Comprehensive Examination
Quantitative Methods
Fall, 2023

Instruction: This exam consists of three parts. You are required to answer all the questions in all the parts.

Grading policy:

1. Each part will be graded separately, and there are four possible results in each part: H (honor pass), P (PhD pass), M (master pass), and F (fail).

2. Each part contains a precise grade determining algorithm.

3. The grades from the three parts will be summarized in the descending order, after which the overall grade will be determined using the algorithm summarized in the table below:

Highest	Middle	Lowest	**Overall**
H	H	H	**H**
H	H	P	**H**
H	H	M	**P**
H	H	F	**M**
H	P	P	**P**
H	P	M	**P**
H	P	F	**M**
H	M	M	**M**
H	M	F	**M**
H	F	F	**F**
P	P	P	**P**
P	P	M	**P**
P	P	F	**M**
P	M	M	**M**
P	M	F	**M**
P	F	F	**F**
M	M	M	**M**
M	M	F	**F**
M	F	F	**F**
F	F	F	**F**

Part I - 203A

Instructions for Part I: Solve every question. For every question in this part, your answer should be numerical; an abstract formula will not be accepted as an answer. Also, you are required to be very specific about the dimension of the a zero vector or zero matrix. If you intend to write a 1×3 zero vector, you should write it $\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$ or $\underset{1\times 3}{0}$. If you simply write "0", it shall be understood to be a scalar.

Grading policy for Part I: Your grade in this part of the exam is based on the total points that you earn. Below is how the grade is determined. In order to avoid any confusion arising from the difference between strict and weak inequalities, the grade assignment algorithm is presented in terms of mathematical inequalities. Note that $<$ denotes a strict inequality, and $\leq$ denotes a weak inequality. Let T denote the total number of points.

1. If $T \geq 25$, you will get H.
2. If $20 \leq T < 25$, you will get P.
3. If $15 \leq T < 20$, you will get M.
4. If $T < 15$, you will get F.

1. (4 pt.) Let $\Phi(\cdot)$ denote the CDF of $N(0,1)$. Let X_n denote a sequence of random variables such that the CDF F_n of X_n is given by
$$F_n(x) = \Phi(nx).$$
What is $\lim_{n\to\infty} E[\cos(X_n)]$? Your answer should be a number.

2. (4 pts.) Let X_1 denote a random sample (of size 1) from $N(0,\theta)$. We have $H_0 : \theta = 1$ and $H_1 : \theta = 3$. You decided to use the Neyman-Pearson test of size 5%. If you observe $X_1 = 4$, do you reject H_0 or not? Your answer should be either Reject or Do not reject. (You may want to use the fact that if $Z \sim N(0,1)$, then $P(Z > 1.645) = 5\%$ and $P(Z > 1.96) = 2.5\%$. Obviously it means that if $X \sim \chi^2(1)$, then $P(X > 1.645^2) = 10\%$ and $P(X > 1.96^2) = 5\%$.)

3. (5 pts.) Suppose that X_n is a sequence of random variables such that $\sqrt{n}(X_n - 1)$ converges in distribution to $N(0,1)$. By the multivariate delta method, we can see that
$$\begin{bmatrix} \sqrt{n}(X_n^3 - \mu_1) \\ \sqrt{n}(\ln(X_n) - \mu_2) \end{bmatrix} \xrightarrow{D} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{bmatrix} \right)$$
for some μ_1, μ_2, $\sigma_{1,1}$, $\sigma_{1,2} = \sigma_{2,1}$, and $\sigma_{2,2}$. What are their numerical values? Your answer should be numbers, not formulae. (Here, the symbol "ln" denotes the natural logarithm.)

4. (5 pts.) Consider a simple situation where the sample size is 1. Let X have the PDF equal to $\lambda \exp(-\lambda x) \cdot 1(x > 0)$. We have $H_0 : \lambda = 1$ vs. $H_1 : \lambda < 1$. Suppose that C is a critical region such that (1) the test of the form "Reject H_0 if $X_1 \in C$" is the uniformly most powerful test; and (2) the probability of rejecting H_0 when $\lambda = 1$ is α. What is the power of your test for when $\lambda = 1/2$ and $\alpha = 1\%$. Your answer should be a number.

5. (6 pts.) Let X_1 denote a random sample (of size 1) from a distribution with PDF equal to $\theta \exp(-\theta x) \cdot 1(x > 0)$. We would like to test $H_0 : \theta = 1$ against $H_1 : \theta \neq 1$.

   (a) (3 pts.) Suppose that $X_1 = 1$. What is the value of the the LR statistic? Your answer should be a number.

   (b) (3 pts.) Do you reject or accept the null at 5% significance level? Your answer should be either Reject or Do not reject. (You may use the approximation $e = 2.7183$. You may want to use the fact that if $Z \sim N(0,1)$, then $P(Z > 1.645) = 5\%$ and $P(Z > 1.96) = 2.5\%$. Obviously it means that if $X \sim \chi^2(1)$, then $P(X > 1.645^2) = 10\%$ and $P(X > 1.96^2) = 5\%$.)

