# Conditional independence

Kaie Kubjas, 23.9.2020

# Agenda

- Conditional independence: definition and axioms
- Conditional independence ideals
- Mostly traditional lecture, one or two short worksheet tasks in Breakout rooms
- At the beginning of the next lecture, we will continue with the same topic: primary decompositions of conditional independence ideals

# Setup

- Random vector $X = (X_1, \ldots, X_m)$
- $X$ takes values in a Cartesian product space $\mathcal{X} = \prod_{i=1}^{m} \mathcal{X}_i$
- We assume that either:
    - $X$ has density $f(x) = f(x_1, \ldots, x_n)$ that is continuous on $\mathcal{X}$, or
    - $\mathcal{X}$ is a finite set and then $f(x)$ is the joint distribution $P(X = x)$

# Setup

- Given $A \subseteq [m]$, let
  - $X_A = (X_a)_{a \in A}$
  - $\mathcal{X}_A = \prod_{a \in A} \mathcal{X}_a$
- Given a partition $A_1 \mid \cdots \mid A_k$ of $[m]$, let $f(x_{A_1}, \ldots, x_{A_k})$ denote $f$ with some variables grouped together

# Marginalization

Def: Let $A \subseteq [m]$. The marginal density $f_A(x_A)$ of $X_A$ is obtained by integrating out $x_{[m]\setminus A}$

$$f_A(x_A) := \int_{x_{[m]\setminus A}} f(x_A, x_{[m]\setminus A}) dx_{[m]\setminus A}$$

for all $x_A$.

# Conditioning

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

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

# Independence

Def: We say that $X_1, \ldots, X_n$ are mutually independent or completely independent if

$$P(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{i=1}^{n} P(X_i \in B_i)$$

for all measurable sets $B_1, \ldots, B_n$ in appropriate spaces.

- Similar characterizations for cumulative distribution function and density function.

# Conditional independence

Def: Let $A, B, C \subseteq [m]$ be pairwise disjoint subsets. We say that $X_A$ is conditionally independent of $X_B$ given $X_C$ if and only if

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

for all $x_A, x_B, x_C$.

- The notation $X_A \perp\!\!\!\perp X_B \,|\, X_C$ (or $A \perp\!\!\!\perp B \,|\, C$) denotes that the random vector $X$ satisfies the conditional independence (CI) statement that $X_A$ is conditionally independent of $X_B$ given $X_C$ .

# Conditional independence

- Let $x_B$ and $x_C$ be such that $f_{B|C}(x_B \mid x_C)$ is defined and positive.

- Assume that $X_A \perp\!\!\!\perp X_B \mid X_C$ holds.

- Then

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

- Given $X_C$, knowing $X_B$ does not give any information about $X_A$.

# Marginal independence

- A statement of the form $X_A \perp\!\!\!\perp X_B := X_A \perp\!\!\!\perp X_B \mid X_\varnothing$ is called a marginal independence statement.

- It corresponds to the factorization of densities

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

- This is the same as the independence of random variables.

# Conditional independence axioms

- Suppose a random vector $X$ satisfies a set of conditional independence statements. Which other conditional independence relations must the same random vector satisfy?

- There is no finite set of axioms from which all conditional independence relations can be deduced.

- There are some easy conditional independence implications, which are called the conditional independence axioms or conditional independence rules.

# Conditional independence axioms

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

- (symmetry) $X_A \perp\!\!\!\perp X_B \,|\, X_C \implies X_B \perp\!\!\!\perp X_A \,|\, X_C$
- (decomposition) $X_A \perp\!\!\!\perp X_{B \cup D} \,|\, X_C \implies X_A \perp\!\!\!\perp X_B \,|\, X_C$
- (weak union) $X_A \perp\!\!\!\perp X_{B \cup D} \,|\, X_C \implies X_A \perp\!\!\!\perp X_B \,|\, X_{C \cup D}$
- (contraction) $X_A \perp\!\!\!\perp X_B \,|\, X_{C \cup D}$ and $X_A \perp\!\!\!\perp X_D \,|\, X_C \implies X_A \perp\!\!\!\perp X_{B \cup D} \,|\, X_C$

# Intersection axiom

Prop (Intersection axiom): Suppose that $f(x) > 0$ for all $x$. Then

$$X_A \perp\!\!\!\perp X_B \,|\, X_{C\cup D} \text{ and } X_A \perp\!\!\!\perp X_C \,|\, X_{B\cup D} \implies X_A \perp\!\!\!\perp X_{B\cup C} \,|\, X_D.$$

- The condition $f(x) > 0$ for all $x$ is stronger than necessary.
- For discrete random variables, precise conditions can be given which guarantee that the intersection axiom holds. This is done using algebra!

