ALGEBRA IN DATA ANALYIS

polynomial systems for statistics and biology

AIDA MARAJ

WHAT IS A POLYNOMIAL SYSTEM?

polynomials: expressions with numbers and variables, that involve only addition, subtraction and multiplication

$x_1 + x_2 + x_3 - 1$ (linear)

$56x_1x_2$ (monomial)

$56x_1x_2 - 8x_3^2$ (binomial)

$2x_1x_2^8 + 3x_1^2 + 6x_2^3x_3^4 - 7$ (the others)

not polynomials: 2^x, $\sin x$, $\frac{1}{y}$, …

$$(p^2, \quad 2p(1-p), \quad (1-p)^2)$$
$$x_{20} \qquad x_{11} \qquad x_{02}$$

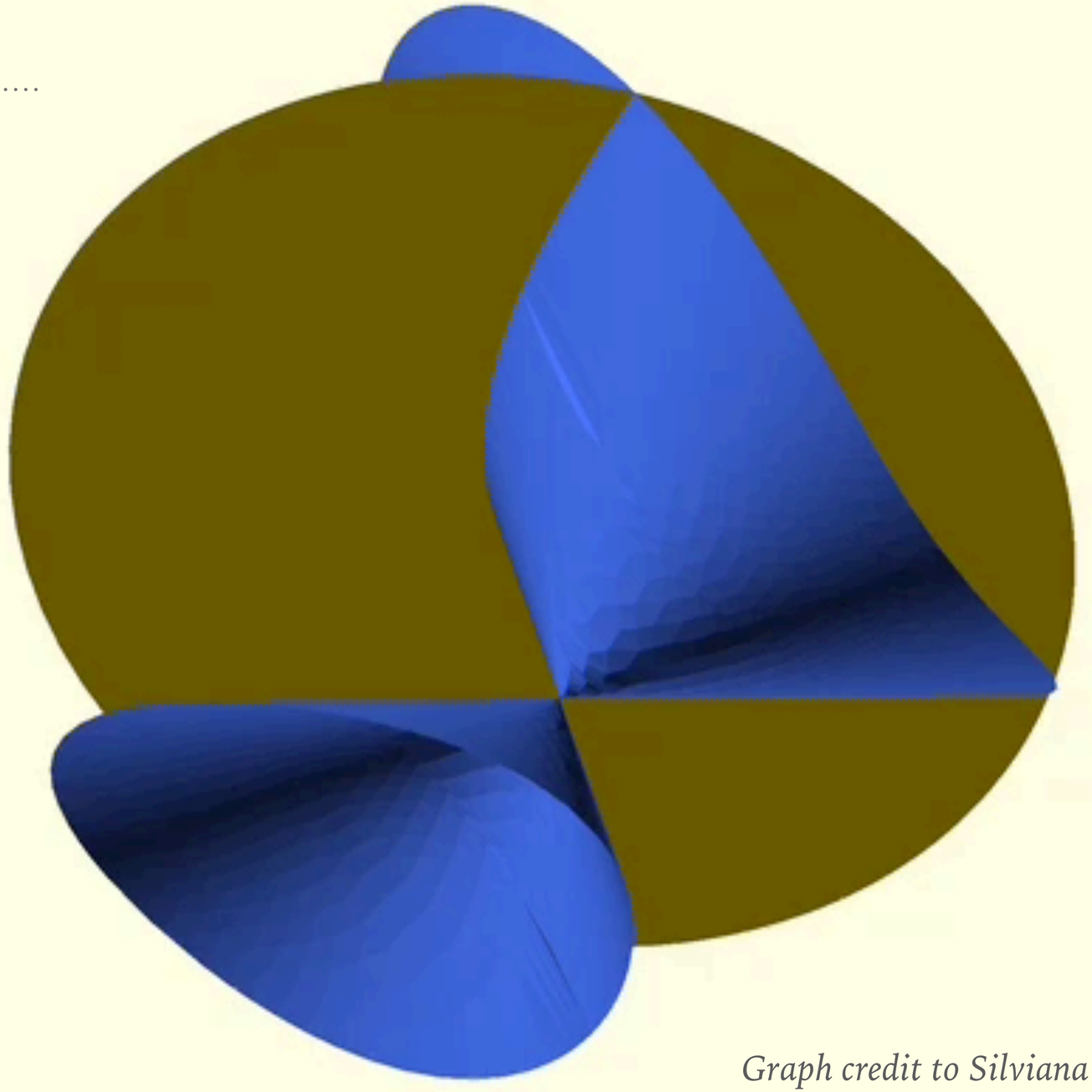
$4x_{20}x_{02} - x_{11}^2$ (Jane's coin flip equation)

