Conditional independence: the algebra behind the models

"Algebraic & Geometric Methods in Statistics"

Sonja Petrović
Created for Math/Stat 561

Jan 23, 2023.

Objective

Understand how to translate conditional independence statements into polynomials, and what these polynomials mean.

Recall conditional independence (CI) from Lecture 3

Definition 4.1.2. Let $A, B, C \subseteq [m]$ be pairwise disjoint. The random vector X_A *is conditionally independent of* X_B *given* X_C if and only if

$$f_{A\cup B|C}(x_A, x_B \mid x_C) = f_{A|C}(x_A \mid x_C) \cdot f_{B|C}(x_B \mid x_C)$$

for all x_A, x_B, and x_C. The notation $X_A \perp\!\!\!\perp X_B | X_C$ is used to denote that the random vector X satisfies the conditional independence statement that X_A is conditionally independent of X_B given X_C. This is often further abbreviated to $A \perp\!\!\!\perp B | C$.

Figure 1: Source: Algebraic statistics, Seth Sullivant, AMS-GSM book

Real life examples[1]

Reminder: The *conditional probability* of A given B is represented by P(A|B). The random variables A and B are said to be **independent** if P(A)= P(A|B) (or alternatively if P(A,B)=P(A) P(B)).

Example 1

Suppose Norman and Martin each toss **separate coins**.

- Let A represent the random variable "Norman's toss outcome", and B represent the random variable "Martin's toss outcome".
- Both A and B have two possible values (Heads and Tails).
- It would be uncontroversial to assume that A and B are independent.

Evidence about B will not change our belief in A.

[1] Credit: Normal Fenton

Example 2

Now suppose both Martin and Norman toss **the same coin**.

- Again A = "Norman's toss outcome", and B = "Martin's toss outcome".
- Assume also that **there is a possibility that the coin in biased towards heads** but we do **not** know this for certain.
- In this case A and B are **not** independent.

Example: observing B = Heads causes us to increase our belief in A = Heads! So P(a|b)>P(b) in the case when a=Heads and b=Heads.

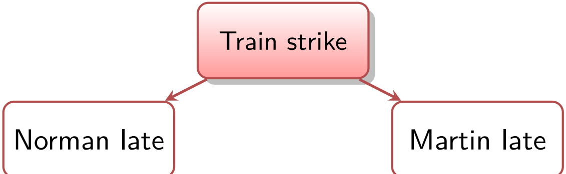
- RVs A and B are both dependent on a separate random variable C, "the coin is biased towards Heads" (which has the values True or False).
- Although A and B are not independent, it turns out that once we know for certain the value of C then any evidence about B cannot change our belief about A.

Specifically: $P(A|C) = P(A|B, C)$, so CI $A \perp\!\!\!\perp B|C$ holds.

Example 3: $A \perp\!\!\!\perp B | C$ see full gif

In many real life situations variables which are believed to be *independent* are actually *only independent conditional* on some other variable.

- Norman and Martin live on opposite sides of the City
- Norman takes the train to work. Martin drives.
- Random variables: A = "Norman late" , B = "Martin late" (true/false)
- $A \perp\!\!\!\perp B$??
  - are you sure? what about fuel shortage?
  - what about ... more traffic on the raod due to a train strike?
- Let C = "train strike".
- Clearly P(A) will increase if C is true; but P(B) will also increase because of extra traffic on the roads.



Example $\mapsto$ homework 2

Discussion of the setup. **[Whiteboard illustration.]**

Consider three **binary** random variables X_1, X_2, X_3, with joint probabilities $P(X_1 = i, X_2 = j, X_3 = 0) = P_{i,j}^{(X_3=0)}$ and $P(X_1 = i, X_2 = j, X_3 = 1) = P_{i,j}^{(X_3=1)}$, with:

$$P^{(X_3=0)} := \begin{pmatrix} 0.05 & 0.15 \\ 0.075 & 0.225 \end{pmatrix}, P^{(X_3=1)} := \begin{pmatrix} 0.125 & 0.125 \\ 0.125 & 0.125 \end{pmatrix}.$$

$\rightarrow$ This is a $2 \times 2 \times 2$ table, similar to the Berkeley admissions example in lecture3 handout.$\leftarrow$

- Find the marginal distribution P_{X_1} of X_1. (Recall that in the discrete case, integration is substituted by summation.)
- Find the conditional distribution $P_{X_2,X_3|X_1}$ of (X_2, X_3) given X_1.
- Is X_2 conditionally independent of X_3 given X_1?
- Is X_1 conditionally independent of X_2 given X_3?

