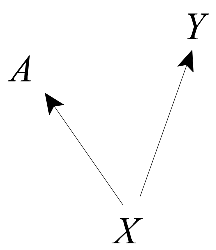
Graphical Models

Graphical displays commonly used to diagram relations among variables

example:



Describe assumptions encoded in diagram

Actually, graphical models provide a rigorous approach

- for describing probabilistic dependencies and independence relationships in nonparametric fashion

- for encoding assumptions, associations about causal relationships among variables

may also be used in deriving nonparametric estimators of causal effects

Will consider formal notation, analysis of graphs graph consists of

- set V of *vertices* or *nodes*
- set E of *edges* or *links* that connect nodes

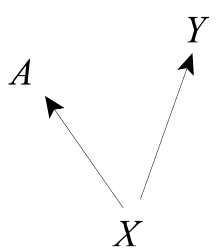
variables connected by an edge are *adjacent*

edges may be

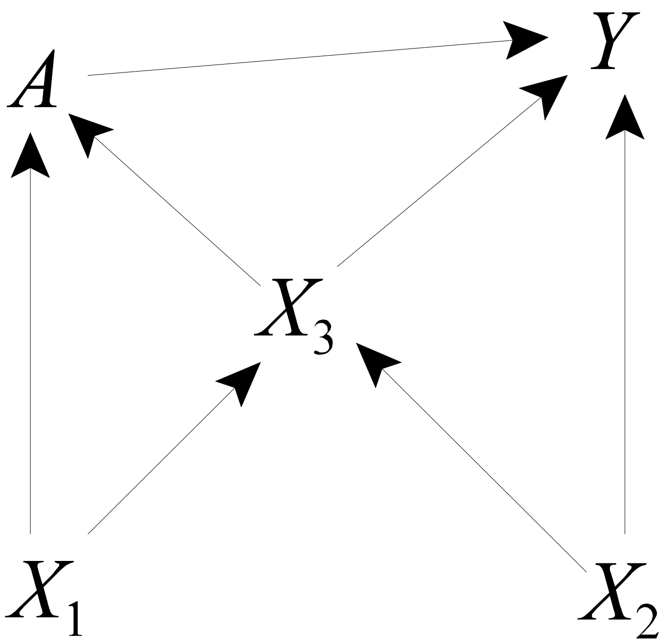
- directed (all edges are marked by a single arrowhead)
- undirected (unmarked links)
- bidirected (arrowheads at both ends; denotes unmeasured common causes)

path: any sequence of edges such that each edge starts with the vertex ending at the previous edge

find paths on graph below



more complicated graph:



A: antihistamine treatment

Y: asthma

X_1: air pollution

X_2: sex

X_3: bronchial reactivity

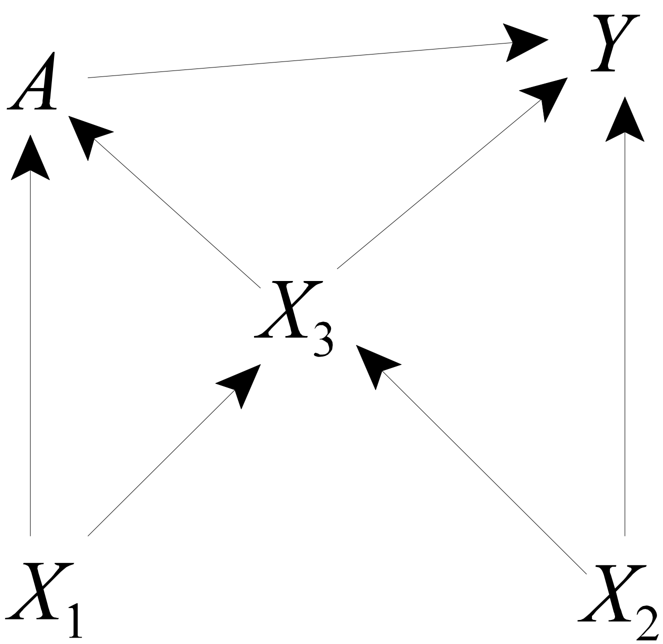
find paths from A to Y

directed graphs: all arrows directed

may include directed cycles (e.g., $A \rightarrow Y$, $Y \rightarrow A$) but not self loops

graph with no cycles: *acyclic*

use terminology of *kinship*



parents

children

ancestors

descendants

family: node and all parents

find each for A

node is

- *root* if no parents
- *sink* if no children

every DAG has at least one root, one sink

Bayesian network:

