- Exercise 1. Suppose that (X_i, Z_i, Y_i): $i = 1, \ldots, n$ are IID random vectors, where X_i takes values in $[0,1]$, and Z_i and Y_i take values in $\{0,1\}$ for $i = 1, \ldots, n$. Suppose that we are interested in estimating f, where

$$f(x) = \ln\left(\frac{P(Z_1 = 1|X_1 = x)}{1 - P(Z_1 = 1|X_1 = x)}\right) \text{ for } x \in [0,1].$$

However, it is assumed that only (X_i, Y_i): $i = 1, \ldots, n$ are observed (Z_is are not observed), and for $x \in [0,1]$ and $z \in \{0,1\}$, we have

$$P(Y_1 = z|X_1 = x, Z_1 = z) = 0.95$$

and

$$P(Y_1 = 1 - z|X_1 = x, Z_1 = z) = 0.05.$$

  (a) Propose an estimator for f. If you use m basis functions to approximate f, choose $m \approx n^{1/3}$.

  (b) Write down a R function that computes the proposed estimator for f based on the observed data.

  (c) Generate data according to the above model with $n = 1000$, $f(x) = \sin(x)$ and each X_i is uniformly distributed on $[0,1]$. Find the (approximate) IMSE of the proposed estimator for f based on 500 simulations.

- Exercise 2. Suppose that (X_i, Z_i, Y_i): $i = 1, \ldots, n$ are IID data, and

$$Y_i = f(X_i, Z_i) + \varepsilon_i$$

for $i = 1, \ldots, n$, where ε_i is of mean zero and standard deviation σ. Suppose that X_i and Z_i take values in $[0,1]$.

  (a) Propose a test for testing H_0: $f(x,z) = f_1(x) + f_2(z)$ for some f_1 and f_2. If you use m basis functions to approximate f_1 or f_2, choose $m \approx n^{1/4}$. If you use J^2 tensor basis functions to approximate f, choose $J \approx 1.5n^{1/5}$.

  (b) Write down a R function that computes the p-value for the proposed test based on the observed data.

  (c) Generate data according to the above model with $n = 1000$, $f(x,z) = \sin(x) + \cos(z)$, X_i and Z_i are independent and uniformly distributed on $[0,1]$, and ε_i is normally distributed with $\sigma = 0.1$. Estimate the probability of rejecting H_0 for the proposed test based on 500 simulations.

  (d) Generate data according to the above model with $n = 1000$, $f(x,z) = \sin(x)\cos(z)$, X_i and Z_i are independent and uniformly distributed on $[0,1]$, and ε_i is normally distributed with $\sigma = 0.1$. Estimate the probability of rejecting H_0 for the proposed test based on 500 simulations.