
Algebraic Geometry Applications in Model Selection

Luis David Garcia–Puente

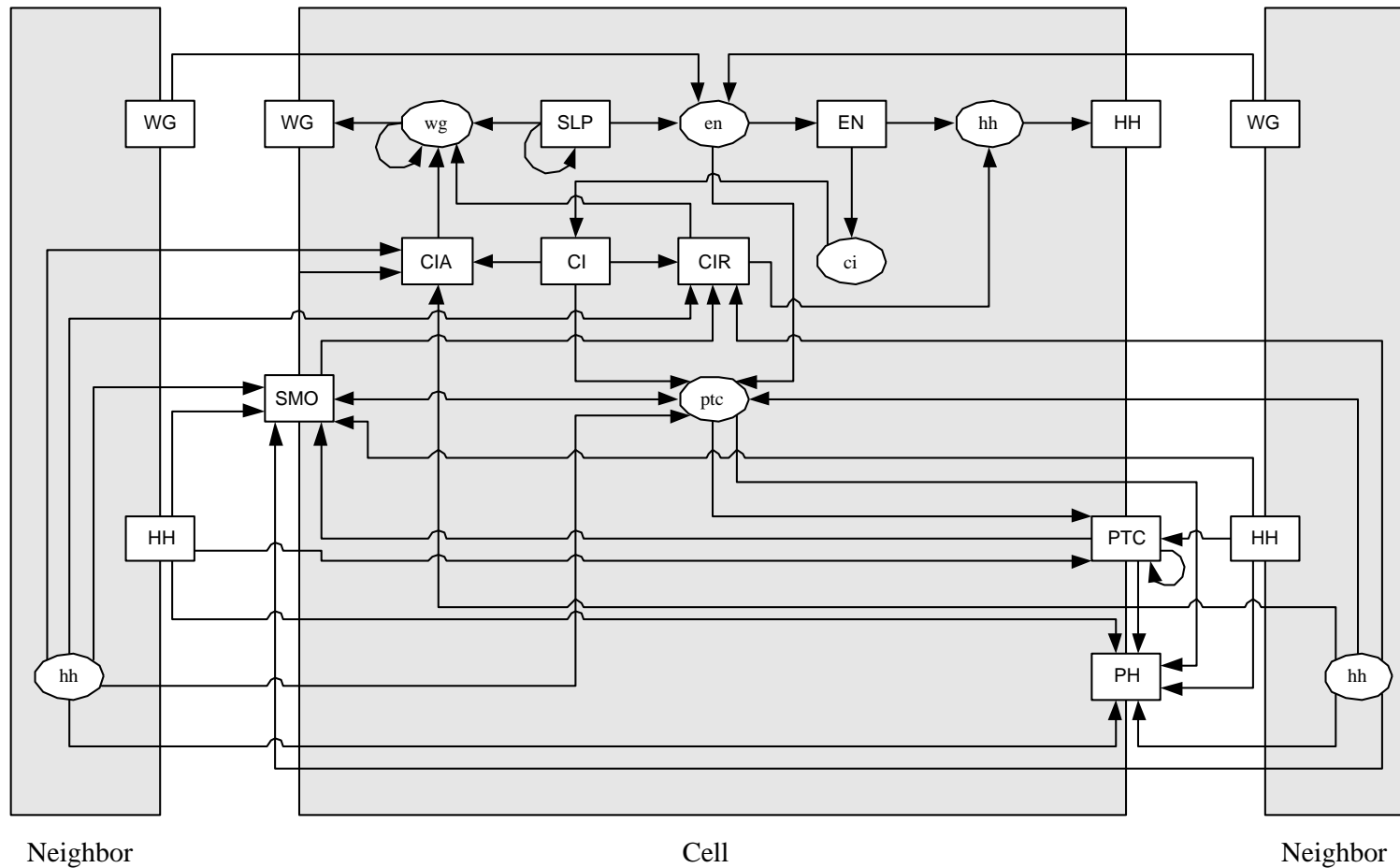
lgarcia@math.vt.edu

Virginia Polytechnic Institute and State University

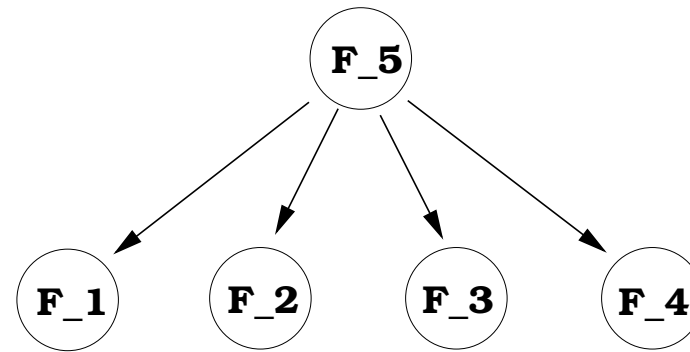
Mathematical Science Research Institute

Texas A&M University

Gene Regulatory Network in the fruit fly *D. melanogaster*



$\square \rightarrow \square$ Protein A regulates the transcription and translation of protein B ,
 $\square \rightarrow \circ$ protein A regulates the transcription of gene (mRNA) b .



● Binary **Random Variables**: $F = \{F_1, F_2, F_3, F_4, F_5\}$.

● Joint Probability **Distribution**:

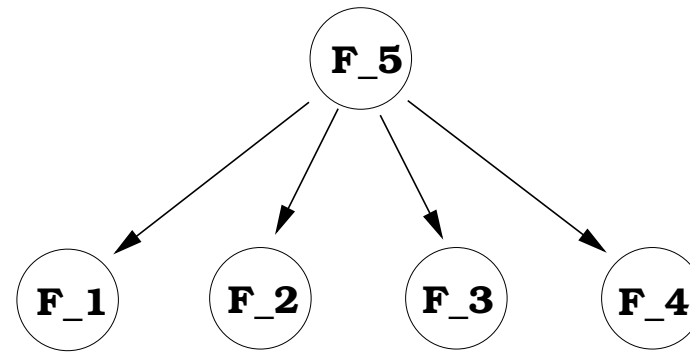
$$p(F_1 = u_1, F_2 = u_2, F_3 = u_3, F_4 = u_4, F_5 = u_5) = \prod_{i=1}^5 p(F_i = u_i | \text{pa}_i).$$

$$p(u_1, u_2, u_3, u_4, u_5) = p(u_1|u_5)p(u_2|u_5)p(u_3|u_5)p(u_4|u_5)p(u_5)$$

