768 Assignment 1*

Your Name and Student ID

2024-08-30

Instructions: *You are to use Quarto Markdown for generating your assignment output file. You begin with the Quarto Markdown script downloaded from A2L, and need to pay attention to information provided via introductory material posted to A2L on working with R, and Quarto Markdown. Having downloaded and added your answers to the Quarto Markdown script, you then are to generate your output file using the "Render" button in the RStudio IDE and, when complete, upload both your Quarto Markdown file and your PDF file to the appropriate folder on A2L.*

Discrete Probability and Cumulative Probability Functions Questions

1. Demonstrate that the unordered kernel estimator of $p(x)$ that uses Aitchison and Aitken's unordered kernel function is *proper* (i.e., it is non-negative and it sums to one over all $x \in \{0, 1, \dots, c-1\}$).

2. Consider the unordered kernel estimator of $p(x)$ that uses Aitchison and Aitken's unordered kernel function.

   i. Express the MSE of $\hat{p}(x)$ in terms of the MSE of $p_n(x)$ and constants Λ_1, Λ_2, and Λ_3 similar to those that were defined in Chapter 1 of the textbook.

   ii. A comparison of the finite sample performance of $\hat{p}(x)$ and that of $p_n(x)$ revolves around the magnitudes of Λ_1, Λ_2, and $\Lambda_3 = p(x)(1 + \Lambda_1)$. Suppose that X has a discrete uniform distribution (i.e., $p(x) = 1/c$ for all $x \in \mathcal{D}$). Express $\operatorname{MSE} \hat{p}(x) - \operatorname{MSE} p_n(x)$ in terms of n, c, and λ and determine its sign for *any* λ.

*M2Studio, arm64, jracine

3. Consider the probability function $p(x)$ for the unordered discrete random variable $X \in \mathcal{D} = \{0, 1, \dots, c-1\}$, where $c \geq 2$ represents the number of unique outcomes. Let $\{X_i\}_{i=1}^n$ represent i.i.d. draws from a distribution with unknown $p(x)$. The kernel estimator of $p(x)$ is given by

$$\hat{p}(x) = \frac{1}{n}\sum_{i=1}^{n} L(X_i, x, \lambda),$$

where $L(\cdot)$ is an unordered kernel function defined by

$$L(X_i, x, \lambda) = \begin{cases} 1 & \text{if } X_i = x \\ \lambda & \text{otherwise,} \end{cases}$$

and where $\lambda \in [0, 1]$.

   i. Derive the bias of this estimator.

   ii. Derive the variance of this estimator.

   iii. Using the SMSE as your criterion, derive the optimal smoothing parameter for this estimator.

   iv. Is this estimator a proper probability function estimator? You must clearly explain and support your answer.

4. Consider an *ordered* random variable with discrete support, $X \in \mathcal{D} = \{0, 1\}$, so that the number of outcomes is $c = 2$. Consider the kernel estimator of $p(x) = \Pr(X = x)$ defined by

$$\hat{p}(x) = \frac{1}{n}\sum_{i=1}^{n} l(X_i, x, \lambda),$$

where $l(\cdot)$ is an *ordered* kernel function defined by

$$l(X_i, x, \lambda) = \lambda^{d_{xi}}/\Lambda_{xi},$$

where $0 \leq \lambda \leq 1$, $d_{xi} = |x - X_i|$, and the normalizing factor $\Lambda_{xi} = \sum_{x \in \mathcal{D}} \lambda^{d_{xi}}$ is tailored to the particular value of $X_i \in \mathcal{D}$.

Presume that you have n independent random draws $\{X_1, X_2, \dots, X_n\}$ from the probability distribution $p(x)$.

   i. How many values can this kernel assume? What is the value of the kernel when $X_i = x$? How about when $X_i \neq x$? Is distance taken into account?

   ii. Derive the bias of this estimator and show that the *leading* bias term is of $O(\lambda)$, which mirrors the result for the unordered case.[1] Presume that $0 \leq \lambda < 1$ so

[1]This will involve approximations that are similar to those we will encounter when we consider kernel-smoothed density estimation, so it will serve you well in what is to come.

that, at a certain point in the proof, you can express $1/(1+\lambda)$ as the infinite series $1-\lambda+\lambda^2-\lambda^3+\ldots$ (hint: first get to the point where you have $E\hat{p}(x) = p(x)+\ldots$ and where you have collected the terms involving $\lambda/(1+\lambda)$, then use this approximation so that you can write $\lambda/(1+\lambda) = \lambda - \lambda^2 + \lambda^3 - \ldots$).

iii. Derive the variance of this estimator up to terms of order $O(\lambda^2)$.

iv. What are the MSE and the SMSE of this estimator? You may leave your expression in terms of Λ_1, Λ_2, and Λ_3 similar to those that were defined in Chapter 1 of the textbook without simplifying any further.

v. What is the optimal smoothing parameter? You may leave your expression in terms of Λ_1, Λ_2, and Λ_3 similar to those that were defined in Chapter 1 of the textbook without simplifying any further.

5. Code up a Monte Carlo simulation that compares the SMSE performance of $p_n(x)$ and $\hat{p}(x)$, where the latter uses Aitchison and Aitken's unordered kernel function with three alternatively chosen smoothing parameters:

    i. The SMSE-optimal λ that uses the *true* (unknown in general) probabilities

    ii. The SMSE-optimal λ that uses plug-in estimates $p_n(x)$ of the probabilities

    iii. The likelihood cross-validated λ

    Run two simulations – one where the probabilities differ substantially across the $x \in \mathcal{D}$ and another where they are the discrete uniform $p(x) = 1/c$. Conduct $M = 1000$ Monte Carlo replications and consider the following probabilities and methods for generating the random samples:

```
## Use one or the other (i.e., comment one out)
## Probabilities differ
p <- c(0.07, 0.13, 0.20, 0.27, 0.33)
## Discrete uniform
p <- c(0.20, 0.20, 0.20, 0.20, 0.20)
## Generate a random sample on the support {0,1,...,c-1}
c <- length(p)
X <- sample(0:(c-1),n,replace=TRUE,prob=p)
```

    For each replication, compute the SMSE via, e.g.,

```
sum((kernel.probability(X,lambda.mse.opt(p))$P-p)^2)
```

    where `lambda.mse.opt(p)` is a function that you have created that accepts a vector of probabilities p and returns λ_{opt} using the formula that we derived. Summarize the SMSEs in the form of boxplots and a tabular summary, as per the simulations in the Practitioner's Corner in Chapter 1 of the textbook. Note that the cross-validated smoothing parameter can be obtained as follows: `npudensbw(~factor(X))$bw` (you need to first install and then load the np package).

Summarize your results in a few sentences, and make sure that you indicate both what you were expecting and what the simulations have revealed. Can you detect a Stein effect at work?