The CI statement is a polynomial in the model probabilities!

Proposition (4.1.6.) & Definition (4.1.7.)

If X is a discrete random vector $X = (X_1, \ldots, X_m)$, then the CI statement $X_A \perp\!\!\!\perp X_B | X_C$ is equivalent to

$$p_{i_A, i_B, i_C, +} \cdot p_{j_A, j_B, i_C, +} - p_{i_A, j_B, i_C, +} \cdot p_{j_A, i_B, i_C, +} = 0$$

for all possible states of the variables i_A, j_A, i_B, j_B, and i_C.
The CI ideal $I_{A \perp\!\!\!\perp B | C}$ is the set of polynomials **generated** by **all** quadratic polynomials above.

The CI statement is a polynomial in the model probabilities!

Proposition (4.1.6.) & Definition (4.1.7.)

If X is a discrete random vector $X = (X_1, \ldots, X_m)$, then the CI statement $X_A \perp\!\!\!\perp X_B | X_C$ is equivalent to

$$p_{i_A,i_B,i_C,+} \cdot p_{j_A,j_B,i_C,+} - p_{i_A,j_B,i_C,+} \cdot p_{j_A,i_B,i_C,+} = 0$$

for all possible states of the variables i_A, j_A, i_B, j_B, and i_C.
The CI ideal $I_{A \perp\!\!\!\perp B|C}$ is the set of polynomials **generated** by **all** quadratic polynomials above.

- Week 1: we wrote the 3-step binary Markov chain model as a (semi)algebraic set: the set of probability distributions satisfying polynomial equations (and inequalities).
- . . . **all** polynomials? how many are there?
- . . . **"generated"**?

Advanced HW (on hw2)

Prove proposition 4.1.6. The outline of the proof is in the book.

Example

Let X_1, X_2, X_3, X_4 be four discrete random variables with the following state spaces: $X_1 \subset \{1, 2, 3\}$, $X_2, X_3, X_4 \subset \{1, 2\}$.

- Interpret: X_1 = gender (M/F/other), X_2 = short hair (1=y/2=n), X_3 = likes soccer (y/n), X_4 = from Brazil (y/n).

$X_1 \perp\!\!\!\perp X_2 | X_3 \iff$

Advanced HW (on hw2)

Prove proposition 4.1.6. The outline of the proof is in the book.

Example

Let X_1, X_2, X_3, X_4 be four discrete random variables with the following state spaces: $X_1 \subset \{1, 2, 3\}$, $X_2, X_3, X_4 \subset \{1, 2\}$.

- Interpret: X_1 = gender (M/F/other), X_2 = short hair (1=y/2=n), X_3 = likes soccer (y/n), X_4 = from Brazil (y/n).

$X_1 \perp\!\!\!\perp X_2 | X_3 \iff p_{1,1,1,+} \cdot p_{2,2,1,+} - p_{1,2,1,+} \cdot p_{2,1,1,+} = 0$

Advanced HW (on hw2)

Prove proposition 4.1.6. The outline of the proof is in the book.

Example

Let X_1, X_2, X_3, X_4 be four discrete random variables with the following state spaces: $X_1 \subset \{1, 2, 3\}$, $X_2, X_3, X_4 \subset \{1, 2\}$.

- Interpret: X_1 = gender (M/F/other), X_2 = short hair (1=y/2=n), X_3 = likes soccer (y/n), X_4 = from Brazil (y/n).

$X_1 \perp\!\!\!\perp X_2 | X_3 \iff p_{1,1,1,+} \cdot p_{2,2,1,+} - p_{1,2,1,+} \cdot p_{2,1,1,+} = 0$

$p_{1,1,1,+} \cdot p_{2,3,1,+} - p_{1,3,1,+} \cdot p_{2,1,1,+} = 0$

$p_{1,2,1,+} \cdot p_{2,3,1,+} - p_{1,3,1,+} \cdot p_{2,2,1,+} = 0$

$\ldots$

Advanced HW (on hw2)

Prove proposition 4.1.6. The outline of the proof is in the book.

