2019 First year Exam – Theory

Statistics
200ABC

June 21, 2019
9:00 to 1:00

Instructions:

There are 7 questions on the examination, each with multiple parts. Select any 5 of them to solve. If you attempt to solve more than 5 questions, you are only to turn in the 5 you want graded. If you turn in partial solutions to more than 5 questions, only 5 will be graded.

Each of the 5 problems you attempt to solve will be worth equal credit, with each accounting for 20% of your final score on this examination.

Your solutions to each of the 5 problems you solve should be written on separate sheets of paper.

DO NOT WRITE ON BOTH SIDES

Label each sheet in the upper right hand corner:

1) with your student id number

2) with the problem number

3) and the page number for that problem.

For example, the labeling on a page might be:

ID#912346378

Problem 3, page 2

THEORY (2019), Problem 1

1. Consider a regression model in which response variables $Y_i, i = 1, \ldots, n$ are presumed to satisfy

$$Y_i = \beta_0 + \beta_1 Z_i + \beta_2 W_i + \epsilon_i$$

with the ϵ_i's independent and identically distributed such that $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

(a) Suppose we wish to implement a *global test* of the model, so that we wish to test the null hypothesis $H_0 : \beta_1 = \beta_2 = 0$ versus the alternative hypothesis that at least one of β_1 or β_2 are non-zero. Let $\widehat{\vec{\beta}}$ denote the ordinary least squares estimator (OLSE) of $\vec{\beta} = (\beta_0, \beta_1, \beta_2)^T$. Assuming σ^2 is known, provide a test statistic that can be used to test the above hypothesis and state the critical value that would be used to provide a level α test of H_0. (Note: You may leave the test statistic in matrix form and you need not provide a numeric critical value, but should say how the critical value would be obtained.)

(b) Suppose that at least one of β_1 or β_2 are non-zero. What is the distribution of the test statistic you provided in part (a)?

(c) In general, σ^2 is unknown and must be estimated. Provide a consistent estimator of σ^2, call it $\widehat{\sigma}^2$, and briefly justify the consistency of your estimator.

(d) If you replace σ^2 with $\widehat{\sigma}^2$ in your test statistic in (a), what is the asymptotic distribution of the resulting test statistic? Justify your answer.

THEORY (2019), Problem 2

Let $X_1, X_2, \ldots$ be independent and identically distributed Uniform(0,1) random variables. Conditional on $X_i = x_i$, $Z_i \sim Bernolli(x_i)$ for $i = 1, 2, \cdots$.

(a) Let

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Discuss the large sample behavior of $\bar{X}_n$.

(b) Find the distribution of $-ln(X_i)$.

(c) Let

$$G_n = (\prod_{i=1}^{n} X_i)^{1/n}.$$

Show that G_n converges in probability to a constant.

(d) Given that $Z_i = 0$, what is the probability of $X_i < 0.5$?

(e) Let

$$\bar{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_n.$$

Discuss the large sample behavior of $\bar{Z}_n$.

THEORY (2019), Problem 3

After examining relevant air pollution data for a certain city in the United States, an environmental scientist postulates that the distribution of the carbon monoxide concentration level X (measured in parts per million, or ppm) above k ppm (where k is a known positive constant) can be accurately modeled by the one-parameter Pareto density function

$$f_X(x) = \frac{\theta k^\theta}{x^{\theta+1}}, \quad 0 < k < x < +\infty, \; \theta > 3.$$

(a) Find an explicit expression of the cumulative distribution function (CDF) $F_X(x)$, and then use this CDF to find the numerical value of $Pr[k+1 < X < k+3 | X > k+1]$ when $k = 1$ and $\theta = 4$.

(b) Develop an explicit expression for $\mu_3 = E\{[X - E(X)]^3\}$, find the limiting value of μ_3 as $\theta \to +\infty$, and explain why this limiting value makes sense even though the above density function is clearly non-symmetric.

(c) After careful thought, this environmental scientist suggests that the distribution of the random variable $Y = \ln(X)$ has more scientific relevance than the distribution of X itself. Develop an explicit expression for the moment generating function $M_Y(t)$ of Y, and then use $M_Y(t)$ directly to find an explicit expression of $E(Y)$.

