Chapter 14: Hidden Variables

May 26th, 2018

Eliana Duarte

OVGU

- A random variable in a model is hidden or latent if it is not measured.

- A random variable in a model is hidden or latent if it is not measured.

- In a hidden variable model this means that the probability densities on the observed random variables are obtained by computing marginals of the joint distribution of a fully observed model.

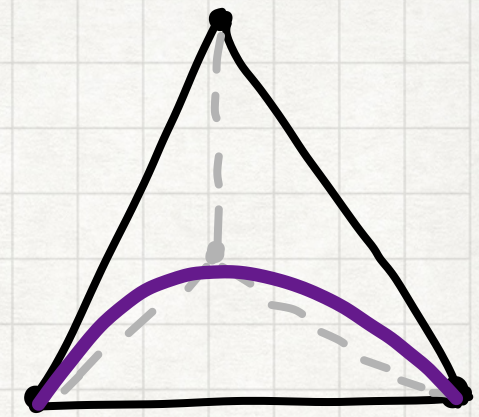
- A random variable in a model is hidden or latent if it is not measured.

- In a hidden variable model this means that the probability densities on the observed random variables are obtained by computing marginals of the joint distribution of a fully observed model.

- Hidden variable models are usually more complex:
  - → semialgebraic description (Example. 14.1.7)
  - → singularities (Proposition 14.1.8)

14.1. MIXTURE MODELS

Example 3.2.9: Binomial Random variable



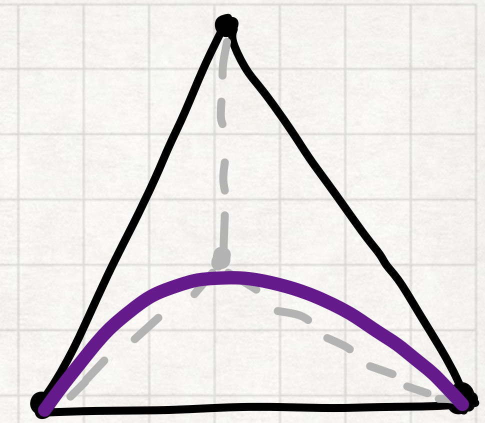
- Consider a biased coin, $P(H) = \theta$, $P(T) = 1 - \theta$

  X = "number of heads in two trials"

  $\mathcal{B}_2 = \{ ((1-\theta)^2, 2(1-\theta)\theta, \theta^2) : 0 \leq \theta \leq 1 \}$ → Parametric Description

  $= \{ (p_0, p_1, p_2) \in \mathbb{R}^3 : p_0 + p_1 + p_2 = 1,\ p_2^2 - 4p_1p_3 = 0,\ p_0, p_1, p_2 \geq 0 \}$ → Implicit Description.

Example 3.2.9: Binomial Random variable



- Consider a biased coin, $P(H)=\theta$, $P(T)=1-\theta$

  $X=$ "number of heads in two trials"

  $\mathcal{B}_2=\{((1-\theta)^2, 2(1-\theta)\theta, \theta^2) : 0\leq\theta\leq 1\}$ → Parametric Description

  $= \{(p_0, p_1, p_2)\in\mathbb{R}^3 : p_0+p_1+p_2=1,\ p_2^2-4p_1p_3=0,\ p_0, p_1, p_2\geq 0\}$ → Implicit Description.

- Suppose we have two coins C_1, C_2, we select one at random, toss the coin twice and record # of heads, Y

$$P(Y=0) = \underbrace{P(C=C_1)}_{\lambda}\underbrace{P(X_{C_1}=0)}_{(1-\theta_1)^2} + \underbrace{P(C=C_2)}_{(1-\lambda)}\underbrace{P(X_{C_2}=0)}_{(1-\theta_2)^2}$$