Example

Let X_1, X_2, X_3, X_4 be four discrete random variables with the following state spaces: $X_1 \subset \{1,2,3\}$, $X_2, X_3, X_4 \subset \{1,2\}$.

- Interpret: X_1 = gender (M/F/other), X_2 = short hair (1=y/2=n), X_3 = likes soccer (y/n), X_4 = from Brazil (y/n).

$X_1 \perp\!\!\!\perp X_2 | X_3 \iff p_{1,1,1,+} \cdot p_{2,2,1,+} - p_{1,2,1,+} \cdot p_{2,1,1,+} = 0$

$p_{1,1,1,+} \cdot p_{2,3,1,+} - p_{1,3,1,+} \cdot p_{2,1,1,+} = 0$

$p_{1,2,1,+} \cdot p_{2,3,1,+} - p_{1,3,1,+} \cdot p_{2,2,1,+} = 0$

$\ldots$

$p_{1,1,2,+} \cdot p_{2,2,2,+} - p_{1,2,2,+} \cdot p_{2,1,2,+} = 0$ and more! !!

- **And** all of these $+$s mean, e.g., $p_{1,1,1,+} = p_{1,1,1,1} + p_{1,1,1,2}$.

Question

Is there an efficient way of (1) encoding these polynomials and (2) generating them for a simple exmaple??

Example Macaulay2 code

```
Macaulay2, version 1.18
i1 : loadPackage "GraphicalModels";
i2 : R = markovRing (3,2,2,2);
o2 =  R
o2 :  PolynomialRing
i3 : rvNames = {gender,hair,soccer,brazil}
o3 =  {gender,hair,soccer,brazil}
o3 :  List
i4 : CIstatements = { {{gender},{hair},{soccer}} }
-- this says gender indep. of hair given soccer
o4 =  {{{gender},{hair},{soccer}}}
o4 :  List
i5 : conditionalIndependenceIdeal(R,CIstatements,rvNames)
```

You can compute this online yourself.. The lines starting with “i” are input lines that you type into the editor to execute them.

$\texttt{ideal}(-p_{1,2,1,1}p_{2,1,1,1} - p_{1,2,1,2}p_{2,1,1,1} - p_{1,2,1,1}p_{2,1,1,2} - p_{1,2,1,2}p_{2,1,1,2} + p_{1,1,1,1}p_{2,2,1,1} + p_{1,1,1,2}p_{2,2,1,1} + p_{1,1,1,1}p_{2,2,1,2} + p_{1,1,1,2}p_{2,2,1,2},$

$-p_{1,2,1,1}p_{3,1,1,1} - p_{1,2,1,2}p_{3,1,1,1} - p_{1,2,1,1}p_{3,1,1,2} - p_{1,2,1,2}p_{3,1,1,2} + p_{1,1,1,1}p_{3,2,1,1} + p_{1,1,1,2}p_{3,2,1,1} + p_{1,1,1,1}p_{3,2,1,2} + p_{1,1,1,2}p_{3,2,1,2},$

$-p_{2,2,1,1}p_{3,1,1,1} - p_{2,2,1,2}p_{3,1,1,1} - p_{2,2,1,1}p_{3,1,1,2} - p_{2,2,1,2}p_{3,1,1,2} + p_{2,1,1,1}p_{3,2,1,1} + p_{2,1,1,2}p_{3,2,1,1} + p_{2,1,1,1}p_{3,2,1,2} + p_{2,1,1,2}p_{3,2,1,2},$

$-p_{1,2,2,1}p_{2,1,2,1} - p_{1,2,2,2}p_{2,1,2,1} - p_{1,2,2,1}p_{2,1,2,2} - p_{1,2,2,2}p_{2,1,2,2} + p_{1,1,2,1}p_{2,2,2,1} + p_{1,1,2,2}p_{2,2,2,1} + p_{1,1,2,1}p_{2,2,2,2} + p_{1,1,2,2}p_{2,2,2,2},$

