DAI Assignment-3

Aditya Neeraje, Balaji Karedla, Moulik Jindal

October 7, 2024

Contents

1 Finding optimal bandwidth

1.1 Part 1

1.1.1 Part (a)

We know that the estimator for distribution function $\hat{f}$ is given by

$$\hat{f}(x) = \sum_{j=1}^{m} \frac{\hat{p}_j}{h} \mathbb{1}\left[x \in B_j\right] \tag{1}$$

Also given, v_j is the number of points falling in the j^{th} bin and $\hat{p}_j = \frac{v_j}{n}$.
From this,

$$\begin{aligned}
\int \hat{f}(x)^2 dx &= \int \left(\sum_{j=1}^{m} \frac{\hat{p}_j}{h} \mathbb{1}\left[x \in B_j\right] \right)^2 dx \\
&= \sum_{j=1}^{m} \int_{B_j} \left(\sum_{j=1}^{m} \frac{\hat{p}_j}{h} \mathbb{1}\left[x \in B_j\right] \right)^2 dx \\
&= \sum_{j=1}^{m} \int_{B_j} \left(\frac{\hat{p}_j}{h} \right)^2 dx \\
&= \sum_{j=1}^{m} \left(\frac{\hat{p}_j}{h} \right)^2 \times h \\
&= \frac{1}{n^2 h} \sum_{j=1}^{m} v_j^2
\end{aligned}$$

1.1.2 Part (b)

The histogram estimator after removing the i^{th} observation is given by

$$\hat{f}_{(-i)}(x) = \sum_{j=1}^{m} \frac{\hat{p}_{j,-i}}{h} \mathbb{1}\left[x \in B_j\right] \tag{2}$$

where $\hat{p}_{j,-i} = \frac{v_{j,-i}}{n-1}$ and $v_{j,-i}$ is the number of points falling in the j^{th} bin after removing the i^{th} observation.
Now,

$$
\begin{aligned}
\sum_{i=1}^{n} \hat{f}_{(-1)}(X_i) &= \sum_{1}^{n} \sum_{j=1}^{m} \frac{\hat{p}_{j,-i}}{h} \mathbb{1}\left[X_i \in B_j\right] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{v_{j,-i}}{(n-1)h} \mathbb{1}\left[X_i \in B_j\right] \\
&= \sum_{j=1}^{m} \sum_{i=1}^{n} \frac{v_{j,-i}}{(n-1)h} \mathbb{1}\left[X_i \in B_j\right] \\
&= \sum_{j=1}^{m} \frac{v_j - 1}{(n-1)h} \mathbb{I}\left[X_i \in B_j\right] \qquad \text{(since whenever } \mathbb{1}[X_i \in B_j] = 1, v_{j,-i} = v_j - 1\text{)} \\
&= \sum_{j=1}^{m} \frac{v_j - 1}{(n-1)h} v_j \\
&= \frac{1}{(n-1)h} \sum_{j=1}^{m} v_j^2 - v_j
\end{aligned}
$$

1.2 Part 2

2 Detecting Anomalous Transactions using KDE

2.1 Designing a custom KDE Class

2.2 Estimating Distribution of Transactions