A new parametrization for binary hidden Markov models

Andrew Critch, UC Berkeley

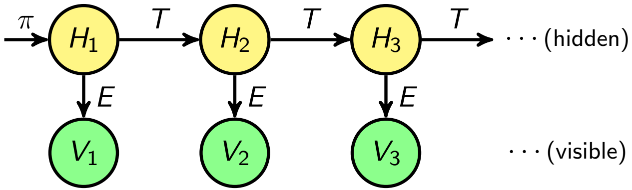
at Pennsylvania State University

June 11, 2012

Please see "Binary hidden Markov models and varieties" [–, 2012], arXiv:1206.0500, for more background on this talk.

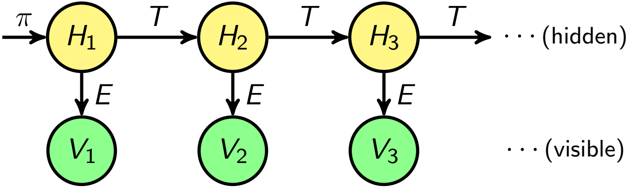
Outline

Introducing Binary Hidden Markov Models



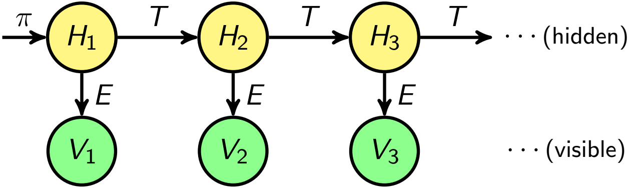
Hidden Markov models are machine learning models with extremely diverse applications, including natural language processing, gesture recognition, genomics, and Kalman filtering of physical measurements. They are highly non-linear models, and just as linear models are amenable to linear algebra techniques, non-linear models are amenable to commutative algebra and algebraic geometry.

Introducing Binary Hidden Markov Models



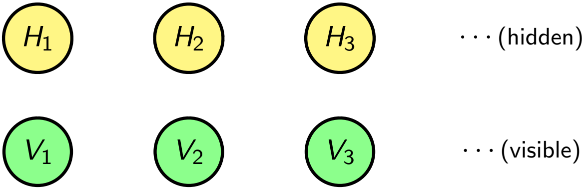
Hidden Markov models are machine learning models with extremely diverse applications, including natural language processing, gesture recognition, genomics, and Kalman filtering of physical measurements. They are highly non-linear models, and just as linear models are amenable to linear algebra techniques, non-linear models are amenable to commutative algebra and algebraic geometry.

Introducing Binary Hidden Markov Models



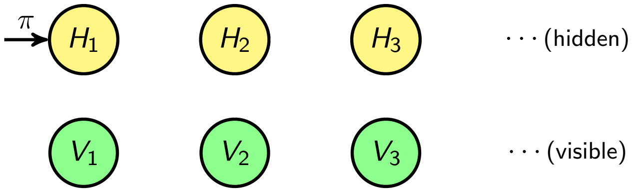
A **Binary Hidden Markov (BHM)** **process** of length n consists of 4 things:

Introducing Binary Hidden Markov Models



(1) A jointly random sequence $(H_1, V_1, H_2, V_2, \ldots, H_n, V_n)$ of binary variables with range $\{0, 1\}$, called hidden nodes and visible nodes,

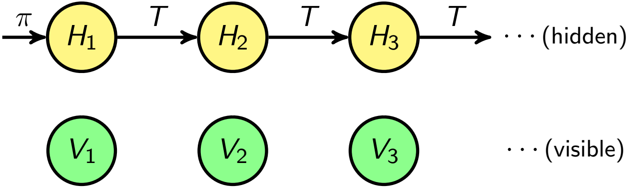
Introducing Binary Hidden Markov Models



(2) A row vector $\pi = \begin{bmatrix}\pi_0, & \pi_1\end{bmatrix}$ which specifies the probability distribution on the first hidden node H_1 by the formula

$$\Pr(H_1 = i) = \pi_i$$

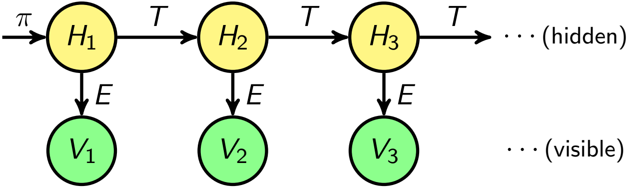
Introducing Binary Hidden Markov Models



(3) A *transition matrix* $T = \begin{bmatrix} T_{00} & T_{01} \\ T_{10} & T_{11} \end{bmatrix}$ specifying conditional "transition" probabilities by the formula

$$\Pr(H_t = j \mid H_{t-1} = i) = T_{ij},$$

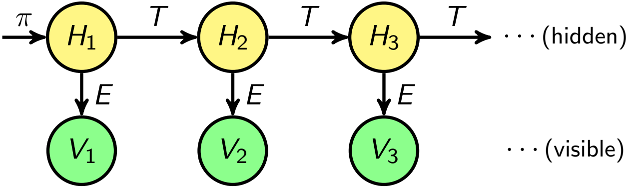
Introducing Binary Hidden Markov Models



(4) An *emission matrix* $E = \begin{bmatrix} E_{00} & E_{01} \\ E_{10} & E_{11} \end{bmatrix}$ specifying conditional "emission" probabilities by the formula

$$\Pr(V_t = j \mid H_t = i) = E_{ij}.$$

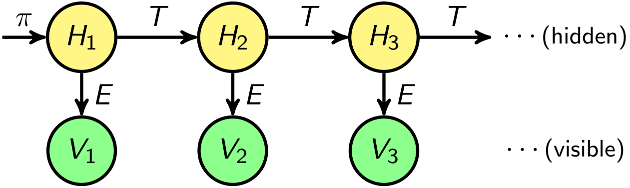
Introducing Binary Hidden Markov Models



Given n, a *parameter vector* $\theta = (\pi, T, E)$ generates a distribution p over the 2^n possible visible sequences $v = (v_1, \ldots v_n)$. We write $p_v = P(V = v|\theta)$, which defines an algebraic map from parameter vectors θ to distributions p:

$$\boldsymbol{\phi_n : \mathbb{C}^5_\theta \to \mathbb{C}^{2^n-1}_p \subseteq \mathbb{P}^{2^n-1}_p}$$

Introducing Binary Hidden Markov Models



We write $\Theta \subseteq \mathbb{C}^5_\theta$ for the classically compact set of those θ whose rows are probability distributions (nonnegative reals summing to 1). The **BHM model** on n nodes, $\mathbf{M_{BHM(n)}}$, is the image $\phi_n(\Theta)$, i.e. the set of visible probability distributions p that can arise from BHM processes as above.