$$\left(\begin{pmatrix} * & * \\ \bullet & \bullet \end{pmatrix}, \begin{pmatrix} * & * \\ \bullet & \bullet \end{pmatrix}, \begin{pmatrix} * & * \\ \bullet & \bullet \end{pmatrix}, \begin{pmatrix} * & * \\ \bullet & \bullet \end{pmatrix}, \begin{pmatrix} * \\ \bullet \end{pmatrix} \right)$$

↓

$2 \times 2 \times 2 \times 2 \times 2$ – table



- Binary **Random Variables**: $F = \{F_1, F_2, F_3, F_4, F_5\}$.
- Joint Probability **Distribution**:

$$p(F_1 = u_1, F_2 = u_2, F_3 = u_3, F_4 = u_4, F_5 = u_5) = \prod_{i=1}^5 p(F_i = u_i | \text{pa}_i).$$

$$p(u_1, u_2, u_3, u_4, u_5) = p(u_1 | u_5) p(u_2 | u_5) p(u_3 | u_5) p(u_4 | u_5) p(u_5)$$

- Number of model parameters $E = 2 + 2 + 2 + 2 + 1 = 9$.
- Number of joint space parameters $D = 2^5 = 32$.
- The image of $\phi : \mathbb{R}^E \longrightarrow \mathbb{R}^D$ contains the set of all joint distributions that **factor** according to G .