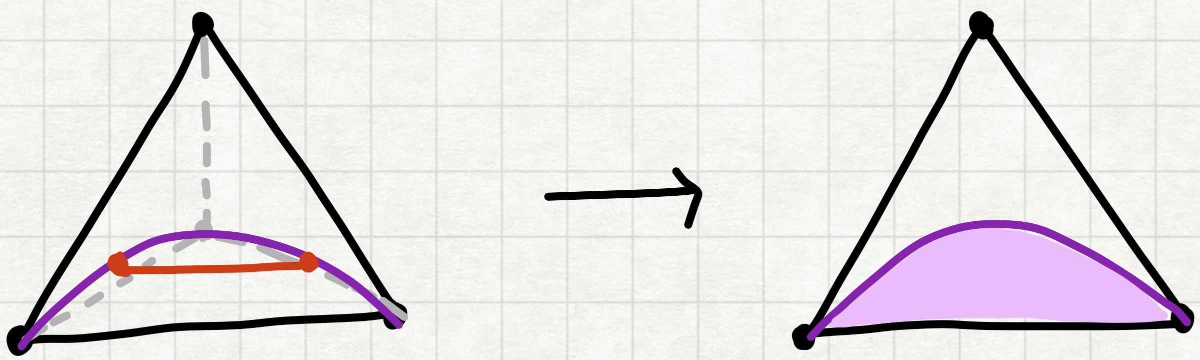
$$= \lambda\cdot(1-\theta_1)^2+(1-\lambda)(1-\theta_2)^2$$

- The distribution of Y is given by

$$\begin{pmatrix} p_0 \\ p_1 \\ p_2 \end{pmatrix} = \lambda \begin{pmatrix} (1-\theta_1)^2 \\ 2\cdot(1-\theta_1)\,\theta_1 \\ \theta_1^2 \end{pmatrix} + (1-\lambda) \begin{pmatrix} (1-\theta_2)^2 \\ 2\cdot(1-\theta_2)\,\theta_2 \\ \theta_2^2 \end{pmatrix}, \qquad \begin{array}{c} 0 \le \lambda \le 1 \\ 0 \le \theta_1 \le 1 \\ 0 \le \theta_2 \le 1 \end{array}$$

$\downarrow$

Parametric Description



$$\Delta_2 \cap \left\{ (p_0, p_1, p_2) : \det \begin{pmatrix} 2p_0 & p_1 \\ p_1 & 2p_2 \end{pmatrix} \geqslant 0 \right\} \rightarrow$$

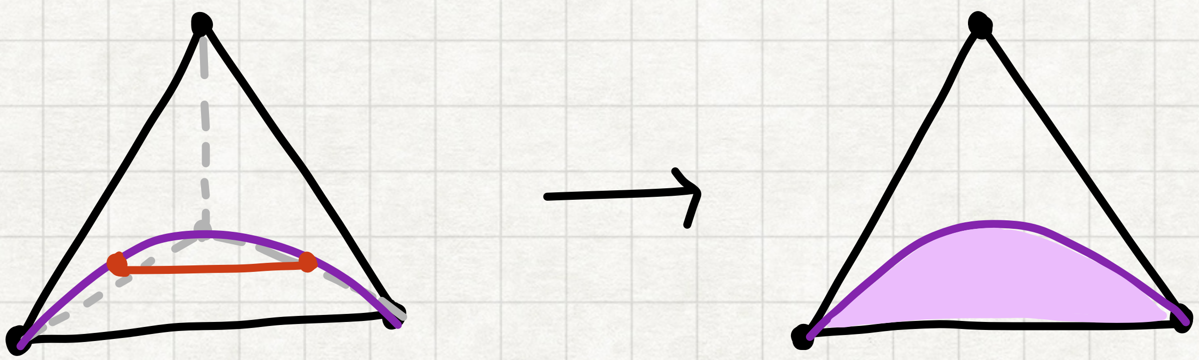
Implicit semialgebraic description

- The distribution of Y is given by

$$\begin{pmatrix} p_0 \\ p_1 \\ p_2 \end{pmatrix} = \lambda \begin{pmatrix} (1-\theta_1)^2 \\ 2\cdot(1-\theta_1)\theta_1 \\ \theta_1^2 \end{pmatrix} + (1-\lambda) \begin{pmatrix} (1-\theta_2)^2 \\ 2\cdot(1-\theta_2)\theta_2 \\ \theta_2^2 \end{pmatrix}, \quad \begin{array}{c} 0 \le \lambda \le 1 \\ 0 \le \theta_1 \le 1 \\ 0 \le \theta_2 \le 1 \end{array}$$