will try to provide probabilistic (and possibly causal) meaning to graph

Graphs afford way of decomposing joint distribution of variables

suppose distribution P on n variables, ordered arbitrarily $X_1, X_2, \ldots, X_n$

can write $pr(X) = pr(X_1, \ldots, X_n) = \prod_j pr(X_j | X_1, \ldots, X_{j-1})$

under assumptions of graph, variable is sensitive only to parents

$pr(X_j | X_1, \ldots, X_{j-1}) = pr(X_j | pa_j)$

PA_j is minimal set of predecessors that renders X_j independent of all other predecessors

If a probability function admits the factorization above relative to a DAG *G*, we say that

- *G* represents *P*
- *G* and *P* are compatible
- *P* is Markov relative to *G*

multiple DAGs or orderings may be compatible with probability distribution *P*

compatibility necessary and sufficient for a DAG *G* to explain empirical data represented by *P*

If generate data at random by recursively generating from probabilities $pr(X_i|PA_i)$, then *P* will be Markov relative to *G*

If *P* Markov relative to *G*, there exists a set of probabilities $pr(X_i|PA_i)$ according to which we can choose the value of value of X_i such that distribution of generated realizations will equal *P*

D-separation

allows one to read all conditional independencies off graph

consider three disjoint sets of nodes/variables: X, Y, and Z

want to test whether $X \perp Y | Z$

need to test whether nodes in Z block all paths from nodes in X to nodes in Y

A path p is d-separated or blocked by a set of nodes Z iff

1. p contains a chain $i \rightarrow m \rightarrow j$ or a fork $i \leftarrow m \rightarrow j$ such that the middle node is in Z

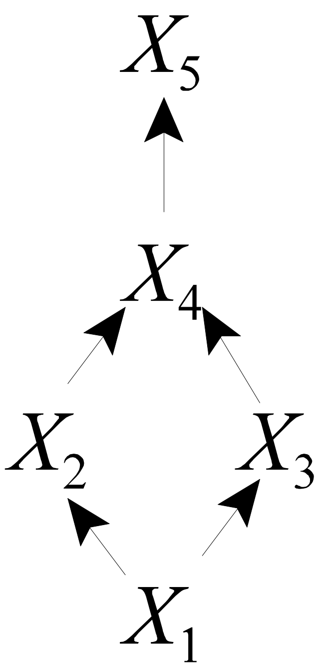
or

2. p contains an inverted fork or collider $i \rightarrow m \leftarrow j$ such that the middle node m is not in Z and that no descendant of m is in Z

A set Z is said to d-separate X from Y iff Z blocks every path from a node in X to a node in Y

Easiest to obtain intuition for d-separation if attribute causal interpretation to graph

Consider following graph first:



X_1: season

X_2: rain

X_3: sprinkler

X_4: wet

X_5: slippery

assumptions characteristic of affluent areas in Southern California

generally true for graphs; assumptions encoded by graph must correspond to subject-matter assumptions

are X_2, X_3 d-separated given

1. Φ?
2. X_1?
3. X_1, X_4?
4. X_5?

1. No

In Southern California, sprinklers may operate daily in summer (dry, hot); not winter (cool, wet); thus, rain, sprinkler associated

2. Yes

Once one knows season, sprinklers operate automatically, unrelated to weather (first approximation)

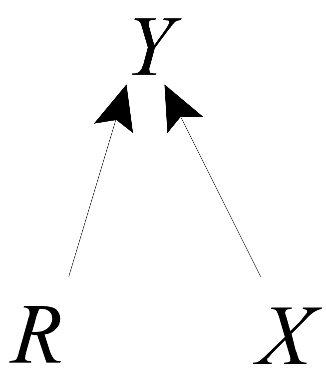
3. No

Conditional on season (e.g., winter) and knowing it is wet, knowing that it has not been raining makes it more likely that a sprinkler has been on

4. No

As in 3, conditioning on slipperiness is like conditioning on wetness; this unblocks path $X_2 \rightarrow X_4 \leftarrow X_3$

Another example:



R-smoking
X-genetics
Y-lung CA

Smoking independent of genetic factors

Both cause lung cancer (how can one say this about genetic factor?)