Introducing Binary Hidden Markov Models

We will see that reparametrizing this model can help with several problems. Let me start by introducing the problem that first got me thinking about this.

Implicitization

Being given the model $\mathbf{M_{BHM(n)}}$ parametrically as $\phi_n(\Theta)$, we would like describe it implicitly:

Problem 1: ideal generation

Exhibit generators for the prime ideal $\mathbf{I_{BHM(n)}}$ of polynomials that vanish on the model $\mathbf{M_{BHM(n)}}$.

Setting these to 0 will yield equations that cut out the model "as well as possible", in that they will define the smallest algerbaic variety containing it, called its Zariski closure.

Implicitization

Being given the model $\mathbf{M}_{\mathbf{BHM(n)}}$ parametrically as $\phi_n(\Theta)$, we would like describe it implicitly:

Problem 1: ideal generation

Exhibit generators for the prime ideal $\mathbf{I}_{\mathbf{BHM(n)}}$ of polynomials that vanish on the model $\mathbf{M}_{\mathbf{BHM(n)}}$.

Setting these to 0 will yield equations that cut out the model "as well as possible", in that they will define the smallest algerbaic variety containing it, called its Zariski closure.

Implicitization

Being given the model $\mathbf{M_{BHM(n)}}$ parametrically as $\phi_n(\Theta)$, we would like describe it implicitly:

Problem 1: ideal generation

Exhibit generators for the prime ideal $\mathbf{I_{BHM(n)}}$ of polynomials that vanish on the model $\mathbf{M_{BHM(n)}}$.

Setting these to 0 will yield equations that cut out the model "as well as possible", in that they will define the smallest algerbaic variety containing it, called its Zariski closure.

Implicitization problems

Previous work on implicitizing general HMMs apply to BHMMs:

- 2005: Bray and Morton found polynomials generating a homogenization of $\mathbf{I}_{\mathbf{BHM(n)}}$ in low degree for small n, and conjecture that for large n, the ideal is generated by quadrics.
- 2008: Schönhuth identifies $\mathbf{M}_{\mathbf{BHM(n)}}$ with a rank-two finitary string process model of length n.
- 2011: Schönhuth exhibits generators for $\mathbf{I}_{\mathbf{BHM(3)}}$ comprising 4 cubic equations using finitary process theory. This method is currently too computationally intensive for $\mathbf{V}_{\mathbf{BHM(4)}}$.

Implicitization problems

Previous work on implicitizing general HMMs apply to BHMMs:

- 2005: Bray and Morton found polynomials generating a homogenization of $\mathbf{I}_{\mathbf{BHM(n)}}$ in low degree for small n, and conjecture that for large n, the ideal is generated by quadrics.
- 2008: Schönhuth identifies $\mathbf{M}_{\mathbf{BHM(n)}}$ with a rank-two finitary string process model of length n.
- 2011: Schönhuth exhibits generators for $\mathbf{I}_{\mathbf{BHM(3)}}$ comprising 4 cubic equations using finitary process theory. This method is currently too computationally intensive for $\mathbf{V}_{\mathbf{BHM(4)}}$.

Implicitization problems

Previous work on implicitizing general HMMs apply to BHMMs:

- 2005: Bray and Morton found polynomials generating a homogenization of $\mathbf{I_{BHM(n)}}$ in low degree for small n, and conjecture that for large n, the ideal is generated by quadrics.
- 2008: Schönhuth identifies $\mathbf{M_{BHM(n)}}$ with a rank-two finitary string process model of length n.
- 2011: Schönhuth exhibits generators for $\mathbf{I_{BHM(3)}}$ comprising 4 cubic equations using finitary process theory. This method is currently too computationally intensive for $\mathbf{V_{BHM(4)}}$.

Implicitization problems

Previous work on implicitizing general HMMs apply to BHMMs:

- 2005: Bray and Morton found polynomials generating a homogenization of $\mathbf{I}_{\mathbf{BHM(n)}}$ in low degree for small n, and conjecture that for large n, the ideal is generated by quadrics.
- 2008: Schönhuth identifies $\mathbf{M}_{\mathbf{BHM(n)}}$ with a rank-two finitary string process model of length n.
- 2011: Schönhuth exhibits generators for $\mathbf{I}_{\mathbf{BHM(3)}}$ comprising 4 cubic equations using finitary process theory. This method is currently too computationally intensive for $\mathbf{V}_{\mathbf{BHM(4)}}$.

Implicitization problems

Method: reparametrization

It turns out Macaulay2 can handle computing generators for $\mathbf{I}_{\mathbf{BHM}(4)}$ if we use a more symbolically efficient parametrization, and the reparametrization itself has other interesting consequences.

