Problem 1

Generate a random sample \((Y_1, X_1), \ldots, (Y_n, X_n)\) iid \(\sim (Y, X)\), where
\[
\begin{pmatrix} Y \\ X \end{pmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix} \right).
\]
Use the cross-validated bandwidth to estimate \(E(Y|X)\) on an equispaced grid. Try this for samples of size \(n \in \{50, 250, 500\}\) and the Epanechnikov and Gaussian kernels.

(i) Plot \(x \mapsto E(Y|X = x)\) and its estimated values on the same graph.
(ii) Report what happens to the estimated values as you vary the bandwidth.
(iii) Experiment with making the evaluation grid finer. See what happens to the plots.
(iv) Submit your code along with the graphs and the time it took for the program to run.

Problem 2

Let \((Y_1, X_1), \ldots, (Y_n, X_n)\) iid \(\sim (Y, X)\), where \(X\) is discrete.

(i) Propose an estimator of \(E(Y|X = x)\), where \(x \in \text{supp}(X)\).
(ii) Show that the estimator is consistent.
(iii) Find its asymptotic distribution.

Problem 3

Let \(\hat{f}(x) = \hat{g}(x)/\hat{p}(x)\) denote the kernel regression estimator of \(E(Y|X = x)\) with \(\hat{g}\) and \(\hat{p}\) as defined in class. Show that if \(x_1, x_2 \in \mathbb{R}^s\) such that \(x_1 \neq x_2\), then \(\text{cov}({\hat{g}(x_1), \hat{p}(x_2)}) = o((nb_n^s)^{-1})\); i.e., when evaluated at different points, the numerator and denominator of \(\hat{f}\) are asymptotically uncorrelated. This result is used in obtaining the joint distribution of \(\hat{f}(x_1)\) and \(\hat{f}(x_2)\).

Problem 4

Let \(\hat{f}(x)\) be the Nadaraya-Watson estimator of \(E(Y|X = x)\). Show that if \(E(Y|X) = \text{constant}\), then \(\hat{f}(x)\) is unbiased for \(EY\). Can you provide some intuition behind this result?

Problem 5

Consider the partially linear model \(Y = \alpha_0 + X'\beta_0 + f^*(Z) + \varepsilon\), where \(E(\varepsilon|X, Z) = 0\). Discuss the conditions under which \(\alpha_0, \beta_0, \) and \(f^*\) are identified.