Homomorphisms and Recursive Factorization

- Let p_u be an **indeterminate** representing $p(u_1, u_2, u_3, u_4, u_5)$.
- Let $\mathbb{R}[D] = \mathbb{R}[p_u \mid u \in \{0, 1\}^5]$.
- Let $a_i = p(F_i = 1 \mid F_5 = 1)$, $b_i = p(F_i = 1 \mid F_5 = 0)$, $t = p(F_5 = 1)$.
- Let $\mathbb{R}[E] = \mathbb{R}[a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4, t]$.
- $\phi : \mathbb{R}^E \rightarrow \mathbb{R}^D$ is specified by $\Phi : \mathbb{R}[D] \rightarrow \mathbb{R}[E]$

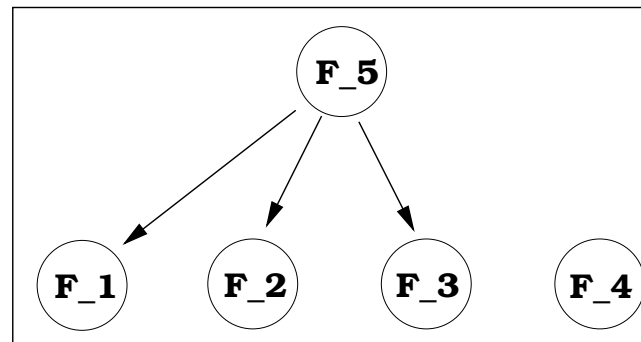
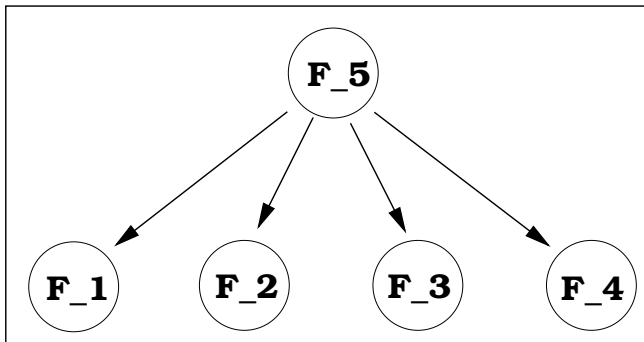
$$p_{u_1 u_2 u_3 u_4 0} \longrightarrow (1 - t) \prod_{i=1}^4 b_i^{u_i} (1 - b_i)^{1-u_i},$$

$$p_{u_1 u_2 u_3 u_4 1} \longrightarrow t \prod_{i=1}^4 a_i^{u_i} (1 - a_i)^{1-u_i}.$$

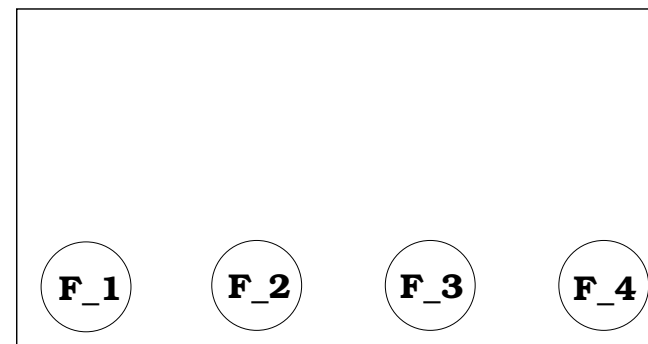
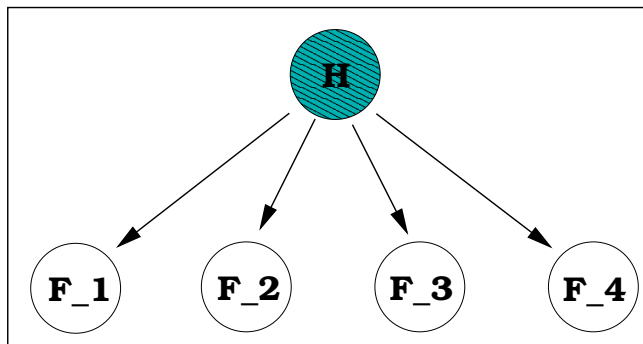
- The variety $V(\ker(\Phi))$ contains the set of all joint probability distributions that **factor** according to G .

