Algebraic Statistics

Pallavy Rai
Roll No: 212305010

Under the guidance of

Dr. Rohith Varma
Assistant Professor
Department of Mathematics

Department of Mathematics
Indian Institute of Technology Palakkad

May 13 2025

Abstract

This presentation explains conditional independence in probability, an important idea in statistics. We show how it's defined using conditional and marginal densities and how it can be written using factorizations. Then, we link these ideas to algebra by using polynomial equations.

We go over key rules like symmetry and the intersection axiom, and use simple examples (like binary variables) to show how algebra helps us find hidden patterns. Finally, we show how graphs can help check when variables are independent, combining ideas from probability, algebra, and graph theory.

Conditional Independence

- Let $X = (X_1, X_2, X_3)$ be a 3-dimensional random vector, where $X_i \in \mathcal{X}_i$ for $i = 1, 2, 3$, and let $f(x_1, x_2, x_3)$ denote its joint density function.
- The **marginal density** of X_1 is given by:

$$f_{X_1}(x_1) = \int_{\mathcal{X}_2} \int_{\mathcal{X}_3} f(x_1, x_2, x_3)\, dx_2\, dx_3$$

- The **conditional density** of (X_1, X_2) given $X_3 = x_3$ is:

$$f_{X_1,X_2|X_3}(x_1, x_2 \mid x_3) = \begin{cases} \dfrac{f(x_1, x_2, x_3)}{f_{X_3}(x_3)} & \text{if } f_{X_3}(x_3) > 0 \\ 0 & \text{otherwise} \end{cases}$$

Definition (Conditional Independence)

We say that X_1 is **conditionally independent** of X_2 given X_3, denoted $X_1 \perp\!\!\!\perp X_2 \mid X_3$, if

$$f_{X_1,X_2|X_3}(x_1, x_2 \mid x_3) = f_{X_1|X_3}(x_1 \mid x_3) \cdot f_{X_2|X_3}(x_2 \mid x_3)$$

Conditional Independence Model

- Let $X = (X_1, \dots, X_m)$ be an m-dimensional random vector taking values in the Cartesian product space

$$\mathcal{X} = \prod_{i=1}^{m} \mathcal{X}_i.$$

- Let $\mathbf{X} \sim f(x) = f(x_1, \dots, x_m)$
- For each subset $A \subseteq [m] := \{1, 2, \dots, m\}$, let $X_A = (X_a)_{a \in A}$ be the subvector of X, and for a partition $A_1 | \dots | A_k$ of $[m]$, we denote the function $f(x)$ as $f(x_{A_1}, \dots, x_{A_k})$.

Definition

Let $A \subseteq [m]$. The marginal density $f_A(x_A)$ of X_A is given by

$$f_A(x_A) = \int_{\mathcal{X}_{[m]\setminus A}} f(x_A, x_{[m]\setminus A})\, d\nu(x_{[m]\setminus A}),$$

where, ν is the product measure on $\mathcal{X}_{[m]\setminus A}$.

Let $A, B \subseteq [m]$ be disjoint subsets and $x_C \in \mathcal{X}_C$. The conditional density of X_A given $X_B = x_B$ is defined as

$$f_{A|B}(x_A \mid x_B) = \begin{cases} \dfrac{f_{A\cup B}(x_A, x_B)}{f_B(x_B)}, & \text{if } f_B(x_B) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Note that $f_{A\cup B}(x_A, x_B) = f_{A\cup B}(x_{A\cup B})$.

Definition (Conditional Independence)

Let $A, B, C \subseteq [m]$ be pairwise disjoint. The random vector X_A is said to be *conditionally independent* of X_B given X_C iff

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

or equivalently,

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

We denote this conditional independence by

$$X_A \perp\!\!\!\perp X_B \mid X_C,$$

often abbreviated as $A \perp\!\!\!\perp B \mid C$.

Marginal Independence

Definition

A statement $X_A \perp\!\!\!\perp X_B$ or $X_A \perp\!\!\!\perp X_B \mid X_\varnothing$ is called *marginal independence*, meaning no conditioning is involved. The corresponding density factorization is

$$f_{A \cup B}(x_A, x_B) = f_A(x_A) f_B(x_B),$$

which is the definition of independence of random variables.

