Written Comprehensive Examination – Theory
Department of Statistics, UC Irvine
Friday, June 19, 2020, 9:00 am to 1:00 pm

- There are **7** questions on the examination. 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 problem should be written on separate sheets of paper. Label each sheet with your student identification number, the problem number, and the page number of that solution written in the upper right hand corner. For example, the labeling on a page may be:

  ID# 912346378
  Problem 2, page 3

- You have **4** hours to complete your solution. Please be prepared to turn in your exam at 1:00pm.

1. Let $N(t),\ t \geq 0$ denote the number of customers entering a store up to time t. Suppose $N(t)$ is a Poisson process with rate $\lambda > 0$ per hour.

   (a) Compute the conditional probability that $N(t_2) = 10$ given $N(t_1) = 3$ where $t_1 < t_2$.

   (b) Compute the conditional probability that $N(t_1) = 3$ given $N(t_2) = 10$ where $t_1 < t_2$.

   (c) Let $t_1 = 1$ (representing 1 hour after opening) and $t_2 = 2$, and suppose $N(t_1) = 3$ and $N(t_2) = 10$. Given this information, write down the likelihood function for λ and compute the MLE $\hat{\lambda}$.

   (d) Suppose, after entering the store, each customer spends a random amount of money with a uniform distribution between a and b $(a < b)$, independently of other customers. Let $S(t)$ denote the total spending by customers who enter the store by time t.

   (i) Compute $E(S(t))$ for fixed $t > 0$.

   (ii) Compute $Var(S(t))$ for fixed $t > 0$.

   (iii) Show that $S(t)/t$ converges in probability to some constant as $t \to \infty$. You need to identify this constant first.

END OF QUESTION (1)

2. Suppose $X_1, X_2, \cdots, X_n \overset{iid}{\sim} Exponential(\lambda)$ with pdf

$$f_X(x) = \lambda e^{-\lambda x} \ , \ x > 0,$$

where $\lambda > 0$. Define $S_n = \sum_{j=1}^{n} X_j$.

(a) Compute the moment generating functions for X_1 and S_n, and use them to show that S_n has a Gamma distribution.

(b) Prove that $E(1/S_n) \geq \lambda/n$.

(c) Find the joint distribution of $(X_1/S_n, 1/S_n)$ using transformations. Are X_1/S_n and $1/S_n$ independent?

(d) Use the delta method to show that $1/S_n$ is asymptotically normal (that is, find constants a_n and b_n such that $((1/S_n) - a_n)/b_n$ converges to $\mathrm{N}(0, 1)$ as $n \to \infty$).

END OF QUESTION (2)

3. Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables having probability density function for $\theta > 2$

$$f_X(x) = \theta x^{-(\theta+1)} \mathbf{1}_{[x>1]}.$$

(a) Find the maximum likelihood estimate $\hat{\theta}$.

(b) What is the asymptotic distribution of $\hat{\theta}$?

(c) Find the uniform minimum variance unbiased estimator of $g(\theta) = 1/\theta$ and derive its distribution.

(d) Find the efficiency of the UMVUE you found in part (c.)

END OF QUESTION (3)

4. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. Poisson random variables having probability mass function

$$p_X(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} 1_{x\in\{0,1,2,3,\ldots\}}.$$

(a) Find a maximum likelihood estimator for λ and and it's asymptotic distribution. Call your estimator $\widehat{\lambda}$.

(b) Find a transformation $g(\cdot)$ such that the asymptotic variance of $g(\widehat{\lambda})$ does not depend on λ.

(c) Find a maximum likelihood estimator for $\theta \equiv \Pr[X_i > 0]$ (ie. the probability that X_i takes on a non-zero count). Call your estimator $\widehat{\theta}$.

(d) Derive the asymptotic distribution of $\widehat{\theta}$.

(e) Define $W_i = 1_{[X_i>0]}, i = 1, 2, \ldots, n$. Find a consistent estimator of θ that is based upon $W_1, W_2, \ldots, W_n$. Call your estimator $\tilde{\theta}$. In larger samples would you prefer $\widehat{\theta}$ or $\tilde{\theta}$ as an estimator of θ. Why?

END OF QUESTION (4)

5. The broken line regression is a special case of nonlinear regression. When the broken points are known, one can fit this model using a multiple linear regression. For example, the following broken line regression extends the simple linear regression by including a known broken point at x_0:

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \epsilon_i \tag{1}$$

where $i = 1, \cdots, n$, $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, and z_i is defined as

$$z_i = \begin{cases} x_i - x_0 & \text{if } x_i > x_0 \\ 0 & \text{if } x_i \le x_0 \end{cases}$$

Our goal is to understand whether there is a broken point at x_0. This can be done by constructing a t-statistic (or equivalent, an F statistic) for $H_0 : \beta_2 = 0$. Please accomplish this goal by answering the following questions. To simply derivations/proofs, let's assume that the x_i's have been centered, i.e., $\sum_{i=1}^n x_i = 0$.

