Combinatorics and Stochasticity for Chemical Reaction Networks

\[ M f = \sum_{x \in \mathbb{N}^S} \frac{z^x}{x!} \langle z'^x e^{z'}, f \rangle = \sum_{x \in \mathbb{N}^S} (z + 1)^x f_x \]
Overview
Chemical reaction networks
Stochasticity
Combinatorics
Formal Semantics
Flowers are theorems

\[
\begin{array}{c}
A \\
\hline
B \\
A \land B \\
\hline
(A \land B) \to C \\
C
\end{array}
\]

\[
\forall n \forall x \forall y \forall z. (n \geq 3 \to x^n + y^n \neq z^n)
\]

Theorems are flowers
Chemical Reaction Networks

How can we model biomolecular systems mathematically?
Chemical reaction networks are abstract models of systems with interacting species.

Continuous models are deterministic

\[ \frac{d[A]}{dt} = 2k_- [B] - 2k_+ [A]^2 \]

\[ \frac{d[B]}{dt} = k_+ [A]^2 - k_- [B] \]

\[ 2A \xrightarrow{k_+} B \]

\[ B \xrightarrow{k_-} 2A \]
Discrete models are stochastic

\[
\frac{dp_{2,2}}{dt} = 12k_+p_{4,1} + 3k_-p_{0,3} - k_+p_{2,2} - 2k_-p_{2,2}
\]
What are discrete, stochastic chemical systems capable of?
If each reaction is in equilibrium with its reverse reaction then the system is in detailed balance.

\[ G_C - G_A - G_B = \ln \frac{k_-}{k_+} \]

\[ \varepsilon_A = e^{-G_A} \]

\[ \varepsilon_B = e^{-G_B} \]

\[ \varepsilon_C = e^{-G_C} \]

\[ k_+ \varepsilon_A \varepsilon_B = k_- \varepsilon_C \]
Detailed balanced systems have stationary multivariate Poisson distributions.

\[ p_{m,n} \propto \frac{\varepsilon_A^m \varepsilon_B^n}{m!n!} \]

Detailed balanced systems have stationary multivariate Poisson distributions.

\[ p_{m,n} \propto \frac{\varepsilon_A^m \varepsilon_B^n}{m!n!} \]

restricted multivariate Poisson distribution

Can detailed balanced systems produce complex distributions?
Botzmann machines are stochastic models of neural networks

\[ E(0,0,1,0;1,0,1) = -w_{3,5} - w_{3,7} + \theta_3 + \theta_5 + \theta_7 \]

\[ p(x) = \frac{e^{-E(x)}}{Z} \]
Botzmann machines are stochastic models of neural networks

\[ E(0, 0, 1, 0; 1, 0, 1) = -w_{2,5} - w_{2,7} - w_{3,5} - w_{3,7} + \theta_3 + \theta_5 + \theta_7 + \theta_2 \]

\[ p(x) = \frac{e^{-E(x)}}{Z} \]
Chemical systems can behave like Boltzmann machines

\[ E_x = -w_{5,7} + \theta_5 + \theta_7 \]

\[ E_y = -w_{3,5} - w_{3,7} - w_{5,7} + \theta_3 + \theta_5 + \theta_7 \]

\[
\ln \frac{k_{y \rightarrow x}}{k_{x \rightarrow y}} = E_y - E_x = -w_{3,5} - w_{3,7} + \theta_3
\]
Chemical systems can produce complex distributions, sample them, and perform inference.
Detailed-balanced systems with hidden species can produce any distribution

Detailed-balanced systems with hidden species can produce any distribution

\[ X_{1,3} \xrightleftharpoons[k_+]{k_-} X_{2,3} + A \]

\[ \ln \frac{k_-}{k_+} = G_{2,3} + G_A - G_{1,3} \]

\[ G_{1,3} = -\ln(1!3!q_{2,3}) - G_A - 3G_B \]

\[ G_{2,3} = -\ln(2!3!q_{2,3}) - 2G_A - 3G_B \]

Chemical reaction networks can approximate arbitrary probability distributions!
Combinatorics

What role does combinatorics play in biology?
Enzymatic Catalysis

Enzyme → Substrate → Product

Scaffold-Aided Enzymaica Catalysis

Enzymatic catalysis on a polymerizing scaffold

The catalytic potential of a complex is its effective concentration of potential enzyme-substrate interactions.
What is the catalytic potential of a cell?
Generating functions encode complex combinatorial patterns as algebraic expressions

\[ x^4 = \begin{array}{cccc}
\bullet & \bullet & \bullet & \bullet \\
\end{array} \]

