasingly abrupt.

12.4.2 Reciprocity, modularity, and structure

The assumption that A_{ij} and A_{ji} are independent is also biologically special. Let

$$\rho = \mathrm{Corr}(A_{ij}, A_{ji}). \tag{12.25}$$

The elliptic law predicts an eigenvalue cloud whose real and imaginary semiaxes are, under the same random-matrix scaling, proportional to $1 + \rho$ and $1 - \rho$. Positive reciprocity makes the cloud more symmetric and expands it along the real axis, whereas negative reciprocity compresses the real axis and expands the imaginary direction. Predator–prey sign structure, for example, can therefore be more stable than an uncorrelated network with the same marginal distribution of interaction strengths.

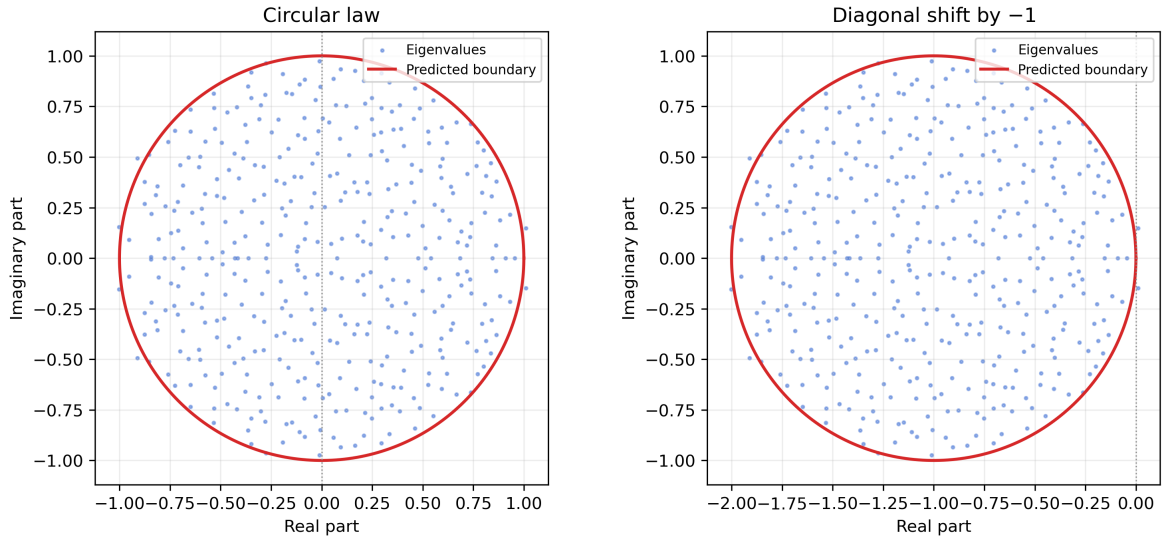

Figure 12.4: Girko's circular law. Eigenvalues of a normalized nonsymmetric random matrix fill a disk (left). Replacing the diagonal by -1 shifts the disk one unit to the left (right), illustrating the geometric origin of May's criterion. The plot is course-generated from Girko (1985) and May (1972).

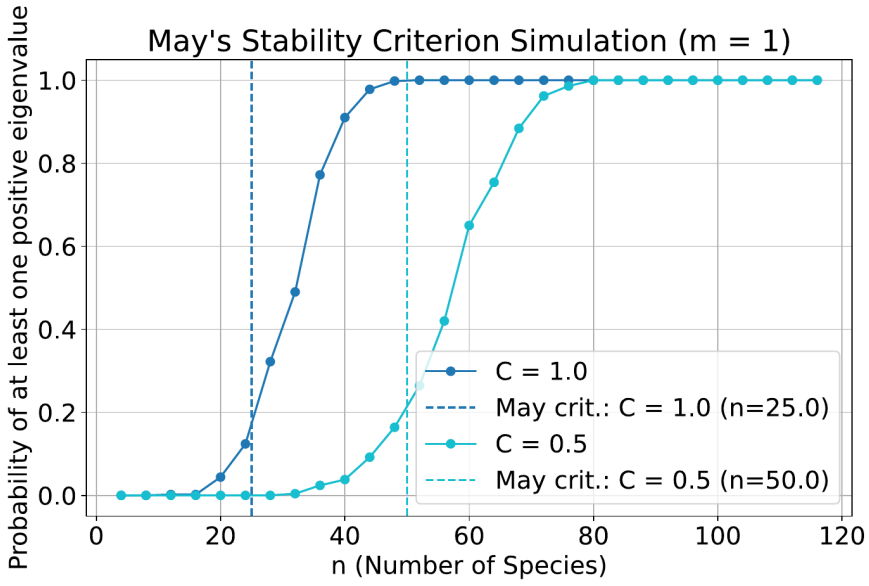

Figure 12.5: Probability that a random community matrix has at least one eigenvalue with positive real part. Dashed lines mark the May thresholds for two connectances. Greater connectance moves the instability transition to a smaller number of species. The plot is course-generated from May (1972).

Modularity provides another route to stability. If n variables split into m independent blocks of size n/m, each block has the approximate criterion

$$\sigma\sqrt{\frac{n}{m}C} < d. \tag{12.26}$$

Perturbations are then confined to smaller subsystems. There is an important bookkeeping caution: turning cross-module interactions into zeros also changes the connectance of the full matrix. A fair comparison must specify whether C is measured within blocks or over all possible pairs.

May's conclusion is not that real large ecosystems must be unstable. It is that a large, dense, strongly interacting *random* network is difficult to stabilize. Real networks are not random: they are sparse, modular, trophically organized, and shaped by ecology and evolution. May's result transforms that observation into a question—what structure allows complex natural systems to remain stable?

Finally, Equation (12.24) is a local statement about one proposed fixed point. If that point is unstable, a nonlinear system may approach a different fixed point, a limit cycle, a chaotic attractor, or a state in which some variables have vanished.

12.5 Predator–prey dynamics

The Lotka–Volterra model, developed independently by Lotka (1920) and Volterra (1926), gives a concrete two-dimensional example of interaction-driven oscillations. Let $x(t)$ be prey abundance and $y(t)$ predator abundance:

$$\frac{dx}{dt} = \alpha x - \beta xy, \tag{12.27}$$

$$\frac{dy}{dt} = \delta xy - \gamma y. \tag{12.28}$$

In the absence of predators, prey grow exponentially at rate α. Encounters remove prey at rate βxy. Predators die at rate γ, while successful encounters increase their population at rate δxy.

The model has the boundary extinction equilibrium $(0, 0)$, which is a saddle: prey increase when introduced without predators. It also has the positive fixed point

$$(x^*, y^*) = \left(\frac{\gamma}{\delta}, \frac{\alpha}{\beta}\right). \tag{12.29}$$

The Jacobian is

$$J(x, y) = \begin{pmatrix} \alpha - \beta y & -\beta x \\ \delta y & \delta x - \gamma \end{pmatrix}. \tag{12.30}$$

At the positive fixed point,

$$J^* = \begin{pmatrix} 0 & -\beta\gamma/\delta \\ \delta\alpha/\beta & 0 \end{pmatrix}, \qquad \lambda_\pm = \pm i\sqrt{\alpha\gamma}. \tag{12.31}$$

Linearization therefore predicts neither decay nor growth. In fact, the nonlinear model has the conserved quantity

$$H(x, y) = \delta x - \gamma \ln\!\left(\frac{x}{x^*}\right) + \beta y - \alpha \ln\!\left(\frac{y}{y^*}\right), \tag{12.32}$$

where $x^* = \gamma/\delta$ and $y^* = \alpha/\beta$. The ratios inside the logarithms are dimensionless; changing their reference scales would add only a constant to H and would not change its level curves. The

conservation law applies only for $x > 0$ and $y > 0$, since it is singular on the extinction boundaries. It can be checked directly:

$$\begin{aligned}\frac{dH}{dt} &= \left(\delta - \frac{\gamma}{x}\right)\frac{dx}{dt} + \left(\beta - \frac{\alpha}{y}\right)\frac{dy}{dt} \\ &= (\delta x - \gamma)(\alpha - \beta y) + (\beta y - \alpha)(\delta x - \gamma) = 0. \end{aligned} \tag{12.33}$$

Therefore every positive initial condition lies on a closed orbit. Prey rise first, predators rise after a delay, abundant predators drive prey down, and predator abundance then declines from lack of food. Historical hare–lynx records helped make this phase-shifted cycle iconic.

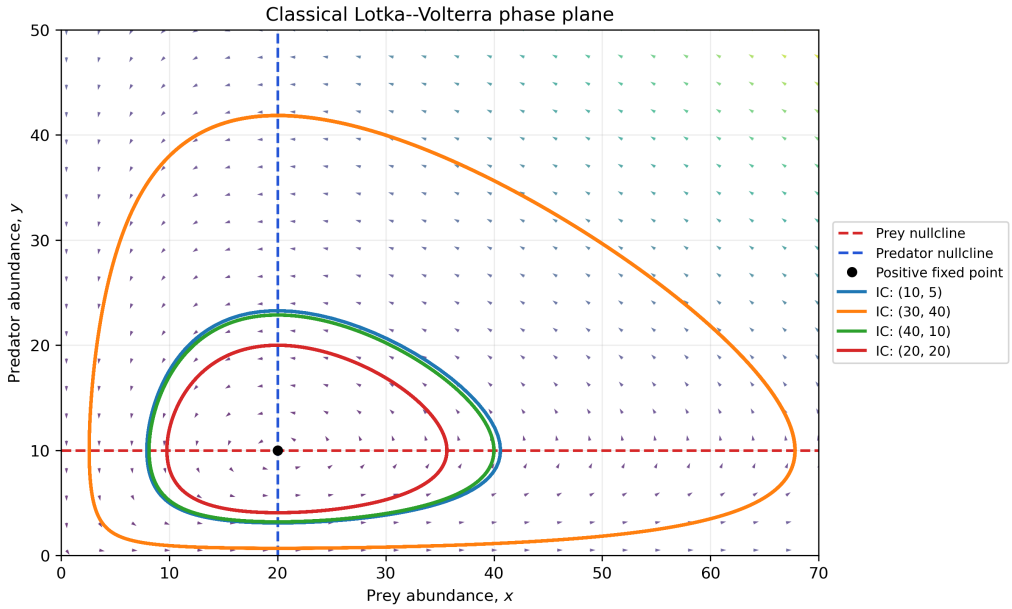

Figure 12.6: Classical Lotka–Volterra phase plane. Red and blue dashed lines are the prey and predator nullclines; their intersection is the positive fixed point. Trajectories follow different closed level sets of $H(x, y)$, so oscillation amplitude depends on the initial condition. The plot is course-generated from Lotka (1920) and Volterra (1926).