Moments and cumulants

These **new coordinates** on $\mathbb{C}_p^{2^n}$ allow faster symbolic computation for with BHMMs in Macaulay2. For indices $I \subseteq [n] = \{1, \ldots, n\}$, we define moments m_I and cumulants k_I by:

$$\begin{aligned} m_I &:= \sum \{p_v \mid v_i = 1 \text{ for all } i \in I\} \\ &= P(V_i = 1 \text{ for all } i \in I), \end{aligned}$$

$$k_I := \text{coefficient of } x^I \text{ in } \left(\log \sum_{I \subseteq \{1 \ldots, n\}} m_I x^I \right)$$

These formulae [Sturmfels and Zwiernik, 2011] define polynomial isomorphisms

$$\mathbb{C}[p_v \mid v \in \{0,1\}^n] \longleftrightarrow \mathbb{C}[m_I \mid I \in [n]] \longleftrightarrow \mathbb{C}[k_I \mid I \in [n]]$$

Moments and cumulants

These **new coordinates** on $\mathbb{C}_p^{2^n}$ allow faster symbolic computation for with BHMMs in Macaulay2. For indices $I \subseteq [n] = \{1, \ldots, n\}$, we define moments m_I and cumulants k_I by:

$$m_I := \sum \{p_v \mid v_i = 1 \text{ for all } i \in I\}$$
$$= P(V_i = 1 \text{ for all } i \in I),$$

$$k_I := \text{coefficient of } x^I \text{ in } \left(\log \sum_{I \subseteq \{1 \ldots, n\}} m_I x^I \right)$$

These formulae [Sturmfels and Zwiernik, 2011] define polynomial isomorphisms

$$\mathbb{C}[p_v \mid v \in \{0,1\}^n] \longleftrightarrow \mathbb{C}[m_I \mid I \in [n]] \longleftrightarrow \mathbb{C}[k_I \mid I \in [n]]$$

Moments and cumulants

These **new coordinates** on $\mathbb{C}_p^{2^n}$ allow faster symbolic computation for with BHMMs in Macaulay2. For indices $I \subseteq [n] = \{1, \ldots, n\}$, we define moments m_I and cumulants k_I by:

$$\begin{aligned} m_I &:= \sum \{p_v \mid v_i = 1 \text{ for all } i \in I\} \\ &= P(V_i = 1 \text{ for all } i \in I), \end{aligned}$$

$$k_I := \text{coefficient of } x^I \text{ in } \left(\log \sum_{I \subseteq \{1 \ldots, n\}} m_I x^I \right)$$

These formulae [Sturmfels and Zwiernik, 2011] define polynomial isomorphisms

$$\mathbb{C}[p_v \mid v \in \{0,1\}^n] \longleftrightarrow \mathbb{C}[m_I \mid I \in [n]] \longleftrightarrow \mathbb{C}[k_I \mid I \in [n]]$$

Moments and cumulants

Examples of moments, with $n = 3$ nodes:

$$\begin{aligned} m_\emptyset &= 1 \\ m_1 &= p_{100} + p_{101} + p_{110} + p_{111} \\ m_{12} &= p_{110} + p_{111} \\ m_{123} &= p_{111} \end{aligned}$$

Examples of cumulants (with any number of nodes):

$$\begin{aligned} k_\emptyset &= 0 \\ k_1 &= m_1 \\ k_{12} &= m_{12} - m_1 m_2 \\ k_{123} &= m_{123} - m_1 m_{23} - m_2 m_{13} - m_3 m_{12} + 2 m_1 m_2 m_3 \end{aligned}$$

New matrix parameters

We introduce **new parameters** $a_0, b, c_0, u, v_0 \in \mathbb{C}$ and write

$$\pi = \frac{1}{2}\begin{bmatrix} 1 - a_0, & 1 + a_0 \end{bmatrix}, \qquad T = \frac{1}{2}\begin{bmatrix} 1 + b - c_0, & 1 - b + c_0 \\ 1 - b - c_0, & 1 + b + c_0 \end{bmatrix},$$

$$E = \begin{bmatrix} 1 - u + v_0, & u - v_0 \\ 1 - u - v_0, & u + v_0 \end{bmatrix}$$

Why this form? Given a BHM process, if we swap the outputs 0 and 1 of the hidden variables H_i, we get a new process that is observationally indistinguishable from it. With the new parameters, this $\mathbb{Z}/2$ action just corresponds to changing the sign of (a_0, c_0, v_0). (The right column of E is made intentionally homogeneous for other reasons.)

New matrix parameters

We introduce **new parameters** $a_0, b, c_0, u, v_0 \in \mathbb{C}$ and write

$$\pi = \frac{1}{2}\begin{bmatrix}1 - a_0, & 1 + a_0\end{bmatrix}, \qquad T = \frac{1}{2}\begin{bmatrix}1 + b - c_0, & 1 - b + c_0 \\ 1 - b - c_0, & 1 + b + c_0\end{bmatrix},$$

$$E = \begin{bmatrix}1 - u + v_0, & u - v_0 \\ 1 - u - v_0, & u + v_0\end{bmatrix}$$

Why this form? Given a BHM process, if we swap the outputs 0 and 1 of the hidden variables H_i, we get a new process that is observationally indistinguishable from it. With the new parameters, this $\mathbb{Z}/2$ action just corresponds to changing the sign of (a_0, c_0, v_0). (The right column of E is made intentionally homogeneous for other reasons.)

Birational parameters

Let

$$\eta_0 = (a_0, b, c_0, u, v_0) \in \mathbb{C}^5_{\eta_0}$$
$$a = a_0 v_0, \quad c = c_0 v_0, \quad v = v_0^2$$
$$\eta = (a, b, c, u, v) \quad \in \mathbb{C}^5_{\eta}$$

Factorization theorem

The map $\psi_n : \mathbb{C}^5_{\eta_0} \to \mathbf{V}_{\mathbf{BHM(n)}}$ factors through the generically $2:1$ map $\mathbb{C}^5_{\eta_0} \to \mathbb{C}^5_{\eta}$ yielding a new parametrization

$$\overline{\psi}_n : \mathbb{C}^5_{\eta} \to \mathbf{V}_{\mathbf{BHM(n)}}$$

Note for geometers: This factorization is finer than the invariant theory quotient by hidden label swapping, which also requires the parameters $a_0^2, a_0 c_0$, and c_0^2 and so does not even embed in $\mathbb{C}^5$.

Birational parameters

Let

$$\eta_0 = (a_0, b, c_0, u, v_0) \in \mathbb{C}^5_{\eta_0}$$

$$a = a_0 v_0, \quad c = c_0 v_0, \quad v = v_0^2$$

$$\eta = (a, b, c, u, v) \quad \in \mathbb{C}^5_{\eta}$$

Factorization theorem

The map $\psi_n : \mathbb{C}^5_{\eta_0} \to \mathbf{V_{BHM(n)}}$ factors through the generically 2 : 1 map $\mathbb{C}^5_{\eta_0} \to \mathbb{C}^5_{\eta}$ yielding a new parametrization

$$\overline{\psi}_n : \mathbb{C}^5_{\eta} \to \mathbf{V_{BHM(n)}}$$

Note for geometers: This factorization is finer than the invariant theory quotient by hidden label swapping, which also requires the parameters $a_0^2, a_0 c_0$, and c_0^2 and so does not even embed in $\mathbb{C}^5$.

