Homework 4 - Maxime Tchibozo (UNI: MT3390)

This assignment required us to choose 2 exercises among the 4. I chose Ex2.6 and Ex3.4

2.6 Exercises

1. Compare instrumental variables with conditioning-basd inference. In particular,

(a) Does instrumental variables help mitigate the curse of dimensionality? How?

Conditioning-based inference is centered around the Robins G-Formula :
$P(Y \mid do(X = x)) = \sum_Z P(Y \mid X, Z)P(Z)$, which requires $P(X, Z) > 0$, meaning we have to observe $X = x$ in each subpopulation of Z. In high dimensions, this is unlikely to be verified, and extrapolation methods/coarsening introduce bias. This is referred to as the "Curse of Dimensionality".

Instrumental variables circumvent the curse of dimensionality when estimating the effect of X on Y by conditioning on an instrument which causes X. This instrument can be a simple binary variable so all of its outcomes are much more likely to be observed. In experiments, we can even set them ourselves.

However, the curse of dimensionality could *subtly* still present in instrumental variables, when verifying the Exclusion Restriction principle: $(Z \perp Y)_{G_{\bar{X}}}$. To verify this assumption, we would have to check that there is no statistical correlation between Z and all possible variables having an effect on Y. This assumption can not be verified if some of these variables are unobserved, so in practice we assume the assumption is true, and avoid the curse of dimensionality altogether.

(b) Causal interpretations come from structural assumptions. What structural assumptions do you make with instrumental variables?

The main assumptions are that

●$Z \rightarrow X$ (this can be controlled in an experiment).

●$(Z \perp Y)_{G_{\bar{X}}}$ (Exclusion restriction).

(c) How much freedom do you have to use different functional forms for the relationship between the causal state and the outcome?

When applying two stage least squares, we only need to measure Z, X and Y so we have a large amount of freedom when modeling the relationship between causal state and outcome.

For other types of IV estimation, for example when using covariance to estimate a treatment effect, we have less freedom when modeling the causal graph. The independence/dependence assumptions between variables from the graph must be empirically verified.

(d) How general is the IV estimator, compared with g-formula based estimators ?

The IV estimator is more general in that it can be applied more easily using only causal state, outome variable, and instrumental variable, whereas g-formula based estimators require us to usually condition on more variables (enough to complete our back-door or front-door sets). However, the Exclusion restriction assumption is a strong assumption, and is harder to verify than backdoor set assumptions.

1. Given the following data generating process,

$Y \sim DY^{(D=1)} + (1 - D)Y^{(D=0)}$

$D \sim ZD^{(Z=1)} + (1 - Z)D^{(Z=0)}$

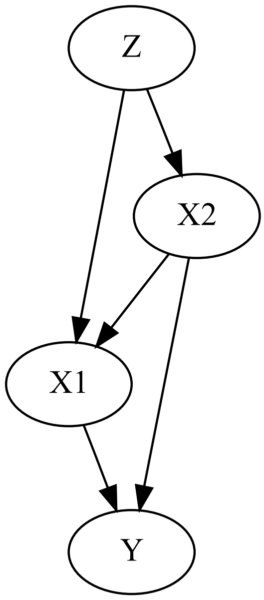
$D^{(Z=1)} \sim max(D^{(Z=0)}, \epsilon_D)$

$D^{(Z=0)} \sim Bin(1, p = logit^{-1}(U))$

$U \sim \mathbf{N}(0, 1)$

In [4]:
```
from graphviz import Digraph
import os
os.environ["PATH"] += os.pathsep + 'C:/Users/Max Tchibozo/Anaconda3/Library/bin/graphviz'



nodelist = ['Z', 'X1', 'X2', 'Y']
edgelist = [('X2', 'X1'), ('X1', 'Y'),
            ('Z', 'X1'),
            ('X2', 'Y'),('Z','X2')]


g = Digraph()
nodes = {}
for node in nodelist:
    g.node(node)

edges = {}
for (parent, child) in edgelist:
    g.edge(parent, child)

g
```

Out[4]:



Z: Weather

X1 : Going to the cinema

X2 : Going outside

Y : School Grades

(b) Can both of these studies be valid? Why or why not?

For the instrumental variable estimator to have low bias, we need a strong correlation between Z (the weather) and X (movie theater attendance) in the first study. A priori, it seems unlikely that there is a strong correlation between these two variables, but we would have to look at data to back up this assumption.

On the other hand, there is probably a strong correlation between Z (the weather) and X (going outside to enjoy the weather). This second study seems likely to have less bias than the first study.