Numerical example:

Smoking	Lung Cancer gene	N	Lung Cancer
No	No	1000	100
	Yes	1000	500
Yes	No	1000	500
	Yes	1000	900

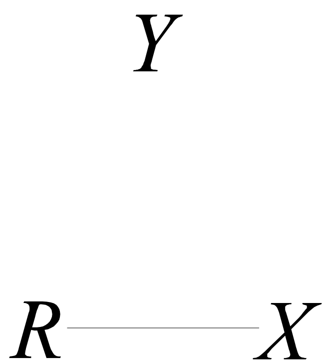
Smoking and lung cancer gene not associated

Conditional on lung cancer ($Y=1$):

	Lung cancer gene	
Smoking	No	Yes
No	100	500
Yes	500	900

If you know someone has lung cancer, knowing that he smokes reduces the probability that he also has the lung cancer gene (from 5/6 to 9/14), since one or the other factor (or both) is responsible for most cases of cancer.

Sometimes can show associations induced by controlling for variable that is a consequence by undirected arc: e.g.,



R-smoking
X-genetics
Y-lung CA

probabilistic implications of d-separation

If sets X and Y are d-separated by Z in a DAG G, then X is independent of Y conditional on Z in every distribution compatible with G

conversely, if X, Y not d-separated by Z, then X and Y are dependent conditional on Z in at least one distribution compatible with G

equivalently,

$(X \perp Y | Z)_G \Rightarrow (X \perp Y | Z)_P$ Whenever G and P compatible

if $(X \perp Y | Z)_P$ holds in all distributions compatible with G, it follows that $(X \perp Y | Z)_G$

Note that it is the direction of arrows in a DAG as well as the absence of arrows which conveys assumptions

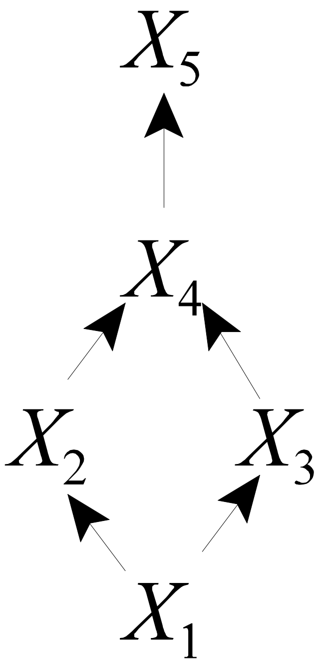
Parental Markov condition

A necessary and sufficient condition for a probability distribution P to be compatible with (Markov relative to) a DAG G is that every variable be independent of all its nondescendants (in G), conditional on its parents

Observational equivalence

2 DAGs are observationally equivalent iff they have the same skeletons and the same sets of v-structures (that is, 2 converging arrows whose tails are not connected by an arrow).

2 variables: $x_1 \to x_2$ and $x_2 \to x_1$ equivalent (i.e., can't deduce causal ordering from association)



X_1: season

X_2: rain

X_3: sprinkler

X_4: wet

X_5: slippery

which of the following will change v-structure?

reverse $x_1 \rightarrow x_2$:

reverse $x_2 \rightarrow x_4$:

reverse $x_1 \rightarrow x_2$: does not change v-structure

reverse $x_2 \rightarrow x_4$: changes v-structure

so far, little discussion of causality with respect to graphs

causal networks:

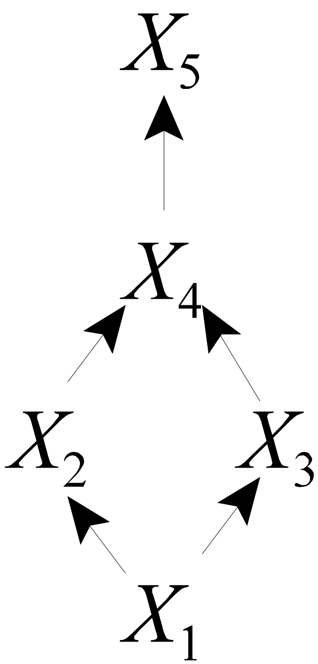
can build graphs without causality to represent conditional independence assumptions

why try to assign causal meaning?

1. Judgments required more meaningful, accessible

   thus, more reliable

Consider following graph:



X_1: season

X_2: rain

X_3: sprinkler

X_4: wet

X_5: slippery

can decompose joint distribution of observables in any order

why model $pr(X_5|X_4,\cdot)$ rather than the reverse?