These are *neutral* cycles, not an attracting biological clock. A small perturbation moves the system to a neighboring orbit rather than returning it to the original one. This fragility is a sign that an important ecological process is missing.

12.6 Resource limitation damps the neutral cycles

Prey cannot grow exponentially forever. Adding logistic growth gives

$$\frac{dx}{dt} = rx\left(1 - \frac{x}{K}\right) - \beta xy, \tag{12.34}$$

$$\frac{dy}{dt} = \delta xy - \gamma y. \tag{12.35}$$

The positive fixed point is

$$x^* = \frac{\gamma}{\delta}, \qquad y^* = \frac{r}{\beta}\left(1 - \frac{\gamma}{\delta K}\right), \tag{12.36}$$

which exists only if $K > \gamma/\delta$. At this point,

$$\operatorname{tr} J^* = -\frac{rx^*}{K} < 0, \qquad \det J^* = \beta\delta x^* y^* > 0. \tag{12.37}$$

For any two-dimensional Jacobian, the characteristic polynomial is

$$\lambda^2 - (\operatorname{tr} J)\lambda + \det J = 0. \tag{12.38}$$

If the roots are real, positive determinant makes their signs agree and negative trace makes both negative. If they are complex conjugates, each has real part $\operatorname{tr} J/2 < 0$. Both eigenvalues therefore have negative real parts. Depending on the discriminant, the approach may be monotonic or oscillatory, but the interior equilibrium is attracting.

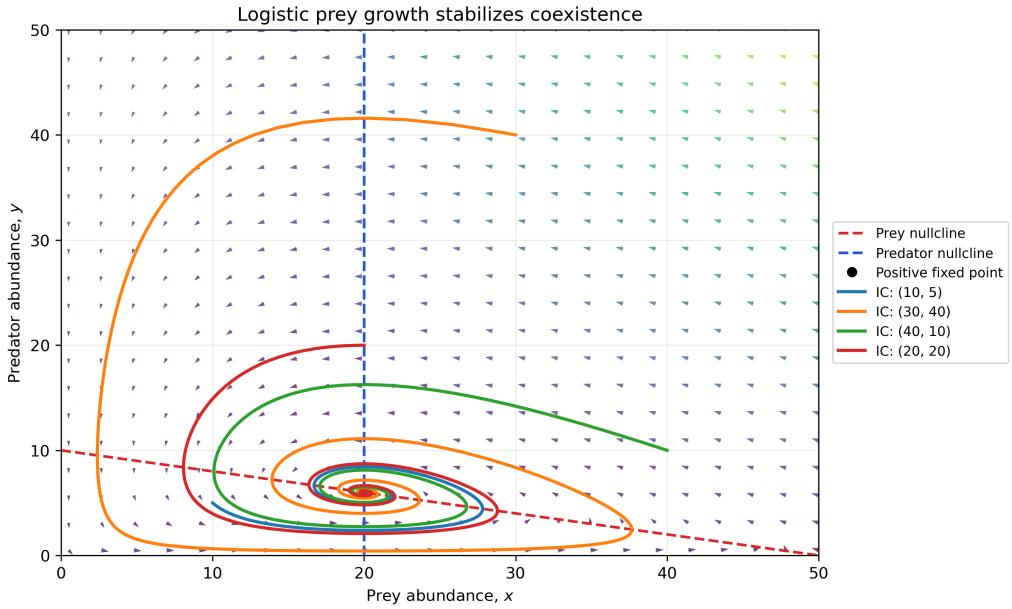

Figure 12.7: Lotka–Volterra dynamics with logistic prey growth. The prey nullcline is now sloped, and trajectories spiral toward a stable coexistence point rather than remaining on neutral closed orbits. The plot is course-generated.

Sustained ecological cycles require additional nonlinear structure. One common extension is the Rosenzweig–MacArthur model, which replaces unlimited mass-action consumption by a saturating functional response:

$$\frac{dx}{dt} = rx\left(1 - \frac{x}{K}\right) - \frac{axy}{1 + ahx}, \tag{12.39}$$

$$\frac{dy}{dt} = b\frac{axy}{1 + ahx} - dy. \tag{12.40}$$

The handling time h limits how rapidly one predator can consume prey. As parameters such as K change, the coexistence fixed point can lose stability through a Hopf bifurcation and an attracting limit cycle can emerge. Unlike the classical Lotka–Volterra orbit, such a cycle has a selected amplitude and attracts nearby trajectories.

Enrichment: deriving the Hopf threshold. Write the saturating consumption rate per predator as

$$g(x) = \frac{ax}{1 + ahx}. \tag{12.41}$$

At a positive fixed point, the predator equation requires $bg(x^*) = d$, so

$$x^* = \frac{d}{a(b - dh)}, \qquad y^* = \frac{brx^*}{d}\left(1 - \frac{x^*}{K}\right). \tag{12.42}$$

Existence therefore requires $b > dh$ and $K > x^*$. At this point the Jacobian has

$$J_{11} = r\left(1 - \frac{2x^*}{K}\right) - g'(x^*)y^*, \quad J_{12} = -g(x^*), \quad J_{21} = bg'(x^*)y^*, \quad J_{22} = 0. \tag{12.43}$$

Because g, g', and y^* are positive, $\det J^* = b\, g(x^*)g'(x^*)y^* > 0$; stability changes when the trace crosses zero. Using

$$\frac{g'(x)}{g(x)} = \frac{1}{x(1 + ahx)} \tag{12.44}$$

and substituting y^* gives

$$\operatorname{tr} J^* = r\,\frac{ahx^* - \dfrac{x^*}{K}(1 + 2ahx^*)}{1 + ahx^*}. \tag{12.45}$$

The trace vanishes at

$$K = K_{\mathrm{H}} = \frac{1}{ah} + 2x^*. \tag{12.46}$$

Thus the coexistence point is locally stable for $x^* < K < K_{\mathrm{H}}$ and unstable for $K > K_{\mathrm{H}}$. Under the usual nondegeneracy conditions, the crossing is a Hopf bifurcation and the standard Rosenzweig–MacArthur model develops an attracting cycle. Increasing the prey carrying capacity can therefore destabilize coexistence, a result known as the paradox of enrichment.

12.7 Generalized Lotka–Volterra communities

For n interacting species, a convenient extension is

$$\frac{dN_i}{dt} = N_i\left[r_i\left(1 - \frac{N_i}{K_i}\right) + \sum_{j \neq i} a_{ij}N_j\right], \qquad i = 1, \ldots, n. \tag{12.47}$$

Here r_i is the intrinsic growth rate and K_i is the carrying capacity of species i in isolation. The coefficient a_{ij} is the per-capita effect of species j on species i: positive values represent facilitation and negative values represent harm. Equivalently, one may write the bracket as $r_i + \sum_j a_{ij}N_j$ by including self-regulation in the interaction matrix with $a_{ii} = -r_i/K_i$.

At a fixed point where every species is present,

$$0 = r_i\left(1 - \frac{N_i^*}{K_i}\right) + \sum_{j \neq i} a_{ij}N_j^*. \tag{12.48}$$

Because the bracket in Equation (12.47) vanishes there, the Jacobian simplifies to

$$J_{ii} = -\frac{r_i N_i^*}{K_i}, \qquad J_{ij} = N_i^* a_{ij} \quad (i \neq j). \tag{12.49}$$

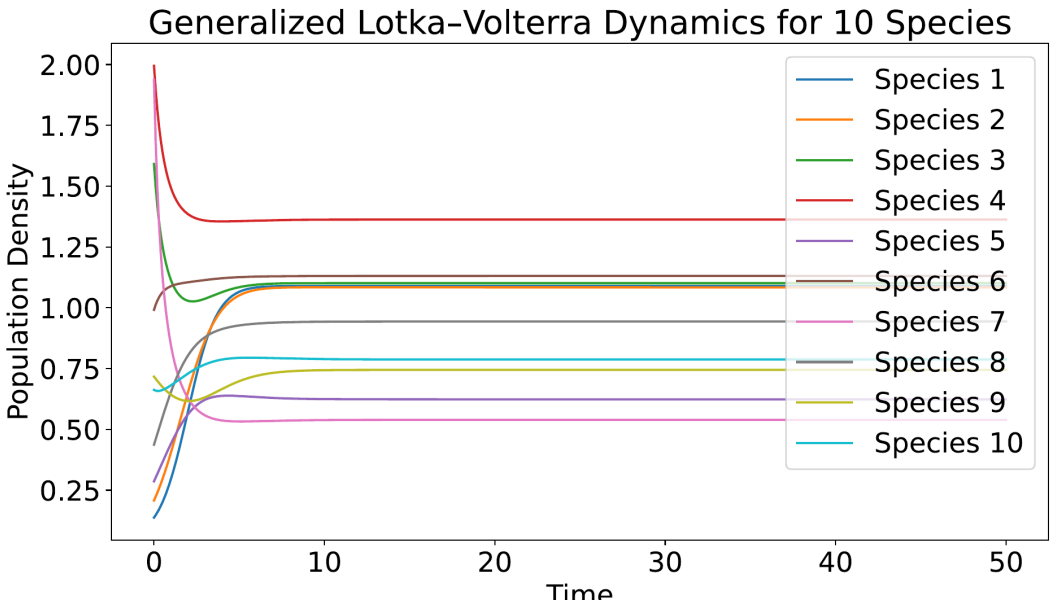

Figure 12.8: A ten-species generalized Lotka–Volterra simulation approaching a stable coexistence state. Transient trajectories differ strongly, but all abundances settle to constant values. The plot is course-generated.