$\downarrow$

Parametric Description



$$\Delta_2 \cap \left\{ (p_0, p_1, p_2) : \det \begin{pmatrix} 2p_0 & p_1 \\ p_1 & 2p_2 \end{pmatrix} \geq 0 \right\} \rightarrow$$

Implicit semialgebraic description

- This model is denoted by $\mathrm{Mixt}^2(\mathcal{B}_2)$ and is the two mixture model of a binomial random variable with two trials.

- Consider a discrete model $\mathcal{M} \subseteq \Delta_{r-1}$, let X be the r.v. modeled by $\mathcal{M}$.

Definition 14.1.1: The K-th mixture model of $\mathcal{M} \subseteq \Delta_{r-1}$ is the family of probability distributions

$$\mathrm{Mixt}^K(\mathcal{M}) = \left\{ \pi_1 P^1 + \cdots + \pi_K P^K : \pi \in \Delta_{K-1}, P^1, \dots, P^K \in \mathcal{M} \right\}$$

- Consider a discrete model $\mathcal{M} \subseteq \Delta_{r-1}$, let X be the r.v. modeled by $\mathcal{M}$.

Definition 14.1.1: The K-th mixture model of $\mathcal{M} \subseteq \Delta_{r-1}$ is the family of probability distributions

$$\mathrm{Mixt}^K(\mathcal{M}) = \{\pi_1 P^1 + \cdots + \pi_K P^K : \pi \in \Delta_{K-1}, P^1, \dots, P^K \in \mathcal{M}\}$$

- Interpretation: → A population is divided into K groups.
  → The i-th group follows a distribution P^i.
  → The hidden variable H is $P(H=i) = \pi_i$

- Consider a discrete model $\mathcal{M} \subseteq \Delta_{r-1}$, let X be the r.v. modeled by $\mathcal{M}$.

Definition 14.1.1: The K-th mixture model of $\mathcal{M} \subseteq \Delta_{r-1}$ is the family of probability distributions

$$\mathrm{Mixt}^K(\mathcal{M}) = \{\pi_1 p^1 + \cdots + \pi_K p^K : \pi \in \Delta_{K-1}, p^1, \dots, p^K \in \mathcal{M}\}$$

- Interpretation: → A population is divided into K groups.
  → The i-th group follows a distribution p^i.
  → The hidden variable H is $P(H=i) = \pi_i$
- In Example 3.2.9. the hidden variable is which coin we choose.
- In a mixture model, we suppose that for some

$\pi = (\pi_1, \dots, \pi_K) \in \Delta_{K-1}$ and $p^1, \dots, p^K \in \mathcal{M}$

$P(H=i) = \pi_i$, $P(X=j \mid H=i) = p_j^i$

$$\Rightarrow P(X=j) = \sum_{i=1}^{K} \pi_i p_j^i$$

Example 14.1.2: Mixture of independence model

Consider two random variables

$\Omega_X = \{\text{Never, sometimes, Frequently}\}$

$X =$ "How much a person watches soccer"

$\Omega_Y = \{\text{Bald, Short, Medium, Long}\}$

$Y =$ "How much hair a person has".

- It is expected that $X \perp\!\!\!\perp Y$, however it is found that people with short hair watch more soccer.
- But within each gender group (men and women) we have $X \perp\!\!\!\perp Y$.
- If we only observe the joint distribution of X and Y, we would observe a distribution in $\mathrm{Mixt}^2(\mathcal{M}_{X \perp\!\!\!\perp Y})$.

14.2. HIDDEN VARIABLE GRAPHICAL MODELS

Example 14.1.3: Independence model in three r.v.

- $\mathcal{M}_{X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3}$, $P(X_1=i_1, X_2=i_2, X_3=i_3) = \underbrace{P(X_1=i_1)}\,\underbrace{P(X_2=i_2)}\,\underbrace{P(X_3=i_3)}$

  $= \alpha_{i_1} \cdot \beta_{i_2} \cdot \gamma_{i_3}$

  $i_1 \in [r_1]$, $i_2 \in [r_2]$, $i_3 \in [r_3]$.

