Causality

Christina Heinze-Deml

1 Lecture 1

1.1 Different types of experiments

Can be viewed in a tree-like structure.

First differentiation is whether a study has controls, or not. Without controls, one can't say very much.

Next, if We do decide to go with controls, what kind of controls are they? They can be contemporary, i.e. happening now, or in the past, i.e. historical data. One has to be wary of confounders in historical data.

If it's contemporary, i.e. a study is being performed in the present, controls may be chosen. If they are not chosen at random, one once again needs to be very careful about confounders.

If the control is randomly chosen, one ought to do double blind trials, where neither the person receiving nor the doctor administering the treatment knows whether the treatment is true or a placebo. This leads us to randomized double-blind trials, the gold standard.

1.2 Internal v.s. External Validity

This is with respect to statements within studies.

For example, in the Salk vaccine trial, in the case where the trial group were people who consented to the vaccine, and the control was randomly drawn from that same group, it is internally correct to make statements about the trial - given that the findings are significant, they are significant for that group.

What may not hold is external validity - is the vaccine equally effective for groups that are different from the treatment group in the trial? External validity, to me, can be thought of as generalization error - do the statements that are true within the study generalize to the broader population etc.

1.3 Simpson's Paradox

Be careful when averaging over groups - sometimes it may be wise to do so, other times not so.

Essentially there may be lurking fixed effects between groups which would change the direction of the association. More on this later I suppose.

2 Lecture 2

2.1 DAG models

Parents of a node are immediate ancestors.

Children are immediate descents.

In general, a DAG decomposes probabilities such that

$$P(X_1, X_2, X_3 \dots X_n) = \prod_{i=0}^{n} P(X_i | X_{\text{parents}(i)})$$

Recall Markov chains - these have that old chestnut, the **Markov property**:

$$X_{i+1}, X_{i+2} \dots \perp\!\!\!\perp X_{i-1} \dots X_1 | X_i$$

In case of DAGs, We have a generalization of this is the **directed graph Markov Property**, which states that

$$X_i \perp\!\!\!\perp X_{\text{Non-descendants of } X_i} | X_{\text{Parents}(X_i)}$$

Since parents insulate the child.

Now, there's also the **Markov blanket**. The idea is that if You condition on the parents of a node, the children of a node, and the children's parents, then the node is independent of everything else no matter what. I think D-separation shows up here.

We need to think a little more about the descendants case, however.

2.2 *d*-Separation

So we have four types of connections (for illustration see page 407 of advanced data analysis from an elementary point of view):

Forward chain, backward chain, fork, collider

The first thing to observe is that *statistical* information flows both ways in a DAG. Observing a parent by definition gives information about the child, but observing the child also gives clues about the parent.

Now We think about the case where Z, the inbetween variable, is not known.

In the forward chain case, X gives information about Z, and knowing stuff about Z tells us information about Y, so $X \not\perp\!\!\!\perp Y$.

In the backward chain case, knowing X makes it's parent Z more likely, and knowing Z is more likely tells us stuff about Y.

In the fork case, where Z is a parent of both X, Y, knowing X still makes Z more likely, which makes Y more likely.

Finally, let's think about the collider.

I can see why knowing something about the collider explains away the other parents of the collider. This I buy.

Yeah I buy the black hole perspective - a thing has two causes, and We observe one and the thing maybe happened, maybe not. Therefore the two

causes are still independent. If We observe the cause and effect, this explains away the other cause.

Now for the case where the middle variable is known - chains block information flow by Markov property, as do forks, but colliders explain away.

So! We are interested in deciding whether variables X, Y are independent given some set of variables S. We say that X, Y are independent given S if every path from X to Y is blocked. A path is called **blocked** if it is not active. A path is **active** if every variable along the path is active. A variable Z is active if, as before:

Z is a collider and in S,

Z is a collider and a descendant of Z is in S,

Z is not a collider and not in Z.

All of which should make sense given the reasoning about chains, forks and colliders.

Conversely then, a variable Z is blocked if

Z is in S and not a collider,

Z is a collider and neither it nor any of it's descendants is in S.

if S blocks all paths between X, Y, then We say that X, Y are **d-separated** by S.

3 Lecture 3

3.1 Do-Calculus

So We have

$$p(y; \text{see}(X = \tilde{x}) = p(y|\tilde{x})$$

Which is just classical stats. To get do calculus We have

$$p(y; \text{do}(X = \tilde{x})$$

That there refers to the distribution of Y given that X has been forced to take a certain value - some sort of intervention has taken place.

3.2 Regime indicator

Is basically do calculus. We introduce some regime variable σ and do

$$p(y; \sigma = \text{whatever})$$

Where whatever is whatever You want. If it's the null symbol, You get back to standard stats. You can say regime a is where You forced X to be some value, and then $p(y; \sigma = a)$ You are talking about the distribution of Y given an intervention on X as specified by σ. The set of regime indicators is S.

Another important bit is what stays the same given different regimes. Distributions that don't change given different regimes are called *stable* or *invariant*.

3.3 Potential Outcomes

Just seem like regime indicators to me.

3.4 Causal Effects

Are formalized by

$$p(y; \sigma = a) \neq p(y; \sigma = b)$$

For different regimes a, b which change some X.
There is also **Average Causal Effect**:

$$E(Y; \mathrm{do}(X = \tilde{x})) - E(Y; \mathrm{do}(X = x'))$$

There are also individual casual effects and their interpretation is unclear to me. They are also cross-world, so. This is all cross-world actually.

