# *Exercises 4: Kernel density estimation*

## Kernel density estimates

### *Histograms*

You're already familiar with the simplest nonparametric density estimate, which is the histogram! Suppose, for simplicity's sake, that we're trying to estimate a density of a random variable $X \sim F$ on the unit interval. Formally speaking, a histogram is a collection of bins $B_j$, $j \in 1, \ldots m$, where

$$B_j = \left[ \frac{j-1}{m}, \frac{j}{m} \right] .$$

Suppose we have a sample $x_1, \ldots, x_n$ from $F$. Let $h = 1/m$ be the bin width, and let $y_j$ be the number of observations in $B_j$. The histogram estimator $\hat{f}(x)$ of the density at point $x$ is

$$\hat{f}(x) = \sum_{j=1}^{m} \frac{\hat{\pi}_j}{h} I(x \in B_j) \, ,$$

where $\hat{\pi}_k = y_j/n$ is the fraction of observations in bin $B_j$, and $I(A)$ is the indicator function of the event $A$.

Let $x$ and $m$ be fixed. Let $B_j$ be the bin containing $x$. Show that

$$\mathrm{E}\{\hat{f}(x)\} = \pi_j/h \quad \text{and} \quad \mathrm{var}\{\hat{f}(x)\} = \frac{\pi_j(1-\pi_j)}{nh^2} \, ,$$

where $\pi_j = \int_{B_j} f(u) du$.