Conditional Independence

Qing Zhou

UCLA Department of Statistics

Stats 212 Graphical Models
Lecture Notes

Outline

1. Definitions of conditional independence (CI)
2. CI in statistical inference
3. Graphoid
4. CI tests

Definitions of conditional independence

Definition: If X, Y, Z are three random variables, we say $X \perp Y \mid Z$ if $\mathbb{P}(X \in A \mid Y, Z)$ is a function of Z only for any measurable set A.

If they admit a joint density (or mass function) f, then

$$X \perp Y \mid Z \Leftrightarrow f_{XY|Z}(x, y|z) = f_{X|Z}(x|z) f_{Y|Z}(y|z).$$

Other equivalent conditions (f as a generic symbol for densities):

- $f(x, y, z) = f(x, z) f(y, z) / f(z)$.
- $f(x|y, z) = f(x|z)$.
- $f(x, z|y) = f(x|z) f(z|y)$.
- $f(x, y, z) = h(x, z) k(y, z)$ for some h, k.
- $f(x, y, z) = f(x|z) f(y, z)$.

CI in statistical inference

CI in statistical inference (Dawid 1979):

- Sufficient and ancillary statistics: Suppose $X \mid \Theta \sim P_\Theta$.
  1. $T = T(X)$ is a sufficient statistic for Θ if $X \perp \Theta \mid T$.
  2. $S = S(X)$ is an ancillary statistic if $S \perp \Theta$.

  Example: $X = (X_1, \ldots, X_n) \mid \mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2)$. Then
  $T_1 = \sum_i X_i$ is sufficient for μ;
  $T_2 = \sum_i (X_i - \bar{X})^2$ is ancillary for μ.
- Model selection: $Y = X\beta + \varepsilon$. If $\text{supp}(\beta) = S$, then $Y \perp (X \setminus X_S) \mid X_S$.

CI in statistical inference

- Parameter identification: $X \mid \Theta, \Phi \sim P_{(\Theta,\Phi)}$. If $X \perp \Phi \mid \Theta$, then Φ is not identifiable.
  Example: Gaussian linear model $Y = X\beta + \varepsilon$ with X not having full column rank.
  Let $\Theta = X\beta \in \text{col}(X)$ and $\Phi = \beta - X^{-}X\beta$ (X^{-} is a g-inverse of X; $XX^{-}X = X$).
  Then $X\Phi = 0$, i.e. $\Phi \in \text{null}(X)$. Thus $Y \perp \Phi \mid (\Theta, \sigma^2)$, i.e. Φ is not identifiable. Note $\dim(\Theta) + \dim(\Phi) = \dim(\beta)$.

CI in statistical inference

- Statistical modeling: Given $\mathbb{P}(X_1, \ldots, X_p)$ with density f and an ordering $(\sigma(1), \ldots, \sigma(p))$, we factorize f

$$f(x) = \prod_{j=1}^{p} f\big(x_{\sigma(j)} \mid x_{\sigma(1)}, \ldots, x_{\sigma(j-1)}\big) = \prod_{j=1}^{p} f\big(x_{\sigma(j)} \mid x_{A_j}\big), \tag{1}$$

  where $A_j \subset \{\sigma(1), \ldots, \sigma(j-1)\}$ is the minimum subset such that (1) holds:

$$X_{\sigma(j)} \perp X_k \mid X_{A_j}, \qquad k \in \{\sigma(1), \ldots, \sigma(j-1)\} \setminus A_j.$$

- Examples: Markov chains, HMMs, etc.

Graphoid

Graphoid axioms (Pearl (1988), §3.1.2.)

CI statement defines a ternary relation: $\langle X, Y \mid Z \rangle$ for $X \perp Y \mid Z$. Suppose X, Y, Z, W are disjoint subsets of random variables from a joint distribution $\mathbb{P}$. Then the CI relation satisfies

(C1) symmetry: $\langle X, Y \mid Z \rangle \Rightarrow \langle Y, X \mid Z \rangle$;

(C2) decomposition: $\langle X, YW \mid Z \rangle \Rightarrow \langle X, Y \mid Z \rangle$;

(C3) weak union: $\langle X, YW \mid Z \rangle \Rightarrow \langle X, Y \mid ZW \rangle$;

(C4) contraction: $\langle X, Y \mid Z \rangle \& \langle X, W \mid ZY \rangle \Rightarrow \langle X, YW \mid Z \rangle$.