Birational parameters

Let
$$\eta_0 = (a_0, b, c_0, u, v_0) \in \mathbb{C}^5_{\eta_0}$$
$$a = a_0 v_0, \quad c = c_0 v_0, \quad v = v_0^2$$
$$\eta = (a, b, c, u, v) \quad \in \mathbb{C}^5_{\eta}$$

Factorization theorem

The map $\psi_n : \mathbb{C}^5_{\eta_0} \to \mathbf{V}_{\mathbf{BHM(n)}}$ factors through the generically $2:1$ map $\mathbb{C}^5_{\eta_0} \to \mathbb{C}^5_{\eta}$ yielding a new parametrization

$$\overline{\psi}_n : \mathbb{C}^5_{\eta} \to \mathbf{V}_{\mathbf{BHM(n)}}$$

Note for geometers: This factorization is finer than the invariant theory quotient by hidden label swapping, which also requires the parameters $a_0^2, a_0 c_0$, and c_0^2 and so does not even embed in $\mathbb{C}^5$.

A factorization theorem

On the moments of the first three nodes, the new parametrization is $\mathbb{C}^5_\eta \to \mathbb{C}^{2^n}_m$ is given by:

$$
\begin{aligned}
m_\emptyset &\mapsto 1 \\
m_1 &\mapsto a + u \\
m_2 &\mapsto ab + c + u \\
m_3 &\mapsto ab^2 + bc + c + u \\
m_{12} &\mapsto abu + ac + au + cu + u^2 + bv \\
m_{13} &\mapsto ab^2u + abc + bcu + b^2v + ac + au + cu + u^2 \\
m_{23} &\mapsto ab^2u + abc + abu + bcu + c^2 + 2cu + u^2 + bv \\
m_{123} &\mapsto ab^2u^2 + 2abcu + abu^2 + bcu^2 + b^2uv + ac^2 + 2acu \\
&\quad + c^2u + au^2 + 2cu^2 + u^3 + abv + bcv + 2buv
\end{aligned}
$$

A factorization theorem

Proof. The theorem relies on the observation that every BHMM lives inside a particular 9-dimensional variety called a *trace variety*, which is the IT quotient of the space of triples of 2×2 matrices under a simultaneous conjugation action by SL_2.

As a quotient, the trace variety is not defined inside any particular ambient space. However, its coordinate ring, a *trace algebra*, was found by Sibirskii [1968] to be generated by 10 elements, which means we can embed the trace variety, and hence *all BHMMs simultaneously*, in $\mathbb{C}^{10}$. The theorem is proven by direct computation in the coordinates of this embedding.

A factorization theorem

Proof. The theorem relies on the observation that every BHMM lives inside a particular 9-dimensional variety called a *trace variety*, which is the IT quotient of the space of triples of 2×2 matrices under a simultaneous conjugation action by SL_2.

As a quotient, the trace variety is not defined inside any particular ambient space. However, its coordinate ring, a *trace algebra*, was found by Sibirskii [1968] to be generated by 10 elements, which means we can embed the trace variety, and hence *all BHMMs simultaneously*, in $\mathbb{C}^{10}$. The theorem is proven by direct computation in the coordinates of this embedding.

Birationality of the new parametrization

Birational Parameter Theorem

The map $\mathbb{C}^5_\eta \to \mathbf{V}_{\mathbf{BHM(n)}}$ is generically injective, and the graph of its birational inverse is given by:

$$b = \frac{m_3 - m_2}{m_2 - m_1} \qquad u = \frac{m_1 m_3 - m_2^2 + m_{23} - m_{12}}{2(m_3 - m_2)}$$

$$a = m_1 - u \qquad c = a - ba + m_2 - m_1$$

$$v = a^2 - \frac{m_1 m_2 - m_{12}}{b}$$

Birationality of the new parametrization

Birational Parameter Theorem

The map $\mathbb{C}^5_\eta \to \mathbf{V}_{\mathbf{BHM(n)}}$ is generically injective, and the graph of its birational inverse is given by:

$$b = \frac{m_3 - m_2}{m_2 - m_1} \qquad u = \frac{m_1 m_3 - m_2^2 + m_{23} - m_{12}}{2(m_3 - m_2)}$$

$$a = m_1 - u \qquad c = a - ba + m_2 - m_1$$

$$v = a^2 - \frac{m_1 m_2 - m_{12}}{b}$$

Birationality of the new parametrization

Proof. These equations can be obtained in Macaulay2 by computing two Gröbner bases of the elimination ideal of the graph of the new parametrization, in Lex monomial order: one with the ordering $[v, c, a, b, u]$, and one with the ordering $[v, c, u, b, a]$.

Each of a, b, c, u and v occurs in the leading term of a some generator in one of these two bases with a simple expression in moments as its leading coefficient. We solve each such generator (set to 0) for the desired parameter.

Birationality of the new parametrization

Proof. These equations can be obtained in Macaulay2 by computing two Gröbner bases of the elimination ideal of the graph of the new parametrization, in Lex monomial order: one with the ordering $[v, c, a, b, u]$, and one with the ordering $[v, c, u, b, a]$.

Each of a, b, c, u and v occurs in the leading term of a some generator in one of these two bases with a simple expression in moments as its leading coefficient. We solve each such generator (set to 0) for the desired parameter.

Generators for $\mathbf{I}_{\mathbf{BHM}(4)}$

Since our new parametrization $\overline{\psi}_4$ is birational, the degree of the equations occuring in computing its kernel is lower than the original parametrization, and Macaulay2 is able to find a generating set for $\mathbf{I}_{\mathbf{BHM}(4)}$ in cumulant coordinates in **under 1 second**. Converting back to homogeneous moment coordinates takes **1.5 hours**.