This is the bridge back to May: local stability is again controlled by the eigenvalues of a community matrix. But the gLV Jacobian is not an arbitrary random matrix. Its rows are scaled by equilibrium abundances, its diagonal reflects self-limitation, and species that go extinct are removed from the surviving community.

The same model family can yield coexistence, extinction of some species, a stable fixed point, a limit cycle, or chaotic dynamics. Stability and feasibility are distinct: a fixed point may be mathematically stable while assigning a negative abundance to one species, in which case it is not biologically admissible. Conversely, extinctions can leave a smaller surviving community whose equilibrium is stable.

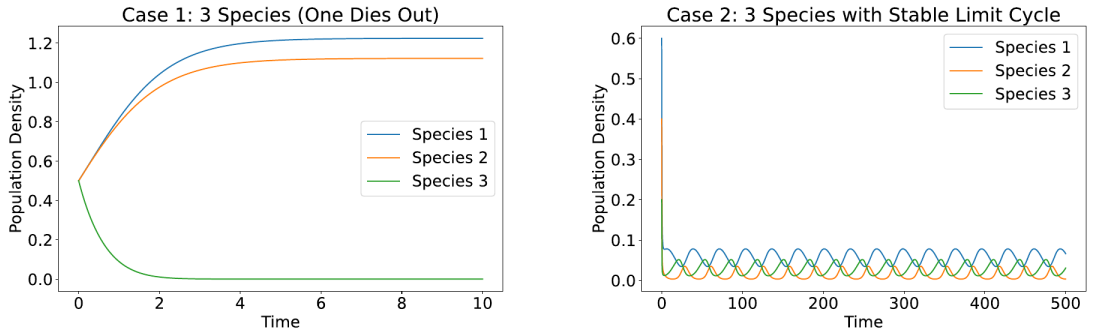

Figure 12.9: Two possible gLV outcomes. One parameter set drives a species to extinction (left); another produces a stable three-species limit cycle (right). The plots are course-generated.

12.7.1 Interaction signs

For a pair of species, the signs of a_{ij} and a_{ji} classify the direct interaction. The ordered pair depends on which species is labeled i, but the biological categories do not.

a_{ij}	a_{ji}	Interaction
$+$	$+$	mutualism: both species benefit
$-$	$-$	competition: both species are harmed
$+$	$-$	predation or parasitism
$+$	0	commensalism: one benefits, one is unaffected
$-$	0	amensalism: one is harmed, one is unaffected
0	0	neutralism: neither has a direct effect

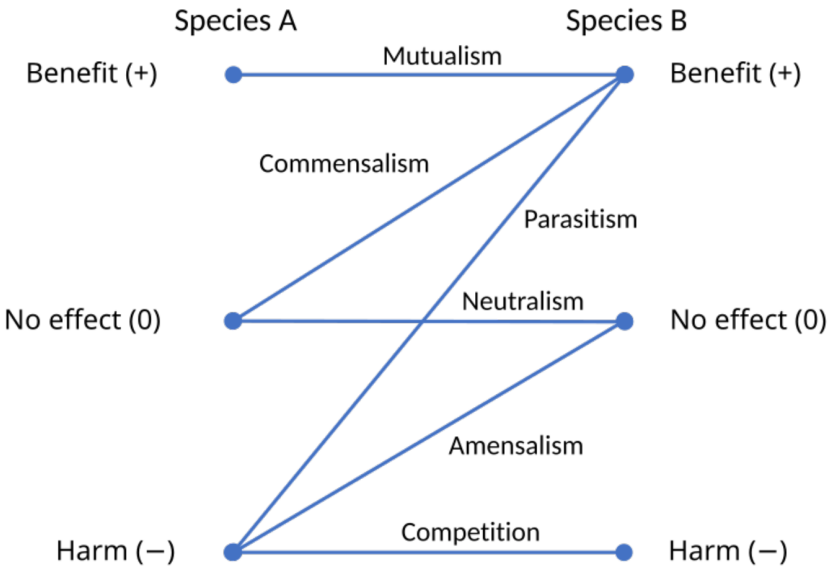

Figure 12.10: The six idealized pairwise relationships obtained from benefit, harm, or no effect on each species. Figure from Wikipedia/Wikimedia Commons.

In real communities, a measured association need not be a direct a_{ij}. Two species can covary because they consume the same resource, respond to the same environmental driver, or are both influenced by a third species. Inferring r_i and a_{ij} is therefore hard: the number of unknowns grows as n^2, abundance measurements are noisy, and environmental conditions vary. Long time series, controlled perturbations, or explicit consumer–resource models are often needed to separate direct interaction from common cause.

12.8 From random interactions to structured dynamics

The combination of May's argument and gLV dynamics suggests a useful synthesis. Randomly chosen interactions may generate a transiently complicated high-dimensional system, but ecological selection, extinction, resource partitioning, and evolutionary change reshape the surviving network.

The effective long-time community can become sparse, modular, and highly structured even if the initial interaction matrix was not.

This is why the complexity–stability question cannot be answered by n alone. Stability depends on interaction strength, connectance, reciprocity, self-regulation, and the fixed point about which the system is linearized. The most useful role of the random-matrix criterion is as a baseline. Deviations from that baseline reveal which biological structures carry the stability.

12.9 Key results

- Delayed negative feedback can turn correction into overshoot and, beyond a threshold, into a sustained oscillator.
- The repressilator uses three nonlinear repression steps plus mRNA–protein dynamics to create an effective delay. Its symmetric fixed point loses stability when the local repression slope exceeds $\chi_c(\beta)$, allowing a tunable cellular rhythm.
- A fixed point of $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$ is locally stable when every eigenvalue of its Jacobian has negative real part.
- For May's random community matrix, the characteristic spectral radius is $\sigma\sqrt{nC}$, giving the approximate stability condition $\sigma\sqrt{nC} < d$.
- Classical Lotka–Volterra predator–prey dynamics have neutral closed orbits; logistic prey growth stabilizes the coexistence point, while richer functional responses can create attracting limit cycles.
- Generalized Lotka–Volterra models connect pairwise interaction networks to ecological outcomes, but feasibility, extinction, structure, and indirect effects must be considered alongside local stability.

13 Stochastic nonlinear dynamics: epidemics and genetic drift

Deterministic models describe the typical motion of a large population, but they do not determine the fate of a process that begins with only a few individuals. An epidemic with $R_0 > 1$ can still disappear before it takes off, and a beneficial mutation can still be lost before selection amplifies it. In both problems, discrete events create fluctuations near an absorbing boundary.

This week develops that common structure in two settings:

- a stochastic SIR epidemic, including early extinction, branching-process theory, and superspreading; and
- the Moran process, including genetic drift, selection, and the probability that a mutation reaches fixation.

The central theme is that a deterministic growth advantage changes a probability of success; it does not guarantee success from a finite initial population.

By the end of this week, students should be able to answer:

- Why can a supercritical epidemic or a beneficial mutation disappear when it begins from only a few individuals?
- How do offspring distributions and transmission heterogeneity determine epidemic extinction probabilities?
- How do drift and selection determine fixation in the Moran process?
- When does mutation supply combine with fixation probability to produce a molecular clock?

13.1 Stochastic epidemic dynamics: demographic variation and superspreading

13.1.1 Deterministic SIR baseline: recall from Week 10

Week 10 introduced the deterministic SIR model with susceptible, infective, and recovered fractions s, i, and r. In time units of the mean infectious period, its dynamics are given by Equations (10.38)–(10.40). The early growth rate is $R_0 s_0 - 1$, the infective fraction peaks at $s = 1/R_0$, and the final susceptible fraction obeys Equation (10.47). This deterministic baseline predicts a continuous trajectory but cannot assign a probability to early extinction when the outbreak begins with only a few infective individuals.

13.1.2 Where and why determinism fails: demographic variation

The deterministic approximation assumes that each compartment is large enough that fluctuations are negligible. This fails in two systematic regimes:

Initiation (small I). When an outbreak is seeded by $I(0) \in \{1, 2, \dots\}$ infectives, chance events dominate: the initial infectives may recover before generating any secondary infections. This implies *early extinction* ("fizzling") even when $R_0 > 1$, a phenomenon absent from the continuous ODE model.

Termination (small I again). Near the end of an outbreak, discreteness matters again: the epidemic ends when I hits exactly zero, whereas the ODE model approaches $i \to 0$ continuously.

These finite-count stochastic effects are often termed *demographic variation*, or demographic noise.

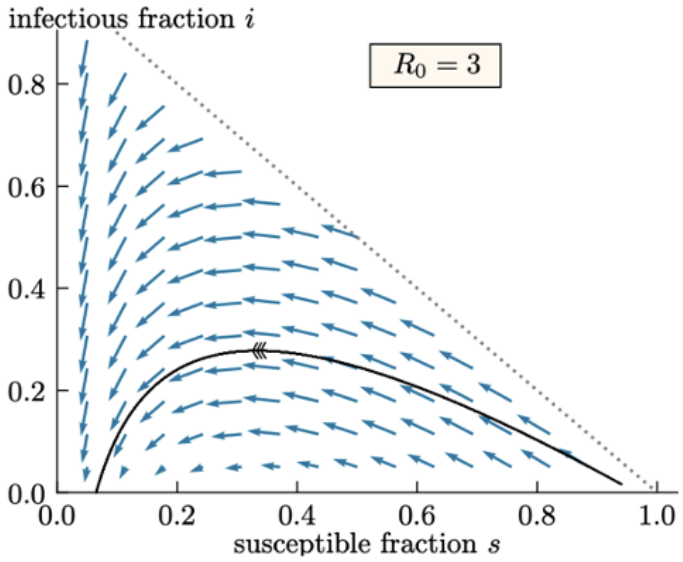

Figure 13.1: Deterministic SIR phase portrait. In the ODE model, trajectories do not "fizzle"; for $R_0 > 1$ they follow the flow to a nontrivial endpoint $s_\infty < 1$ on the $i = 0$ boundary (Nelson 2022).