ZEROS OF A POLYNOMIAL SYSTEM

$4x_1x_3 - x_2^2 = 0$ (1,2,1)

$x_1 + x_2 + x_3 - 1 = 0$ (1,0,0)

$(p^2, \quad 2p(1-p), \quad (1-p)^2)$



Graph credit to Silviana Amethyst

DATA ANALYSIS VIA POLYNOMIALS

		Previous Infection		
Take Aspirin Reg.	**Age**	Varicella	Influenza	Gastroenteritis
Yes	Child	19	11	2
	Teen	9	10	1
No	Child	400	100	74
	Teen	304	88	50

*Belay, Bresee, Holman, Khan, Shahriari, and Schonberger. Reye's syndrome in the united states from 1981 through 1997. New England Journal of Medicine, 340(18):1377–1382, 1999.

Is there any relation between type of infection, taking aspirin regularly, and age?

DATA ANALYSIS VIA POLYNOMIALS

		Previous Infection		
Take Aspirin Reg.	**Age**	Varicella	Influenza	Gastroenteritis
Yes	Child	19/1070	11/1070	2/1070
	Teen	9/1070	10/1070	1/1070
No	Child	400/1070	100/1070	74/1070
	Teen	304/1070	88/1070	50/1070

*Belay, Bresee, Holman, Khan, Shahriari, and Schonberger. Reye's syndrome in the united states from 1981 through 1997. New England Journal of Medicine, 340(18):1377–1382, 1999.

Is there any relation between type of infection, taking aspirin regularly, and age?

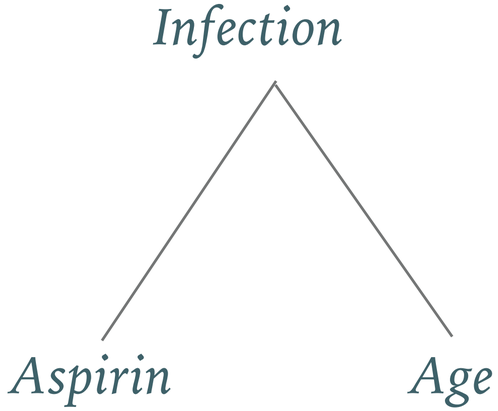
DATA ANALYSIS VIA POLYNOMIALS

		Previous Infection		
Take Aspirin Reg.	**Age**	Varicella	Influenza	Gastroenteritis
Yes	Child	19/1070	11/1070	2/1070
	Teen	9/1070	10/1070	1/1070
No	Child	400/1070	100/1070	74/1070
	Teen	304/1070	88/1070	50/1070

*Belay, Bresee, Holman, Khan, Shahriari, and Schonberger. Reye's syndrome in the united states from 1981 through 1997. New England Journal of Medicine, 340(18):1377–1382, 1999.

Is there any relation between type of infection, taking aspirin regularly, and age?

$$P(Inf = V, Asp = Y, Age = C) = P(Inf = V, Asp = Y) * P(Inf = V, Age = C)$$



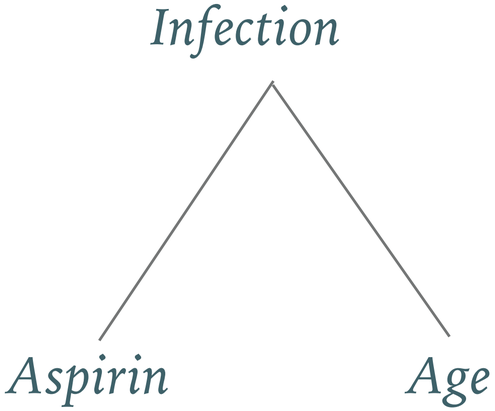
DATA ANALYSIS VIA POLYNOMIALS

		Previous Infection		
Take Aspirin Reg.	**Age**	Varicella	Influenza	Gastroenteritis
Yes	Child	19/1070	11/1070	2/1070
	Teen	9/1070	10/1070	1/1070
No	Child	400/1070	100/1070	74/1070
	Teen	304/1070	88/1070	50/1070

*Belay, Bresee, Holman, Khan, Shahriari, and Schonberger. Reye's syndrome in the united states from 1981 through 1997. New England Journal of Medicine, 340(18):1377–1382, 1999.

Is there any relation between type of infection, taking aspirin regularly, and age?

