### 1. Covariate Shift Problem

Data-generating process (DGP):

- There are two stratified data:
  
  \[(X_i, Y_i) \sim p(x, y), \quad \hat{Y}_i \sim q(x, y),\]
  
  where \(X_i, Y_i \in \mathbb{R}^d\) and \(\hat{Y}_i \in \mathbb{R}\). \(\hat{Y}_i\) is unobservable.

- Observations:
  
  \[(\hat{X}_i, \hat{Y}_i)_{i=1}^{m} \sim p(x, y), \quad \{\hat{Y}_i\}_{i=1}^{m} \sim q(x).\]

- Furthermore, we put the following assumption on the conditional pdf:
  
  \[p(y|x) = p(y|x)p(x), \quad q(y|x) = p(y|x)q(x).\]

  - \(p(y|x)\) is invariant across the two data.
  - \(p(x)\) and \(q(x)\) can be changed.
  - \(p(x)\) and \(q(x)\) have a common support.

- This setting is called learning under covariate shift.

### 2. Definition of Correct Model

**Linear model:**

- Assume a linear model of \(E[Y|X]\) as \(Z^T(X)\beta^*\).

- \(Z()\) is a mapping from \(X_i\) to some linear models.

**Definition of correct model**

- Our model specification is defined from the viewpoint of prediction.

- Parameter that minimizes the MSE over \(p(x, y)\) is defined as
  \[a_0 = \arg\min_{a} E_{p(x,y)}[\|y - Z^T(X_i)b\|^2].\]

- Parameter that minimizes the MSE over \(q(x, y)\) is defined as
  \[\gamma_0 = \arg\min_{\gamma} E_{q(x,y)}[\|\gamma - Z^T(X_i)b\|^2].\]

- If \(a_0 = \gamma_0\), the model is specified correctly.

- If \(a_0 \neq \gamma_0\), the model is misspecified.

- By using this definition, consider the following hypothesis:
  
  \[\mathcal{H}_0: a_0 = \gamma_0 \text{ and } \mathcal{H}_1: a_0 \neq \gamma_0\]

- If \(\mathcal{H}_0\) is rejected, the model specification is incorrect.

### 3. Covariate Shift Adaptation

- However, we cannot observe \(\hat{Y}_i\).

- Let us define a parameter estimated from \((\hat{X}_i, \hat{Y}_i)_{i=1}^{m}\) as
  \[\hat{a}_0 = \arg\min_{a} E_{\hat{p}(x,y)}[\|\hat{y} - Z^T(\hat{X}_i)b\|^2],\]
  
  where \(E_{\hat{p}(x,y)}\) denotes the sample average of the samples from \(p(x, y)\).

- Then, for \(\{\hat{X}_i\}_{i=1}^{m}\), we define the following estimator:
  \[\hat{\gamma} = \arg\min_{\gamma} E_{\hat{p}(x,y)}[\|\hat{y} - Z^T(\hat{X}_i)b\|^2]\]
  
  \[= \arg\min_{\gamma} E_{\hat{p}(x,y)}[\|\hat{y} - Z^T(\hat{X}_i)b\|^2] \frac{q(x)}{p(x)}\]

- Thus, we approximate \(E_{q(x,y)}\) by using \(E_{\hat{p}(x,y)}\) and \(q(x)/p(x)\).

- Let us denote the density ratio \(\frac{q(x)}{p(x)}\) by \(r^*(x)\).

- We can estimate the density ratio with machine learning methods, e.g., uLSIF (Kanamori et al. (2012)).

### 4. Double/Debiased Least Squares Estimator

- Consider the asymptotic distribution of \(\hat{\gamma}\).

  - The density ratio is estimated by machine learning methods.

  - The estimator does not satisfy Donsker’s condition.

- We use double/debiased machine learning to avoid this problem.

  - An estimator \(\hat{p}\) with a doubly robust form.

  - Cross-fitting.

- Doubly robust estimator of the MSE over \(q(x, y)\):
  
  \[E_{\hat{p}(x,y)}[\|\hat{y} - Z^T(\hat{X}_i)b\|^2] \approx E_{\hat{p}(x,y)}[\|\hat{y} - Z^T(\hat{X}_i)b\|^2] = E_{\hat{p}(x,y)}[\|\hat{y} - Z^T(\hat{X}_i)b\|^2] = E_{\hat{p}(x,y)}[\|\hat{y} - Z^T(\hat{X}_i)b\|^2].\]

- \(\hat{f}(x)\) is some consistent estimator of \(f^*(x) = E[Y|X].\)

- \(\hat{f}(x)\) is some consistent estimator of \(r^*(x) = \frac{q(x)}{p(x)}\).

- We construct the empirical MSE by using cross-fitting.

  - Then, if \(n = m \rightarrow \sqrt{n}(\hat{\gamma} - \gamma^*) = \mathcal{N}(0, \Sigma)\)

### 5. Hypothesis Testing

- We construct the test statistics to investigate the hypothesis
  
  \[\mathcal{H}_0: a_0 = \gamma_0 \text{ and } \mathcal{H}_1: a_0 \neq \gamma_0\]

- A standard choice is to use Wald statistics.

- We can construct Wald statistics by using the estimators \(\hat{a}\) and \(\hat{\gamma}\).

- The Wald statistics follows \(\chi^2(k)\) distribution.

  - \(k\) is the dimension of the linear model.

- We conduct hypothesis testing using the test statistics.

  - If the null hypothesis reject, we can say that model is misspecified.

### References