# Discrete conditional independence models

- A vector of discrete random variables $X_1, \ldots, X_m$
- $X_j$ takes values in $[r_j]$
- $X$ takes values in $\mathscr{R} = \prod_{j=1}^{m} [r_j]$
- For $A \subseteq [m]$, $\mathscr{R}_A = \prod_{a \in A} [r_a]$

# Discrete conditional independence models

Prop: If $X$ is a discrete random vector, then the conditional independence statement $X_A \perp\!\!\!\perp X_B \,|\, X_C$ holds if and only if

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

for all $i_A, j_A \in \mathscr{R}_A$, $i_B, j_B \in \mathscr{R}_B$ and $i_C \in \mathscr{R}_C$.

- The notation $p_{i_A, i_B, i_C, +}$ denotes the probability $P(X_A = i_A, X_B = i_B, X_C = i_C)$ which can be written as

$$p_{i_A, i_B, i_C, +} = \sum_{j_{[m]\setminus A\cup B\cup C} \in \mathscr{R}_{[m]\setminus A\cup B\cup C}} p_{i_A, i_B, i_C, j_{[m]\setminus A\cup B\cup C}}.$$

# Conditional independence ideal

Def: The conditional independence ideal $I_{A \perp\!\!\!\perp B|C}$ is generated by the polynomials

$p_{i_A,i_B,i_C,+} \cdot p_{j_A,j_B,i_C,+} - p_{i_A,j_B,i_C,+} \cdot p_{j_A,i_B,i_C,+}$ for all $i_A, j_A \in \mathscr{R}_A, i_B, j_B \in \mathscr{R}_B$ and $i_C \in \mathscr{R}_C$.

Def: The probability simplex in $\mathbb{R}^{\mathscr{R}}$ is

$$\Delta_{\mathscr{R}} = \left\{ p \in \mathbb{R}^{\mathscr{R}} : \sum_{i \in \mathscr{R}} p_i = 1, p_i \geq 0 \text{ for all } i \right\}.$$

Def: The discrete conditional independence model is $\mathscr{M}_{A \perp\!\!\!\perp B|C} := V_\Delta(I_{A \perp\!\!\!\perp B|C}) \subseteq \Delta_{\mathscr{R}}$.

Example: Let $m = 2$ and consider the ordinary independence statement $X_1 \perp\!\!\!\perp X_2$. Then

$$I_{1 \perp\!\!\!\perp 2} = \langle p_{i_1,j_1} p_{i_2,j_2} - p_{i_1,j_2} p_{i_2,j_1} : i_1, i_2 \in [r_1], j_1, j_2 \in [r_2] \rangle.$$

# Conditional independence ideal

- If $\mathscr{C} = \{X_{A_1} \perp\!\!\!\perp X_{B_1} \,|\, X_{C_1}, X_{A_2} \perp\!\!\!\perp X_{B_2} \,|\, X_{C_2}, \ldots\}$ is a set of conditional independence statements, then the conditional independence ideal is defined as

$$I_{\mathscr{C}} = \sum_{A \perp\!\!\!\perp B | C \in \mathscr{C}} I_{A \perp\!\!\!\perp B | C}.$$

- The model $\mathscr{M}_{\mathscr{C}} := V_{\Delta}(I_{\mathscr{C}}) \subseteq \Delta_{\mathscr{R}}$ consists of all probability distributions that satisfy all the conditional independence statements in $\mathscr{C}$.

# Next time

- Statistics primer
- Pretask:
  - Read Chapters 5.1-5.2 (pages 99-104)
  - Answer the following questions:
    1. What is the difference between probability and statistics?
    2. Can a model be parametric and implicit?
    3. The book uses $X_1, \ldots, X_m$ and $X^{(1)}, \ldots, X^{(n)}$. What is the difference between the two notations?