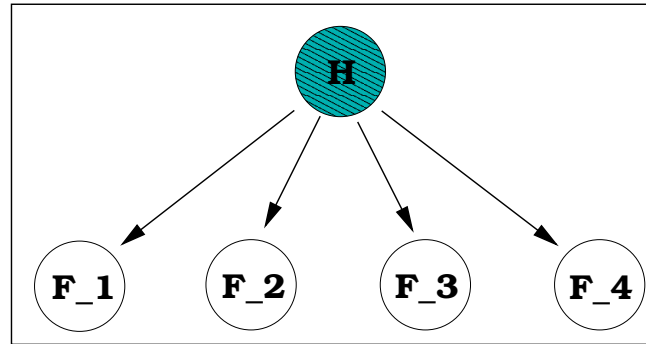
- Choose the appropriate **model** M that **best fits** a given set of observations D .

F_1	F_2	F_3	F_4	F_5
0	1	0	1	1
0	0	0	1	1
1	0	1	0	0
0	1	0	0	0
0	0	0	0	0
0	1	0	1	1



F_1	F_2	F_3	F_4
0	1	0	1
0	0	0	1
1	0	1	0
0	1	0	0
0	0	0	0
0	1	0	1

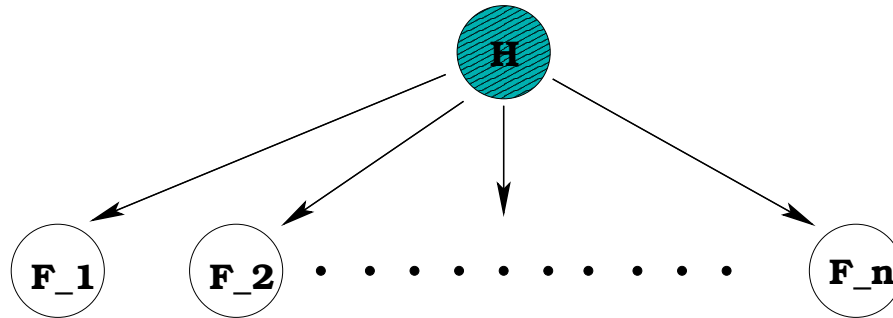




$$\begin{aligned}
 p(u_1, u_2, u_3, u_4) &= \sum_{l=1}^2 p(u_1, u_2, u_3, u_4, l) \\
 &= \sum_{l=1}^2 p(u_1|l)p(u_2|l)p(u_3|l)p(u_4|l)p(l).
 \end{aligned}$$

$$\Phi' : \mathbb{R}[p_{u_1 u_2 u_3 u_4}] \rightarrow \mathbb{R}[a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4, t]$$

$$p_{u_1 u_2 u_3 u_4} \longrightarrow t \prod_{i=1}^4 a_i^{u_i} (1 - a_i)^{1-u_i} + (1 - t) \prod_{i=1}^4 b_i^{u_i} (1 - b_i)^{1-u_i}.$$



- H is the **hidden** variable, and its levels $1, 2, \dots, r$ are called the **classes**.
- The observed random variables F_1, \dots, F_n are the **features** of the model.
- $\ker(\Phi)$ is the ideal of the **join** of r copies of the **Segre** variety

$$S_{r_1, r_2, \dots, r_n} := \mathbb{P}^{r_1-1} \times \mathbb{P}^{r_2-1} \times \dots \times \mathbb{P}^{r_n-1} \subset \mathbb{P}^{r_1 r_2 \dots r_n - 1}.$$

- The **naive Bayesian network** with r classes and n features corresponds to the r -th **secant variety** of a Segre product of n projective spaces

$$V(\ker(\Phi')) = S_{r_1, r_2, \dots, r_n}^r.$$

Bayesian Approach to Model Selection

Choose M that **maximizes** the **marginal likelihood**:

$$p(D|M) = \int_{\Omega} e^{N\mathcal{L}(Y_D|\omega)} \mu(\omega) d\omega.$$

- Ω denotes the domain of the model parameters ω .
- $\mu(\omega)$ is the **prior parameter density**, $N = |D|$.
- Y_D is the averaged sufficient statistics.
- \mathcal{L} is the **log-likelihood function** of M .