\[ \frac{d}{dx} x^4 = 4x^3 = \begin{array}{cccc}
\bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet \\
\end{array} + \begin{array}{cccc}
\bullet & \bullet & \bullet & \bullet \\
\end{array} + \begin{array}{cccc}
\bullet & \bullet & \bullet & \bullet \\
\end{array} + \begin{array}{cccc}
\bullet & \bullet & \bullet & \bullet \\
\end{array} + \begin{array}{cccc}
\bullet & \bullet & \bullet & \bullet \\
\end{array} \\

f = 1 + x + x^2 + x^3 + \cdots = 1 + \begin{array}{cccc}
\bullet & \\
\bullet & \\
\bullet & \\
\end{array} + \begin{array}{cccc}
\bullet & \bullet & \\
\bullet & \bullet & \\
\end{array} + \begin{array}{cccc}
\bullet & \bullet & \bullet & \\
\bullet & \bullet & \bullet & \\
\end{array} + \cdots \\

= 1 + \begin{array}{cccc}
\bullet & \\
\bullet & \\
\bullet & \\
\end{array} f = 1 + xf = \frac{1}{1-x} \\

\frac{df}{dx} = \frac{1}{(1-x)^2} = f^2 = (1+\begin{array}{cccc}
\bullet & \\
\bullet & \\
\bullet & \\
\end{array} + \begin{array}{cccc}
\bullet & \bullet & \\
\bullet & \bullet & \\
\end{array} + \cdots) \begin{array}{cccc}
\bullet & \\
\bullet & \\
\bullet & \\
\end{array} (1+\begin{array}{cccc}
\bullet & \\
\bullet & \\
\bullet & \\
\end{array} + \begin{array}{cccc}
\bullet & \bullet & \\
\bullet & \bullet & \\
\end{array} + \cdots) \\

We can use generating functions to encode the class of all possible complexes

\[ W = \alpha b^3 \sigma^3 a^2 b^3 s^4 \]

\[ = a + b + s + \alpha a s + \beta b s + \sigma s^2 + \alpha \beta a b s + \cdots \]

\[ = a + b + \sum_{n=1}^{\infty} \sigma^{n-1} s^n (1 + \alpha a)^n (1 + \beta b)^n \]

\[ = a + b + \frac{s(1 + \alpha a)(1 + \beta b)}{1 - \sigma s(1 + \alpha a)(1 + \beta b)} \]
We can take derivatives of the complex generating function to express conservation laws.

\[ x \frac{d}{dx} x^4 = 4x^4 \]

\[ = \square \square \square \square + \square \square \square \square + \square \square \square \square + \square \square \square \square \]

\[ s \frac{\partial}{\partial s} \begin{array}{c}
\text{complex generating function}
\end{array} = \begin{array}{c}
\text{complex generating function}
\end{array} + \begin{array}{c}
\text{complex generating function}
\end{array} + \begin{array}{c}
\text{complex generating function}
\end{array} \]

\[ t_A = a \frac{\partial W}{\partial a} \quad t_B = b \frac{\partial W}{\partial b} \quad t_S = s \frac{\partial W}{\partial s} \]
We can take derivatives of the complex generating function to extract the catalytic potential.
Catalytic potential is sensitive to polymerization parameters

The stochasticity of the system can also be analyzed using generating functions.

\[ Z = e^W \quad \text{ensemble generating function} \]

\[ = \sum_{(t_A, t_B, t_C) \in \mathbb{N}^3} Z_{t_A, t_B, t_C} a^{t_A} b^{t_B} c^{t_C} \]

\[ Z_{1,1,1} = \]

28
The discrete system has a polymer of **maximum length**
Formal Semantics

Can we interpret each variable formally rather than numerically?
The chemical master equation is a probability flux equation

\[ \frac{dp_x}{dt} = \sum_{r \in R} k_r \frac{(x + \rho_r - \pi_r)!}{(x - \pi_r)!} p_{x + \rho_r - \pi_r} - k_r \frac{x!}{(x - \rho_r)!} p_x \]
Probability generating functions can encode complex probability distributions in algebraic expressions.

\[
f = \sum_{(m,n) \in \mathbb{N}^2} p_{m,n} z_A^m z_B^n
\]

\[
= p_{0,0} + p_{1,0} + p_{0,1} + p_{2,0} + p_{1,1} + p_{0,2} + \cdots
\]

\[
f = e^{\lambda_A (z_A - 1) + \lambda_B (z_B - 1)} = \sum_{(m,n) \in \mathbb{N}} \left( e^{-\lambda_A} \frac{\lambda_A^m}{m!} \right) \left( e^{-\lambda_B} \frac{\lambda_B^n}{n!} \right) z_A^m z_B^n
\]

\[
\propto + + + + + + + + \cdots
\]
Reaction operators can be encoded as partial differential operators

$$z_C \partial_A \partial_B (z_A^2 z_B^2) = 4 z_A z_B z_C$$

$$z_C \partial_A \partial_B \begin{array}{c}
\text{1} \\
\text{2}
\end{array} + \begin{array}{c}
\text{1} \\
\text{2}
\end{array} + \begin{array}{c}
\text{1} \\
\text{2}
\end{array} + \begin{array}{c}
\text{1} \\
\text{1}
\end{array}$$