Theorem (solution to problems 1)

In moment or probability coordinates, the homogeneous ideal $\mathbf{I}_{\mathbf{BHM}(4)}$ is minimally generated by 21 homogeneous quadrics and 29 homogeneous cubics.

Generators for $\mathbf{I}_{\mathbf{BHM}(4)}$

Since our new parametrization $\overline{\psi}_4$ is birational, the degree of the equations occuring in computing its kernel is lower than the original parametrization, and Macaulay2 is able to find a generating set for $\mathbf{I}_{\mathbf{BHM}(4)}$ in cumulant coordinates in **under 1 second**. Converting back to homogeneous moment coordinates takes **1.5 hours**.

Theorem (solution to problems 1)

In moment or probability coordinates, the homogeneous ideal $\mathbf{I}_{\mathbf{BHM}(4)}$ is minimally generated by 21 homogeneous quadrics and 29 homogeneous cubics.

Generators for $\mathbf{I}_{\mathbf{BHM}(4)}$

Since our new parametrization $\overline{\psi}_4$ is birational, the degree of the equations occuring in computing its kernel is lower than the original parametrization, and Macaulay2 is able to find a generating set for $\mathbf{I}_{\mathbf{BHM}(4)}$ in cumulant coordinates in **under 1 second**. Converting back to homogeneous moment coordinates takes **1.5 hours**.

Theorem (solution to problems 1)

In moment or probability coordinates, the homogeneous ideal $\mathbf{I}_{\mathbf{BHM}(4)}$ is minimally generated by 21 homogeneous quadrics and 29 homogeneous cubics.

What the generators look like

In probability coordinates, the generators had the following sizes:

21 quadrics: 8, 8, 12, 14, 16, 21, 24, 24, 26, 26, 28, 32, 32, 41, 42, 43, 43, 44, 45, 72, 72 terms.

29 cubics: 32, 43, 44, 44, 44, 52, 52, 56, 56, 61, 69, 71, 74, 76, 78, 81, 99, 104, 109, 119, 128, 132, 148, 157, 176, 207, 224, 236, 429 terms.

In moment coordinates, they are much shorter:

21 quadrics: 4, 4, 4, 4, 6, 6, 6, 6, 6, 6, 6, 6, 8, 8, 8, 8, 8, 10, 10, 10, 17 terms.

29 cubics: 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 8, 8, 8, 8, 10, 10, 10, 10, 10, 12, 12, 13, 14, 16, 18, 21, 27, 35 terms.

What the generators look like

In probability coordinates, the generators had the following sizes:

21 quadrics: 8, 8, 12, 14, 16, 21, 24, 24, 26, 26, 28, 32, 32, 41, 42, 43, 43, 44, 45, 72, 72 terms.

29 cubics: 32, 43, 44, 44, 44, 52, 52, 56, 56, 61, 69, 71, 74, 76, 78, 81, 99, 104, 109, 119, 128, 132, 148, 157, 176, 207, 224, 236, 429 terms.

In moment coordinates, they are much shorter:

21 quadrics: 4, 4, 4, 4, 6, 6, 6, 6, 6, 6, 6, 6, 8, 8, 8, 8, 8, 10, 10, 10, 17 terms.

29 cubics: 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 8, 8, 8, 8, 10, 10, 10, 10, 10, 12, 12, 13, 14, 16, 18, 21, 27, 35 terms.

What the generators look like

The shortest quadric and cubic generators are:

$$g_{2,1} = m_{23}m_{13} - m_2 m_{134} - m_{13}m_{12} + m_1 m_{124}$$
$$g_{3,1} = m_{12}^3 - 2m_1 m_{12} m_{123} + m_\emptyset m_{123}^2 + m_1^2 m_{1234} - m_\emptyset m_{12} m_{1234}$$

Note that these are *also* homogeneous with respect to the **number of subscripts** in each term. In fact...

What the generators look like

The shortest quadric and cubic generators are:

$$g_{2,1} = m_{23}m_{13} - m_2 m_{134} - m_{13}m_{12} + m_1 m_{124}$$
$$g_{3,1} = m_{12}^3 - 2m_1 m_{12} m_{123} + m_\emptyset m_{123}^2 + m_1^2 m_{1234} - m_\emptyset m_{12} m_{1234}$$

Note that these are *also* homogeneous with respect to the **number of subscripts** in each term. In fact...

Bi-homogeneity of $\mathbf{I}_{\mathbf{BHM(n)}}$

Bihomogeneity Theorem

In **moment coordinates** $\mathbf{I}_{\mathbf{BHM(n)}}$ is always bihomogeneous, with the second grading given by $deg(m_I) = size(I)$.

Geometrically, this means that $\mathbf{V}_{\mathbf{BHM(n)}}$ is invariant under a generically free action of $(\mathbb{C}^*)^2$.

Warning: This is **not** true for the grading $deg(p_I) = size(I)$!

Bi-homogeneity of $\mathbf{I}_{\mathbf{BHM(n)}}$

Bihomogeneity Theorem

In **moment coordinates** $\mathbf{I}_{\mathbf{BHM(n)}}$ is always bihomogeneous, with the second grading given by $deg(m_I) = size(I)$.

Geometrically, this means that $\mathbf{V}_{\mathbf{BHM(n)}}$ is invariant under a generically free action of $(\mathbb{C}^*)^2$.

Warning: This is **not** true for the grading $deg(p_I) = size(I)$!