6. (6 pts.) Suppose that the conditional distribution of Y given X is Uniform$(0, X)$, and X has a uniform(0,1) distribution. Hint: Some (not all) students may find it easier to work with the joint PDF of (X, Y), which is equal to $\frac{1}{x}1(0 < y < x) \cdot 1(0 < x < 1)$. Some students may find it easier to work with $f_{Y|X}(y|x) = \frac{1}{x}1(0 < y < x)$ and $f_X(x) = 1(0 < x < 1)$.

   (a) (2 pts.) What is $E[Y]$? Your answer must be a number.

   (b) (2 pts.) What is $\text{Var}(Y)$? Your answer must be a number.

   (c) (2 pts.) Are X and Y independent? Your answer must be either Yes or No.

Part II - 203B

Instruction for Part II: Solve every question.

Grading policy for Part II: Your grade in this part of the exam is based on the total points that you earn. Below is how the grade is determined. In order to avoid any confusion arising from the difference between strict and weak inequalities, the grade assignment algorithm is presented in terms of mathematical inequalities. Note that $<$ denotes a strict inequality, and $\leq$ denotes a weak inequality. Let T denote the total number of points.

1. If $T \geq 80$, you will get H.
2. $60 \leq T < 80$, you will get P.
3. $45 \leq T < 60$, you will get M.
4. If $T < 45$, you will get F.

Question 1 (50 points) Let $\{Y_i, X_i\}_{i=1}^n$ be an i.i.d. sample with $Y_i \in \mathbf{R}$ and $X_i \in \mathbf{R}^d$ satisyfying

$$Y_i = X_i'\beta_0 + \varepsilon_i \qquad E[(\varepsilon - E[\varepsilon])(X - E[X])] = 0. \tag{1}$$

In what follows please be careful that we have assumed that the covariance between ε and X equals zero, but not necessarily that $E[\varepsilon X] = 0$.

(a) (10 points) Does model (1) imply that $E[\varepsilon] = 0$? Justify your answer.

(b) (10 points) Show that under our assumptions we must have the restriction

$$E[\{(Y_i - E[Y]) - (X_i - E[X])'\beta_0\}(X - E[X])] = 0.$$

(c) (10 points) Can the moment restriction in part (b) identify β_0 if X_i contains a constant? Why or why not? Justify your answer.

(d) (10 points) Let $\bar{Y}_n \equiv \frac{1}{n}\sum_{i=1}^n Y_i$ and $\bar{X}_n \equiv \frac{1}{n}\sum_{i=1}^n X_i$, and define the estimator

$$\hat{\beta}_n \equiv \arg\min_{b \in \mathbf{R}^d} \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y}_n - (X_i - \bar{X}_n)'b)^2.$$

Establish the consistency of $\hat{\beta}_n$ to β_0 clearly stating any assumptions you employ.

(e) (10 points) Another researcher is concerned that you did not include a constant in (1). He instead prefers the more traditional regression

$$(\tilde{\alpha}_n, \tilde{\beta}_n) \equiv \arg\min_{a \in \mathbf{R}, b \in \mathbf{R}^d} \frac{1}{n}\sum_{i=1}^n (Y_i - a - X_i'b)^2.$$

How does his estimator $\tilde{\beta}_n$ compare to your estimator $\hat{\beta}_n$ from part (c)? Justify your answer. (**Hint:** Frisch-Waugh-Lovell Theorem).

Question 2 (50 points) Let $Z \in \{0,1\}$ be an instrument, $D \in \{0,1\}$ a treatment, and Y an observable outcome. Throughout, assume a LATE framework in which there are two potential outcome $(Y(0), Y(1))$, two potential treatment assignments $(D(0), D(1))$, and assume that the observable D and Y are determined according to

$$D = D(0) + Z(D(1) - D(0)) \qquad Y = Y(0) + D(Y(1) - Y(0));$$

i.e. we observe the potential outcome corresponding to actual treatment status, and the potential treatment assignment corresponding to the realization of Z. Further assume that $(Y(0), Y(1), D(0), D(1))$ are independent of Z, the monotonicity condition that

$$P(D(1) \geq D(0)) = 1, \tag{2}$$

and that we have available an i.i.d. sample $\{Y_i, D_i, Z_i\}_{i=1}^n$ of (Y, D, Z).

(a) (10 points) Show that under the stated assumptions we must have

$$P(D = 1|Z = 1) = P(D(1) = 1) \qquad P(D = 1|Z = 0) = P(D(0) = 1).$$

(b) (10 points) Use part (a) to argue that if the montonicity assumption (i.e. equation (2)) is correct, then the following restriction must hold

$$P(D = 1|Z = 1) - P(D = 1|Z = 0) \geq 0. \tag{3}$$

(c) (10 points) In order to check whether the monotonicity assumption is correct, we compute the following sample analogue to the quantities in (3):

$$\frac{\sum_{i=1}^n D_i Z_i}{\sum_{i=1}^n Z_i} - \frac{\sum_{i=1}^n D_i(1 - Z_i)}{\sum_{i=1}^n (1 - Z_i)} \tag{4}$$

(here recall that $D \in \{0,1\}$ and $Z \in \{0,1\}$). Carefully derive the asymptotic distribution of the estimator in (4).

