Hicham Janati - hicham.janati@telecom-paris.fr

# TD2: Density estimators

Let $\mathbf{X}$ be a random variable in $\mathbf{P}([0,1],\mathcal{L})$ where $\mathcal{L}$ is the Lebesgue measure. We assume that $\mathbf{P_X}$ admits a density with respect to $\mathcal{L}$ denoted by $f$. The purpose of this problem is to estimate $f$ given samples from $\mathbf{X}$. We assume that $f \in \mathcal{C}^2([0,1])$.

## 1 Histograms

Consider $n$ independently and identically distributed samples $\mathbf{X}_1,\dots,\mathbf{X}_n$. We cut $[0,1]$ into $m$ equal boxes $B_1,\dots,B_m$ of size $h \stackrel{\text{def}}{=} \frac{1}{m}$ where:

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

For each $x \in [0,1]$, we consider approximating $f(x)$ using a count of the samples in its associated box $\widehat{p_j}$ defined by:

$$\widehat{p_j} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{B_j}(\mathbf{X}_i), \quad \widehat{f_n(x)} = \sum_{j=1}^{m} \frac{\widehat{p_j}}{h}\mathbb{1}_{B_j}(x)$$

1. Compute the bias and the variance of the estimator $\hat{p_j}$ as a function of $p_j \stackrel{\text{def}}{=} \int_{B_j} f$.
2. Let $x \in B_j$. Show that:
$$\mathbb{E}(\widehat{f_n(x)}) = \frac{p_j}{h}, \quad \mathbb{V}(\widehat{f_n(x)}) = \frac{p_j(1-p_j)}{nh^2}$$
3. For $x \in B_j$, show that the bias of the estimator can be written as:
$$b(x) \stackrel{\text{def}}{=} \mathbb{E}(\widehat{f_n}(x)) - f(x) = f'(x)\left((j-\frac{1}{2})h - x\right) + \mathcal{O}(h^2)$$
4. Let $x_j$ be the mid-point of $B_j$ given by $(j-\frac{1}{2})h$. Show that:
$$f'(x) = f'(x_j) + \mathcal{O}(h).$$
5. Conclude:
$$\int_0^1 b^2 \mathrm{d}x = \frac{h^2}{12}\int_0^1 f'^2 + o(h^2).$$
6. Show that:
$$\int_0^1 \mathbb{V}(\widehat{f_n})(x)\mathrm{d}x = \frac{1}{nh} + \mathcal{O}(\frac{1}{n})$$
7. What happens to the bias and to the variance as $h \to 0$ ?
8. Deduce that the mean integrated squared error (MISE) of the estimator verifies:
$$\mathcal{R}(\widehat{f_n}, f) \stackrel{\text{def}}{=} \mathbb{E}\left[\int_0^1 (\widehat{f_n} - f)^2\right] = \frac{h^2}{12}\int_0^1 f'^2 + \frac{1}{nh} + o(h^2) + o(\frac{1}{nh})$$
9. Suggest a theoretical target $h^\star$.