Written Comprehensive Examination - Theory

Department of Statistics, UC Irvine

Friday, June 17, 2022, 9:00am to 1:00pm

- 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. Only write on one side of each sheet 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.

Problem 1. Conditional probabilities.

(a) Prove or disapprove: If $P(A|B) = P(A|B^c)$, then A and B are independent.

(b) Now, let $B_1, \ldots, B_n$ be mutually disjoint, and let $B = \cup_{j=1}^{n} B_j$. Suppose $P(B_j) > 0$ and $P(A|B_j) = p$ for $j = 1, \ldots, n$. Show that $P(A|B) = p$.

(c) A supplier of a certain testing device claims that his device has high reliability with $P(A|B) = P(A^c|B^c) = 0.95$, where $A = \{\text{device indicates component is faulty}\}$ and $B = \{\text{component is faulty}\}$. You hope to use the device to locate the faulty components in a large batch of components of which 5% are faulty. What is $P(B|A)$?

(d) Continue from (c). Now suppose you want $P(B|A) = 0.9$. Let $p = P(A|B) = P(A^c|B^c)$. How large does p have to be? (keep two decimal points.)

(End of Problem 1)

Problem 2. Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Define

$$\begin{aligned}
\bar{X}_k &= \frac{1}{k}\sum_{i=1}^{k} X_i, \\
\bar{X}^*_{n-k} &= \frac{1}{n-k}\sum_{i=k+1}^{n} X_i, \\
\bar{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i, \\
S_k^2 &= \frac{1}{k-1}\sum_{i=1}^{k}(X_i - \bar{X}_k)^2, \\
S^{*2}_{n-k} &= \frac{1}{n-k-1}\sum_{i=k+1}^{n}(X_i - \bar{X}_{n-k})^2, \\
S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.
\end{aligned}$$

Provide proper arguments to answer the following questions.

(a) What is the distribution of $\sigma^{-2}[(k-1)S_k^2 + (n-k-1)S^{*2}_{n-k}]$?;

(b) What is the distribution of $\frac{1}{2}(\bar{X}_k + \bar{X}^*_{n-k})$?

(c) What is the distribution of $\sigma^{-2}(X_i - \mu)^2$?

(d) What is the distribution of S_k^2/S^{*2}_{n-k}?

(e) What is the distribution of $(\bar{X} - \mu)/(S/\sqrt{n})$?

(End of Problem 2)

Problem 3. Let $X_1, \ldots, X_n$ be a random sample from the density

$$f(x;\theta) = \theta(1+x)^{-(1+\theta)} 1_{(0,\infty)}(x), \;\; \theta > 0.$$

.

(a) Estimate θ by the method of moments (assuming $\theta > 1$.)

(b) Find the maximum likelihood estimator of $1/\theta$.

(c) Find a complete and sufficient statistic if one exists.

(d) Find the Cramer-Rao lower bound for unbiased estimators of $1/\theta$.

(e) Find the UMVUE of $1/\theta$ if such exists.

(f) Find the UMVUE of θ if such exists.

(End of Problem 3)

Problem 4 Let $X_1, \ldots, X_n$ be a random sample from the uniform distribution over the interval $(\theta, \theta + 1)$. To test $H_0 : \theta = 0$ versus $H_1 : \theta > 0$, the following test was used:

$$\text{Reject } H_0 \text{ if and only if } Y_n \geq 1 \text{ or } Y_1 \geq k,$$

where Y_1 and Y_n are the first and the n-th order statistics, respectively, and k is a constant.

(a) Determine k so that the test will have size α.

(b) Find the power function of the test you obtained in part (a).

(c) Prove or disprove: If k is selected so that the test has size α, then the given test is uniformly most powerful of size α.

(End of Problem 4)

Problem 5.

Consider the following simple form of the discriminant analysis model for bivariate data with binary X and continuous Y. Suppose X is Bernoulli with $\Pr(X=1) = 1 - \Pr(X=0) = p \in (0,1)$, and Y given $X = j$ is normal with mean μ_j, variance $\sigma^2 > 0,\ j = 0, 1$.

(a) Find the *marginal* mean and variance of Y.

(b) Suppose we observe a random sample $(x_i,\, y_i),\ i = 1, \ldots, n$ on X and Y. Derive maximum likelihood estimators of $(p,\, \mu_0,\, \mu_1,\, \sigma^2)$.

(c) Consider the estimator $\tilde{\mu}_1 \equiv \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i}$. Show that $\tilde{\mu}_1$ is consistent for μ_1 as $n \to \infty$.

