DS-GA 3001: Applied Statistics (Fall 2023-24)
Final, Thursday December 21st

Instructions:

- You have **110 minutes**, 4:00PM - 5:50PM
- The exam has 3 problems, totaling 100 points (+5 bonus points).
- Please answer each problem in the space below it.
- You are allowed to carry the textbook, your own notes and other course related material with you. Electronic devices are not allowed.
- Please read the problems carefully.
- Unless otherwise specified, you are required to provide explanations of how you arrived at your answers.
- You can use previous parts of a problem even if you did not solve them.
- The problems may not be arranged in an increasing order of difficulty. If you get stuck, it might be wise to try other problems first.
- Good luck and enjoy!

Full name: ____

N number: ____

1. Short questions. *(40 points)*

Provide a short answer to each of the questions. Each question is worth 10 points.

(a) In semiparametric statistics, let $s^{\theta}_{(\theta_0,\eta_0)}(y)$ and $s^{\eta}_{(\theta_0,\eta_0)}(y)$ be the score functions for the target parameter θ and nuisance parameter η, respectively. Write down the definition of the efficient score function $s^{\text{eff}}_{(\theta_0,\eta_0)}(y)$ for θ, and show that

$$\mathbb{E}_{(\theta_0,\eta_0)}[s^{\text{eff}}_{(\theta_0,\eta_0)}(y)] = 0.$$

(b) Explain the *unconfoundedness* in the potential outcome model of causal inference. Propose a real-life scenario where this assumption is violated.

(c) Consider the nonparametric regression problem on $[0, 1]$, and below we list several estimators covered in class. Which of the following operations will *increase* the bias (and consequently *decrease* the variance)?

i. increase the bandwidth h in the Nadaraya–Watson estimator;
ii. increase the polynomial degree k in the local polynomial regression;
iii. increase the regularization parameter λ in cubic smoothing spline regression;
iv. increase the number of kept terms m in the Fourier projection estimator;
v. increase the threshold t in the wavelet soft-thresholding estimator.

Write Y (Yes) or N (No) for each operation, without explanations.

(d) For a class of continuous functions $\{\phi_i(x)\}_{i\geq 1}$ on $[0,1]$, write down the definition of these functions being *orthonormal*. Show that if $\{\phi_i(x)\}_{i\geq 1}$ and $\{\psi_j(y)\}_{j\geq 1}$ are two orthonormal classes of functions on $[0,1]$, then the class of bivariate functions $\{f_{i,j}(x,y) = \phi_i(x)\psi_j(y)\}_{i,j\geq 1}$ are orthonormal on $[0,1]\times[0,1]$.

2. Estimation of causal functionals. *(30 points)*

Consider the following model for causal inference: let X be the covariate, $W \in \{0, 1\}$ be the binary indicator of treatment with $\mathbb{E}[W \mid X = x] = e(x)$, and Y be the observed outcome with $\mathbb{E}[Y \mid X = x] = \mu(x)$. The target is to estimate the causal functional

$$\psi = \mathbb{E}[\mathsf{Cov}(W, Y \mid X)],$$

while treating $(e(x), \mu(x))$ as nuisance parameters.

Throughout this problem the following covariance definitions will be useful:

$$\begin{aligned}\mathsf{Cov}(W, Y \mid X) &= \mathbb{E}[WY \mid X] - \mathbb{E}[W \mid X]\mathbb{E}[Y \mid X] \\ &= \mathbb{E}[(W - \mathbb{E}[W \mid X])(Y - \mathbb{E}[Y \mid X]) \mid X].\end{aligned}$$

(a) Show that

$$f_{(\psi,e,\mu)}(X, W, Y) = WY - e(X)\mu(X) - \psi$$

is an estimating function, i.e. $\mathbb{E}[f_{(\psi,e,\mu)}(X, W, Y)] = 0$. *(10 points)*

(b) Show that

$$g_{(\psi,e,\mu)}(X, W, Y) = (W - e(X))(Y - \mu(X)) - \psi$$

is also an estimating function, i.e. $\mathbb{E}[g_{(\psi,e,\mu)}(X, W, Y)] = 0$. *(10 points)*

(c) Show that $g_{(\psi,e,\mu)}$ is doubly robust, i.e. for any nuisance estimates $(\widehat{e}, \widehat{\mu})$, it holds that

$$\mathbb{E}[g_{(\psi,\widehat{e},\mu)}(X, W, Y)] = 0,$$
$$\mathbb{E}[g_{(\psi,e,\widehat{\mu})}(X, W, Y)] = 0.$$

(10 points; hint: it might be easier to work on the difference $\mathbb{E}[g_{(\psi,\widehat{e},\mu)} - g_{(\psi,e,\mu)}]$*.)*

3. Nonparametric functional estimation. *(30 points + 5 bonus points)*

Let f be an unknown density on $[0,1]$, and we observe i.i.d. $X_1, \cdots, X_n \sim f$. Instead of estimating f itself, suppose that our target is to estimate the quadratic functional

$$Q(f) = \int_0^1 f(x)^2 dx.$$

(a) Given an estimator $\widehat{f} \geq 0$ for f, write down the plug-in estimator $\widehat{Q}$ for $Q(f)$. Suppose that $\int_0^1 |\widehat{f}(x) - f(x)| dx \leq \varepsilon$, and $\max\{f(x), \widehat{f}(x)\} \leq L$ for all $x \in [0,1]$. Show that your above estimator $\widehat{Q}$ satisfies

$$|\widehat{Q} - Q(f)| \leq C\varepsilon,$$

for a constant $C > 0$ depending only on L. *(10 points)*

(b) Given an estimator $\widehat{f}$, another plug-in estimator is defined as

$$\widehat{Q}_1 = \frac{2}{n}\sum_{i=1}^{n} \widehat{f}(X_i) - \int_0^1 \widehat{f}(x)^2 dx.$$

Show that

$$\mathbb{E}[\widehat{Q}_1] = Q(f) - \int_0^1 (\widehat{f}(x) - f(x))^2 dx.$$

We assume that $\widehat{f}$ is independent of $(X_1, \cdots, X_n)$ (e.g. constructed from another sample via sample splitting) and thus treated as *fixed*. *(10 points)*

(c) Under the setting of (b), suppose that $\int_0^1 (\widehat{f}(x) - f(x))^2 dx \le \varepsilon^2$, and $0 \le \widehat{f}(x) \le L$ for all $x \in [0, 1]$. By analyzing the bias and variance separately, show that

$$\mathbb{E}[(\widehat{Q}_1 - Q(f))^2] \le \varepsilon^4 + \frac{C}{n},$$

for a constant $C > 0$ depending only on L.

(d) Now suppose the target is to estimate the entropy functional

$$h(f) = -\int_0^1 f(x) \log f(x) dx,$$

where log is the natural logarithm. Given an estimator $\widehat{f}$ independent of $(X_1, \cdots, X_n)$, propose an estimator $\widehat{h}$ of $h(f)$ in a similar spirit to (b). Justify your answer. *(5 bonus points)*