13.1.3 Stochastic SIR as a Markov jump process (Gillespie direct method)

In the stochastic SIR model, the state is integer-valued (S, I) (with $R = N - S - I$). Two events occur with state-dependent propensities (hazards):

Infection. $(S, I) \to (S - 1, I + 1)$ with propensity

$$a_{\text{inf}}(S, I) = \beta \frac{SI}{N}. \tag{13.1}$$

Recovery. $(S, I) \to (S, I - 1)$ with propensity

$$a_{\text{rec}}(S, I) = \gamma I. \tag{13.2}$$

The total propensity is

$$a_0(S, I) = a_{\text{inf}}(S, I) + a_{\text{rec}}(S, I). \tag{13.3}$$

Gillespie direct method. Given the current state (S, I) at time t:

1. Draw $u_1, u_2 \sim \text{Uniform}(0, 1)$.
2. Sample the waiting time to the next event:

$$\Delta t = -\frac{\ln u_1}{a_0(S, I)}. \tag{13.4}$$

3. Choose the event type: infection if $u_2 < a_{\text{inf}}/a_0$, otherwise recovery.
4. Update (S, I) and advance time by Δt.
5. Stop when $I = 0$ (extinction) or at a prescribed terminal time.

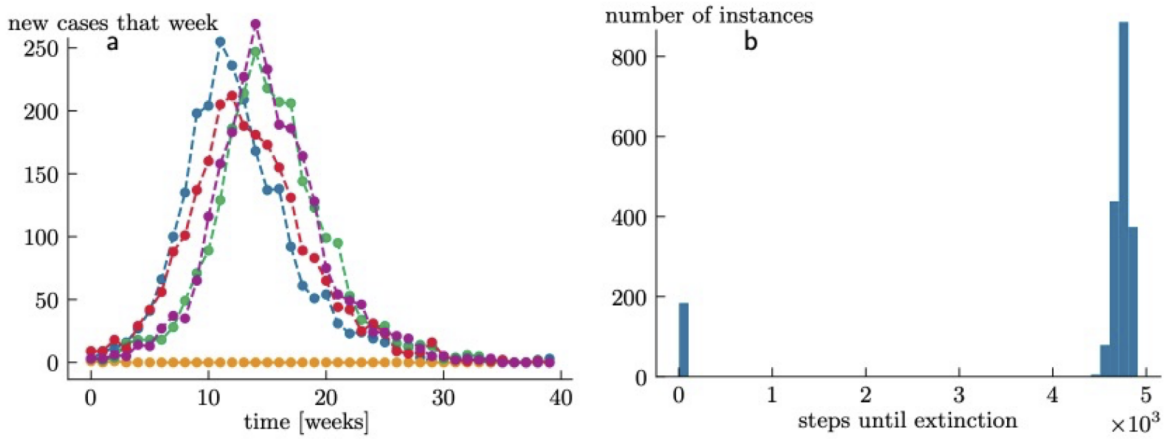

Figure 13.2: Stochastic SIR simulations with $N = 2816$, $I(0) = 3$, and $R_0 = 2.2$. Some realizations fizzle almost immediately, while successful outbreaks follow similar macroscopic time courses (left). The distribution of extinction times is strongly bimodal (right) (Nelson 2022).

Representative parameters. Example simulations use $N = 2816$ and initial $I(0) = 3$, illustrating both rapid extinctions and successful outbreaks for R_0 values such as 2.2 and 1.3.

13.1.4 Secondary infections per individual: a Poisson–exponential mixture

A useful individual-level random variable is the *attributed infection count* ℓ_{inf}: the number of secondary infections caused by a given infective early in an outbreak.

In the early phase $S \approx N$, infection events generated by one infective are approximately a Poisson process. Work in dimensionless time $\tau = \gamma t$, so that recoveries occur at unit rate and the infection rate is approximately R_0. Let T be the infectious period in τ-time. Under a constant recovery hazard,

$$T \sim \text{Exponential}(1). \tag{13.5}$$

Conditional on T, the number of secondary infections is

$$\ell_{\text{inf}} \mid T \sim \text{Poisson}(R_0 T). \tag{13.6}$$

This is the same Poisson–exponential mixture derived for translational bursts in Week 8; compare Equations (8.57) and (8.58). Here the recovery rate and infection rate play the roles of the mRNA-decay and protein-production rates. Reusing that result with mean burst size R_0 gives the geometric law

$$\Pr(\ell = l) = \frac{1}{1 + R_0} \left(\frac{R_0}{1 + R_0} \right)^l, \qquad l = 0, 1, 2, \ldots \tag{13.7}$$

with

$$\mathbb{E}[\ell] = R_0, \qquad \text{Var}(\ell) = R_0(1 + R_0). \tag{13.8}$$

This is the zero-based convention for the geometric distribution. It describes the number of failures before a success and is the distribution of $J - 1$ for the trial-index variable J introduced in Section 2.4. Thus $\log \Pr(\ell)$ is exactly linear in ℓ under this null; observed data may only approximately follow that semilog straight line.

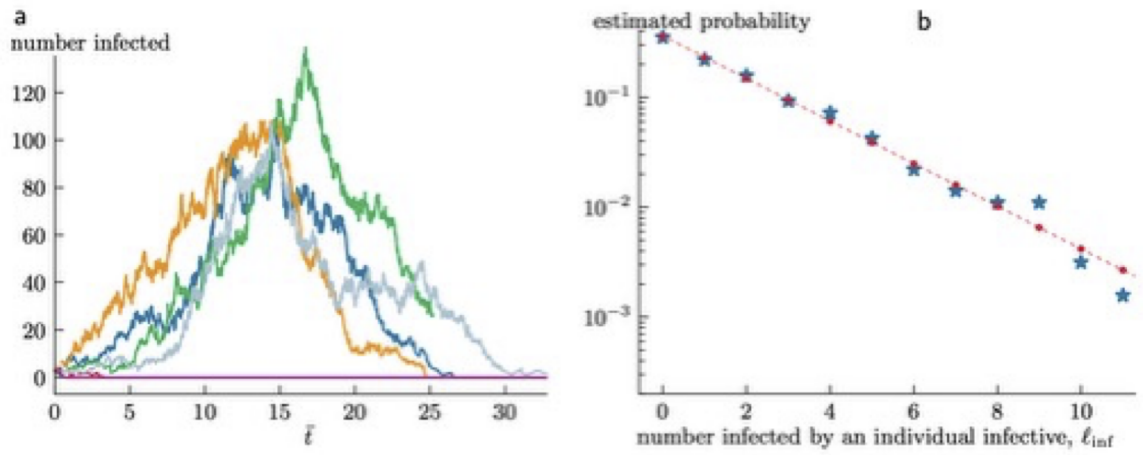

Figure 13.3: Stochastic SIR dynamics at $R_0 = 1.3$. Representative outbreaks include both rapid extinction and successful epidemic curves (left). The distribution of secondary infections per infective is approximately geometric under the Poisson–exponential null model (right) (Nelson 2022).

13.1.5 Branching-process interpretation and fizzle probability

Extinction probability: general theory. Let Z_n be the number of infectives in generation n of the branching process; generation zero contains the initially infected individual. Let the offspring generating function be

$$G(z) \equiv \mathbb{E}[z^{\ell}] = \sum_{l=0}^{\infty} \Pr(\ell = l)\, z^{l}. \tag{13.9}$$

If the process starts with $Z_0 = 1$, define the extinction probability

$$q \equiv \Pr\big(\exists\, n \text{ such that } Z_n = 0 \,\big|\, Z_0 = 1\big). \tag{13.10}$$

Condition on the first-generation size $Z_1 = \ell$. If $\ell = l$, then the process splits into l independent lineages, each of which goes extinct with probability q; hence the conditional extinction probability is q^{l}. Averaging over ℓ gives

$$q = \sum_{l=0}^{\infty} \Pr(\ell = l)\, q^{l} = \mathbb{E}[q^{\ell}] = G(q), \tag{13.11}$$

and the relevant solution is the smallest $q \in [0, 1]$. The extinction event grows as successive generations are observed, which is why the smallest fixed point is the physical extinction probability.

For $Z_0 = k$ initial infectives and independent lineages, extinction requires all k lineages to go extinct, so

$$q_k = q^{k}. \tag{13.12}$$

A fundamental threshold is controlled by the mean offspring number $m = \mathbb{E}[\ell]$. Note that $G(1) = 1$ always, so $q = 1$ is always a solution. Also,

$$G'(1) = \mathbb{E}[\ell] = m. \tag{13.13}$$

For a nondegenerate offspring law, meaning $\Pr(\ell = 1) < 1$, convexity of G implies that if $m \leq 1$, there is no fixed point below one and extinction is certain: $q = 1$. The excluded deterministic case

$\ell = 1$ with probability one persists forever without growing or becoming extinct. If $m > 1$, then $G'(1) > 1$, so $G(z)$ crosses the line z at an additional point $q \in (0, 1)$, yielding $q < 1$ and a positive probability of "takeoff" (non-extinction).

In the SIR initiation regime, $m = \mathbb{E}[\ell] = R_0$, reproducing the familiar threshold $R_0 = 1$ but now with a quantitative *probability* of fizzling even when $R_0 > 1$.

Extinction probability for the geometric offspring law. For the geometric distribution in Equation (13.7), let $p = 1/(1 + R_0)$ and $1 - p = R_0/(1 + R_0)$. Its generating function is

$$G(z) = \sum_{l=0}^{\infty} p(1-p)^l z^l = \frac{p}{1-(1-p)z} = \frac{1}{1+R_0-R_0 z}. \tag{13.14}$$

Substituting into (13.11) yields

$$q = \frac{1}{1+R_0-R_0 q} \quad \Longrightarrow \quad R_0 q^2 - (1+R_0)q + 1 = 0. \tag{13.15}$$

This quadratic has roots $q = 1$ and $q = 1/R_0$. Selecting the smallest root in $[0, 1]$,

$$q = \begin{cases} 1, & R_0 \le 1, \\ \frac{1}{R_0}, & R_0 > 1. \end{cases} \tag{13.16}$$

Thus, for $R_0 > 1$, the probability of successful takeoff from a single introduction is $1 - q = 1 - 1/R_0$. If there are k initial infectives, the fizzle probability is $q_k = (1/R_0)^k$.