(a) Let's first fit a reduced model by ignoring z_i's. Show that the least squares estimate (LSE) of β_0 and β_1 are, respectively,

$$\hat{\beta}_0 = \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$$

Please keep in mind that x_i's have been centered, i.e., $\sum_{i=1}^n x_i = 0$

(b) Let H be the hat matrix in the reduced model (see (a)). Show that

$$H = \frac{\mathbf{1}\mathbf{1}^T}{n} + \frac{xx^T}{\sum_{i=1}^n x_i^2},$$

where $\mathbf{1}$ is the $n \times 1$ vector of 1's and $x = (x_1, \cdots, x_n)^T$.

(c) If the full model, i.e., the broken linear regression model in (1), is true, is $\hat{\beta}_1$ unbiased for β_1? Please explain your answer.

(d) Show that the LSE of β_2 is

$$\hat{\beta}_2 = \frac{z^T(I - H)Y}{z^T(I - H)z},$$

where $z = (z_1, \cdots, z_n)^T$ and $\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$.

(e) Derive a test statistic for $H_0 : \beta_2 = 0$. Be sure to provide all the details such as degrees of freedom. Necessary justifications should also be included. Note, the parameter σ^2 is unknown.

END OF QUESTION (5)

6. Please answer the following questions.

(a) F statistics in linear models can be constructed using quadratic forms. Consider $Y = X\beta + \epsilon$. Assume that Y is an $n \times 1$ vector and X is an $n \times p$ design matrix with rank p. Explain how you can use quadratic forms to construct an F-statistic for $H_0 : \beta = 0$. Be sure to (1) provided necessary assumptions (e.g., distributional assumptions) and (2) justify your answer.

(b) State the Gauss-Markov Theorem. Be sure to provide all necessary conditions. Proof is not needed.

(c) Describe the problem of simultaneous confidence intervals in multiple linear regressions.

(d) Please describe two methods for constructing simultaneous confidence intervals.

END OF QUESTION (6)

7. (a) Let X and Y be two independent random variables with densities f and g, respectively, both symmetric about 0. Let G be the distribution function corresponding to the density function g. Show that for any $\lambda \in \mathbf{R}$, the following

$$2G(\lambda y)f(y) \qquad y \in \mathbf{R} \tag{1}$$

is a density function in $\mathbf{R}$.

(b) If the pdf $f(y)$ and the distribution function $G(y)$ in (1) are the standard normal density function $\phi(y)$ and the standard normal distribution function $\Phi(y)$, then the density in (1) is called the *skew-normal pdf* with parameter λ, with short-hand notation SN(λ).
Let Z_1 and Z_2 be two independent standard normal random variables, and let $\lambda \in \mathbf{R}$. Show that

$$X = \frac{\lambda}{\sqrt{1+\lambda^2}}\,|Z_1| + \frac{1}{\sqrt{1+\lambda^2}}\,Z_2$$

follows a SN(λ) distribution.

(c) Show that if $X \sim SN(\lambda)$ and $Z \sim N(0,1)$, then $|X|$ and $|Z|$ are identically distributed.

(d) Show that if $X_1 \sim SN(\lambda_1)$ and $X_2 \sim SN(\lambda_2)$ are independent and admit the representation in (b), $X_1 + X_2$ is **not** necessarily distributed according to a skew-normal distribution with the representation in (b).
[***Hint:*** You might want to start by deriving the mgf of $X \sim SN(\lambda)$ in (b).]

END OF QUESTION (7)

Table 1: Common distributions and densities.

Distribution	Notation	Density
Bernoulli	Bern(θ)	$f(y\|\theta) = \theta^y(1-\theta)^{1-y}$
Binomial	Bin(n, θ)	$f(y\|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$
Multinomial	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	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	Pois(θ)	$f(y\|\theta) = \theta^y e^{-\theta}/y!$
Exponential	Exp(θ)	$f(y\|\theta) = \theta e^{-\theta y}I_{(0,\infty)}(y)$
Gamma	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 Gamma($n/2, 1/2$)
Weibull	Weib(α, θ)	$f(y\|\theta) = \theta\alpha y^{\alpha-1}\exp\left(-\theta y^{\alpha}\right)I_{(0,\infty)}(\theta)$
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	Cauchy(θ)	same as $t(1, \theta, 1)$
Dirichlet	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