$-p_{1,2,2,1}p_{3,1,2,1} - p_{1,2,2,2}p_{3,1,2,1} - p_{1,2,2,1}p_{3,1,2,2} - p_{1,2,2,2}p_{3,1,2,2} + p_{1,1,2,1}p_{3,2,2,1} + p_{1,1,2,2}p_{3,2,2,1} + p_{1,1,2,1}p_{3,2,2,2} + p_{1,1,2,2}p_{3,2,2,2},$

$-p_{2,2,2,1}p_{3,1,2,1} - p_{2,2,2,2}p_{3,1,2,1} - p_{2,2,2,1}p_{3,1,2,2} - p_{2,2,2,2}p_{3,1,2,2} + p_{2,1,2,1}p_{3,2,2,1} + p_{2,1,2,2}p_{3,2,2,1} + p_{2,1,2,1}p_{3,2,2,2} + p_{2,1,2,2}p_{3,2,2,2})$

Class work:

Determine why these are correct. worksheet 2.

Homework 1, problem 3 [due today]

Example 3.1.6 (Marginal independence)**.** The (marginal) independence statement $X_1 \perp\!\!\!\perp X_2$, or equivalently, $X_1 \perp\!\!\!\perp X_2 \mid X_\emptyset$, amounts to saying that the matrix

$$\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r_2} \\ p_{21} & p_{22} & \cdots & p_{2r_2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r_11} & p_{r_12} & \cdots & p_{r_1r_2} \end{pmatrix}$$

has rank one. The independence ideal $I_{1\perp\!\!\!\perp 2}$ is generated by the 2×2-minors:

$$I_{1\perp\!\!\!\perp 2} = \left\langle p_{i_1i_2}p_{j_1j_2} - p_{i_1j_2}p_{i_2j_1} \mid i_1, j_1 \in [r_1], i_2, j_2 \in [r_2] \right\rangle.$$

For marginal independence, we already saw these quadratic binomial constraints in Chapter 1. $\square$

Figure 2: From the Lectures on Algebraic Statistics book

Algebraic varieties and polynomial ideals

We have already seen these in Lecture 1, and in homework 1 (problem 4). Here is a brief overview of what you need to know.

- A variety is the solution set to a simultaneous system of polynomial equations.
- If I is an ideal[2], then V (I) is the variety defined by the vanishing of *all* polynomials in I.
- Hilbert basis theorem: even if I is infinite (it is!), there exists a finite basis for every I.

Important questions for statistics [digest next 2 slides]:

- what are points in a variety?
- how do you check if a point is on a variety?
- what if you are given an observation of 3 binary random variables, can you use polynomials to check some CI statements?

[2] an *ideal* is the infinite set of polynomial combination of some generating set.

Recall Example 1.1.2 from the book: 3-step Markov chain.

$p_{ijk} = P(X_1 = i, X_2 = j, X_3 = k)$ and $P(X_3 = k | X_1 = i, X_2 = j) = \frac{p_{ijk}}{p_{ij+}}$.

You verified that a probability distribution, represented by a vector of probabilities $p = (p_{000}, p_{001}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}, p_{111}) \in \mathbb{R}^8$, being in this model is *equivalent to* the following four conditions:

$$p_{ijk} \geq 0 \text{ for all } i, j, k \in \{0, 1\}, \qquad \sum_{i,j,k} p_{ijk} = 1,$$

$$p_{000}p_{101} - p_{001}p_{100} = 0, \text{ and } p_{010}p_{111} - p_{011}p_{110} = 0.$$

- In this example: **what is the variety**?

Recall Example 1.1.2 from the book: 3-step Markov chain.

$p_{ijk} = P(X_1 = i, X_2 = j, X_3 = k)$ and $P(X_3 = k | X_1 = i, X_2 = j) = \frac{p_{ijk}}{p_{ij+}}$.

You verified that a probability distribution, represented by a vector of probabilities $p = (p_{000}, p_{001}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}, p_{111}) \in \mathbb{R}^8$, being in this model is *equivalent to* the following four conditions:

$$p_{ijk} \geq 0 \text{ for all } i, j, k \in \{0, 1\}, \qquad \sum_{i,j,k} p_{ijk} = 1,$$

$$p_{000}p_{101} - p_{001}p_{100} = 0, \text{ and } p_{010}p_{111} - p_{011}p_{110} = 0.$$

- In this example: **what is the variety**?
- Is the point $(1/8, 1/8, \ldots, 1/8)$ **on** this variety? (That is, is this joint probability vector in the model?)
  - [Verify right now, by hand.]