Interpretation and limits of validity. Equation (13.16) makes the "fizzle" phenomenon explicit: even in a supercritical setting $R_0 > 1$, a finite fraction of introductions die out through chance recovery before transmission. The approximation is valid primarily during the initiation phase, when I is small and $S \approx N$. Once I becomes appreciable, depletion of susceptibles and correlations between infectives invalidate the independence and stationarity assumptions. Conditional on avoiding early extinction, the dynamics then cross over to the deterministic SIR trajectory.

13.1.6 Superspreading and overdispersion

The geometric law in Equation (13.7) is a principled null model: overdispersion arises solely from the random infectious period. Real outbreaks, however, may show substantially heavier tails because of superspreading.

A simple diagnostic compares the observed variance to the geometric identity

$$\mathrm{Var}(\ell) = \mathbb{E}[\ell]\big(\mathbb{E}[\ell] + 1\big). \tag{13.17}$$

For example, values such as $\mathbb{E}[\ell] \approx 0.9$ but $\mathrm{Var}(\ell) \approx 16$ are far larger than $0.9 \times 1.9 \approx 1.71$, indicating overdispersion beyond the geometric null.

13.1.7 Minimal superspreader extension: two infective classes

A minimal mechanistic extension introduces two classes of infectives:

- **Regular** infectives with reproduction number $R_0^{(r)}$.

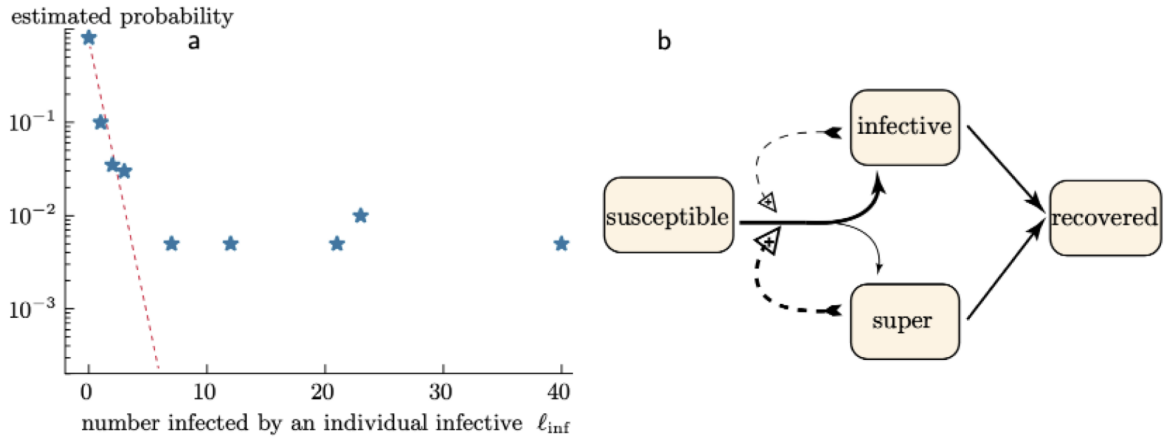

Figure 13.4: Evidence for superspreading (long-tailed ℓ_{inf}) and a minimal heterogeneous model in which a small fraction of individuals have much larger transmission potential. The outbreak data (left) are from the CDC SARS report; the heterogeneous model (right) is from Nelson (2022).

- **Superspreaders** with reproduction number $R_0^{(s)} \gg R_0^{(r)}$.

Let p be the probability that a newly infected individual is a superspreader. In the early phase, the offspring distribution becomes a mixture,

$$\Pr(\ell = l) = (1-p)\,\Pr_{R_0^{(r)}}(\ell = l) + p\,\Pr_{R_0^{(s)}}(\ell = l), \tag{13.18}$$

where $\Pr_{R_0}(\ell = l)$ can be taken as the geometric law (13.7) within each class. The aggregate mean offspring number is

$$R_{\text{eff}} = (1-p)R_0^{(r)} + pR_0^{(s)}. \tag{13.19}$$

This mean retains the invasion threshold $R_{\text{eff}} > 1$, but it does not by itself determine the fizzle probability. If each newly infected individual independently draws its class, the mixture generating function is

$$G_{\text{mix}}(z) = \frac{1-p}{1 + R_0^{(r)} - R_0^{(r)} z}\,\frac{p}{1 + R_0^{(s)} - R_0^{(s)} z}, \tag{13.20}$$

and the extinction probability is the smallest solution of $q = G_{\text{mix}}(q)$. It is not the weighted average of the two within-class extinction probabilities. Persistent or inherited transmission types would instead require a multitype branching process.

Why mixtures inflate variance. Even if each class individually follows the geometric null, heterogeneity between classes adds variance and produces heavier tails. A small p can dominate tail behavior if $R_0^{(s)}$ is large.

Representative parameters. An example scenario uses $p = 0.02$ (2%) superspreaders and an approximately $20\times$ larger R_0 in the superspreader class; below we take $R_0^{(s)} = 25$. Such heterogeneity can markedly increase both outbreak severity, conditional on success, and run-to-run variability.

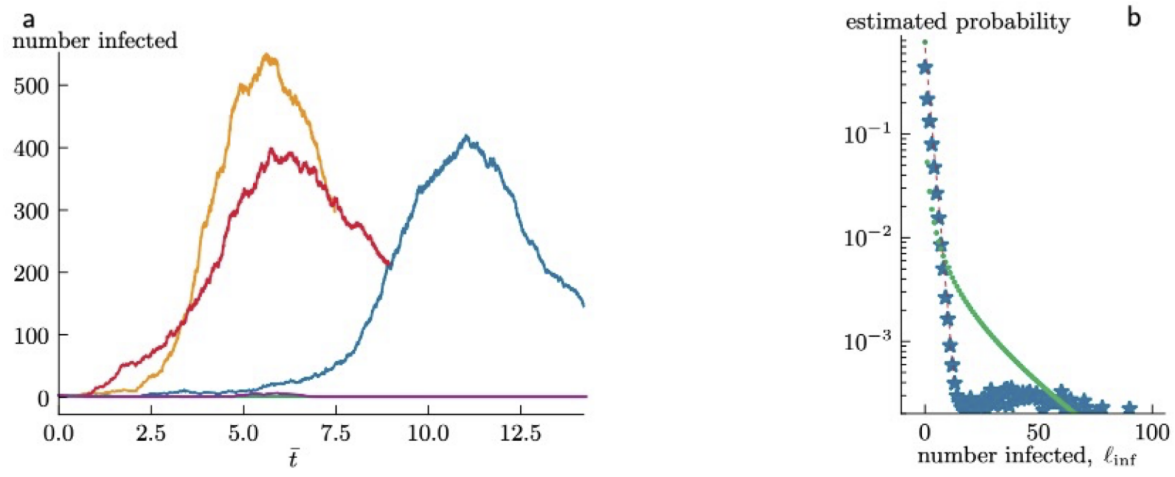

Figure 13.5: Superspreaders increase outbreak variability and lengthen the tail of the offspring distribution. A small high-transmission class is compared with the baseline stochastic SIR model (Nelson 2022).

Explicit overdispersion calculation: variance-to-mean ratio (Fano factor). A compact way to quantify "superspreading" is the variance-to-mean ratio

$$\text{VMR} \;\equiv\; \frac{\text{Var}(\ell)}{\mathbb{E}[\ell]}, \tag{13.21}$$

also known as the Fano factor. For a two-class mixture (regular and superspreader), let the class be $C \in \{r, s\}$ with $\Pr(C = s) = p$, $\Pr(C = r) = 1 - p$. Write class-conditional means and variances as

$$m_r = \mathbb{E}[\ell \mid C = r], \quad v_r = \text{Var}(\ell \mid C = r), \qquad m_s = \mathbb{E}[\ell \mid C = s], \quad v_s = \text{Var}(\ell \mid C = s).$$

Then

$$\mathbb{E}[\ell] = (1 - p)m_r + pm_s \;\equiv\; m. \tag{13.22}$$

Applying the law of total variance from Equation (8.74) to the class variable C gives

$$\begin{aligned} \text{Var}(\ell) &= \mathbb{E}\big[\text{Var}(\ell \mid C)\big] + \text{Var}\big(\mathbb{E}[\ell \mid C]\big) \\ &= (1 - p)v_r + pv_s \;+\; (1 - p)(m_r - m)^2 + p(m_s - m)^2. \end{aligned} \tag{13.23}$$

If, as in the basic SIR initiation model, each class has a geometric offspring law with mean $R_0^{(j)}$, then $m_j = R_0^{(j)}$ and $v_j = m_j(1 + m_j)$ for $j \in \{r, s\}$. Substitution into Equation (13.23) gives an explicit closed form for $\text{Var}(\ell)$ and hence VMR.