Bi-homogeneity of $\mathbf{I}_{\mathbf{BHM(n)}}$

Proof. The the parametrization $\mathbb{C}^5_{\eta_0} \to \mathbb{C}^{2^n}_m$ can be shown to be homogeneous with respect to a grading where

- $deg(a_0) = deg(b) = deg(c_0) = 0$,
- $deg(\mathbf{u}) = deg(\mathbf{v_0}) = 1$
- $deg(m_I) = size(I)$

Recall that E was written somewhat differently from π and T; this was precisely to achieve homogeneity of the parametrization:

$$\pi = \frac{1}{2}\begin{bmatrix}1 - a_0, & 1 + a_0\end{bmatrix}, \qquad T = \frac{1}{2}\begin{bmatrix}1 + b - c_0, & 1 - b + c_0 \\ 1 - b - c_0, & 1 + b + c_0\end{bmatrix},$$

$$E = \begin{bmatrix}1 - u + v_0, & \mathbf{u} - \mathbf{v_0} \\ 1 - u - v_0, & \mathbf{u} + \mathbf{v_0}\end{bmatrix}$$

Bi-homogeneity of $\mathbf{I_{BHM(n)}}$

Proof. The the parametrization $\mathbb{C}^5_{\eta_0} \to \mathbb{C}^{2^n}_m$ can be shown to be homogeneous with respect to a grading where

- $deg(a_0) = deg(b) = deg(c_0) = 0$,
- $deg(\mathbf{u}) = deg(\mathbf{v_0}) = 1$
- $deg(m_I) = size(I)$

Recall that E was written somewhat differently from π and T; this was precisely to achieve homogeneity of the parametrization:

$$\pi = \frac{1}{2}\left[1 - a_0, \quad 1 + a_0\right], \qquad T = \frac{1}{2}\begin{bmatrix} 1 + b - c_0, & 1 - b + c_0 \\ 1 - b - c_0, & 1 + b + c_0 \end{bmatrix},$$

$$E = \begin{bmatrix} 1 - u + v_0, & \mathbf{u} - \mathbf{v_0} \\ 1 - u - v_0, & \mathbf{u} + \mathbf{v_0} \end{bmatrix}$$

Bi-homogeneity of $\mathbf{I}_{\mathbf{BHM(n)}}$

Application: finding low-degree generators

We can now apply the block-diagonalization techniques of Bray and Morton [2005] to find *all* generators of $\mathbf{M}_{\mathbf{BHM(n)}}$ up to any finite degree.

N.B. Bray and Morton's original approach relaxed the parameter constraint $\pi_0 + \pi_1 = 1$ to obtain a smaller ideal that was homogeneous with respect to $deg(p_I) = size(I)$. This is why they did not find the four cubics shown by Schönhuth [2011] to generate $\mathbf{I}_{\mathbf{BHM(3)}}$.

Bi-homogeneity of $\mathbf{I_{BHM(n)}}$

Application: finding low-degree generators

We can now apply the block-diagonalization techniques of Bray and Morton [2005] to find *all* generators of $\mathbf{M_{BHM(n)}}$ up to any finite degree.

N.B. Bray and Morton's original approach relaxed the parameter constraint $\pi_0 + \pi_1 = 1$ to obtain a smaller ideal that was homogeneous with respect to $deg(p_I) = size(I)$. This is why they did not find the four cubics shown by Schönhuth [2011] to generate $\mathbf{I_{BHM(3)}}$.

Problem 2: model membership testing

Given an observed distribution $p \in \mathbb{C}_p^{2^n}$, how can we determine whether p could arise from a binary hidden Markov process, i.e., whether $p \in \mathbf{M_{BHM(n)}}$?

Solution: a semialgebraic membership test

Apply the birational parametrization inverse $\overline{\psi}_n^{-1}$.

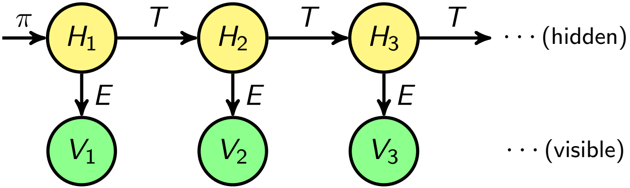
If $\overline{\psi}_n^{-1}(p)$ is undefined, we reduce to checking membership to one of two easily understood submodels of $\mathbf{M_{BHM(n)}}$, which I call INID and EBHMM.

Otherwise, we have $\overline{\psi}_n^{-1}(p) = (a, b, c, u, v)$ and v is non-zero. We choose v_0 to be either square root of v_0 of v to obtain matrices $\theta = (\pi, T, E)$, and then

$$p \in \mathbf{M_{BHM(n)}} \iff \theta \in \Theta \text{ and } \phi_n(\theta) = p$$

Problem 3

Given a BHM process, what algebraic expressions in the entries of π, T, and E can be measured based on observable data alone?



Rational parameters (skip to save time)

Consider any algebraic statistical model $\Theta \subseteq \mathbb{C}^k \xrightarrow{\phi} \mathbb{C}^n$. Usually Θ is Zariski dense in $\mathbb{C}^k$, and therefore Zariski irreducible.

A **parameter** is any function $s : \Theta \to \mathbb{C}$. A parameter is *rational* if it is the restriction of a rational function $\mathbb{C}^k \to \mathbb{C}$.

For example, in BHMM, any expression like

$$\frac{\pi_1 + 2E_{01} - c_0^3}{T_{11} - a^2 + b + u}$$

is a rational parameter. Such parameters form a field, $K \simeq \mathbb{C}(a_0, b, c_0, u, v_0)$, by Zariski density of Θ. In this talk, all parameters are rational.

Rational parameters (skip to save time)

Consider any algebraic statistical model $\Theta \subseteq \mathbb{C}^k \xrightarrow{\phi} \mathbb{C}^n$. Usually Θ is Zariski dense in $\mathbb{C}^k$, and therefore Zariski irreducible.

A **parameter** is any function $s : \Theta \to \mathbb{C}$. A parameter is *rational* if it is the restriction of a rational function $\mathbb{C}^k \to \mathbb{C}$.

For example, in BHMM, any expression like

$$\frac{\pi_1 + 2E_{01} - c_0^3}{T_{11} - a^2 + b + u}$$

is a rational parameter. Such parameters form a field, $K \simeq \mathbb{C}(a_0, b, c_0, u, v_0)$, by Zariski density of Θ. In this talk, all parameters are rational.

Rational parameters (skip to save time)

Consider any algebraic statistical model $\Theta \subseteq \mathbb{C}^k \xrightarrow{\phi} \mathbb{C}^n$. Usually Θ is Zariski dense in $\mathbb{C}^k$, and therefore Zariski irreducible.

A **parameter** is any function $s : \Theta \to \mathbb{C}$. A parameter is *rational* if it is the restriction of a rational function $\mathbb{C}^k \to \mathbb{C}$.