If the joint density of $\mathbb{P}$ wrt a product measure is positive and continuous, then

(C5) intersection: $\langle X, Y \mid ZW \rangle \& \langle X, W \mid ZY \rangle \Rightarrow \langle X, YW \mid Z \rangle$.

In the above, $YW := Y \cup W$.

Graphoid

Any ternary relation $\langle A, B \mid C \rangle$ that satisfies (C1) to (C4) is called a *semi-graphoid*. If (C5) also holds, then it is called a *graphoid*.

Examples of graphoid:

1. Conditional independence of $\mathbb{P}$ (positive and continous).
2. Graph separation in undirected graph: $\langle X, Y \mid Z \rangle$ means nodes Z separate X and Y, i.e. $X - Z - Y$.
3. Partial orthogonality: Let X, Y, Z be disjoint sets of linearly independent vectors in $\mathbb{R}^n$. $\langle X, Y \mid Z \rangle$ means $P_Z^{\perp} X$ is orthogonal to $P_Z^{\perp} Y$. Here $P_Z^{\perp} X = (I_n - P_Z) X$ is the residual after projecting X onto span(Z).

Graph separation provides an intuitive graphical interpretation for the CI axioms.

Graphoid

Example application of CI in causal inference:

- Treatment X, outcome Y. Let I indicates each individual, $I = 1, \ldots, n$. Want to test if $Y \perp X \mid I$ (untestable).
- Suppose $Z = Z(I)$ is a set of sufficient covariates such that $Y \perp I \mid (X, Z)$. Then

$$Y \perp X \mid I \Leftrightarrow Y \perp X \mid Z \text{ (testable based on data)} \quad (2)$$

- Proof outline:
  Note $Y \perp X \mid I \Leftrightarrow Y \perp X \mid (I, Z)$ because $Z = Z(I)$.
  $\Leftarrow$: Sufficient set and RHS of (2) imply $Y \perp (I, X) \mid Z$ by (C4) and thus $Y \perp X \mid (I, Z)$ by (C3).
  $\Rightarrow$: Sufficient set and LHS ($Y \perp X \mid (I, Z)$) imply $Y \perp (X, I) \mid Z$ by (C5) and thus $Y \perp X \mid Z$ by (C2).

CI tests

Conditional independence tests ($H_0 : X \perp Y \mid S$):

- Gaussian data: partial correlation $\text{cor}(X, Y \mid S) = 0$.
  1. Sample covariance matrix $\widehat{\Sigma}$ from data columns of (X, Y, S).
  2. $\widehat{\Omega} = (\omega_{ij}) \leftarrow \widehat{\Sigma}^{-1}$ and $\widehat{\rho}_{XY|S} = -\omega_{12}/\sqrt{\omega_{11}\omega_{22}}$.
  3. Fisher z-transformation,

$$z(X, Y|S) = \frac{1}{2} \log \left(\frac{1 + \widehat{\rho}_{XY|S}}{1 - \widehat{\rho}_{XY|S}} \right)$$

and $\sqrt{n - |S| - 3} \cdot z(X, Y|S) \mid H_0 \sim \mathcal{N}(0, 1)$.

CI tests

Conditional independence tests ($H_0 : X \perp Y \mid S$):

- Discrete data: G^2 or χ^2 test for conditional independence.

$$G^2(X, Y; S = s) = 2 \sum_{x,y} O_{xys} \log(O_{xys}/E_{xys}),$$

$$G^2(X, Y; S) = \sum_{s} G^2(X, Y; S = s) \mid H_0 \sim \chi^2_{(|X|-1)(|Y|-1)|S|},$$

E_{xys}: expected counts under H_0; O_{xys}: observed counts.

References I

A.P. Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 41:1–31, 1979.

Judea Pearl. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann, 1988.