

ECON 7020  
Philip Shaw  
Problem Set 1  
Due date: Feb. 29, 2024

**Problem 1.** Consider the nonparametric frequency estimator for discrete random variables below:

$$p_n(x) = \frac{1}{n} I(X_i = x) \quad (1)$$

- Prove that this estimator is consistent for a general population probability function  $p(x)$ . State carefully where each of the required assumptions are needed.
- Using the Lindeberg-Levy CLT, prove that the  $p_n(x) \stackrel{d}{\sim} N(p(x), \frac{p(x)(1-p(x))}{n})$ . How would you estimate the variance in practice?
- Now write a script file that implements the estimator  $p_n(x)$  at any point  $x$  taking as inputs the data and the point  $x$ . Name this function `udens.R`.<sup>1</sup>
- Using the file `gcb2010coded.csv`, estimate the probability function across different income groups. Graph your estimated probability function. Explain your findings.
- Now compute 95% confidence intervals across the support of income using the expression  $p_n(x) \pm 1.96se(p_n(x))$ .

---

<sup>1</sup>Hint: modify the `cdffun.R` file and then rename it to `udens.R`.