- Then $\mathrm{Mixt}^K(\mathcal{M}_{X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3})$ consists of convex combinations of these probability distributions

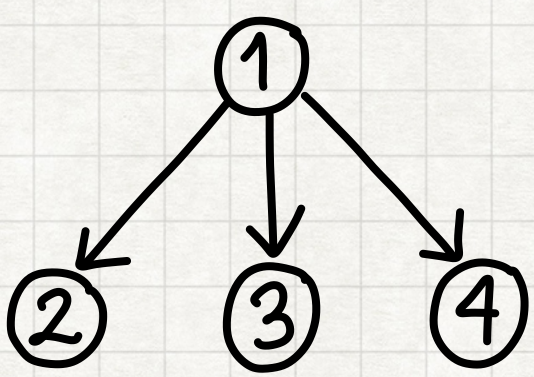
$$p_{i_1 i_2 i_3} = \sum_{h=1}^{K} \pi_h \alpha_{hi_1} \beta_{hi_2} \gamma_{hi_3}, \quad \pi \in \Delta_{K-1}.$$

Example 14.1.3: Independence model in three r.v.

- $\mathcal{M}_{X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3}$, $P(X_1 = i_1, X_2 = i_2, X_3 = i_3) = \underbrace{P(X_1 = i_1)}\underbrace{P(X_2 = i_2)}\underbrace{P(X_3 = i_3)}$
  $= \alpha_{i_1} \cdot \beta_{i_2} \cdot \gamma_{i_3}$
  $i_1 \in [r_1]$, $i_2 \in [r_2]$, $i_3 \in [r_3]$.
- Then $\mathrm{Mixt}^K(\mathcal{M}_{X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3})$ consists of convex combinations of these probability distributions

$$p_{i_1 i_2 i_3} = \sum_{h=1}^{K} \pi_h \alpha_{h i_1} \beta_{h i_2} \gamma_{h i_3} , \quad \pi \in \Delta_{K-1}.$$

Example 14.2.1: Claw Tree



- Let X_1 be a hidden variable.
- $p_{i_2 i_3 i_4} = \sum_{i_1=1}^{r_1} \pi_{i_1} \alpha_{i_1 i_2} \beta_{i_1 i_3} \gamma_{i_1 i_4}$, where

- $\alpha_{i_1 i_2} = P(X_2 = i_2 \mid X_1 = i_1)$, $\beta_{i_1 i_2} = P(X_3 = i_3 \mid X_1 = i_1)$, $\gamma_{i_1 i_4} = P(X_4 = i_4 \mid X_1 = i_1)$
- $X_2 \perp\!\!\!\perp (X_3, X_4) \mid X_1$, $X_3 \perp\!\!\!\perp (X_2, X_4) \mid X_1$, $X_4 \perp\!\!\!\perp (X_2, X_3) \mid X_1$

Hidden variable models under multivariate normal distribution

- $X \sim \mathcal{N}(\mu, \Sigma)$ a random normal vector, $(\mu, \Sigma) \in \mathbb{R}^m \times PD_m$
- For $A \subseteq [m]$, $X_A \sim \mathcal{N}(\mu_A, \Sigma_{A,A})$

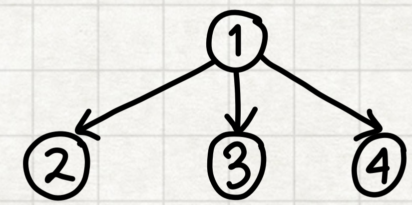
Prop 14.2.2: Let $\mathcal{M} \subseteq \mathbb{R}^m \times PD_m$ be an algebraic exponential family with vanishing ideal $I = I(\mathcal{M}) \subseteq \mathbb{R}[\Sigma]$. Let $H \sqcup O = [m]$ be a partition into hidden variables H and observed variables O. The hidden variable model consists of all marginal distributions on the variables X_O. $I_{hidden} = I \cap \mathbb{R}[\Sigma_{O,O}]$.