Numerical example. Take $p = 0.02$ superspreaders and $R_0^{(s)} = 25$. If their reproduction number is $20\times$ that of regular infectives, then $R_0^{(r)} = 25/20 = 1.25$. Then

$$m = (1 - p)R_0^{(r)} + pR_0^{(s)} = 0.98 \cdot 1.25 + 0.02 \cdot 25 = 1.725, \tag{13.24}$$

$$v_r = R_0^{(r)}\big(1 + R_0^{(r)}\big) = 1.25 \cdot 2.25 = 2.8125, \tag{13.25}$$

$$v_s = R_0^{(s)}\big(1 + R_0^{(s)}\big) = 25 \cdot 26 = 650. \tag{13.26}$$

The between-class contribution is large:

$$(m_r - m)^2 = (1.25 - 1.725)^2 = 0.225625, \qquad (m_s - m)^2 = (25 - 1.725)^2 = 541.725625.$$

Therefore

$$\begin{aligned} \mathrm{Var}(\ell) &= 0.98\,(2.8125 + 0.225625) + 0.02\,(650 + 541.725625) \\ &= 0.98 \cdot 3.038125 + 0.02 \cdot 1191.725625 \\ &\approx 2.97736 + 23.83451 = 26.81188, \end{aligned} \tag{13.27}$$

and the variance-to-mean ratio is

$$\mathrm{VMR} = \frac{\mathrm{Var}(\ell)}{\mathbb{E}[\ell]} \approx \frac{26.8119}{1.725} \approx 15.5. \tag{13.28}$$

For comparison, a *single-class* geometric offspring model with the same mean $m = 1.725$ would have $\mathrm{VMR} = 1 + m \approx 2.725$. Thus, even a small superspreader fraction can inflate the dispersion by a factor of approximately $15.5/2.7 \approx 5.7$, with the dominant contribution coming from between-class heterogeneity rather than within-class geometric variability.

The same matched-mean comparison changes extinction. Solving Equation (13.20) with these parameters gives

$$q_{\mathrm{mix}} \simeq 0.748, \qquad q_{\mathrm{homogeneous}} = \frac{1}{m} \simeq 0.580. \tag{13.29}$$

Thus heterogeneity can make introductions more likely to fizzle even though the occasional successful outbreaks are driven by a much longer transmission tail.

13.2 Population genetics and evolutionary dynamics

13.2.1 Motivation and conceptual link to epidemic "fizzling"

A newly arising mutation is typically introduced at very low copy number. Even if it is beneficial, it can be lost by chance because reproduction and death are discrete random events in a finite population. This is directly analogous to epidemic initiation: even when $R_0 > 1$, an infection chain can terminate early because the first few infectives may recover before transmitting. The Moran process has two absorbing states, mutation loss and fixation. In the epidemic branching approximation, extinction is absorbing, whereas "takeoff" means escaping the small-number regime and entering a macroscopic outbreak. In both cases, the key question is the probability that a rare lineage avoids early extinction. The serial-dilution experiments discussed in Week 10 make this contrast concrete. Once an evolved type is abundant, its deterministic fitness advantage can be measured by direct competition with its ancestor. At the instant the mutation first appears, however, it is a single lineage whose fate is controlled by discrete birth and death events.

13.2.2 The Moran process: a minimal finite-population model of drift and selection

The Moran process (Moran 1958) describes a haploid population of constant size N. At the locus being modeled, each individual carries one allele, either wild-type A or mutant a. Let

$$X_m \in \{0, 1, \ldots, N\} \tag{13.30}$$

be the number of mutants after replacement event m. The states $X = 0$, mutation loss, and $X = N$, mutation fixation, are absorbing.

The fixation calculation below uses the event-indexed discrete-time Moran chain, obtained by repeating the following elementary event:

1. **Birth (reproduction):** choose one individual to reproduce, with probability proportional to its fitness.
2. **Death (replacement):** choose one individual uniformly at random to be removed.
3. The offspring of the reproducing individual replaces the removed individual, keeping population size fixed at N.

Thus each event changes X by at most ± 1: the mutant count increases by one, decreases by one, or stays the same. Self-transitions affect elapsed time but not fixation probability. A continuous-time Moran process is obtained by letting replacement events occur according to a Poisson clock. Changing the overall clock rate changes fixation times but not fixation probabilities.

Fitness parameterization. Assign fitness 1 to wild-type A, and $1 + s$ to mutant a, where s is the selection coefficient. Define the relative fitness ratio

$$r \equiv 1 + s. \tag{13.31}$$

Fitness weights must be positive, so $r > 0$, equivalently $s > -1$. When $s = 0$ (so $r = 1$), the process is neutral drift.

13.2.3 Transition probabilities per event

Condition on the current mutant count $X = i$. There are i mutants and $N - i$ wild-types.

Probability of choosing a mutant to reproduce. Because reproduction is proportional to fitness, the probability that the reproducing individual is mutant equals

$$\Pr(\text{mutant reproduces} \mid X = i) = \frac{r\, i}{r\, i + (N - i)}. \tag{13.32}$$

Probability of choosing a wild-type to die. Death is uniform over individuals, so

$$\Pr(\text{wild-type dies} \mid X = i) = \frac{N - i}{N}, \qquad \Pr(\text{mutant dies} \mid X = i) = \frac{i}{N}. \tag{13.33}$$

Upward and downward steps. A net increase $i \to i + 1$ happens when a mutant reproduces and a wild-type dies:

$$\Pr(i \to i + 1) = \frac{r\, i}{r\, i + (N - i)} \cdot \frac{N - i}{N}. \tag{13.34}$$

A net decrease $i \to i - 1$ happens when a wild-type reproduces and a mutant dies:

$$\Pr(i \to i - 1) = \frac{N - i}{r\, i + (N - i)} \cdot \frac{i}{N}. \tag{13.35}$$

Events in which the reproducer and the removed individual have the same type leave i unchanged and do not affect fixation probability.

A particularly useful quantity is the *step bias*:

$$\frac{\Pr(i \to i + 1)}{\Pr(i \to i - 1)} = r, \tag{13.36}$$

which is constant, independent of i. This algebraic simplification is the main reason the Moran model yields a clean finite-N fixation formula.

13.2.4 Fixation probability: statement, proof, and weak-selection limit

Let

$$\rho_i \;\equiv\; \Pr(\text{eventual fixation at } X = N \mid X_0 = i) \tag{13.37}$$

be the fixation probability starting from $i \in \{0, 1, \ldots, N\}$ mutants.

Step 1: a recursion for ρ_i. Condition on the next Moran event starting from state $X = i$. Using the one-step transition probabilities $\Pr(i \to i+1)$ and $\Pr(i \to i-1)$ from Equations (13.34)–(13.35), we have

$$\rho_i = \Pr(i \to i + 1)\, \rho_{i+1} + \Pr(i \to i - 1)\, \rho_{i-1} + \Pr(i \to i)\, \rho_i. \tag{13.38}$$

Rearranging gives the standard harmonicity condition

$$0 \;=\; \Pr(i \to i + 1)\, (\rho_{i+1} - \rho_i) \;+\; \Pr(i \to i - 1)\, (\rho_{i-1} - \rho_i), \qquad i = 1, \ldots, N - 1. \tag{13.39}$$

Boundary conditions are

$$\rho_0 = 0, \qquad \rho_N = 1. \tag{13.40}$$

Step 2: reduce to a first-order difference equation using the step bias. Define increments $\Delta_i \equiv \rho_i - \rho_{i-1}$ for $i = 1, \dots, N$. Then Equation (13.39) is equivalent to

$$\Pr(i \to i+1)\,\Delta_{i+1} \;=\; \Pr(i \to i-1)\,\Delta_i. \tag{13.41}$$

Using the Moran step bias in Equation (13.36), $\Pr(i \to i+1)/\Pr(i \to i-1) = r$, we obtain the constant-ratio recursion

$$\Delta_{i+1} \;=\; \frac{1}{r}\,\Delta_i \qquad \Longrightarrow \qquad \Delta_i \;=\; \Delta_1\, r^{-(i-1)}. \tag{13.42}$$

Step 3: impose boundary conditions to obtain ρ_i. Since $\rho_0 = 0$, we can write

$$\rho_i \;=\; \sum_{k=1}^{i} \Delta_k \;=\; \Delta_1 \sum_{k=0}^{i-1} r^{-k}. \tag{13.43}$$

Imposing $\rho_N = 1$ gives

$$1 \;=\; \Delta_1 \sum_{k=0}^{N-1} r^{-k} \qquad \Longrightarrow \qquad \Delta_1 \;=\; \left(\sum_{k=0}^{N-1} r^{-k} \right)^{-1}. \tag{13.44}$$

Substituting into Equation (13.43) yields the exact finite-N Moran fixation probability

$$\rho_i \;=\; \frac{\sum_{k=0}^{i-1} r^{-k}}{\sum_{k=0}^{N-1} r^{-k}} \;=\; \frac{1 - r^{-i}}{1 - r^{-N}}, \qquad (r \neq 1), \tag{13.45}$$

Neutral case ($r = 1$): explicit proof. When $r = 1$, the recursion in Equation (13.41) reduces to $\Delta_{i+1} = \Delta_i$, so all increments are equal: $\Delta_i \equiv \Delta$. Then $\rho_i = i\Delta$ from Equation (13.43). Using $\rho_N = 1$ gives $1 = N\Delta$, hence

$$\rho_i = \frac{i}{N}, \qquad (r = 1). \tag{13.46}$$

This makes the symmetry intuition precise: under neutrality, fixation probability equals initial frequency.

Weak selection: first-order expansion for $|Ns| \ll 1$. Write $r = 1 + s$. Because the powers extend to N, a uniform weak-selection expansion requires $|Ns| \ll 1$, not merely $|s| \ll 1$. Using

$$\ln r = s - \frac{s^2}{2} + O(s^3), \qquad r^{-k} = 1 - ks + \frac{k(k+1)}{2} s^2 + O(k^3 s^3), \tag{13.47}$$

the numerator and denominator of Equation (13.45) become

$$1 - r^{-i} = is - \frac{i(i+1)}{2} s^2 + O(i^3 s^3), \tag{13.48}$$

$$1 - r^{-N} = Ns - \frac{N(N+1)}{2} s^2 + O(N^3 s^3). \tag{13.49}$$

Expanding their ratio to first order in s gives

$$\rho_i \;=\; \frac{i}{N}\left(1 + \frac{s}{2}(N - i)\right) + O\big((Ns)^2\big) \;=\; \frac{i}{N} + \frac{s}{2}\,\frac{i(N-i)}{N} + O\big((Ns)^2\big). \tag{13.50}$$

In particular, for a single new mutant ($i = 1$),

$$\rho_1 \;=\; \frac{1}{N} + \frac{s}{2}\,\frac{N-1}{N} + O\big((Ns)^2\big)\,. \tag{13.51}$$

Equations (13.50)–(13.51) quantify how weak selection perturbs the neutral baseline: the correction is largest for intermediate i, maximal near $i = N/2$, and vanishes at the absorbing boundaries $i = 0, N$. The remainder is written in the combined small parameter Ns, making explicit that an $O(s^2)$ remainder with fixed N would not be uniform as population size changes.