END OF QUESTION (3)

THEORY (2019), Problem 4

The concentration Y of lead in the blood of children of age x is postulated to have the density

$$f_Y(y|x,\beta) = (\beta x)^{-1}e^{-y/(\beta x)}, \; y>0, \; x>0, \; \beta>0,$$

where x is a deterministic value (the so-called fixed design in contrast to the random design where X is random). For n pairs of independent observations $(x_1, Y_1), \dots, (x_n, Y_n)$, consider the following:

(a) Derive explicit expressions for the MLE $\hat{\beta}_{ML}$ of β as well as for its mean and variance given $(x_1, \dots, x_n)$.

(b) Derive explicit expressions for the ordinary least squares estimator $\hat{\beta}_{LS}$ of β, which minimizes $Q = \sum_{i=1}^n [Y_i - E(Y_i)]^2$, as well as for its mean and variance given $(x_1, \dots, x_n)$.

(c) Derive explicit expressions for the method of moments estimator $\hat{\beta}_{MM}$ of β, which solves the equation $E(\bar{Y}) = \bar{Y}$, as well as for its mean and variance given $(x_1, \dots, x_n)$.

(d) Which of the above three estimators would you recommend? (Be precise and thorough in your statistical reasoning.)

(e) Now consider a random design with iid observations $(X_1, Y_1), \dots, (X_n, Y_n)$ for a population with conditional density

$$f_{Y|X}(y|x;\beta) = (\beta x)^{-1}e^{-y/(\beta x)}, \; y>0, \; x>0, \; \beta>0,$$

and X follows an unknown distribution $f_X(x)$ that does not depend on β. Derive the asymptotic distribution of the MLE of β.

END OF QUESTION (4)

THEORY (2019), Problem 5

Suppose that $Y = (Y_1, \cdots, Y_n)^T$ is an $n \times 1$ random vector and

$$Y = \mathbf{1}\beta_0 + X\beta + \epsilon$$

where X is a $n \times p$ design matrix with rank p, β is a $p \times 1$ vector of coefficient parameters, $\mathbf{1}$ is an $n \times 1$ vector of $1's$, ϵ is a multivariate normal random vector with mean $\mathbf{0}_{n\times 1}$ and variance-covariance matrix $\sigma^2 I$. You may assume that $\mathbf{1}$ is linearly independent with X.

(a) Show that the least square estimate of β is $\hat{\beta} = [X^T(I - \frac{J}{n})X]^{-1}X^T(I - \frac{J}{n})Y$, where J is an $n \times n$ matrix of $1's$, i.e., $J = \mathbf{1}\mathbf{1}^T$.

(b) The total sum of squares in Y is defined as $SSTO = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$, where $\bar{Y}$ is the average of the $Y_i's$. Show that $SSTO = Y^T(I - \frac{J}{n})Y$.

(c) Show that $E[Y^T AY] = \sigma^2 tr(A) + E[Y]^T A\, E[Y]$. Use this result to find the expected value of $SSTO$ and express your answer in terms of β, σ^2, X, and necessary constants.

(d) The result in (c) suggests that the variance in Y is contributed by two sources, one of which can be measured by the residual sum of squares: $RSS = (Y - \hat{Y})^T(Y - \hat{Y})$. Show that $RSS = Y^T(I - \frac{J}{n} - \tilde{X}(\tilde{X}\tilde{X}^T)^{-1}\tilde{X}^T)Y$ where $\tilde{X} = (I - \frac{J}{n})X$.

(e) Let $SSX = SSTO - RSS$. Use RSS, SSX, and necessary constants to construct a random variable that follows an F-distribution when $\beta = \mathbf{0}$. Be sure to justify your answer.

THEORY (2019), Problem 6

Consider two independent random variables X and Y with the same geometric distribution with density $p(1-p)^{x-1}$, $0 < p < 1$, $x = 1, 2, \ldots$, where p is the success probability.

(a) Show that $Pr(X = Y) = \frac{p}{2-p}$.

(b) Show that $Pr(X \geq 2Y) = \frac{1-p}{3-3p+p^2}$.

(c) Prove that $\min\{X, Y\}$ and $X - Y$ are independent random variables.

END OF QUESTION (6)

THEORY (2019), Problem 7

7. Consider a random sample $X_1, \ldots, X_n$ from a distribution with the density

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)$$

where $x > 0$, $\alpha > 0$, and $\beta > 0$.

(a) Find the MLE of $\theta = (\alpha, \beta)$.