Hidden variable models under multivariate normal distribution

- $X \sim \mathcal{N}(\mu, \Sigma)$ a random normal vector, $(\mu,\Sigma) \in \mathbb{R}^m \times PD_m$
- For $A \subseteq [m]$, $X_A \sim \mathcal{N}(\mu_A, \Sigma_{A,A})$

Prop 14.2.2: Let $\mathcal{M} \subseteq \mathbb{R}^m \times PD_m$ be an algebraic exponential family with vanishing ideal $I = I(\mathcal{M}) \subseteq \mathbb{R}[\Sigma]$. Let $H \sqcup O = [m]$ be a partition into hidden variables H and observed variables O. The hidden variable model consists of all marginal distributions on the variables X_O. $I_{\text{hidden}} = I \cap \mathbb{R}[\Sigma_{O,O}]$.

Example 14.2.3: Gaussian Claw tree



$$\Sigma = \begin{array}{cc} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} & \\ \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{pmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \end{array}$$

- In $\mathbb{R}[\Sigma] = \mathbb{R}[\sigma_{11}, \ldots, \sigma_{44}]$

$$I(\mathcal{M}_G) = \langle \sigma_{11}\sigma_{23} - \sigma_{12}\sigma_{13},\ \sigma_{11}\sigma_{24} - \sigma_{12}\sigma_{14},\ \sigma_{13}\sigma_{24} - \sigma_{14}\sigma_{23},\ \sigma_{11}\sigma_{34} - \sigma_{13}\sigma_{14},\ \sigma_{12}\sigma_{34} - \sigma_{14}\sigma_{23} \rangle$$

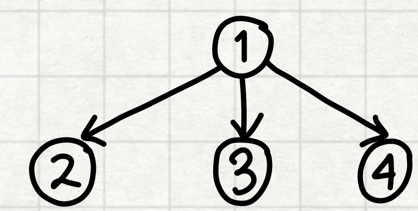
- If X_1 is hidden, the vanishing ideal is $I(\mathcal{M}_G) \cap \mathbb{R}[\sigma_{22}, \sigma_{23}, \ldots, \sigma_{44}] = \langle 0 \rangle$

Hidden variable models under multivariate normal distribution

- $X \sim \mathcal{N}(\mu, \Sigma)$ a random normal vector, $(\mu,\Sigma) \in \mathbb{R}^m \times PD_m$
- For $A \subseteq [m]$, $X_A \sim \mathcal{N}(\mu_A, \Sigma_{A,A})$

Prop 14.2.2: Let $\mathcal{M} \subseteq \mathbb{R}^m \times PD_m$ be an algebraic exponential family with vanishing ideal $I = I(\mathcal{M}) \subseteq \mathbb{R}[\Sigma]$. Let $H \sqcup O = [m]$ be a partition into hidden variables H and observed variables O. The hidden variable model consists of all marginal distributions on the variables X_O. $I_{hidden} = I \cap \mathbb{R}[\Sigma_{O,O}]$.

Example 14.2.3: Gaussian Claw tree



$$\begin{array}{cc} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \Sigma = & \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{pmatrix} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \end{array}$$

- In $\mathbb{R}[\Sigma] = \mathbb{R}[\sigma_{11}, \ldots, \sigma_{44}]$

$$I(\mathcal{M}_G) = \langle \sigma_{11}\sigma_{23} - \sigma_{12}\sigma_{13},\ \sigma_{11}\sigma_{24} - \sigma_{12}\sigma_{14},\ \sigma_{13}\sigma_{24} - \sigma_{14}\sigma_{23},\ \sigma_{11}\sigma_{34} - \sigma_{13}\sigma_{14},\ \sigma_{12}\sigma_{34} - \sigma_{14}\sigma_{23} \rangle$$