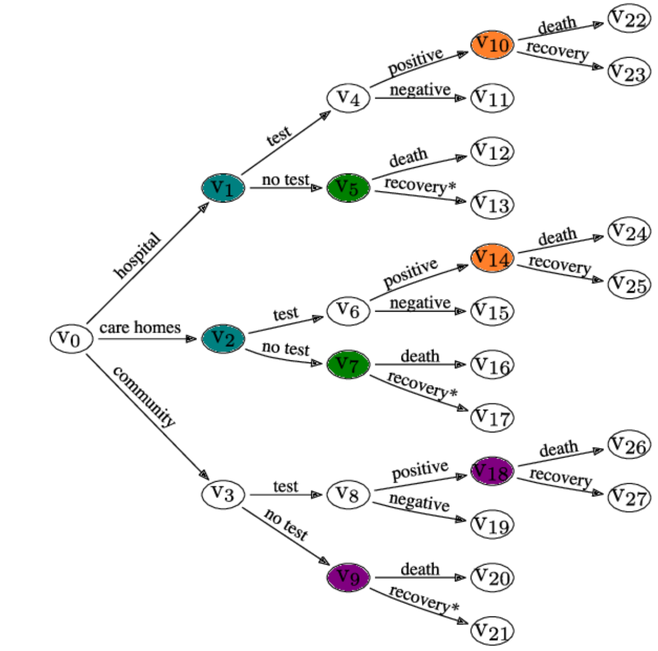
$$P(Inf = V, Asp = Y, Age = C) = P(Inf = V, Asp = Y) * P(Inf = V, Age = C)$$

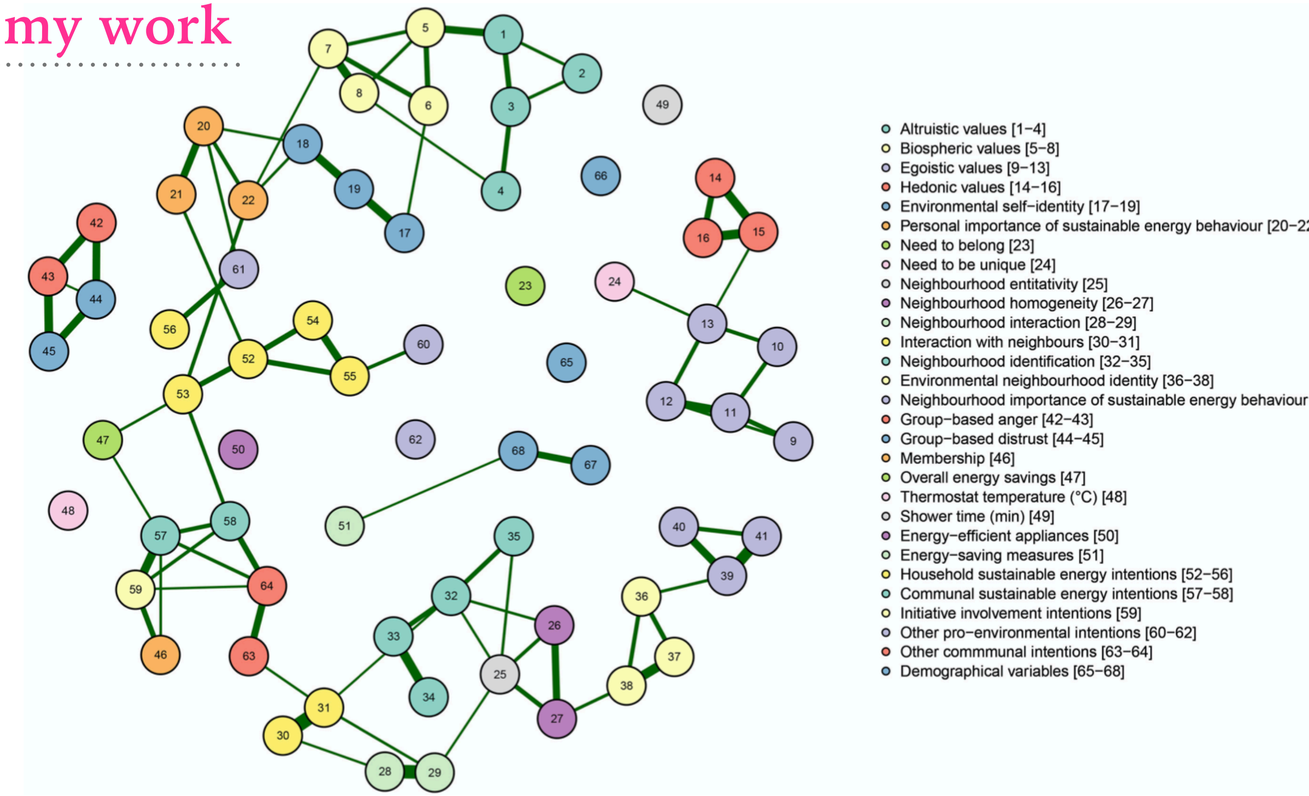


$$x_{111}x_{122} - x_{121}x_{112} = 0$$
$$x_{211}x_{222} - x_{221}x_{212} = 0$$
$$x_{311}x_{322} - x_{321}x_{312} = 0$$

Persi Diaconis and Bernd Sturmfels. **Algebraical algorithms for sampling from conditional distributions**. Annals of statistics, 26(1):363–397, 1998.