$$\theta_u = t \prod_{i=1}^n a_i^{u_i} (1 - a_i)^{1-u_i} + (1 - t) \prod_{i=1}^n b_i^{u_i} (1 - b_i)^{1-u_i}.$$

$$p(D|M) = \int_{(0,1)^{2n+1}} e^{N \sum_u Y_u \ln \theta_u(\omega)} \mu(\omega) d\omega.$$

Asymptotic Approximation for the Marginal Likelihood

Theorem (Watanabe 2001, Geiger and Rusakov 2002)

Let $I(N) = \int_{W_\epsilon} e^{-Nf(w)} \mu(w) dw$ where W_ϵ is some closed ϵ -box around w_0 , which is a minimum point of f in W_ϵ , and $f(w_0) = 0$. Assume that f and μ are analytic functions, $\mu(w_0) \neq 0$. Then,

$$\ln I(N) = \lambda_1 \ln N + (m_1 - 1) \ln \ln N + O(1)$$

where the rational number $\lambda_1 < 0$ and m_1 are the largest pole and its multiplicity of the analytic continuation of

$$J(\lambda) = \int_{f(w) < \epsilon} f(w)^\lambda \mu(w) dw \quad \operatorname{Re}(\lambda) > 0$$

Resolution Theorem [Atiyah 1970]

Let $f(w)$ be a real analytic function defined in a neighborhood of $0 \in \mathbb{R}^d$. Then there exists an open set W that contains 0, a real analytic manifold U , and a proper analytic map $g : U \longrightarrow W$ such that:

1. $g : U \setminus U_0 \longrightarrow W \setminus W_0$ is an isomorphism, where $W_0 = f^{-1}(0)$ and $U_0 = g^{-1}(W_0)$.
2. For each point $p \in U$ there are local analytic coordinates (u_1, \dots, u_d) centered at p so that, locally near p ,

$$f(g(u_1, \dots, u_d)) = a(u_1, \dots, u_d) u_1^{k_1} \cdots u_d^{k_d},$$

where $k_i \geq 0$ and $a(u)$ is an analytic function with analytic inverse $1/a(u)$.

$$I[N, Y_D] = \int_{(0,1)^{2n+1}} e^{N \sum_u Y_u \ln \theta_u(\omega)} \mu(\omega) d\omega.$$

Assume the following conditions

1. The density $\mu(\omega)$ is bounded and bounded away from zero on Ω .
2. The statistics $Y_D = (Y_1, \dots, Y_{2^n})$ satisfy $Y_i > 0$.
3. There exists N_0 such that Y_D equals the limiting statistics Y for all $N \geq N_0$.

$$I[N, Y_D] = \int_{(0,1)^{2n+1}} e^{N \sum_u Y_u \ln \theta_u(\omega)} \mu(\omega) d\omega.$$

Then for $n \geq 3$ as $N \rightarrow \infty$:

• If $Y \in S^2 \setminus S'$ (regular point)

$$\ln I[N, Y_D] = N \ln P(Y|\omega_{ML}) - \frac{2n+1}{2} \ln N + O(1),$$

• If $Y \in S' \setminus S''$ (type 1 singularity)

$$\ln I[N, Y_D] = N \ln P(Y|\omega_{ML}) - \frac{2n-1}{2} \ln N + O(1),$$

• If $Y \in S''$ (type 2 singularity)

$$\ln I[N, Y_D] = N \ln P(Y|\omega_{ML}) - \frac{n+1}{2} \ln N + O(1),$$

The **naive Bayesian network** with r classes and n features corresponds to the r -th **secant variety** of a Segre product of n projective spaces

$$V(\ker(\Phi')) = S_{r_1, r_2, \dots, r_n}^r.$$

The prime ideal $\ker(\Phi')$ of any naive Bayes model M with 2 classes and n features is generated by the 3×3 -subdeterminants of any two-dimensional table obtained by flattening the n -dimensional table $(p_{u_1 u_2 \dots u_n})$.

- Since the **independence and non-independence constraints** on the distribution over the observable variables vary from one model to another, they can be used to distinguished between models.
- Since these **constraints** are over the observable variables, their fit to data can be measured directly with some specially designed statistical tests.