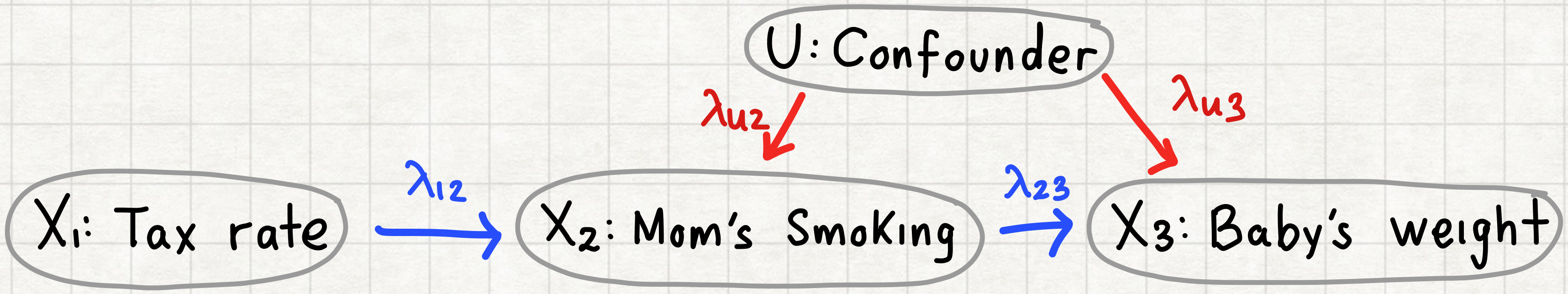
- If X_1 is hidden, the vanishing ideal is $I(\mathcal{M}_G) \cap \mathbb{R}[\sigma_{22}, \sigma_{23}, \ldots, \sigma_{44}] = \langle 0 \rangle$.

☆ The model is characterized by

$$\sigma_{jk}(\sigma_{ii}\sigma_{jk} - \sigma_{ij}\sigma_{ik}) \geq 0$$

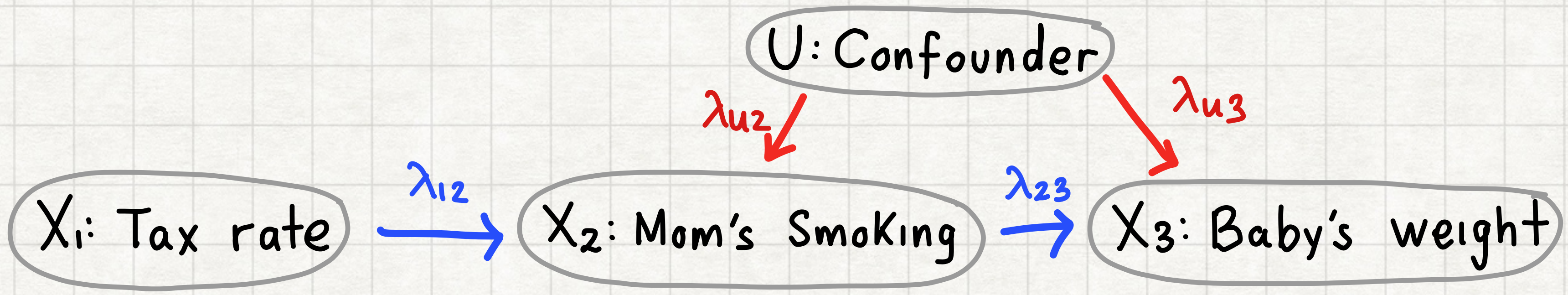
Example 14.2.7: Instrumental variables

Does a mother smoking during pregnancy harm the baby?



Example 14.2.7: Instrumental variables

Does a mother smoking during pregnancy harm the baby?



- The recursive factorization property for Gaussian DAGs translates this into the structural equations

$$
\begin{aligned}
X_1 &= \lambda_{01} + \varepsilon_1 \\
X_2 &= \lambda_{02} + \lambda_{12} X_1 + \lambda_{u2} U + \varepsilon_2 \\
X_3 &= \lambda_{03} + \lambda_{23} X_2 + \lambda_{u3} U + \varepsilon_3 \\
U &= \lambda_{0u} + \varepsilon_4
\end{aligned}
$$

☆ We are interested in the coefficient λ_{23}.