Recall Example 1.1.2 from the book: 3-step Markov chain.

$p_{ijk} = P(X_1 = i, X_2 = j, X_3 = k)$ and $P(X_3 = k | X_1 = i, X_2 = j) = \frac{p_{ijk}}{p_{ij+}}$.

You verified that a probability distribution, represented by a vector of probabilities $p = (p_{000}, p_{001}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}, p_{111}) \in \mathbb{R}^8$, being in this model is *equivalent to* the following four conditions:

$$p_{ijk} \geq 0 \text{ for all } i, j, k \in \{0, 1\}, \qquad \sum_{i,j,k} p_{ijk} = 1,$$

$$p_{000}p_{101} - p_{001}p_{100} = 0, \text{ and } p_{010}p_{111} - p_{011}p_{110} = 0.$$

- In this example: **what is the variety**?
- Is the point $(1/8, 1/8, \ldots, 1/8)$ **on** this variety? (That is, is this joint probability vector in the model?)
  - [Verify right now, by hand.]
- Find an example of a point on the variety, which is a point in this model.
  - [Discussion of HW 1.b).]

Varieties

Points on the model $\leftrightarrow$ points on the variety.

What are model *ideals*?

If $p_{000}p_{101} - p_{001}p_{100} = 0$ and $p_{010}p_{111} - p_{011}p_{110} = 0$ both hold for a probability vector $p = (p_{000}, p_{001}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}, p_{111}) \in \mathbb{R}^8$, then **what other polynomial equations also hold**?

Varieties

Points on the model $\leftrightarrow$ points on the variety.

What are model *ideals*?

If $p_{000}p_{101} - p_{001}p_{100} = 0$ and $p_{010}p_{111} - p_{011}p_{110} = 0$ both hold for a probability vector $p = (p_{000}, p_{001}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}, p_{111}) \in \mathbb{R}^8$, then **what other polynomial equations also hold**?

- *All* such equations vanish on the points of the model.
  - [Discussion.]
- The (infinite) collection of all of these polynomials is the ideal of this statistical model.
- It is an equivalent description of the model.
  - [$\rightarrow$ algebraic geometry.]

Ideals

The (defining) ideal of a model is the set of all polynomials that simultaneously vanish on all points in the model.

Luckily, there is a finite basis for each ideal.

How to combine several CI statements?

Sum of ideals.

Example 3.1.10. Let X_1, X_2, X_3, X_4 be binary random variables, and consider the conditional independence model

$$\mathcal{C} = \{1 \perp\!\!\!\perp 3 \mid \{2,4\}, 2 \perp\!\!\!\perp 4 \mid \{1,3\}\}.$$

These are the conditional independence statements that hold for the graphical model associated to the four cycle graph with edges $\{12, 23, 34, 14\}$; see Section 3.2. The conditional independence ideal is generated by eight quadratic binomials:

$$\begin{aligned} I_{\mathcal{C}} &= I_{1 \perp\!\!\!\perp 3 \mid \{2,4\}} + I_{2 \perp\!\!\!\perp 4 \mid \{1,3\}} \\ &= \langle p_{1111}p_{2121} - p_{1121}p_{2111}, p_{1112}p_{2122} - p_{1122}p_{2112}, \\ &\quad p_{1211}p_{2221} - p_{1221}p_{2211}, p_{1212}p_{2222} - p_{1222}p_{2212}, \\ &\quad p_{1111}p_{1212} - p_{1112}p_{1211}, p_{1121}p_{1222} - p_{1122}p_{1221}, \\ &\quad p_{2111}p_{2212} - p_{2112}p_{2211}, p_{2121}p_{2222} - p_{2122}p_{2221}\rangle . \end{aligned}$$

Figure 3: From the Lectures on Algebraic Statistics book

Appendix

Here are some additional examples you may wish to explore, to familiarize yourself with conditional independence:

- This is where the Martin&Normal example came from.
- This informal website has some additional interesting examples.
- Here is a set of slides with several real-world examples of CI random variables.

License

This document is created for Math/Stat 561, Spring 2023, at Illinois Tech.

The first example is from Kaie Kubjas' course. Other online sources are cited throughout.