13.2.5 A general birth–death fixation formula

The Moran calculation is a special case of a broader result. Consider any one-dimensional birth–death process on $i = 0, 1, \ldots, N$, with absorbing endpoints. Let α_i be the probability of an upward step $i \to i+1$ and β_i the probability of a downward step $i \to i-1$. Assume $\alpha_i > 0$ for every interior state used in the ratio below, with $\beta_i \geq 0$. Self-transitions do not affect which boundary is eventually reached. Define the local bias ratio

$$\gamma_i = \frac{\beta_i}{\alpha_i}. \tag{13.52}$$

The same increment argument used above gives

$$\rho_i = \frac{1 + \displaystyle\sum_{j=1}^{i-1} \prod_{k=1}^{j} \gamma_k}{1 + \displaystyle\sum_{j=1}^{N-1} \prod_{k=1}^{j} \gamma_k}, \qquad i = 1, \ldots, N-1. \tag{13.53}$$

For the Moran process, $\gamma_i = 1/r$ is independent of i, so both sums are geometric and Equation (13.53) reduces to Equation (13.45). When fitness depends on frequency, spatial position, or ecological state, the γ_i need not be constant, but the same formula still applies as long as the process remains one dimensional with nearest-neighbor steps. If an interior upward probability vanishes, the ratio representation becomes singular and that state is a barrier to fixation at N; the state space should then be split at the barrier and the boundary-value problem solved on each reachable component.

13.2.6 Mutation supply, fixation, and the molecular clock

Fixation probability alone does not determine how rapidly evolution proceeds; mutations must first arise. Let u be the total neutral mutation probability per reproductive event into the class of alternatives whose substitution rate is being calculated. One Moran generation consists of approximately N replacement events. A population of size N therefore supplies new mutants at rate approximately Nu per generation. If each new mutant fixes with probability ρ_1, successful substitutions occur at rate

$$R = Nu\rho_1. \tag{13.54}$$

Equivalently, the waiting time to the next successful substitution is approximately exponential with mean $1/(Nu\rho_1)$ when successful mutations are rare and independent. This is the origin–fixation, or weak-mutation, regime: a mutant lineage is normally lost or fixed before another relevant lineage competes with it. When mutation supply is high, several lineages can segregate simultaneously,

clonal interference or recombination can change their fixation probabilities, and $Nu\rho_1$ is no longer a complete substitution rate model.

For a neutral mutant, Equation (13.46) gives $\rho_1 = 1/N$. Therefore

$$R = Nu\frac{1}{N} = u. \tag{13.55}$$

A larger population produces more neutral mutants, but each has a proportionally smaller chance of fixation. The two effects cancel. If the mutation rate is approximately constant, neutral substitutions accumulate at an approximately constant rate—the *molecular clock*.

This result gives the population-genetic interpretation of the substitution rate μ in the Jukes–Cantor model of Section 9.7. The Markov chain in Week 9 describes changes already fixed along a lineage. Equation (13.54) explains how the underlying mutation process and finite-population drift determine the rate of those substitutions. In our Jukes–Cantor convention, μ is the total substitution rate away from the current base. If $u_{i\to j}$ denotes the neutral mutation probability from base i to a particular alternative j per reproductive event, then in general $\mu_i = \sum_{j\neq i} u_{i\to j}$. JC69 assumes that this total is the same for every starting base, $\mu_i = \mu$, and that each of the three target bases has rate $\mu/3$. This distinction between total and target-specific rates is the clock convention used here.

Pedagogical takeaway. The Moran model makes the drift–selection competition explicit. A beneficial mutation with $s > 0$ has an enhanced fixation probability, but it still frequently dies out when introduced at low copy number because the process must avoid early downward fluctuations before selection can reliably amplify it. This is the evolutionary analog of epidemic "fizzling" in the initiation phase.

13.3 The shared absorbing-state picture

The epidemic and evolutionary models use different biological language but the same mathematical logic. The number of infective individuals or mutants changes through discrete birth–death events. Zero is absorbing in both systems. The Moran process has a second absorbing boundary at fixation, while the epidemic branching approximation separates early extinction from takeoff into a macroscopic outbreak.

In each case, the deterministic threshold describes the direction of average growth away from the boundary. The exact stochastic calculation answers the more relevant finite-number question: starting from i individuals, what is the probability of reaching the successful outcome before the process is absorbed at zero?

13.4 Key results

- Demographic fluctuations matter most at epidemic initiation and termination, when the number of infective individuals is small.
- An exponentially distributed infectious period mixed with Poisson transmission gives a geometric offspring distribution.
- For that geometric branching process, the early-extinction probability from one introduction is $q = 1$ for $R_0 \leq 1$ and $q = 1/R_0$ for $R_0 > 1$.
- Heterogeneity in individual transmission adds a between-class variance and produces the long tails associated with superspreading.
- In the Moran process, neutral fixation probability equals initial frequency, $\rho_i = i/N$.

- With mutant relative fitness r, the exact fixation probability is $\rho_i = (1 - r^{-i})/(1 - r^{-N})$ for $r > 0$ and $r \neq 1$.
- For a general one-dimensional birth–death process, fixation is determined by the product of the local downward-to-upward bias ratios along the state space.
- Mutation supply and fixation combine to give substitution rate $R = Nu\rho_1$; for neutral mutations, $\rho_1 = 1/N$ and therefore $R = u$, the molecular-clock result.

Bibliography

Anscombe, F. J. (Feb. 1973). "Graphs in Statistical Analysis". In: *The American Statistician* 27.1, pp. 17–21. ISSN: 0003-1305. DOI: 10.1080/00031305.1973.10478966.

Becskei, A. and L. Serrano (June 2000). "Engineering Stability in Gene Networks by Autoregulation". In: *Nature* 405.6786, pp. 590–593. ISSN: 1476-4687. DOI: 10.1038/35014651.

Berg, H. and E. Purcell (Nov. 1977). "Physics of Chemoreception". In: *Biophysical Journal* 20.2, pp. 193–219. ISSN: 00063495. DOI: 10.1016/S0006-3495(77)85544-6.

Cronin, C. A., W. Gluba, and H. Scrable (June 2001). "The Lac Operator-Repressor System Is Functional in the Mouse". In: *Genes & Development* 15.12, pp. 1506–1517. ISSN: 0890-9369, 1549-5477. DOI: 10.1101/gad.892001.

Drescher, H., S. Weiskirchen, and R. Weiskirchen (Nov. 2021). "Flow Cytometry: A Blessing and a Curse". In: *Biomedicines* 9.11, p. 1613. ISSN: 2227-9059. DOI: 10.3390/biomedicines9111613.

Elowitz, M. B. and S. Leibler (Jan. 2000). "A Synthetic Oscillatory Network of Transcriptional Regulators". In: *Nature* 403.6767, pp. 335–338. ISSN: 1476-4687. DOI: 10.1038/35002125.

Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain (Aug. 2002). "Stochastic Gene Expression in a Single Cell". In: *Science* 297.5584, pp. 1183–1186. DOI: 10.1126/science.1070919.

Epstein, W., S. Naono, and F. Gros (Aug. 1966). "Synthesis of Enzymes of the Lactose Operon during Diauxic Growth of Escherichiacoli". In: *Biochemical and Biophysical Research Communications* 24.4, pp. 588–592. ISSN: 0006-291X. DOI: 10.1016/0006-291X(66)90362-7.

Erez, A., T. A. Byrd, R. M. Vogel, G. Altan-Bonnet, and A. Mugler (Feb. 2019). "Universality of Biochemical Feedback and Its Application to Immune Cells". In: *Physical Review E* 99.2, p. 022422. DOI: 10.1103/PhysRevE.99.022422.

Erez, A., J. G. Lopez, B. G. Weiner, Y. Meir, and N. S. Wingreen (Sept. 2020). "Nutrient Levels and Trade-Offs Control Diversity in a Serial Dilution Ecosystem". In: *eLife* 9. Ed. by D. Weigel, A. Sanchez, A. Succurro, and L. Dai, e57790. ISSN: 2050-084X. DOI: 10.7554/eLife.57790.

Erez, A., R. Mukherjee, and G. Altan-Bonnet (2019). "Quantifying the Dynamics of Hematopoiesis by In Vivo IdU Pulse-Chase, Mass Cytometry, and Mathematical Modeling". In: *Cytometry Part A* 95.10, pp. 1075–1084. ISSN: 1552-4930. DOI: 10.1002/cyto.a.23799.

Erez, A., R. Vogel, A. Mugler, A. Belmonte, and G. Altan-Bonnet (2018). "Modeling of Cytometry Data in Logarithmic Space: When Is a Bimodal Distribution Not Bimodal?" In: *Cytometry Part A* 93.6, pp. 611–619. ISSN: 1552-4930. DOI: 10.1002/cyto.a.23333.

Gardner, T. S., C. R. Cantor, and J. J. Collins (Jan. 2000). "Construction of a Genetic Toggle Switch in Escherichia Coli". In: *Nature* 403.6767, pp. 339–342. ISSN: 1476-4687. DOI: 10.1038/35002131.

Gillespie, D. T. (Dec. 1976). "A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions". In: *Journal of Computational Physics* 22.4, pp. 403–434. ISSN: 0021-9991. DOI: 10.1016/0021-9991(76)90041-3.

Girko, V. L. (Jan. 1985). "Circular Law". In: *Theory of Probability & Its Applications* 29.4, pp. 694–706. ISSN: 0040-585X. DOI: 10.1137/1129095.

Golding, I., J. Paulsson, S. M. Zawilski, and E. C. Cox (Dec. 2005). "Real-Time Kinetics of Gene Activity in Individual Bacteria". In: *Cell* 123.6, pp. 1025–1036. ISSN: 0092-8674. DOI: 10.1016/j.cell.2005.09.031.