(b) Find the Fisher information $I(\theta)$.

(c) Write down the asymptotic distribution of θ.

(d) Derive the Cramer-Rao lower bound for the variance of unbiased estimators of β.

(e) Provide a large sample confidence interval for β.

(f) Given α, find the conjugate prior for β and write down the corresponding posterior distribution.

Table 1: Common distributions and densities.

Distribution	Notation	Density
Bernoulli	$\text{Bern}(\theta)$	$f(y\|\theta) = \theta^y(1-\theta)^{1-y}$
Binomial	$\text{Bin}(n,\theta)$	$f(y\|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$
Multinomial	$\text{Multi}(n;\theta_1,\theta_2,\ldots,\theta_K)$	$f(y\|\theta) = \frac{n!}{y_1!y_2!\ldots y_K!}\theta_1^{y_1}\theta_2^{y_2}\cdots\theta_K^{y_K}$
Beta	$\text{Beta}(a,b)$	$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}I_{(0,1)}(\theta)$
Uniform	$U(a,b)$	$p(\theta) = \frac{I_{(a,b)}(\theta)}{b-a}$
Poisson	$\text{Pois}(\theta)$	$f(y\|\theta) = \theta^y e^{-\theta}/y!$
Exponential	$\text{Exp}(\theta)$	$f(y\|\theta) = \theta e^{-\theta y}I_{(0,\infty)}(y)$
Gamma	$\text{Gamma}(a,b)$	$p(\theta) = [b^a/\Gamma(a)]\theta^{a-1}e^{-b\theta}I_{(0,\infty)}(\theta)$
Chi-squared	$\chi^2(n)$	Same as $\text{Gamma}(n/2, 1/2)$
Weibull	$\text{Weib}(\alpha,\theta)$	$f(y\|\theta) = \theta\alpha y^{\alpha-1}\exp\left(-\theta y^{\alpha}\right)I_{(0,\infty)}(\theta)$
Normal	$N(\theta, 1/\tau)$	$f(y\|\theta,\tau) = (\sqrt{\tau/2\pi})\exp\left[-\tau(y-\theta)^2/2\right]$
Student's t	$t(n,\theta,\sigma)$	$f(y\|\theta) = [1+(y-\theta)^2/n\sigma^2]^{(n+1)/2}$
		$\times\, \Gamma[(n+1)/2]/\Gamma(n/2)\sigma\sqrt{n\pi}$
Cauchy	$\text{Cauchy}(\theta)$	same as $t(1,\theta,1)$
Dirichlet	$\text{Dirichlet}(a_1,a_2,a_3)$	$p(\theta) = \Gamma(a_1+a_2+a_3)/\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)$
		$\times\, \theta_1^{a_1-1}\theta_2^{a_2-1}(1-\theta_1-\theta_2)^{a_3-1}$
		$\times\, I_{(0,1)}(\theta_1)I_{(0,1)}(\theta_2)I_{(0,1)}(1-\theta_1-\theta_2)$

Table 2: Means, Modes, and Variances.

Distribution	Mean	Mode	Variance
Bern(θ)	θ	0 if $\theta < .5$ 1 if $\theta > .5$	$\theta(1-\theta)$
Bin(n, θ)	$n\theta$	integer closest to $n\theta$	$n\theta(1-\theta)$
Beta(a, b)	$a/(a+b)$	$(a-1)/(a+b-2)$ if $a > 1, b \geq 1$	$ab/(a+b)^2(a+b+1)$
$U(a, b)$	$.5(a+b)$	everything a to b	$(b-a)^2/12$
Pois(θ)	θ	integer closest to θ	θ
Exp(θ)	$1/\theta$	0	$1/\theta^2$
Gamma(a, b)	a/b	$(a-1)/b$ if $a > 1$	a/b^2
$\chi^2(n)$	n	$n-2$ if $n > 2$	$2n$
Weib(α, θ)	$\Gamma[(\alpha+1)/\alpha]/\theta$	$[(\alpha-1)/\alpha]^{1/\alpha}/\theta$	$\Gamma[(\alpha+2)/\alpha]-\mu^2$
$N(\theta, 1/\tau)$	θ	θ	$1/\tau$
$t(n, \theta, \sigma)$	θ if $n \geq 2$	θ	$\sigma^2 n/(n-2)$ if $n \geq 3$
Cauchy(θ)	Undefined	θ	Undefined