Intuitively easier to decompose in order of what influences what

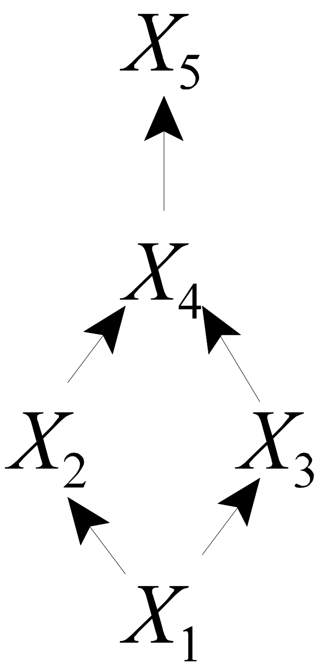
judgments anchored to fundamental building blocks of our knowledge (wetness influences slipperiness, not vice versa)

thus more natural

2. Easy to see how system would respond to changes

ask what happens when there is an intervention what would happen to graph if disabled sprinkler (describe new graph)

old graph



X_1: season

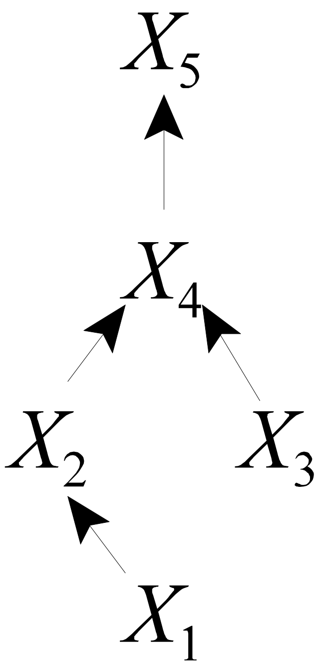
X_2: rain

X_3: sprinkler

X_4: wet

X_5: slippery

other mechanisms, probability judgments would stay in place



X_1: season

X_2: rain

X_3: sprinkler

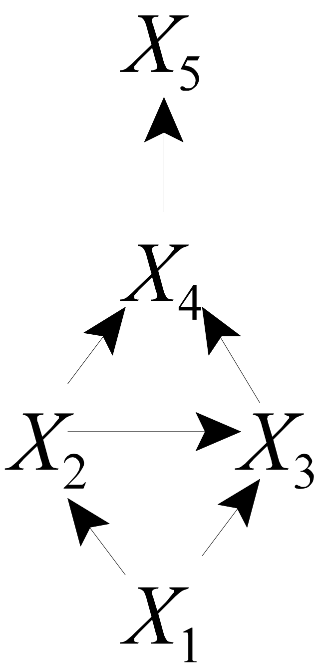
X_4: wet

X_5: slippery

assumption: system is organized modularly

system will react to intervention purely through local changes, propagated through the system through the various pathways still intact

what if changed system so turned off sprinkler when it rains?



X_1: season

X_2: rain

X_3: sprinkler

X_4: wet

X_5: slippery

Assumption: system is stable

Modular organization of knowledge

assume change is local

allows us to predict effect of intervention

Definitions:

Let $P(v)$ be a probability distribution on a set V of variables
Let $P_x(v)$ denote the distribution resulting from intervention $do(X=x)$ that sets a subset X of variables to constants.
Let $\boldsymbol{P}_*$ denote the set of all interventional distributions $P_x(v), X \subseteq V$. $\boldsymbol{P}_*$ Includes $P(v)$, which denotes no intervention.

A DAG is said to be a causal Bayesian network compatible with $\boldsymbol{P}_*$ iff the following condition hold for every $p_x \in \boldsymbol{P}_*$:

1. $P_x(v)$ is Markov relative to G
2. $P_x(v_i|pa_i)=P(v_i|pa_i)$ for all $V_i \in X$ whenever v_i consistent with $X=x$
3. $P_x(v_i|pa_i)=P(v_i|pa_i) \forall V_i \notin X$ Whenever pa_i consistent with $X=x$

A causal Bayesian network allows us to compute the distribution distribution $P_x(v)$ resulting from any intervention $do(X=x)$ as a truncated factorization:

$P_x(v) = \prod_{i|V_i \notin X} P(v_i|pa_i)$ for all v consistent with x

2 properties

1. For all i, $P(v_i|pa_i) = P_{pa_i}(v_i)$; the conditional probability resulting from

   setting a variable's parents is the same as observing the same value of the variable's parents

   similar to ignorability conditions for a single variable

2. For all i and every subset of variables disjoint of $\{V_i, PA_i\}$, we have

   $P_{pa_i,s}(v_i) = P_{pa_i}(v_i)$

If set a variable's parents and other variables, only setting variable's parents has effect