☆ $\mathrm{Cov}(X_1, X_3) = \lambda_{23}\,\mathrm{Cov}(X_1, X_2)$

- $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$ are indep. with zero mean
- The coefficients λ_{ij} are unknown parameters

$$X_1 = \lambda_{01} + \varepsilon_1$$

$$X_2 = \lambda_{02} + \lambda_{12} X_1 + \lambda_{u2} U + \varepsilon_2 \qquad \tilde{\varepsilon}_2 = \lambda_{u2} U + \varepsilon_2$$

$$X_3 = \lambda_{03} + \lambda_{23} X_2 + \lambda_{u3} U + \varepsilon_3 \quad , \qquad \tilde{\varepsilon}_3 = \lambda_{u3} U + \varepsilon_3$$

$$U = \lambda_{0u} + \varepsilon_4$$

$$
\begin{aligned}
X_1 &= \lambda_{01} & & & &+ \varepsilon_1 \\
X_2 &= \lambda_{02} &+ \lambda_{12} X_1 &+ \lambda_{u2} U &&+ \varepsilon_2 \\
X_3 &= \lambda_{03} &+ \lambda_{23} X_2 &+ \lambda_{u3} U &&+ \varepsilon_3 \\
U &= \lambda_{0u} & & & &+ \varepsilon_4
\end{aligned}
\qquad , \qquad
\begin{aligned}
\tilde{\varepsilon}_2 &= \lambda_{u2} U + \varepsilon_2 \\
\tilde{\varepsilon}_3 &= \lambda_{u3} U + \varepsilon_3
\end{aligned}
$$

$$\Downarrow$$

$$
\begin{aligned}
X_1 &= \lambda_{01} & &+ \varepsilon_1 \\
X_2 &= \lambda_{02} &+ \lambda_{12} X_1 &+ \tilde{\varepsilon}_2 \\
X_3 &= \lambda_{03} &+ \lambda_{23} X_2 &+ \tilde{\varepsilon}_3
\end{aligned}
\qquad ,
$$

$$
\begin{aligned}
\omega_{23} &:= \mathrm{Cov}[\tilde{\varepsilon}_2, \tilde{\varepsilon}_3] \\
&= \mathrm{Cov}[\lambda_{u2} U + \varepsilon_2, \lambda_{u3} U + \varepsilon_3] \\
&= \lambda_{u2} \lambda_{u3} \mathrm{Var}[U] \neq 0 .
\end{aligned}
$$

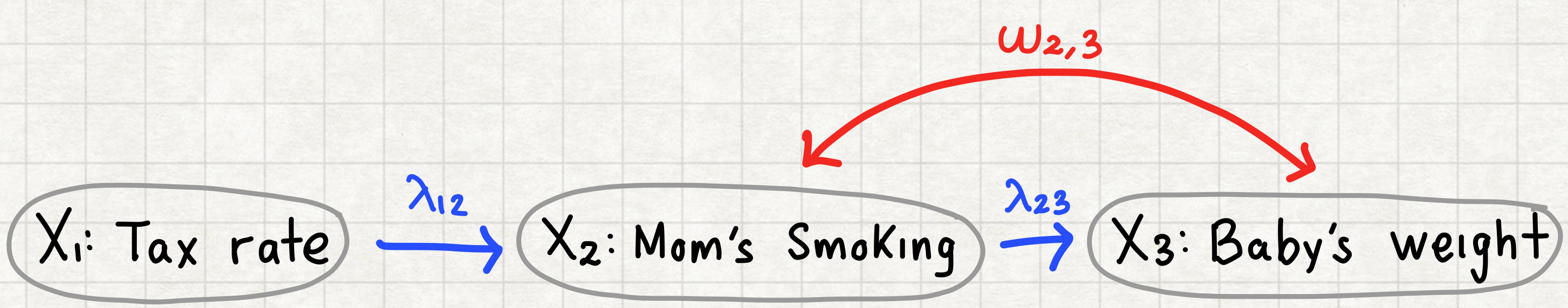
$$
\begin{aligned}
X_1 &= \lambda_{01} && + \varepsilon_1 \\
X_2 &= \lambda_{02} + \lambda_{12} X_1 + \lambda_{u2} U && + \varepsilon_2 \\
X_3 &= \lambda_{03} + \lambda_{23} X_2 + \lambda_{u3} U && + \varepsilon_3 \\
U &= \lambda_{0u} && + \varepsilon_4
\end{aligned}
\quad , \quad
\begin{aligned}
\tilde{\varepsilon}_2 &= \lambda_{u2} U + \varepsilon_2 \\
\tilde{\varepsilon}_3 &= \lambda_{u3} U + \varepsilon_3
\end{aligned}
$$