Inference Rule of Conditional Independence

Proposition

Let $A, B, C, D \subseteq [m]$ be pairwise disjoint subsets. Then

1. ***Symmetry:***

$$X_A \perp\!\!\!\perp X_B \mid X_C \quad \Rightarrow \quad X_B \perp\!\!\!\perp X_A \mid X_C.$$

2. ***Decomposition:***

$$X_A \perp\!\!\!\perp X_{B \cup D} \mid X_C \quad \Rightarrow \quad X_A \perp\!\!\!\perp X_B \mid X_C.$$

3. ***Weak Union:***

$$X_A \perp\!\!\!\perp X_{B \cup D} \mid X_C \quad \Rightarrow \quad X_A \perp\!\!\!\perp X_B \mid X_{C \cup D}.$$

4. ***Contraction:***

$$X_A \perp\!\!\!\perp X_B \mid X_{C \cup D} \text{ and } X_A \perp\!\!\!\perp X_D \mid X_C \quad \Rightarrow \quad X_A \perp\!\!\!\perp X_{B \cup D} \mid X_C.$$

Intersection Axiom

Proposition

*(**Intersection axiom**) If $f_X(x) > 0$ for all x, then*

$$X_A \perp\!\!\!\perp X_B \mid X_{C \cup D} \text{ and } X_A \perp\!\!\!\perp X_C \mid X_{B \cup D} \Rightarrow X_A \perp\!\!\!\perp X_{B \cup C} \mid X_D.$$

Theorem

Let $(\Omega, \mathcal{F}, \mathbb{P})$ *be the probability space. If* $\mathbf{X} = (X_1, \ldots, X_m)$ *is a discrete random vector, then*

$$X_A \perp\!\!\!\perp X_B \mid X_C \iff$$

$$P(X_A = x_A, X_B = x_B, X_C = x_C) \cdot \mathbb{P}(X_A = x'_A, X_B = x'_B, X_C = x_C)$$
$$= P(X_A = x_A, X_B = x'_B, X_C = x_C) \cdot \mathbb{P}(X_A = x_A\prime, X_B = x_B, X_C = x_C) \ldots (1)$$

where $A, B, C \subseteq \Omega$ *such that* $A \cap B \cap C = \varnothing$*,*

$$X_A = (X_i)_{i \in A}, \quad X_B = (X_j)_{j \in B}, \quad X_C = (X_k)_{k \in C}$$

Definition (Conditional Independence Ideal)

The ideal of *conditional independence* is the ideal generated by the quadratic polynomial of Equation (1) of the above theorem.

Example

- Let $m = 2$ and $X_1 \perp\!\!\!\perp X_2$, then:

$$I_{1\perp\!\!\!\perp 2} = \langle p_{i_1 j_1} p_{i_2 j_2} - p_{i_1 j_2} p_{i_2 j_1} : i_1, i_2 \in [n_1], j_1, j_2 \in [n_2] \rangle.$$

- If $m = 3$ and $X_1 \perp\!\!\!\perp X_2 \mid X_3$, then:

$$I_{1\perp\!\!\!\perp 2|3} = \langle p_{i_1 j_1 k} p_{i_2 j_2 k} - p_{i_1 j_2 k} p_{i_2 j_1 k} : k \in [n_3], i_1, i_2 \in [n_1], j_1, j_2 \in [n_2] \rangle.$$

Let $\mathcal{E} = \{A_1 \perp\!\!\!\perp B_1 \mid C_1, A_2 \perp\!\!\!\perp B_2 \mid C_2, \ldots\}$ be a set of all the conditional independence statements. Then:

$$I_{\mathcal{E}} := \sum_{A \perp\!\!\!\perp B | C \in \mathcal{E}} I_{A \perp\!\!\!\perp B | C},$$

which is generated by all the quadratic polynomials of the ideal $I_{A_i \perp\!\!\!\perp B_i | C_i}$.

Example (Binary Contraction Axiom)

Let $\mathcal{E} = \{1 \perp\!\!\!\perp 2 \mid 3,\ 2 \perp\!\!\!\perp 3\}$.