3.5 Intervention Graphs

One way of going about it is just to add a regime node. The DAG is called an intervention DAG if the regime node is a source node (i.e. no incoming edges), the original DAG is a valid DAG (and We're just adding a regime node to the original valid DAG) and d-separation between regime node and any node in the DAG implies lack of causation - changes in regime node won't affect d-separated stuff.

Sidenote: **markov equivalence** of two DAGS is: suppose You have some DAG. Any old DAG. Now You can write down the conditional independence equations for that DAG, basically d-separation. Now switch the directions of the edges around a bit. If You switched the edge directions but the conditional independence equation is the same, then You have two DAGs which are markov equivalent.

3.6 Causal DAGs

Just change the interpretation of edges to be causal. So d-separated things are causally independent, and do-calculus forces probabilities to be one.

4 Lecture 4

Okay so none of the above was necessary lol.

Main bit about causal DAGs - the causal structure is assumed. So when You perform a "do" operation, You just swap out the value of that node in the DAG and replace the probabilities on that edge to 1 for the case You are interested in. Same for calculations.

So when writing down decomposed joint distribution, the probability of the parent vanishes and probability of child becomes the conditional distribution.

4.1 Simpson's paradox

How to pick a part when to condition on variables.

Well, recall that if You perform a do on a causal DAG, it's equivalent to removing incoming edges to that node.

So in the case of gender, We perform a do and now the graph has treatment $\rightarrow$ recovery and gender $\rightarrow$ recovery.

But, We want just treatment $\rightarrow$ recovery. So We condition based on gender to remove it.

In the case of blood pressure, there is no confounder, just treatment affect two variables. Performing a do still gives You all the information You need. I think.

5 Lecture 5

Okay so two variables share a confounder if there exists a path from confounder to both of them.

SEMs are just more nuanced ways to write DAGs. Parents become arguments of functions, children become stuff that depends on the output. Really just a causal DAG but each node gets a function that shows how the node depends on the parents.

Usually linear, in which case they are called linear structural equation models.

Clicker Q - true, false, true, true.

There are different kinds of interventions:

Surgical just sets a value

Shifting just adds noise.

Imperfect cuts incoming edges and replaces with some constant + noise.

5.1 Total Causal Effect

So there exists a total causal effect between X and Y if

Performing do X=x1 and X=x2 yields different distributions for Y.

Performing P(Y—do(X=x1)) $\neq$ P(Y)

5.2 Finding causal structure

Total average causal effect: look at all paths between two variables and add the weights. That the path method.

We can also **regress** to find the total average causal effect. To find effect of X1 on X4, regress with X4 as target and X1 as predictor, plus the parents of X1.

Adjustment set is the set such that regression on X plus adjustment set yields valid total causal effects.

6 Exercises

6.1 Series 1

Q1 a)

Evidence suggests it does but does not prove it - the problem is that the "examined" group of women opted in to the checks, and therefore they may be more health-oriented by default.

In other words, the treatment group is not randomized due to the choice involved, and so causal relationships can't be derived (though they may be suggested).

Q1 b)

Because the groups have identical make-up. Sure, the treatment group was asked to engage more actively in checkups for breast cancer, but this fact does not change their likelihood of dying from an illness other than breast cancer (or maybe changes the odds very slightly).

Q1 c)

Feels like a trick question - the refused group consists more of poorer women, and their rate of breast cancer is higher. The rate is lower amongst the examined group, which are supposed to be richer.

Ah the control group. So the idea is that taking just the poors gives You a rate of 1.5, but taking a mix gives You 2.0, so the poors are less affected.

Q1 d)

Since rich people suffer from fewer diseases, and the examined group consists primarily of the rich, this relationship between wealth and disease accounts for the disparity.

Q2 a)

Observational. Don't know if You could call it a study.

Q2 b)

Yes.

Q2 c)

Yes.

Q2 d)

Not without further interpretation.

One, if the gene is random, it should appear equally in controlling and non-controlling families.

But the presence of the obesity gene may make parents more controlling, so.

Q2 e)

Children's large weight makes parents more controlling in order to get the child to lose weight.

Q2 f)

Depends on the underlying causal structure. Overall, no.

Q3 a)

Non-descendants of D: A.

Q2 b)

Okay so to be separated from F We need to have C. But knowing C opens the collider ACD, so, can't be separated.

Q2 c)

Aren't A and D separated by default? Just don't mess with it.

Otherwise, C, F cannot be present, ever. Other than that, all nodes are game - G, B, E, H.

For the second path, We can do B and then the chain is blocked so E, H, G are all fine.

So either B and any subset of GBEH, or any subset of GH

Q4 a)

Suppose the opposite - two nodes are connected without the use of a collider. But since they share no ancestors, and there are no colliders, they also cannot share descendants, and so this contradicts our assumption that they are connected.

Q4 b)

Either the two nodes are disconnected and are therefore d-separated since no path exists.

Or they are connected by a collider, but neither the collider nor any descendents of the collider are known, so that's also fine.

6.2 Series 2

Q1 a)

Okay so the path method.

Paths from 2 to 4 are: $2 \to 3 \to 4$ and $2 \to 4$.

So the sum is just 3*5+4=19.

Q2 a)

C is standard gaussian, E is gaussian with mean 0 and variance 17.

Q2 b)

Okay We are not doing the conditioning lol.

Conditioning $E|C$ gives mean of eight and variance one.

$C|do(E = 2)$ leaves C untouched, and likewise the other way around.

Q3 a)

A