

First Year Examination
Department of Statistics, University of Florida
May 7, 2010, 8:00 am - 12:00 noon

Instructions:

1. You have four hours to answer questions in this examination.
2. You must show your work to receive credit.
3. Questions 1 through 5 are the “theory” questions and questions 6 through 10 are the “applied” questions. You must do exactly four of the theory questions and exactly four of the applied questions
4. **Write your answers to the theory questions on the blank paper provided. Write only on one side of the paper, and start each question on a new page.**
5. **Write your answers to the applied questions on the exam itself.**
6. While the 10 questions are equally weighted, some questions are more difficult than others.
7. The parts within a given question are not necessarily equally weighted.
8. You are allowed to use a calculator.

The following abbreviations and terminology are used throughout:

- ANOVA = analysis of variance
- CRD = completely randomized design
- iid = independent and identically distributed
- mgf = moment generating function
- ML = maximum likelihood
- MOM = method of moments
- pdf = probability density function
- pmf = probability mass function
- UMP = uniformly most powerful
- $\mathbb{Z}^+ = \{0, 1, 2, 3, \dots\}$
- $\mathbb{N} = \{1, 2, 3, \dots\}$
- $\mathbb{R}^+ = (0, \infty)$

You may use the following facts/formulas without proof:

Beta density: $X \sim \text{Beta}(\alpha, \beta)$ means X has pdf

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} I_{(0,1)}(x)$$

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

Gamma density: $X \sim \text{Gamma}(\alpha, \beta)$ means X has pdf

$$f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} I_{(0,\infty)}(x)$$

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

Negative Binomial mass function: $X \sim \text{NB}(r, p)$ means X has pmf

$$P(X = x) = \binom{r+x-1}{x} p^r (1-p)^x I_{\mathbb{Z}^+}(x),$$

where $p \in (0, 1)$ and $r \in \mathbb{N}$. Also, $E(X) = r(1-p)/p$ and $\text{Var}(X) = r(1-p)/p^2$.

Iterated Expectation Formula: $E(X) = E[E(X|Y)]$.

Iterated Variance Formula: $\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}[E(X|Y)]$.

1. Suppose we have two urns. Urn I contains three yellow balls, four red balls, and two green balls. Urn II contains one yellow ball, two green balls, and four purple balls. Consider a two-stage experiment in which we randomly draw three balls from Urn I and move them to Urn II, and then we randomly draw one ball from (the new) Urn II.

(a) Define two events as follows:

$$A = \{ \text{Two yellow balls and one green ball are moved from Urn I to Urn II} \}$$

and

$$B = \{ \text{A green ball is drawn from Urn II} \} .$$

Find the probabilities of these two events. (Express all numerical answers in this problem as ratios of integers or sums of ratios of integers - do not use any decimals.)

(b) Are A and B independent?

(c) Find the probability that at least two of the balls moved from Urn I to Urn II were yellow, given that the ball drawn from Urn II was yellow.

2. Let (X, Y) be a bivariate random vector with joint pdf

$$f_{X,Y}(x, y) = c \left[I_{(0,1)}(x) I_{(-1,0)}(y) - x I_{(-1,0)}(x) I_{(0,1)}(y) \right] ,$$

where c is an unknown normalizing constant. Let $S \in \mathbb{R}^2$ denote the support of the density $f_{X,Y}(x, y)$.

(a) Find the value of c .

(b) Let $A \subset \mathbb{R}^2$ denote the triangle whose vertices are at the points $(-1/2, 0)$, $(1/2, -1/3)$ and $(1/2, 2)$. Find $P((X, Y) \in A)$.

(c) Calculate the marginal pdf of X .

(d) Calculate the conditional pdf of Y given $X = x$.

(e) Find the conditional expectation of Y given that $X = -1/4$.

(f) Find a second joint pdf, call it $f_{X,Y}^*(x, y)$, that has exactly the same marginals as $f_{X,Y}(x, y)$, but whose support, call it S^* , has twice as large an area as S .

(g) Find a third joint pdf, call it $\tilde{f}_{X,Y}(x, y)$, that has exactly the same x -marginal as $f_{X,Y}(x, y)$, but whose support, call it \tilde{S} , satisfies S

3. Suppose that the conditional pmf of Z given θ is

$$P(Z = z | \theta) = \theta (1 - \theta)^z I_{\mathbb{Z}^+}(z),$$

and that, marginally, $\theta \sim \text{Beta}(\alpha, \beta)$.

- Find the marginal pmf of Z .
- Derive a formula for $E[\theta^a(1 - \theta)^b]$ that involves only α, β, a, b and the gamma function. Does this formula hold for all $(a, b) \in \mathbb{R}^2$?
- Calculate the marginal mean and variance of Z , assuming that $\alpha > 2$. (Hint: Use iterated expectation and variance.)

For the remainder of this problem, assume that Z_1, \dots, Z_n are iid with common pmf given by

$$P(Z = z; \alpha) = \frac{\alpha^2 (\alpha + 1) \Gamma(\alpha) (z + 1)!}{\Gamma(\alpha + z + 3)} I_{\mathbb{Z}^+}(z),$$

where $\alpha > 2$.

- Find the MOM estimator of α and call it $\tilde{\alpha}_n = \tilde{\alpha}_n(z)$.
- Show that

$$\sqrt{n}(\tilde{\alpha}_n - \alpha) \xrightarrow{d} \mathbf{N}(0, v(\alpha)),$$

and identify the function $v(\cdot)$.