Polynomial approaches to statistical models in my work





Polynomial approaches to discrete models

Taylor Brysiewicz and Aida Maraj. **Lawrence lifts, matroids, and maximum likelihood degrees**. 2023.

Jane Ivy Coons, Joseph Cummings, Benjamin Hollering, and Aida Maraj. **Generalized cut polytopes for binary hierarchical models**. *Algebraic Statistics, 2023*

Christiane Görgen, Aida Maraj, and Lisa Nicklasson. **Staged tree models with toric structure.** *Journal of Symbolic Computation*, 113:242–268, 2022.

Aida Maraj. **Algebraic and geometric properties of hierarchical models**. *PhD dissertation–University of Kentucky*, 2020

Aida Maraj and Uwe Nagel. **Equivariant Hilbert series for hierarchical models.** *Algebraic Statistics*, 12(1):21– 42, 2021.

Polynomial approaches to multivariate Gaussian models

Tobias Boege, Jane Ivy Coons, Christopher Eur, Aida Maraj, and Frank Röttger. **Reciprocal maximum likelihood degrees of Brownian motion tree models.** *Le Matematiche*, 76(2):383 – 398, 2021.

Jane Ivy Coons, Shelby Cox, Aida Maraj, and Ikenna Nometa. **Maximum Likelihood Degrees of Brownian Motion Tree Models: Star Trees and Root Invariance,** 2024

Jane Ivy Coons, Aida Maraj, Pratik Misra, and Miruna-Stefana Sorea. **Symmetrically colored Gaussian graphical models with toric vanishing ideals.** *SIAM Journal on Applied Algebra and Geometry*, 7(1):133–158, 2023.

Authors listed alphabetically

Polynomial approaches in theory of evolution dates back to 1893

III. *Contributions to the Mathematical Theory of Evolution.*

By KARL PEARSON, *University College, London.*

Communicated by Professor HENRICI, *F.R.S.*

Received October 18,—Read November 16, 1893.

[PLATES 1—5.]

CONTENTS.

…to explain the asymmetry in data measured from a population of Naples' crabs, believing it was possible that two subpopulations of crabs were present in the sample — Gaussian mixture. He proposes method of moments.

$$
\begin{aligned}
\alpha_1 + \alpha_2 &= 1 \\
\alpha_1\mu_1 + \alpha_2\mu_2 &= 0 \\
\alpha_1(\mu_1^2 + \sigma_1^2) + \alpha_2(\mu_2^2 + \sigma_2^2) &= m_2 \\
\alpha_1(\mu_1^3 + 3\mu_1\sigma_1^2) + \alpha_2(\mu_2^3 + 3\mu_2\sigma_2^2) &= m_3 \\
\alpha_1(\mu_1^4 + 6\mu_1^2\sigma_1^2 + 3\sigma_1^4) + \alpha_2(\mu_2^4 + 6\mu_2^2\sigma_2^2 + 3\sigma_2^4) &= m_4 \\
\alpha_1(\mu_1^5 + 10\mu_1^3\sigma_1^2 + 15\mu_1\sigma_1^4) + \alpha_2(\mu_2^5 + 10\mu_2^3\sigma_2^2 + 15\mu_2\sigma_2^4) &= m_5.
\end{aligned}
$$

THEORY OF POLYNOMIAL SYSTEMS = ALGEBRAIC GEOMETRY

algebra software: Julia, Macaulay2, Bertini, SageMath, Oscar

a trick: invertible linear change of variables

$$x_1x_2 - x_3x_4 - x_5x_4 + x_1x_6 = x_1(x_2 + x_6) - (x_3 + x_5)x_4 = p_1p_2 - p_3p_4$$

Relevant theory and applications in my past work

1. Tobias Boege, Jane Ivy Coons, Christopher Eur, Aida Maraj, and Frank Röttger. **Reciprocal maximum likelihood degrees of Brownian motion tree models**. *Le Matematiche*, 76(2):383 – 398, 2021.
2. Jane Ivy Coons, Shelby Cox, Aida Maraj, and Ikenna Nometa. **Maximum Likelihood Degrees of Brownian Motion Tree Models: Star Trees and Root Invariance,** 2024
3. Christiane Görgen, Aida Maraj, and Lisa Nicklasson. **Staged tree models with toric structure.** *Journal of Symbolic Computation*, 113:242–268, 2022.
4. Aida Maraj and Arpan Pal. **Symmetry Lie algebras of varieties with applications to algebraic statistics**, 2023.
5. Thomas Kahle and Julian Vill. **Efficiently deciding if an ideal is toric after a linear coordinate change**. 2024.

