Université Hassiba Benbouali Chlef
Faculté des Sciences Exactes & Informatique
Département de Mathématiques

Année universitaire 2021/2022
Master II : Math.Appli & Stat
Module : Stat Non Parametrique

Fiche TD $N =^\circ$ **03(b) :** Estimation NP de la fonction densité de probabilité

Exercice 01:
In the case when the true density is Uniform(0,1) calculate the exact bias of the histogram.
Exercice 02:
Prove that if $\widehat{f}_X(x)$ is the Kernel density estimator, then

$$var(\widehat{f}_X(x)) = \frac{1}{n}\left((K_h^2 * f_X)(x) - (K_h * f_X)^2(x)\right)$$

where $(f * g)(x) = \int f(x-y)g(y)dy$

Exercice 03:
Let $(X_1, ..., X_n)$ be a random sample from a distribution on $\mathbb{R}$ with Lebesgue density $2^{-1}(1-\theta^2)e^{\theta x - |x|}$, where $\theta \in (-1,1)$ is unknown.

1. Show The cumulative distribution function ?
2. Show that the median of the distribution of X_1 is given by $m(\theta) = (1-\theta)^{-1}\log(1+\theta)$ when $\theta > 0$ and $m(\theta) = -m(-\theta)$ when $\theta < 0$.
3. Show that the mean of the distribution of X_1 is $\mu(\theta) = 2\theta/(1-\theta^2)$.

Exercice 04:

1. If $K(t) = \frac{15}{16}(1-x^2)^2$; $|x| \leq 1$ Find $\int K^{(2)}(x)dx$ and $\int x^2K(x)dx$? if $f''(x) = -1$ find h^*?
2. The efficiency of a kernel $K(.)$ is defined as:

$$eff(K) = \frac{3}{5\sqrt{5}}\left(\int t^2K(t)dt\right)^{(-1/2)}\left(\int K(t)^2dt\right)^{(-1)}$$

   Determine the efficiencies for the following kernels:

   a. Biweight: $K(t) = \frac{15}{16}(1-t^2)^2$; $|t| \leq 1$.
   b. Triangular: $K(t) = (1-|t|)$; $|t| \leq 1$.
   c. Normal: $K(t) = \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$; $t \in \mathbb{R}$
   d. Rectangular: $K(t) = \frac{1}{2}$; $|t| \leq 1$.

Exercice 05:
Which of the following serve as kernel functions for a density estimator?
Prove your assertion one way or the other.

a. $K(x) = \mathbf{1}_{(-1<x<l)/2}$,
b. $K(x) = \mathbf{1}_{(0<x<1)}$,

c. $K(x) = 1/x$,
d. $K(x) = \frac{3}{2}(2x+1)(1-2x)\mathbf{1}_{-\frac{1}{2}<x<\frac{1}{2}}$,
e. $K(x) = 0.75(1-x^2)\mathbf{1}_{(-1<x<1)}$,

Exercice 06:

A natural estimate of the derivative of a density $f'(x)$ is the derivative of a kernel estimate of the density; that is,

$$\hat{f}'(x) = \frac{}{nh^2}\sum_{i=1}^{n} K'\left(\frac{x - X_i}{h}\right)$$

(assuming differentiability of K). Calculations similar to those leading to h^* of $\hat{f}$ imply that the optimal bandwidth is $O(n^{-1/7})$, with optimal AMISE of order $O(n^{-1/7})$. Compare "reasonable" choices of h for estimation of f' . Are the density derivative estimates less precisely determined than the density estimates, as the asymptotics would suggest?

Exercice 06: Calculate the exact values of $\displaystyle\int K^2(u)du$ and $\displaystyle\int u^2K(u)du$ for the Gaussian, Epanechnikov and Quartic kernels.

Exercice 07: Multivariate Density Estimation

Kernel density estimation can be easily generalized from univariate to multivariate data, in theory if not always in practice The general form of the estimator is

$$\hat{f}(x) = \frac{1}{n|H|}\sum_{i=1}^{n} K_d\left(\frac{x - X_i}{H^{-1}}\right)$$

where $|H|$ is the absolute value of the determinant of the matrix H. Here $K_d : \mathbb{R} \longrightarrow \mathbb{R}$ is the kernel function, often taken to be a d-variate probability density function, and if is a nonsingular $d \times d$ bandwidth matrix A popular technique for generating K_d from a univariate kernel K is by using a product kernel,

$$K_d(u) = \prod_{i=1}^{n} K(u_j).$$

using multivariate Taylor Series expansions Assume that all second partial derivatives of f are piecewise continuous and square integrable, and that the kernel K_d satisfies the usually conditions

Define $h > 0$ and the $d \times d$ matrix A to satisfy $H = hA$, where A has unit determinant Then, if $h \longrightarrow 0$ and $nh_d \longrightarrow \infty$ as $n \longrightarrow \infty$, show that the AMISE has the form

where $\bigtriangledown^2 f(u)$ is the $d \times d$ Hessian matrix,

$$\bigtriangledown^2 f(u) = \frac{\partial f(u)}{\partial u_i \partial u_j}$$

The optimal H is not generally available in closed form, but AMISE shows that h should be taken to be $O(n^{-1/(d+4)})$,