$1 \perp\!\!\!\perp 2 \mid 3$ and $2 \perp\!\!\!\perp 3 \Rightarrow 2 \perp\!\!\!\perp \{1,3\}$ (by contraction and symmetry)

Considering this in case of binary random variables,

$$I_{\mathcal{E}} = \langle p_{111}p_{221} - p_{121}p_{211}, p_{112}p_{222} - p_{122}p_{212},\\ (p_{111} + p_{211})(p_{122} + p_{222}) - (p_{112} + p_{212})(p_{121} + p_{221})\rangle$$

On computing the primary decomposition of the example in software (named `Singular`), we get:

$$I_{\mathcal{E}} = I_{2 \perp\!\!\!\perp \{1,3\}} \cap \langle p_{112}p_{222} - p_{122}p_{212},\ p_{111} + p_{211},\ p_{121} + p_{221}\rangle\\ \cap \langle p_{111}p_{221} - p_{121}p_{211},\ p_{122} + p_{222},\ p_{112} + p_{212}\rangle$$

Hence,

$$Z(I_{\mathcal{E}}) = Z(I_{2 \perp\!\!\!\perp \{1,3\}}) \cup Z(K) \cup Z(L)$$

where,

$$\begin{aligned} L &= \langle p_{111}p_{221} - p_{121}p_{211},\ p_{122} + p_{222},\ p_{112} + p_{212} \rangle \\ K &= \langle p_{112}p_{222} - p_{122}p_{212},\ p_{111} + p_{211},\ p_{121} + p_{221} \rangle \end{aligned}$$

Now consider the second component of primary decomposition of $I_{\mathcal{E}}$, K. Consider the matrix:

$$P = \begin{bmatrix} p_{111} & p_{112} & p_{211} & p_{212} \\ p_{121} & p_{122} & p_{221} & p_{222} \end{bmatrix}$$

whose rows are indexed by $[r_2]$ and column are indexed by $[r_1] \times [r_3]$.

Since $p_{111} = p_{211} = p_{121} = p_{221} = 0$, we get:

$$P = \begin{bmatrix} 0 & p_{112} & 0 & p_{212} \\ 0 & p_{122} & 0 & p_{222} \end{bmatrix}$$

But if $p_{112}p_{222} - p_{122}p_{212} = 0$, then all 2×2 minors of the matrix P are zero. Hence,

$$Z(K) \cap \Delta_7 \subseteq Z(I_{2 \perp\!\!\!\perp \{1,3\}}) \cap \Delta_7$$

Similarly, we can prove for L,

$$Z(L) \cap \Delta_7 \subseteq Z(I_{2 \perp\!\!\!\perp \{1,3\}}) \cap \Delta_7$$

Hence, with the help of primary decomposition and some probabilistic reasoning, we can verify the conditional independence implication.

Construction of graph, G_f

To state the strongest possible version, we can associate a graph G_f to the density function $f(x_1, x_2, x_3)$. The graph G_f has a vertex set

$$V = \{(i_2, i_3) \in [r_2] \times [r_3] \mid f_{X_2,X_3}(i_2, i_3) > 0\}.$$

Two vertices (i_2, i_3) and (j_2, j_3) are connected by an edge in G_f if $i_2 = j_2$ or $i_3 = j_3$.

Sufficient condition for Intersection Axiom

Theorem

Let X be a discrete random variable with density function f satisfying the conditional independence (CI) statements

$$X_1 \perp\!\!\!\perp X_2 \mid X_3 \quad \textit{and} \quad X_1 \perp\!\!\!\perp X_3 \mid X_2.$$

If G_f is a connected graph, then it follows that

$$X_1 \perp\!\!\!\perp (X_2, X_3).$$

Conclusion and Takeaways

- Conditional independence is a powerful concept connecting probability, algebra, and graph theory.
- We saw how CI statements translate into algebraic conditions (polynomial ideals).

Reference

1. Algebraic Statistics, by Seth Sullivant, Graduate Studies in Mathematics, Vol. 194, American Mathematical Society, Providence, RI, 2018, xiii+409 pp., ISBN 978- 1-4704-3517-2

Thank you