For example, in BHMM, any expression like

$$\frac{\pi_1 + 2E_{01} - c_0^3}{T_{11} - a^2 + b + u}$$

is a rational parameter. Such parameters form a field, $K \simeq \mathbb{C}(a_0, b, c_0, u, v_0)$, by Zariski density of Θ. In this talk, all parameters are rational.

Rational parameters (skip to save time)

Consider any algebraic statistical model $\Theta \subseteq \mathbb{C}^k \xrightarrow{\phi} \mathbb{C}^n$. Usually Θ is Zariski dense in $\mathbb{C}^k$, and therefore Zariski irreducible.

A **parameter** is any function $s : \Theta \to \mathbb{C}$. A parameter is *rational* if it is the restriction of a rational function $\mathbb{C}^k \to \mathbb{C}$.

For example, in BHMM, any expression like

$$\frac{\pi_1 + 2E_{01} - c_0^3}{T_{11} - a^2 + b + u}$$

is a rational parameter. Such parameters form a field, $K \simeq \mathbb{C}(a_0, b, c_0, u, v_0)$, by Zariski density of Θ. In this talk, all parameters are rational.

Kinds of identifiability (skip to save time)

A parameter $s \in K$ is **(set-theoretically) identifiable** if for all $\theta, \theta' \in \Theta$, $\phi(\theta) = \phi(\theta')$ implies $s(\theta) = s(\theta')$. This means we can determine the value of $s(\theta)$ from the observables $\phi(\theta)$. In other words, $s = \sigma \circ \phi$ for some set-theoretic function $\sigma : \phi(\Theta) \to \mathbb{C}$.

Identifiability is a very widely application notion, e.g. in

- Chemical reaction networks: Craciun and Pantea [2008]
- Epidemiology: Meshkat, Eisenberg, and DiStefano [2009]
- Causal inference: Sullivant, Garcia-Puente, and Spielvogel [2010]

Set theoretic identifiability is a very restrictive condition, and for applications some weaker notions are just as good:

Kinds of identifiability (skip to save time)

A parameter $s \in K$ is **(set-theoretically) identifiable** if for all $\theta, \theta' \in \Theta$, $\phi(\theta) = \phi(\theta')$ implies $s(\theta) = s(\theta')$. This means we can determine the value of $s(\theta)$ from the observables $\phi(\theta)$. In other words, $s = \sigma \circ \phi$ for some set-theoretic function $\sigma : \phi(\Theta) \to \mathbb{C}$.

Identifiability is a very widely application notion, e.g. in

- Chemical reaction networks: Craciun and Pantea [2008]
- Epidemiology: Meshkat, Eisenberg, and DiStefano [2009]
- Causal inference: Sullivant, Garcia-Puente, and Spielvogel [2010]

Set theoretic identifiability is a very restrictive condition, and for applications some weaker notions are just as good:

Kinds of identifiability (skip to save time)

A parameter $s \in K$ is **(set-theoretically) identifiable** if for all $\theta, \theta' \in \Theta$, $\phi(\theta) = \phi(\theta')$ implies $s(\theta) = s(\theta')$. This means we can determine the value of $s(\theta)$ from the observables $\phi(\theta)$. In other words, $s = \sigma \circ \phi$ for some set-theoretic function $\sigma : \phi(\Theta) \to \mathbb{C}$.

Identifiability is a very widely application notion, e.g. in

- Chemical reaction networks: Craciun and Pantea [2008]
- Epidemiology: Meshkat, Eisenberg, and DiStefano [2009]
- Causal inference: Sullivant, Garcia-Puente, and Spielvogel [2010]

Set theoretic identifiability is a very restrictive condition, and for applications some weaker notions are just as good:

Kinds of identifiability

We say that a rational parameter $s \in K$ is

rationally identifiable if $s = \sigma \circ \phi$ for some rational map $\sigma : \phi(\Theta) \to \mathbb{C}$. This notion is used without a name by Sullivant, Garcia-Puente, and Spielvogel [2010].

generically identifiable if there is a (relatively) Zariski dense open subset $U \subseteq \Theta$ such that $s|_U = \sigma \circ \phi|_U$ for some set-theoretic function $\sigma : \phi(U) \to \mathbb{C}$.

algebraically identifiable if there is a polynomial function $g(p, q) := \sum_i g_i(p_1, \ldots, p_n)q^i$ on $\phi(\Theta) \times \mathbb{C}$ of degree $d > 0$ in q (so that g_d is not identically 0 on $\phi(\Theta)$) such that $g(\phi(\theta), s(\theta)) = 0$ for all $\theta \in \Theta$ (and hence all $\theta \in \mathbb{C}^k$).

Kinds of identifiability

We say that a rational parameter $s \in K$ is

rationally identifiable if $s = \sigma \circ \phi$ for some rational map $\sigma : \phi(\Theta) \to \mathbb{C}$. This notion is used without a name by Sullivant, Garcia-Puente, and Spielvogel [2010].

generically identifiable if there is a (relatively) Zariski dense open subset $U \subseteq \Theta$ such that $s|_U = \sigma \circ \phi|_U$ for some set-theoretic function $\sigma : \phi(U) \to \mathbb{C}$.

algebraically identifiable if there is a polynomial function $g(p, q) := \sum_i g_i(p_1, \ldots, p_n)q^i$ on $\phi(\Theta) \times \mathbb{C}$ of degree $d > 0$ in q (so that g_d is not identically 0 on $\phi(\Theta)$) such that $g(\phi(\theta), s(\theta)) = 0$ for all $\theta \in \Theta$ (and hence all $\theta \in \mathbb{C}^k$).