(d) Now consider a Bayesian approach based on the random sample in part (b). Assume the joint prior

$$f(p, \mu_0, \mu_1, \sigma^2) \propto p^{1/2}(1-p)^{1/2}.$$

Write down the joint posterior of $(p,\, \mu_0,\, \mu_1,\, \sigma^2)$ and compute the marginal posterior of p given observed data.

(e) In the setting of part (d) derive the conditional posterior of (μ_0, μ_1) given (p, σ^2) (and observed data). Identify this well known distribution by name and parameter.

(End of Problem 5)

Problem 6.

Let $Y \in \mathbf{R}^n$ be the vector of n observations. Let $X_{n\times p}$ be the design matrix in the linear model

$$Y = X\beta + \epsilon,$$

where $\beta \in \mathbf{R}^p$ is the vector of coefficients , and $\epsilon \in \mathbf{R}^n$ is the vector of random errors. Answer the following questions.

(a) When $rank(X) < p$, are we able to estimate $a^T\beta$ for any arbitrary $a \in \mathbf{R}^p$? Justify your answer.

(b) We discussed the Gauss-Markov theorem in class. What are the minimal assumptions of ϵ do we need in the Gauss-Markov theorem?

(c) The Gauss-Markov theorem is about BLUE. Explain what BLUE is (proof is not required).

For the rest of questions, we assume that $rank(X) = p$. Let's consider the problem of predicting a new observation Y_0 at $x_0 \in \mathbf{R}^p$, i.e., $Y_0 = x_0^T\beta + \epsilon_0$, where

- $E(\epsilon_0) = 0$ and $Var(\epsilon_0) = \sigma^2$.
- $E(\epsilon) = \mathbf{0}_{n\times 1}$ and $Cov(\epsilon) = \sigma^2\mathbf{I_n}$
- ϵ_0 is independent of ϵ.

(d) Let $\hat{Y}_0 = x_0^T\hat{\beta}$, where $\hat{\beta}$ is the OLSE from $Y = X\beta + \epsilon$. Show that $E(\hat{Y}_0) = E(Y_0)$, i.e., $\hat{Y}_0$ is an unbiased predictor of Y_0.

(e) It is obvious that the $\hat{Y}_0$ defined in (d) is linear in Y. Now, consider another linear unbiased predictor $\tilde{Y}_0 = d^TY$. Show that d must satisfy $d^TX = x_0^T$.

(f) Prove that the $\hat{Y}_0$ in (d) is the best linear unbiased predictor (BLUP) of Y_0 where "best" is defined in terms of minimum variance.

(End of Problem 6)

Problem 7.

Consider the following linear model

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

where Y is an $n \times 1$ random vector, β_0 is the intercept (a scalar), $\beta \in \mathbf{R}^p$, $\mathbf{1}$ is the vector of n 1's, X is an $n \times p$ design matrix with rank p, and $\epsilon \sim N(0, \sigma^2 \mathbf{I_n})$. We would like to assess the null hypothesis $H_0 : \beta = \mathbf{0}$.

(a) Show that the model can be re-parameterized to

$$Y = \mathbf{1}\alpha_0 + X_c\beta + \epsilon.,$$

where $X_c = (\mathbf{I_n} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)X$, i.e., the column-centered design matrix. Be sure to express α_0 as a function of β_0 and β.

(b) Show that the OLSE of β is $\hat{\beta} = (X_c^T X_c)^{-1} X_c^T Y$.

(c) Show that $\hat{\beta}_0 = \bar{Y} - \frac{\mathbf{1}^T X}{n}\hat{\beta}$, where $\bar{Y} = \mathbf{1}^T Y/n$.

(d) Is $\hat{\beta}$ independent of $\bar{Y}$? Justify your answer.

(e) Derive the distribution of $\hat{\beta}$.

(f) Show that the residual sum of squares is $RSS = Y^T(\mathbf{I_n} - \frac{1}{n}\mathbf{1}\mathbf{1}^T - X_c(X_c^T X_c)^{-1} X_c^T)Y$.

(g) Find the distribution of RSS/σ^2. Be sure to provide the degrees of freedom of the distribution.

(h) Construct an F-statistic for $H_0 : \beta = 0$. Feel free to use the results in (a)-(g).

(End of Problem 7)

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	$t(n,\theta,\sigma)$	$f(y\|\theta) = \left[1+(y-\theta)^2/n\sigma^2\right]^{(n+1)/2}$
		$\times\,\Gamma[(n+1)/2]/\Gamma(n/2)\sigma\sqrt{n\pi}$
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