Authors listed alphabetically

THEORY OF POLYNOMIAL SYSTEMS = ALGEBRAIC GEOMETRY

algebra software: Julia, Macaulay2, Bertini, SageMath, Oscar

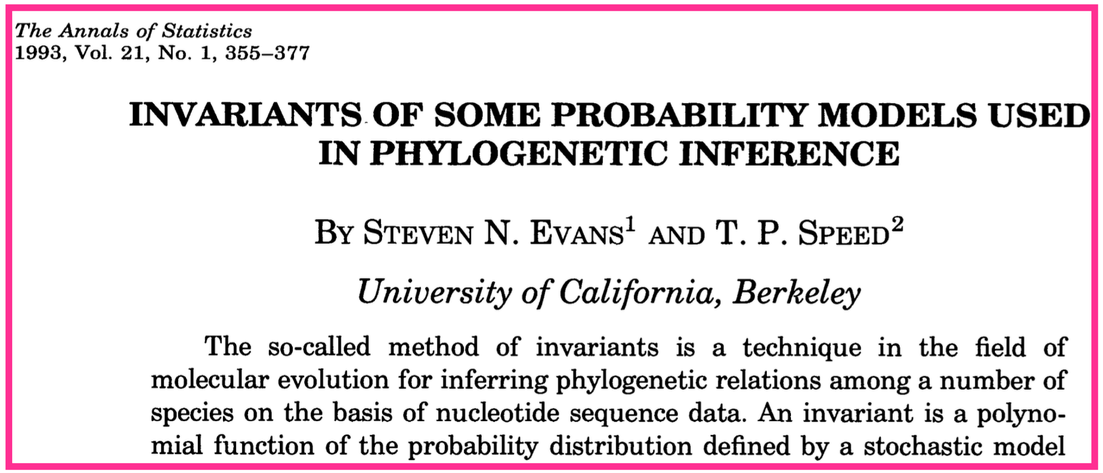
a trick: invertible linear change of variables

$$x_1x_2 - x_3x_4 - x_5x_4 + x_1x_6 = x_1(x_2 + x_6) - (x_3 + x_5)x_4 = p_1p_2 - p_3p_4$$

Relevant theory and applications in my past work

1. Tobias Boege, Jane Ivy Coons, Christopher Eur, Aida Maraj, and Frank Röttger. **Reciprocal maximum likelihood degrees of Brownian motion tree models**. *Le Matematiche*, 76(2):383 – 398, 2021.
2. Jane Ivy Coons, Shelby Cox, Aida Maraj, and Ikenna Nometa. **Maximum Likelihood Degrees of Brownian Motion Tree Models: Star Trees and Root Invariance,** 2024
3. Christiane Görgen, Aida Maraj, and Lisa Nicklasson. **Staged tree models with toric structure.** *Journal of Symbolic Computation*, 113:242–268, 2022.
4. Aida Maraj and Arpan Pal. **Symmetry Lie algebras of varieties with applications to algebraic statistics**, 2023.
5. Thomas Kahle and Julian Vill. **Efficiently deciding if an ideal is toric after a linear coordinate change**. 2024.

Authors listed alphabetically



The Annals of Statistics
1993, Vol. 21, No. 1, 355–377

INVARIANTS OF SOME PROBABILITY MODELS USED IN PHYLOGENETIC INFERENCE

BY STEVEN N. EVANS[1] AND T. P. SPEED[2]

University of California, Berkeley

The so-called method of invariants is a technique in the field of molecular evolution for inferring phylogenetic relations among a number of species on the basis of nucleotide sequence data. An invariant is a polynomial function of the probability distribution defined by a stochastic model

THE ALGEBRA IN DATA ANALYSIS GROUP

at MPI CBG and CSBD will:

- investigate statistical and biological questions through the lens of algebra, geometry and combinatorics, and
- develop new math motivated by this perspective

THE ALGEBRA IN DATA ANALYSIS GROUP

at MPI CBG and CSBD will:

- investigate statistical and biological questions through the lens of algebra, geometry and combinatorics, and
- develop new math motivated by this perspective

THANK YOU!