Durham University

# EXAMINATION PAPER

| Examination Session: | Year: | Exam Code: |
|---|---|---|
| May/June | 2024 | MATH4391-WE01 |

**Title:** Nonparametric statistics IV

| | | |
|---|---|---|
| Time: | 2 hours | |
| Additional Material provided: | | |
| Materials Permitted: | | |
| Calculators Permitted: | No | Models Permitted: Use of electronic calculators is forbidden. |

| | |
|---|---|
| Instructions to Candidates: | Answer all questions.<br>Section A is worth 40% and Section B is worth 60%. Within each section, all questions carry equal marks.<br>Students must use the mathematics specific answer book. |

| | |
|---|---|
| **Revision:** | |

**ED01/2024**

## SECTION A

**Q1** Suppose the density function of $X$ is $f$ which has its support on $[0, 1]$. A histogram partitions the support of $f$ into several bins; the count of observations in each bin is then a density estimate. We assume the following $m$ bins, which define a partition:

$$B_1 = \left[0, \frac{1}{m}\right), B_2 = \left[\frac{1}{m}, \frac{2}{m}\right), \ldots, B_{m-1} = \left[\frac{m-2}{m}, \frac{m-1}{m}\right), B_m = \left[\frac{m-1}{m}, 1\right].$$

In general, for a given point $x \in B_\ell$, the density estimator using the histogram is

$$\hat{f}(x) = \frac{\#\text{ of observations in } B_\ell}{nh},$$

where $h = 1/m$ and $n$ is the number of observations.

(a) Find the forms of $\mathrm{E}(\hat{f}(x))$ and $\mathrm{Var}(\hat{f}(x))$, respectively.

(b) We assume that the function $f$ is Lipschitz continuous over an interval $B_\ell$ i.e. there exists a positive constant $\gamma_\ell$ such that $|f(x) - f(y)| < \gamma_\ell |x - y|$ for all $x, y \in B_\ell$. Under this assumption, show the following result.

$$\mathrm{MSE}(\hat{f}(x)) \leq \gamma_\ell^2 h^2 + \frac{f(\xi_\ell)}{nh}, \qquad (1)$$

where $\xi_\ell \in B_\ell$.

(c) Find the value $h^*$ of the smoothing parameter that minimizes the upper bound given in (1) and find the rate of convergence in terms of $n$ of the histogram estimator under $h^*$. Compare this rate of convergence to that of the Cramér-Rao lower bound for parametric estimators.

**Q2** Suppose the density function of $X$ is $f$ and consider the observations $x_1, \ldots, x_n$. We assume that the $\ell^{\text{th}}$-order kernel $K_{[\ell]}$ satisfies the following:

$$\int K_{[\ell]}(u)du = 1, \quad \int u^j K_{[\ell]}(u)du = 0, \quad j = 1, \ldots, \ell - 1, \quad \text{and} \quad \int u^\ell K_{[\ell]}(u)du \neq 0,$$

where $\ell \geq 1$ is an integer. Suppose that the kernel density estimator $\hat{f}_n$ of $f$ is now based on an $\ell^{\text{th}}$-order kernel $K_{[\ell]}$ and the density $f$ has an $\ell^{\text{th}}$ continuous square integrable derivative. Let $h > 0$ denote a bandwidth for kernel density estimation.

(a) Show that at a given data point $x_0$

$$\mathrm{E}(\hat{f}_n(x_0)) = f(x_0) + \left\{\frac{\mu_\ell(K_{[\ell]})}{\ell!}\right\} h^\ell f^{(\ell)}(x_0) + o(h^\ell),$$

where $\mu_\ell(K_{[\ell]}) = \int x^\ell K_{[\ell]}(x)dx$.

(b) Using the result in (a) and the following fact:

$$\mathrm{Var}(\hat{f}_n(x_0)) \leq \frac{1}{nh} f(x_0) \int K^2_{[\ell]}(y)dy + o\Big(\frac{1}{nh}\Big),$$

find the expression for the Asymptotic Integrated MSE (AIMSE).