Kinds of identifiability

We say that a rational parameter $s \in K$ is

rationally identifiable if $s = \sigma \circ \phi$ for some rational map $\sigma : \phi(\Theta) \to \mathbb{C}$. This notion is used without a name by Sullivant, Garcia-Puente, and Spielvogel [2010].

generically identifiable if there is a (relatively) Zariski dense open subset $U \subseteq \Theta$ such that $s|_U = \sigma \circ \phi|_U$ for some set-theoretic function $\sigma : \phi(U) \to \mathbb{C}$.

algebraically identifiable if there is a polynomial function $g(p, q) := \sum_i g_i(p_1, \ldots, p_n) q^i$ on $\phi(\Theta) \times \mathbb{C}$ of degree $d > 0$ in q (so that g_d is not identically 0 on $\phi(\Theta)$) such that $g(\phi(\theta), s(\theta)) = 0$ for all $\theta \in \Theta$ (and hence all $\theta \in \mathbb{C}^k$).

Kinds of identifiability

We say that a rational parameter $s \in K$ is

rationally identifiable if $s = \sigma \circ \phi$ for some rational map $\sigma : \phi(\Theta) \to \mathbb{C}$. This notion is used without a name by Sullivant, Garcia-Puente, and Spielvogel [2010].

generically identifiable if there is a (relatively) Zariski dense open subset $U \subseteq \Theta$ such that $s|_U = \sigma \circ \phi|_U$ for some set-theoretic function $\sigma : \phi(U) \to \mathbb{C}$.

algebraically identifiable if there is a polynomial function $g(p, q) := \sum_i g_i(p_1, \ldots, p_n)q^i$ on $\phi(\Theta) \times \mathbb{C}$ of degree $d > 0$ in q (so that g_d is not identically 0 on $\phi(\Theta)$) such that $g(\phi(\theta), s(\theta)) = 0$ for all $\theta \in \Theta$ (and hence all $\theta \in \mathbb{C}^k$).

Parameter classification problem

Problem 3'

Which BHMM parameters are identifiable in each sense?

Lemma

For any algebraic model $\Theta \subseteq \mathbb{C}^k \xrightarrow{\phi} \mathbb{C}^n$, if Θ is Zariski irreducible, then the sets of rationally, generically, and algebraically identifiable parameters are all fields.

Proof: The main idea is to actually be working with the Zariski topology on Θ.

Andrew Critch, critch@math.berkeley.edu

Parameter classification problem

Problem 3'

Which BHMM parameters are identifiable in each sense?

Lemma

For any algebraic model $\Theta \subseteq \mathbb{C}^k \xrightarrow{\phi} \mathbb{C}^n$, if Θ is Zariski irreducible, then the sets of rationally, generically, and algebraically identifiable parameters are all fields.

Proof: The main idea is to actually be working with the Zariski topology on Θ.

Parameter classification problem

Problem 3'

Which BHMM parameters are identifiable in each sense?

Lemma

For any algebraic model $\Theta \subseteq \mathbb{C}^k \xrightarrow{\phi} \mathbb{C}^n$, if Θ is Zariski irreducible, then the sets of rationally, generically, and algebraically identifiable parameters are all fields.

Proof: The main idea is to actually be working with the Zariski topology on Θ.

Parameter classification problem

Call these fields K_{ri}, K_{gi}, and K_{ai}. Sullivant et al. [2010] showed that for rational parameters, generic identifiability implies algebraic identifiability, so for any irredicibly parametrized model we have a series of field extensions

$$K_{ri} \subseteq K_{gi} \subseteq K_{ai} \subseteq K$$

Theorem (solution to problem 3)

For $\mathbf{M_{BHM(n)}}$ where $n \geq 3$,

$$\mathbb{C}(a, b, c, u, v) = K_{ri} = K_{gi} \subsetneq K_{ai} = \mathbb{C}(a_0, b, c_0, u, v_0)$$

Andrew Critch, critch@math.berkeley.edu

Parameter classification problem

Call these fields K_{ri}, K_{gi}, and K_{ai}. Sullivant et al. [2010] showed that for rational parameters, generic identifiability implies algebraic identifiability, so for any irredicibly parametrized model we have a series of field extensions

$$K_{ri} \subseteq K_{gi} \subseteq K_{ai} \subseteq K$$

Theorem (solution to problem 3)

For $\mathbf{M_{BHM(n)}}$ where $n \geq 3$,

$$\mathbb{C}(a, b, c, u, v) = K_{ri} = K_{gi} \subsetneq K_{ai} = \mathbb{C}(a_0, b, c_0, u, v_0)$$

Parameter classification problem

Call these fields K_{ri}, K_{gi}, and K_{ai}. Sullivant et al. [2010] showed that for rational parameters, generic identifiability implies algebraic identifiability, so for any irredicibly parametrized model we have a series of field extensions

$$K_{ri} \subseteq K_{gi} \subseteq K_{ai} \subseteq K$$

Theorem (solution to problem 3)

For $\mathbf{M_{BHM(n)}}$ where $n \geq 3$,

$$\mathbb{C}(a, b, c, u, v) = K_{ri} = K_{gi} \subsetneq K_{ai} = \mathbb{C}(a_0, b, c_0, u, v_0)$$

\end{talk}[Thank you!]

Andrew Critch, critch@math.berkeley.edu

Bibliography I

N. Bray and J. Morton. Equations defining hidden Markov models. In *Algebraic Statistics for Computational Biology*, chapter 11. Cambridge Univerisy Press, 2005.

G. Craciun and C. Pantea. Identifiability of chemical reaction networks. *Math. Chem.*, 44(1):244 – 259, 2008.

N. Meshkat, M. Eisenberg, and J. J. DiStefano. An algorithm for finding globally identifiable parameter combinations of nonlinear ode models using Gröbner bases. *Mathematical Biosciences*, 222(2):61 – 72, 2009.

A. Schönhuth. Equations for hidden Markov models. arXiv:0901.3749, 2008.

A. Schönhuth. Generic identification of binary-valued hidden Markov processes. arXiv:1101.3712, 2011.

K. Sibirskii. Algebraic invariants for a set of matrices. *Siberian Mathematical Journal*, 9:115–124, 1968. ISSN 0037-4466.

Bibliography II

B. Sturmfels and P. Zwiernik. Binary cumulant varieties, 2011. arXiv:1103.0153.

S. Sullivant, L. D. Garcia-Puente, and S. Spielvogel. Identifying causal effects with computer algebra. Proceedings of the 26th Conference of Uncertainty in Artificial Intelligence, 2010.