Adjustment

(a) prior to intervention

(b) after $do(X = x)$

Figure 1: Data generating process

To estimate the causal effect of X on Y, we need to condition on Z so as to block the backdoor path $X \leftarrow Z \rightarrow Y$.

$$P(Y|do(X = x)) = \sum_{z \in \mathcal{Z}} P(Y|X = x, Z = z)P(Z)$$

With a binary treatment X we can also write:

$$ATE = E(Y|do(X = 1)) - E(Y|do(X = 0))$$

This assmumes Z is fully observable, ie there are no unobservable variables, U that cause both X and Y.

Covariate Shift

Seperate from the problem of adjustment, lets define covariate shift.

- We have training data $d_{train} = \{(x_1, y_1), ..., (x_n, y_n)\}$ sampled from $P(X, Y) = P(X)P(Y|X)$.
- The test data $d_{test} = \{(x'_1, y'_1), ..., (x'_{n'}, y'_{n'})\}$ is assumed to be sampled from $Q(X)P(Y|X)$, where $Q(X) \neq P(X)$. As usual for test data, we observe only the inputs $\{x'_1, ..., x'_{n'}\}$
- We wish to use the training data to learn a hypothosis/function $h : \mathcal{X} \rightarrow \mathcal{Y}$ that minimises $\sum_{i=1}^{n'} \mathcal{L}(h(x'_i, y'_i))$, where $\mathcal{L}$ is some loss.

The covariate shift assumption is that $P(Y|X)$ has not changed.

Although we have assumed the underlying mapping, $P(Y|X)$, we are trying to learn has not changed between the test and the train data we still need to adapt standard prediction algorithms as otherwise (through a preference for 'simple' functions) we may select a function that fits well where $P(X)$ is high but performs badly where $Q(X)$ is high. There are some situations that require no adaptation. For example, if the true function is linear and we fit an (unregularized) linear model, then model that is in expectation optimal on the training data is also optimal for the test data. Another interesting example is if the model is a Gaussian process .

Covariate shift in the adjustment problem

In the adjustment problem, we have observational training data, $d_{train} = \{(x_1, z_1, y_1), ...(x_n, z_n, y_n)\} \sim P(Z, X)P(Y|Z, X)$. If we intervene and set $X = x$ then the data will be sampled from $P(Z)\delta(X - x)P(Y|Z, X)$. Note that the covariate shift assumption is satisfied in this scenario, $P(Y|Z, X)$ does not change between the train and test settings. What has changed is the joint distribtion $P(X, Z)$.