(d) (10 points) Propose an estimator for the asymptotic variance of the estimator in equation (4). You do not need to formally establish its consistency.

(e) (10 points) Use the results of pats (b)-(d) to propose a test of the monotonicity assumption in (2). You do not need to establish formal results, but you should clearly outline exactly how to compute the test if we want the probability of a Type I error to be α.

Part III - 203C

Instruction for Part III: Solve every question. The total number of points is 100. Allocate your time wisely!

Let T denote the total number of points.

1. If $T \geq 85$, you will get H.
2. $60 \leq T < 85$, you will get P.
3. $45 \leq T < 60$, you will get M.
4. If $T < 45$, you will get F.

Question 1 (60 points)

Consider a model specified as

$$Y = m\left(X\ \varepsilon\right)$$

where $(Y, X) \in R^2$ is observed, ε is unobserved and distributed independently of X, and the function $m : R \to R$ is continuously differentiable and strictly increasing. Denote the cumulative distribution of ε by F_ε and denote the cumulative distribution of (Y, X) by $F_{Y,X}$. Assume that $F_{Y,X}$ is known, F_ε is unknown, m is unknown and all these functions are continuously differentiable.

(a; 10) Obtain an expression for the conditional distribution of Y given $X = x$ in terms of only m, the distribution of ε, and values of observable variables.

(b; 10) Obtain an expression for the conditional expectation of Y given $X = x$ in terms of only m, the distribution of ε, and values of observable variables.

(c; 5) Suppose that ε is *not* distributed independently of X. Denote the distribution of (ε, X) by $\widetilde{F}_{\varepsilon,X}$. Obtain expressions analogous to those in (a) and (b).

Assume now that $m(t) = \gamma\ t$ for some $\gamma > 0$ of unknown value. Assume also that ε is distributed independently of X with a $N\left(\mu, \sigma^2\right)$ distribution and that X is distributed with a Uniform distribution on the segment $[2, 4]$.

(d; 5) Obtain an expression for the conditional expectation of Y given $X = x$ and an expression for the conditional distribution of Y given $X = x$ in terms of only γ, the parameters of the distribution of (ε, X), values of observable variables, and known functions.

(e; 5) Determine what parameters, if any, are identified without normalizing others. Provide proofs or counterexamples.

(f; 5) After imposing, if needed, normalizations, specify how to estimate the identified parameters by a Maximum Likelihood Method using N independent observations $\{(Y_i, X_i)\}_{i=1,\ldots,N}$

(e; 10) Sketch a proof that the estimators that you specified in (f) are consistent and asymptotically Normal.

(f; 10) Suppose that the distribution of X were only known to be Uniform on a segment $[\alpha - 1, \alpha + 1]$ for some $\alpha > 1$. Would your answer to (e) change, and if so, how? Can α be identified? If α can be identified, what method would you use to estimate α? What properties such estimator for α would have? Justify your answers.

Question 2 (40 points)

Students in certain schools have to choose between programs A, B, or C. Suppose that the utility of student i for program j ($j = A,\ B, C$) is given by

$$U_j^i = \alpha_j + \gamma_j'\ S_i + \beta'\ X_j + \varepsilon_j^i$$

where α_j, γ_j, and β are parameters of unknown values, ε_j^i is a variable of unobserved value corresponding to student i and program j, S_i is a vector of observed characteristics of student i, and X_j is a vector of observed characteristics of program j. Suppose that $(\varepsilon_A^i, \varepsilon_B^i, \varepsilon_C^i)$ is distributed independently of (S_i, X_A, X_B, X_C) with a cummulative distribution function $F_{\varepsilon_A^i, \varepsilon_B^i, \varepsilon_C^i}$. Assume that the program each student chooses is the one that maximizes his/her utility.

(a; 10) Provide an expression for the probability that student i chooses program A, given $(S_i, X_A, X_B, X_C) = (s_i, x_A, x_B, x_C)$, in terms of the given functions.

(b; 15) Suppose that the distribution of $(\varepsilon_A^i, \varepsilon_B^i, \varepsilon_C^i)$ is $N(\mu, \Omega)$. Analyze the identification of the parameters of the distribution of $(\varepsilon_A^i, \varepsilon_B^i, \varepsilon_C^i)$ and of the utility functions. Are there any conditions on (S, X_A, X_B, X_C) that need to be satisfied?

(c; 5) Would your answer to (b) change if β were allowed to be different across programs? Explain.

(d; 10) How would you test the hypothesis that $\beta = 0$ versus the alternative that $\beta \neq 0$ using a sample of N independent observations on the choices made by N students. Explain in detail.