

### 3.5 Interventional

We can modify an SEM by setting e.g.  $Z_k = a$ . This is called a perfect intervention. The new SEM determines a new distribution for  $Z$ .

Expectations on probabilities under the new distribution are written by adding " $do(Z_k = a)$ " e.g.  $E(Z_j | do(Z_k = a))$ . Not necessarily true that  $E(Z_j | do(Z_k = a)) = E(Z_j | Z_k = a)$ .

Example cont After intervention  $do(Z_2 = 1)$  (everyone is forced to go to the catch up lectures), we have a new SEM:

$$Z_3 = \varepsilon_3 \sim \text{Bern}(\frac{1}{4}) \quad Z_2 = 1$$

$$Z_1 = \mathbb{1}_{\{\varepsilon_1(1+Z_3) > \frac{1}{2}\}} \quad \varepsilon_1 \sim U[0, 1]$$

$$\text{Thus } P(Z_1 = 1 | do(Z_2 = 1)) = \frac{3}{4} \times \frac{1}{4} + \frac{1}{2} \times \frac{3}{4} = \frac{9}{16}$$

$$\begin{aligned} P(Z_1 = 1 | Z_2 = 1) &= \sum_{j \in \{0, 1\}} P(Z_1 = 1 | Z_2 = 1, Z_3 = j) P(Z_3 = j | Z_2 = 1) \\ &= \sum_{j \in \{0, 1\}} P(Z_1 = 1 | Z_2 = 1, Z_3 = j) \frac{P(Z_2 = 1 | Z_3 = j) P(Z_3 = j)}{P(Z_2 = 1)} \\ &= \frac{16}{9} \left( \frac{3}{4} \times \frac{3}{4} \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{2} \times \frac{3}{4} \right) = \frac{7}{12} \neq \frac{9}{16} \end{aligned}$$

### 3.6 Markov properties on DAGs

Let  $P$  be the distribution of  $Z$  and suppose it has density  $f$ .

Def Given a DAG  $G$ , say  $P$  satisfies the

(i) the Markov factorisation property w.r.t.  $G$  if

$$f(Z_1, \dots, Z_p) = \prod_{k=1}^p f(Z_k | Z_{\text{pa}(k)})$$

(ii) the global Markov property w.r.t.  $G$  if  $\forall$  disjoint  $A, B, S \subseteq \{1, \dots, p\}$  ( $AB \neq \emptyset$ )

$A, B$  d-separated by  $S \Rightarrow Z_A \perp\!\!\!\perp Z_B | Z_S$  (if  $S = \emptyset$  then  $Z_A \perp\!\!\!\perp Z_B$ )

Theorem 3.2 If  $P$  has a density  $f$  w.r.t. a product measure then (i) & (ii) above are equivalent.

We will often use "Markov" to mean "global Markov".

Prop 33 Let  $P$  be the distribution given by an SEM with DAG  $G$ . Then  $P$  satisfies the Markov factorization property w.r.t.  $G$ .

Proof: Let  $\pi$  be a topological ordering of  $G$  and write  $\tau = \pi^{-1}$ . We can write

$$f(z_1, \dots, z_p) = f(z_{\tau(p)}) f(z_{\tau(p-1)} | z_{\tau(p-2)}, \dots, z_{\tau(1)}) \dots f(z_{\tau(2)} | z_{\tau(1)})$$

Now  $pa(\tau(k)) \subseteq \{\tau(1), \dots, \tau(k-1)\}$ . Since  $z_{\tau(k)}$  is a function of  $z_{pa(\tau(k))}$  and independent  $\varepsilon_{\tau(k)}$ , we know that  $f(z_{\tau(k)} | z_{\tau(k-1)}, \dots, z_{\tau(1)}) = f(z_{\tau(k)} | z_{pa(\tau(k))})$ .  
Now substitute into eqn above.  $\square$

### 3.7 Causal structure learning

Given  $P$ , how can we find the DAG of the SEM that generated it?

#### 3.7.1 Three obstacles

##### Causal minimality

If  $P$  is generated by an SEM with DAG  $G$ , then  $P$  is Markov w.r.t.  $G$ .

Conversely, if  $P$  is Markov w.r.t. DAG  $G$ , then there is an SEM with DAG  $G$  that could have generated  $P$ .

Indeed take

$$Z_j = h_j(Z_{pa(j)}, \varepsilon_j) \text{ where } h_j(\tilde{Z}_{pa(j)}, e) = F_{Z_j | Z_{pa(j)}}^{-1}(\tilde{Z}_{pa(j)})(e)$$

$$\varepsilon_1, \dots, \varepsilon_p \sim U[0, 1]$$

$F_{Z_j | Z_{pa(j)}} = \tilde{Z}_{pa(j)}$  is the distribution  $f^u$  of  $Z_j | Z_{pa(j)} = \tilde{Z}_{pa(j)}$ .

But  $P$  will be Markov w.r.t. many DAGs (e.g. any DAG whose skeleton is the complete graph).

So e.g.  $Z_1 \perp\!\!\!\perp Z_2$  can be expressed

$$Z_1 = 0 \times Z_2 + \varepsilon_1, \quad Z_2 = \varepsilon_2$$

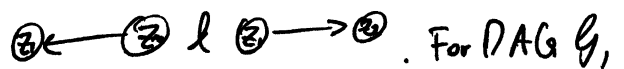


Def<sup>n</sup>  $P$  satisfies causal minimality w.r.t. DAG  $G$  if it is Markov w.r.t.  $G$  but not Markov w.r.t. any proper subgraph of  $G$ .

Markov equivalent DAGs

Two different DAGs may satisfy the same collection of d-separations, e.g.

let  $\mathcal{M}(G) = \{ \text{distributions } P : P \text{ is Markov w.r.t. } G \}$ .



Def<sup>n</sup> Two DAGs are Markov equivalent if  $\mathcal{M}(G_1) = \mathcal{M}(G_2)$ .