$$\Downarrow$$

$$
\begin{aligned}
X_1 &= \lambda_{01} && + \varepsilon_1 \\
X_2 &= \lambda_{02} + \lambda_{12} X_1 && + \tilde{\varepsilon}_2 \\
X_3 &= \lambda_{03} + \lambda_{23} X_2 && + \tilde{\varepsilon}_3
\end{aligned}
\quad ,
$$

$$
\begin{aligned}
\omega_{23} &:= \mathrm{Cov}[\tilde{\varepsilon}_2, \tilde{\varepsilon}_3] \\
&= \mathrm{Cov}[\lambda_{u2} U + \varepsilon_2, \lambda_{u3} U + \varepsilon_3] \\
&= \lambda_{u2} \lambda_{u3} \mathrm{Var}[U] \neq 0 .
\end{aligned}
$$

☆ The error terms are now correlated



Mixed graph representing an instrumental variable model.

LINEAR STRUCTURAL EQUATION MODELS

- Let $G=(V, B, D)$, $V=\{\text{vertices}\}$, $B=\{\text{bidirected edges}\}$, $D=\{\text{directed edges}\}$.

  G is called a mixed graph.

LINEAR STRUCTURAL EQUATION MODELS

- Let $G=(V,B,D)$, $V=\{\text{vertices}\}$, $B=\{\text{bidirected edges}\}$, $D=\{\text{directed edges}\}$.

  G is called a mixed graph.

- $PD(B)=\{\Omega\in PD_m : \omega_{ij}=0 \text{ if } i\neq j \text{ and } i\leftrightarrow j\notin B\}$
- $\mathbb{R}^D=\{\Lambda\in\mathbb{R}^{m\times m} : \Lambda_{ij}=\lambda_{ij} \text{ if } i\to j\in D \text{ and } \Lambda_{ij}=0 \text{ otherwise}\}$
- Let $\epsilon\sim\mathcal{N}(0,\Omega)$ where $\Omega\in PD(B)$.
- For $j\in V$ define the random variables X_j,

$$X_j=\sum_{k\in pa(j)}\lambda_{kj}X_k+\epsilon_j$$

LINEAR STRUCTURAL EQUATION MODELS

- Let $G=(V,B,D)$, $V=\{\text{vertices}\}$, $B=\{\text{bidirected edges}\}$, $D=\{\text{directed edges}\}$.

  G is called a mixed graph.

- $PD(B)=\{\Omega\in PD_m : \omega_{ij}=0 \text{ if } i\neq j \text{ and } i\leftrightarrow j\notin B\}$
- $\mathbb{R}^D=\{\Lambda\in\mathbb{R}^{m\times m} : \Lambda_{ij}=\lambda_{ij} \text{ if } i\to j\in D \text{ and } \Lambda_{ij}=0 \text{ otherwise}\}$
- Let $\epsilon\sim\mathcal{N}(0,\Omega)$ where $\Omega\in PD(B)$.
- For $j\in V$ define the random variables X_j,

$$X_j=\sum_{k\in pa(j)}\lambda_{kj}X_k+\epsilon_j$$

Prop 14.2.8: Let $G=(V,B,D)$ be a mixed graph. Let $\Omega\in PD(B)$, and $\epsilon\sim\mathcal{N}(0,\Omega)$ and $\Lambda\in\mathbb{R}^D$. Then the random vector X is a multivariate normal r.v. with covariance matrix $\Sigma=(I_d-\Lambda)^{-T}\Omega(I_d-\Lambda)^{-1}$

- $\mathcal{M}_G=\{(I_d-\Lambda)^{-T}\Omega(I_d-\Lambda)^{-1} : \Omega\in PD(B),\ \Lambda\in\mathbb{R}^D\}$

Thank you.