(c) The AIMSE-optimal bandwidth is given as follows:

$$h^{\text{AIMSE}} = \left[\frac{(\ell!)^2 \int K^2_{[\ell]}(y)dy}{2n\ell\{\mu_\ell(K_{[\ell]})\}^2 \int (f^{(\ell)}(x))^2 dx}\right]^{1/(2\ell+1)}.$$

Find the rate of convergence of the kernel density estimator in terms of $n$ under $h^{\text{AIMSE}}$.

## SECTION B

**Q3** Observations $(x_i, y_i)$ follow a nonparametric regression model

$$y_i = r(x_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where $\mathrm{E}(\varepsilon_i) = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma^2$, and the $\varepsilon_i$ are uncorrelated. The design points are equally spaced with $0 < x_1 < x_2 < \cdots < x_n < 1$.

(a) Let $\hat{r}(x)$ be the running mean (local average) estimate of $r$ at $x$ using a bandwidth of $\lambda$ (corresponding to a neighbourhood of width $2\lambda$). Give an explicit expression for $\hat{r}(x)$, and compute the *bias* and the *variance* of $\hat{r}(x)$.

(b) The *risk* is defined as

$$R(\hat{r}) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[\{\hat{r}(x_i) - r(x_i)\}^2\right].$$

Show that $R(\hat{r}) = B^2 + V$, where

$$B^2 = \frac{1}{n}\sum_{i=1}^{n} \{r(x_i) - \mathrm{E}[\hat{r}(x_i)]\}^2, \quad V = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}[\hat{r}(x_i)].$$

(c) Assume that $r(x)$ is sufficiently smooth and that $|r'(x)|$ is bounded on $[0, 1]$. The running mean estimate of $r$ is *mean-square consistent* if $R(\hat{r}) \to 0$ as $n \to \infty$. Show that this is the case if certain conditions on $\lambda$ and $n\lambda$ are satisfied, and state these conditions clearly.

(d) Give an example of an explicit dependence of $\lambda$ on $n$ that satisfies these conditions.

(e) Define the cross-validation loss, $\mathrm{CV}(\lambda)$, and explain in detail how you could use cross-validation to choose an appropriate value of $\lambda$ for smoothing a specific data set.

**Q4** **4.1** (a) A researcher makes the following first- and second-order belief specifications over $Y = (Y_1, Y_2, Y_3)^T$:

$$\mathrm{E}(Y) = (0,3,3)^T \text{ and } \mathrm{Var}(Y) = \begin{pmatrix} 5 & -1 & 2 \\ -1 & 3 & -1 \\ 2 & -1 & 3 \end{pmatrix}.$$

Using Bayes linear updating, find the researcher's adjusted expectation and variance for $Y_1$ after observing both $Y_2 = 2$ and $Y_3 = 4$.

(b) If the researcher instead had simply changed their minds about the values for $Y_2$ and $Y_3$ such that their new beliefs for those were

$$\mathrm{E}_n\left[(Y_2, Y_3)^T\right] = (2,4)^T \text{ and } \mathrm{Var}_n\left[(Y_2, Y_3)^T\right] = \begin{pmatrix} 1 & -0.2 \\ -0.2 & 1 \end{pmatrix},$$

what would the effect be on their beliefs about $Y_1$ in contrast to your answer to (a)? Note that you do not need to do any calculations here.

**4.2** (a) On holiday, I buy a bag of sweets of a kind I've never tried before. The first six sweets I pull out of the bag have the following wrapper colours: red, yellow, red, red, green, yellow.
Using Hill's circular $A_{(n)}$ assumption, and nonparametric predictive inference, find the upper and lower probability that the next sweet drawn from the bag has a red wrapper, if I know there are only three colours of wrapper in the bag.
You may assume sweets are an infinite population, and that each bag contains sweets drawn randomly from that population.

(b) Consider the set of probability distributions consistent with the intervals $I = \{[0.2, 0.4], [0.1, 0.5], [0.4, 0.8]\}$.

(i) Explain why $I$ is not a reachable set.

(ii) Using an appropriate expression, convert $I$ into a reachable set, $I'$.