For both studies, the exclusion restriction would require the weather to have no effect on grades other than through movie attendence (study 1) or going outside (study 2). It seems reasonable to assume that the weather affects grades mainly through students going outside. So for study 2, using an IV estimator could be valid. On the other hand, for study 1 it would be impossible, even if we remove the effect of weather on people going to the cinema, the weather would still affect grades by influencing whether people go outside or not.

3.4 Exercises

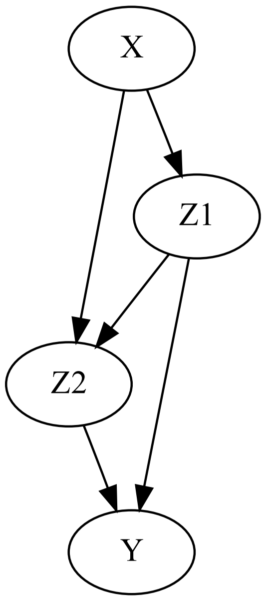
1. For a causal state X, mechanism Z, and outcome Y, why do we require P(Z,X) >0 for all Z and where P(Z) > 0 and P(X) > 0, respectively?

If we have P(Z)>0 and P(X)>0, P(X,Z)>0 means that we need to observe all combinations of X and Z. The Front-Door criterion relies on X having an effect on Y ONLY through Z. If there are values of Z for which X has no effect on Y, and Z is caused by X, then Z must be deterministic in some cases. This is impossible.

1. Can you use the front-door criterion with the following graph? Explain.

In [6]:
```
nodelist = ['X', 'Z1', 'Z2', 'Y']
edgelist = [('Z1', 'Z2'), ('Z1', 'Y'),
            ('X', 'Z1'),
            ('Z2', 'Y'),('X','Z2')]


g = Digraph()
nodes = {}
for node in nodelist:
    g.node(node)

edges = {}
for (parent, child) in edgelist:
    g.edge(parent, child)

g
```

Out[6]:



Let $Z = \{Z_1, Z_2\}$.

- Z intercepts all directed paths from X to Z.
- There is no unblocked back-door path from X to Z.
- All back-door paths from Z to Y are blocked by X

Z verifies the Front-door criterion.

1. Write a linear data-generating proces for the graph in figure 3-3, where every structural equation is linear in the parents.

$X \sim \mathbf{B}(\frac{1}{2})$

$Z_1 = \beta_{Z_1} X$

$Z_2 = \beta_{Z_2} X + \beta_{Z_1 Z_2} Z_2$

$Y = \mathbf{N}(\gamma_{Z_1} Z_1 + \gamma_{Z_2} Z_2)$

(a) What portion of the average treatment effect of X on Y is from Z_1? and Z_2?

$E(Y \mid do(X = x)) = \sum_{Z_1, Z_2} P(Z_1, Z_2 \mid X) \sum_{x'} \mathbf{E}(Y \mid X = x', Z_1 = z_1, Z_2 = z_2) P(X = x')$

$= \sum_{Z_1, Z_2} P(Z_1, Z_2 \mid X) \sum_{x'} (\gamma_{Z_1} Z_1 + \gamma_{Z_2} Z_2)$ by integrating over x' and using $\mathbf{E}(Y \mid X = x', Z_1 = z_1, Z_2 = z_2) = \gamma_{Z_1} Z_1 + \gamma_{Z_2} Z_2$

$= \gamma_{Z_1} \sum_{Z_1} P(Z_1 \mid X) Z_1 + \gamma_{Z_2} \sum_{Z_2} P(Z_2 \mid X) Z_2$

$= \gamma_{Z_1} \mathbf{E}(Z_1 \mid X) + \gamma_{Z_2} \mathbf{E}(Z_2 \mid X)$

We can directly evaluate the effect of the ATE of X on Y from Z_1 and the effect from Z_2.

(b) Can you see how to read these partial effects off from the data-generating process?

$\gamma_{Z_1} \sum_{Z_1} P(Z_1 \mid X) Z_1 = \gamma_{Z_1} \beta_{X_1} x$

and $\gamma_{Z_2} \sum_{Z_2} P(Z_2 \mid X) Z_2 = \gamma_{Z_2} E(Z_2 \mid X) = \gamma_{Z_2} (\beta_{Z_2} X + \beta_{Z_1 Z_2} Z_1)$.

We could have directly read the partial effects of from the data-generating process by simply multiplying the factors together:

For X through Z_1 and its direct path to Y : $\gamma_{Z_1} \beta_{Z_1}$

For X through Z_2 and its direct path to Y : $\gamma_{Z_2} \beta_{Z_2}$

For X through Z_1 and its indirect path to Y through X_2 : $\gamma_{Z_2} \beta_{Z_1} \beta_{Z_1 Z_2}$.

```
In [ ]:
```