$$A_r \ f = k_r (z_C - z_A z_B) \partial_A \partial_B \ f$$

The **stochastic dynamics** of a system can be obtained by applying reaction operators recursively

\[
\sum_{r \in R} A_r = A \\
\frac{df}{dt} = Af
\]

\[
f = f_0 + \int_0^t Af \, dt
\]

\[
\int_0^t A \, dt = R \\
f = f_0 + Rf
\]

\[
= f_0 + Rf_0 + RRf_0 + RRf_0 + \cdots
\]

\[
= f_0 + t Af_0 + \frac{t^2}{2} A Af_0 + \frac{t^3}{6} AAF_0 + \cdots
\]

\[
= e^{tA}f_0 = E f_0
\]
The stochastic dynamics can be expressed as a path integral in terms of waiting operators.

\[ R = \rightarrow - \downarrow \]

\[ f = f_0 + RF \]

\[ f + \downarrow f = f_0 + Ff \]

\[ W = 1 - \downarrow + \downarrow\downarrow - \downarrow\downarrow\downarrow + \cdots = \frac{1}{1 + \downarrow} \]

\[ f = Wf_0 + WFf \]

\[ = Wf_0 + W\rightarrow Wf_0 + W\rightarrow W\rightarrow Wf_0 + \cdots \]

\[ Wf_0 = e^{-tA^-}f_0 \]
A system has a stationary Poisson distribution if, and only if, it is complex-balanced

\[ k_1 \varepsilon_A \varepsilon_B = k_2 \varepsilon_C \varepsilon_D + k_3 \varepsilon_A \varepsilon_C \]

\[ \mathcal{A} e^{\varepsilon_A z_A + \varepsilon_B z_B + \varepsilon_C z_C + \varepsilon_D z_D} = 0 \]

\[ p_{i,j,k,l} \propto \frac{\varepsilon_A^i \varepsilon_B^j \varepsilon_C^k \varepsilon_D^l}{i! j! k! l!} \]

\[ 0 = \mathcal{A} e^{\varepsilon z} = e^{\varepsilon z} \sum_{c \in C} z^c \left( \sum_{r : \pi_r = c} k_r \varepsilon^\rho_r - \sum_{r : \rho_r = c} k_r \varepsilon^c \right) \]

\[ \sum_{r : \pi_r = c} k_r \varepsilon^\rho_r = \sum_{r : \rho_r = c} k_r \varepsilon^c \]
Factorial moments are expectation values of permutations

\[ \mu_3 = (3 \cdot 2 \cdot 1)p_3 + (4 \cdot 3 \cdot 2)p_4 + (5 \cdot 4 \cdot 3)p_5 + \cdots = \sum_{n \geq 3} \frac{n!}{(n - 3)!} p_n \]

\[ p_n = e^{-\lambda} \frac{\lambda^n}{n!} \quad \mu_n = \lambda^n \]

\[ \frac{d\mu_x}{dt} = \sum_{r \in R} k_r \left( \sum_{y \leq \pi_r} \binom{\pi_r}{y} \frac{x!}{(x - y)!} \mu_{x-y+\pi_r} - \sum_{y \leq \rho_r} \binom{\rho_r}{y} \frac{x!}{(x - y)!} \mu_{x-y+\rho_r} \right) \]

The generating functions of factorial moments and probability have a simple relationship

\[ m(z) = f(z+1) \quad f = e^{\lambda(z-1)} \quad m = e^{\lambda z} \]

\[(z + 1)^3 = 1 + 3z + 3z^2 + z^3\]

The generating functions of factorial moments and probability have a simple relationship
The Barrier of Objects

“In Nature, interaction involves objects directly and never by a numerical value describing them. Stepping outside of conventional dynamical systems requires taking this observation seriously.”

W. Fontana & L. Buss, 1992

Acknowledgements

Erik Winfree

Walter Fontana
Acknowledgements

William Poole
Manoj Gopalkrishnan
Thomas Ouldridge
David Anderson
Daniele Cappelletti
Robert Johnson
Fangzhou Xiao
Damien Woods
Chris Thachuck
Dave Doty
Stephan Badelt
Frits Dannenberg
Héctor Medina
Acknowledgements

Manuel Razo
Porfirio Quintero
Enrique Amaya
Manolo Flores
David Larios
Alejandro Granados
Jorge Castellanos

Santa Fe Institute
Acknowledgements
Acknowledgements
No Soy Poeta

Dicen que no soy poeta
tiene razón quien lo diga
Tan solo escribo versos
para bendecir a Dios
o elogiar a una hormiga

Pascual Ortiz Saucedo