Hill, M. O. (1973). "Diversity and Evenness: A Unifying Notation and Its Consequences". In: *Ecology* 54.2, pp. 427–432. ISSN: 1939-9170. DOI: 10.2307/1934352.

Jost, L. (2006). "Entropy and Diversity". In: *Oikos* 113.2, pp. 363–375. ISSN: 1600-0706. DOI: 10.1111/j.2006.0030-1299.14714.x.

Jukes, T. H. and C. R. Cantor (Jan. 1969). "CHAPTER 24 - Evolution of Protein Molecules". In: *Mammalian Protein Metabolism*. Ed. by H. N. Munro. Academic Press, pp. 21–132. ISBN: 978-1-4832-3211-9. DOI: 10.1016/B978-1-4832-3211-9.50009-7.

Kessler, D. A. and H. Levine (July 2013). "Large Population Solution of the Stochastic Luria–Delbrück Evolution Model". In: *Proceedings of the National Academy of Sciences* 110.29, pp. 11682–11687. DOI: 10.1073/pnas.1309667110.

Kimura, M. (June 1980). "A Simple Method for Estimating Evolutionary Rates of Base Substitutions through Comparative Studies of Nucleotide Sequences". In: *Journal of Molecular Evolution* 16.2, pp. 111–120. ISSN: 1432-1432. DOI: 10.1007/BF01731581.

Leemis, L. M. and J. T. McQueston (Feb. 2008). "Univariate Distribution Relationships". In: *The American Statistician* 62.1, pp. 45–53. ISSN: 0003-1305. DOI: 10.1198/000313008X270448.

Leinster, T. and C. A. Cobbold (2012). "Measuring Diversity: The Importance of Species Similarity". In: *Ecology* 93.3, pp. 477–489. ISSN: 1939-9170. DOI: 10.1890/10-2402.1.

Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler (Dec. 1991). "Long-Term Experimental Evolution in Escherichia Coli. I. Adaptation and Divergence During 2,000 Generations". In: *The American Naturalist* 138.6, pp. 1315–1341. DOI: 10.1086/285289.

Lopez, J. G., Y. Hein, and A. Erez (Apr. 2024). "Grow Now, Pay Later: When Should a Bacterium Go into Debt?" In: *Proceedings of the National Academy of Sciences* 121.16, e2314900121. DOI: 10.1073/pnas.2314900121.

Lopez, J. A. and A. Erez (2026). "Mathematical Modelling and Intuition in Microbiology: A Perspective". In: *Environmental Microbiology* 28.4. e70266 1612994, e70266. ISSN: 1462-2920. DOI: 10.1111/1462-2920.70266.

Lotka, A. J. (July 1920). "Analytical Note on Certain Rhythmic Relations in Organic Systems". In: *Proceedings of the National Academy of Sciences* 6.7, pp. 410–415. DOI: 10.1073/pnas.6.7.410.

Lozupone, C. and R. Knight (Dec. 2005). "UniFrac: A New Phylogenetic Method for Comparing Microbial Communities". In: *Applied and Environmental Microbiology* 71.12, pp. 8228–8235. ISSN: 0099-2240, 1098-5336. DOI: 10.1128/AEM.71.12.8228-8235.2005.

Luria, S. E. and M. Delbrück (Nov. 1943). "MUTATIONS OF BACTERIA FROM VIRUS SENSITIVITY TO VIRUS RESISTANCE". In: *Genetics* 28.6, pp. 491–511. ISSN: 1943-2631. DOI: 10.1093/genetics/28.6.491.

Mandelbrot, B. (Sept. 1974). "A Population Birth-and-Mutation Process, I: Explicit Distributions for the Number of Mutants in an Old Culture of Bacteria". In: *Journal of Applied Probability* 11.3, pp. 437–444. ISSN: 0021-9002, 1475-6072. DOI: 10.2307/3212688.

May, R. M. (Aug. 1972). "Will a Large Complex System Be Stable?" In: *Nature* 238.5364, pp. 413–414. ISSN: 1476-4687. DOI: 10.1038/238413a0.

Measles Outbreak —- Netherlands, April 1999–January 2000 (2000). URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm4914a2.htm.

Mills, F. C., M. L. Johnson, and G. K. Ackers (Nov. 1976). "Oxygenation-Linked Subunit Interactions in Human Hemoglobin: Experimental Studies on the Concentration Dependence of Oxygenation Curves". In: *Biochemistry* 15.24, pp. 5350–5362. ISSN: 0006-2960. DOI: 10.1021/bi00669a023.

Monod, J. (1942). *Recherches Sur La Croissance Des Cultures Bacteériennes*. Hermann and Cie.

Moran, P. a. P. (Jan. 1958). "Random Processes in Genetics". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 54.1, pp. 60–71. ISSN: 1469-8064, 0305-0041. DOI: 10.1017/S0305004100033193.

Nelson, P. (2022). *Physical Models of Living Systems: Probability, Simulation, Dynamics*. Chiliagon Science. ISBN: 978-1-7375402-4-3.

Novick, A. and M. Weiner (July 1957). "Enzyme Induction as an All-or-None Phenomenon*". In: *Proceedings of the National Academy of Sciences* 43.7, pp. 553–566. DOI: 10.1073/pnas.43.7.553.

Ozbudak, E. M., M. Thattai, H. N. Lim, B. I. Shraiman, and A. van Oudenaarden (Feb. 2004). "Multistability in the Lactose Utilization Network of Escherichia Coli". In: *Nature* 427.6976, pp. 737–740. ISSN: 1476-4687. DOI: 10.1038/nature02298.

Pace, H. C., P. Lu, and M. Lewis (Mar. 1990). "Lac Repressor: Crystallization of Intact Tetramer and Its Complexes with Inducer and Operator DNA." In: *Proceedings of the National Academy of Sciences* 87.5, pp. 1870–1873. DOI: 10.1073/pnas.87.5.1870.

Perelson, A. S., A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho (Mar. 1996). "HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time". In: *Science* 271.5255, pp. 1582–1586. DOI: 10.1126/science.271.5255.1582.

Potvin-Trottier, L., N. D. Lord, G. Vinnicombe, and J. Paulsson (Oct. 2016). "Synchronous Long-Term Oscillations in a Synthetic Gene Circuit". In: *Nature* 538.7626, pp. 514–517. ISSN: 1476-4687. DOI: 10.1038/nature19841.

Rao, C. R. (Feb. 1982). "Diversity and Dissimilarity Coefficients: A Unified Approach". In: *Theoretical Population Biology* 21.1, pp. 24–43. ISSN: 0040-5809. DOI: 10.1016/0040-5809(82)90004-1.

Rényi, A. (1961). "On Measures of Entropy and Information". In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California.

Rosenfeld, N., M. B. Elowitz, and U. Alon (Nov. 2002). "Negative Autoregulation Speeds the Response Times of Transcription Networks". In: *Journal of Molecular Biology* 323.5, pp. 785–793. ISSN: 0022-2836. DOI: 10.1016/S0022-2836(02)00994-4.

Rosenfeld, N., J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz (Mar. 2005). "Gene Regulation at the Single-Cell Level". In: *Science* 307.5717, pp. 1962–1965. DOI: 10.1126/science.1106914.

Rossi-Fanelli, A. and E. Antonini (Oct. 1958). "Studies on the Oxygen and Carbon Monoxide Equilibria of Human Myoglobin". In: *Archives of Biochemistry and Biophysics* 77.2, pp. 478–492. ISSN: 0003-9861. DOI: 10.1016/0003-9861(58)90094-8.

Severe Acute Respiratory Syndrome — Singapore, 2003 (2003). URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5218a1.htm.

Shannon, C. E. (1948). "A Mathematical Theory of Communication". In: *The Bell system technical journal* 27.3, pp. 379–423. ISSN: 0005-8580.

Stephens, G. J., B. Johnson-Kerner, W. Bialek, and W. S. Ryu (Apr. 2008). "Dimensionality and Dynamics in the Behavior of C. Elegans". In: *PLOS Computational Biology* 4.4, e1000028. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1000028.

Thaiss, C. A., M. Levy, T. Korem, L. Dohnalová, H. Shapiro, D. A. Jaitin, E. David, D. R. Winter, M. Gury-BenAri, E. Tatirovsky, T. Tuganbaev, S. Federici, N. Zmora, D. Zeevi, M. Dori-Bachash, M. Pevsner-Fischer, E. Kartvelishvily, A. Brandis, A. Harmelin, O. Shibolet, Z. Halpern, K. Honda, I. Amit, E. Segal, and E. Elinav (Dec. 2016). "Microbiota Diurnal Rhythmicity Programs Host Transcriptome Oscillations". In: *Cell* 167.6, 1495–1510.e12. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2016.11.003.

Vogel, R. M., A. Erez, and G. Altan-Bonnet (Sept. 2016). "Dichotomy of Cellular Inhibition by Small-Molecule Inhibitors Revealed by Single-Cell Analysis". In: *Nature Communications* 7.1, p. 12428. ISSN: 2041-1723. DOI: 10.1038/ncomms12428.

Volterra, V. (Oct. 1926). "Fluctuations in the Abundance of a Species Considered Mathematically1". In: *Nature* 118.2972, pp. 558–560. ISSN: 1476-4687. DOI: 10.1038/118558a0.

Weiss, R. A. (May 1993). "How Does HIV Cause AIDS?" In: *Science* 260.5112, pp. 1273–1279. DOI: 10.1126/science.8493571.

Wigner, E. P. (1955). "Characteristic Vectors of Bordered Matrices With Infinite Dimensions". In: *Annals of Mathematics* 62.3, pp. 548–564. ISSN: 0003-486X. DOI: 10.2307/1970079. JSTOR: 1970079.

Yildiz, A., J. N. Forkey, S. A. McKinney, T. Ha, Y. E. Goldman, and P. R. Selvin (June 2003). "Myosin V Walks Hand-Over-Hand: Single Fluorophore Imaging with 1.5-Nm Localization". In: *Science* 300.5628, pp. 2061–2065. DOI: 10.1126/science.1084398.