Notes:

implicitly, all variables on graph seem subject to intervention

can use graph, along with rules, to compute marginal distributions of what would happen under intervention; subject of chapter 3 in book, *Biometrika* paper

Truncated factorization: $P_x(v) = \prod_{i|V_i \notin X} P(v_i|pa_i)$ can also be written

$$P_x(v) = \prod_{i|V_i \notin X} P(v_i|pa_i) = \prod_i P(v_i|pa_i) / \prod_{i|V_i \in X} P(v_i|pa_i) = P(v) / \prod_{i|V_i \in X} P(v_i|pa_i)$$

what does this look like?

Inverse probability of treatment weighted estimates

Conditions for confounding and choice of sufficient sets of confounders:

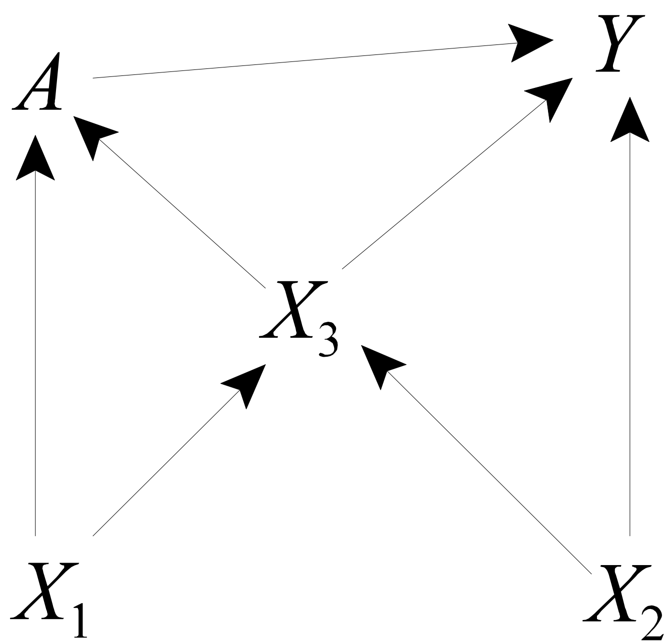
For intervention A, outcome Y, other variables X

remove all arrows pointing out of A:

call resulting graph $G_{\underline{A}}$

can control for confounding using usual standardization formula if

$(A \perp Y | X)_{G_{\underline{A}}}$



A: antihistamine treatment

Y: asthma

X_1: air pollution

X_2: sex

X_3: bronchial reactivity

Is X_3 sufficient to control confounding?

No. Although X_3 blocks several paths from A to Y, it does not block (actually unblocks) the path $A \leftarrow X_1 \rightarrow X_3 \leftarrow X_2 \rightarrow Y$

What are the sets of variables sufficient to control confounding?

$$\{X_1, X_3\}, \{X_2, X_3\}, \{X_1, X_2, X_3\}$$

Bottom line: definition of confounders is not unique

Variables not affected by exposure

Useful as method for describing associations among variables, causal assumptions

Assumptions must be considered critically

Less useful in deriving practical statistical methods for data analysis

Alternative view: related to structural equations models

Each node in graph can be described by functional relationship

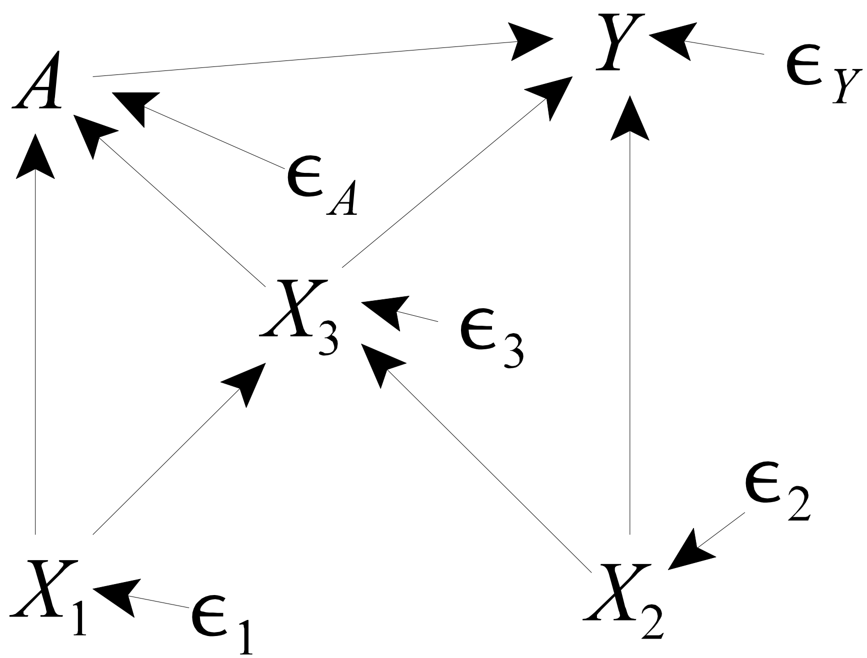
$X_i = f\{pa_i, \epsilon_i)$

ϵ_i: error term, unknown

often assumed independent of other error terms

corresponds to graph with no unmeasured common causes like adding additional nodes to graphs

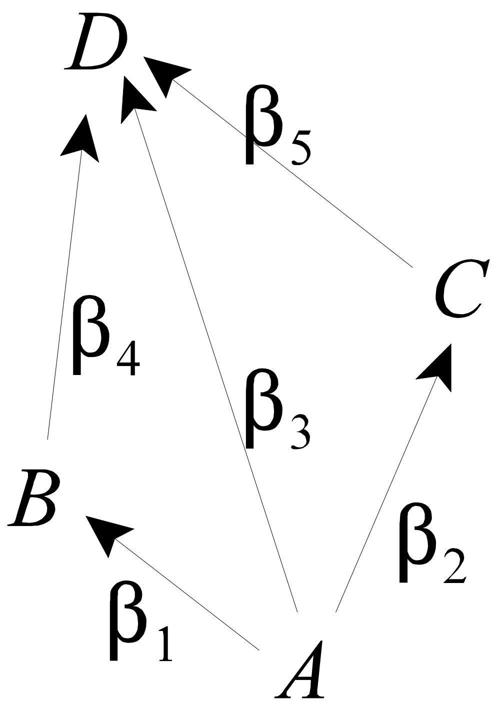
augmented graph



each variable is a deterministic function of its inputs

In path diagrams, typically assume that all variables are normally distributed, linear model for effect of each variable on others

$$X_i = f\{pa_i, \epsilon_i) = PA_i \beta + \epsilon_i$$



have regressions of each variable on parents:

$$A = \beta_{0A} + \epsilon_A$$
$$B = \beta_{0B} + A\beta_1 + \epsilon_B$$
$$C = \beta_{0C} + A\beta_2 + \epsilon_C$$
$$D = \beta_{0D} + A\beta_3 + B\beta_4 + C\beta_5 + \epsilon_D$$

Compute effect of *A* on D

In path diagram with all exogenous errors, can compute pathwise effects, total effect as sum of pathwise effects

$$
\begin{aligned}
B &= \beta_{0B} + \hat{A}\beta_1 + \epsilon_B \\
C &= \beta_{0C} + \hat{A}\beta_2 + \epsilon_C \\
D &= \beta_{0D} + \hat{A}\beta_3 + B\beta_4 + C\beta_5 + \epsilon_D \\
&= \beta_{0D} + \hat{A}\beta_3 + (\beta_{0B} + \hat{A}\beta_1 + \epsilon_B)\beta_4 + (\beta_{0C} + \hat{A}\beta_2 + \epsilon_C)\beta_5 + \epsilon_D \\
&= (\beta_{0D} + \beta_{0B}\beta_4 + \beta_{0C}\beta_5) + \hat{A}(\beta_3 + \beta_1\beta_4 + \beta_2\beta_5) + (\epsilon_B\beta_4 + \epsilon_C\beta_5 + \epsilon_D) \\
&= \beta_{0D}^{*} + \hat{A}\beta_A^{*} + \epsilon_D^{*}
\end{aligned}
$$

effect of A on D: $\beta_A^{*} \equiv \beta_3 + \beta_1\beta_4 + \beta_2\beta_5$

usefulness of formula depends on assumptions of uncorrelatedness of errors

what is direct effect of A on D? Indirect effect?

Direct effect: β_3

Indirect effect: $\beta_1\beta_4+\beta_2\beta_5$

Will discuss direct/indirect effects at more length later