4. Let X_1, \dots, X_n be iid random variables from the pdf

$$f(x; \theta) = \frac{2x}{\theta} e^{-x^2/\theta} I_{\mathbb{R}^+}(x)$$

where $\theta > 0$.

- What kind of parameter is θ : location, scale or neither? (Explain your answer.)
- Derive a formula for EX^p that holds for any $p > -2$.
- Find a function of n , call it $g(n)$, such that $g(n) \sum_{i=1}^n X_i^4$ is unbiased for θ^2 .
- Find the ML estimator of θ^2 .
- Find the Cramér-Rao Lower Bound for the variance of an unbiased estimator of θ^2 .
- Find the best unbiased estimator of θ^2 .

5. Suppose that Y_1, \dots, Y_n are iid $\text{NB}(r, q)$.

(a) Derive the mgf of Y_1 .

(b) Derive the distribution of $\sum_{i=1}^n Y_i$.

(c) Suppose we have four *identical* quarters and we would like to test $H_0 : p \leq 1/2$ vs. $H_1 : p > 1/2$, where p is the unknown probability that any one of the quarters comes up heads when it is flipped. We decide to perform an experiment. Each quarter will be picked up and flipped repeatedly until the first head occurs. Let X_i denote the number of *tails* that occur before the first head occurs when the i th quarter is flipped, $i = 1, 2, 3, 4$. Use the X_i s to construct a UMP size 0.1875 test of H_0 vs. H_1 .

Q.7. Derive the expected mean squares for treatments and error for the balanced 1-Way ANOVA with g

treatments and n replicates per treatment: $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ $\sum_{i=1}^g \alpha_i = 0$ $\varepsilon_{ij} \sim NID(0, \sigma^2)$

Q.8. A multiple regression model is fit based on $n=30$ individuals, relating Y to X_1 and X_2 (the model also contains an intercept). The coefficient of determination is $R^2 = 0.90$. Note: $\sum_{i=1}^{30} (Y_i - \bar{Y})^2 = 17652$.

p.8.a. Complete the Analysis of Variance and test $H_0: \beta_1=\beta_2=0$ at the 0.05 significance level.

Source	df	SS	MS	F	F(.05)
Regression					
Residual					
Total					

p.8.b. Based on the least squares estimate of the parameter vector, and $(X'X)^{-1}$ given below, test $H_0: \beta_1=\beta_2$ at the 0.05 significance level

$(X'X)^{-1}$				Beta-hat
0.1714	-0.0143	-0.0167		96.37
-0.0143	0.0029	0.0000		5.29
-0.0167	0.0000	0.0042		5.36

Q.9. An experiment is conducted as a Randomized Complete Block Design with 3 treatments applied to 8 blocks. The model fit is given below, as well as the data:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad i = 1, \dots, 3 \quad j = 1, \dots, 8 \quad \sum_{i=1}^3 \alpha_i = 0 \quad \beta_j \sim NID(0, \sigma_\beta^2) \quad \varepsilon_{ij} \sim NID(0, \sigma^2) \quad COV(\beta_j, \varepsilon_{ij}) = 0 \quad \forall i, j, j'$$

	Trt1	Trt2	Trt3	Mean
Block1	10	12	14	
Block2	16	18	17	
Block3	8	10	9	
Block4	20	22	24	
Block5	19	21	20	
Block6	17	17	17	
Block7	18	16	26	
Block8	4	12	17	
Mean				

$$\sum_{i=1}^3 \sum_{j=1}^8 (y_{ij} - \bar{y}_{..})^2 = 688$$

p.9.a. Complete the following ANOVA table, and test whether treatment effects exist at the 0.05 significance level. $H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$

Source	df	SS	MS	F	F(.05)
Trts					
Blocks					
Error					
Total					

p.9.b. Use Bonferroni's method to obtain simultaneous 95% confidence intervals for all pairs of differences among treatment means.

This question is based on the following regression model, and \mathbf{X} is of full column rank (no linear dependencies among predictor variables). $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ $\mathbf{X} \equiv n \times p'$ $\boldsymbol{\beta} \equiv p' \times 1$ $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$. For a symmetric matrix \mathbf{A} :

Given: $\frac{d(\mathbf{a}'\mathbf{x})}{d\mathbf{x}} = \mathbf{a}$ $\frac{d(\mathbf{x}'\mathbf{A}\mathbf{x})}{d\mathbf{x}} = 2\mathbf{A}\mathbf{x}$ $E(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = \text{tr}(\mathbf{A}\mathbf{V}_Y) + \boldsymbol{\mu}_Y' \mathbf{A}\boldsymbol{\mu}_Y$

Cochran's Theorem: Suppose \mathbf{Y} is distributed as follows with nonsingular matrix \mathbf{V} :

$\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{V})$ $r(\mathbf{V}) = n$ then if $\mathbf{A}\mathbf{V}$ is idempotent :

$\mathbf{Y}'\left(\frac{1}{\sigma^2} \mathbf{A}\right)\mathbf{Y}$ is distributed non - central χ^2 with : (a) $df = r(\mathbf{A})$ and (b) Noncentrality parameter : $\Omega = \frac{1}{2\sigma^2} \boldsymbol{\mu}' \mathbf{A}\boldsymbol{\mu}$

Q.10.a. Derive the least squares estimator for $\boldsymbol{\beta}$

Q.10.b. Obtain $\frac{SS(\text{Model})}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n \left(\hat{Y}_i\right)^2$ and $\frac{SS(\text{Residual})}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n e_i^2$ in matrix form (quadratic forms in \mathbf{Y}).

Obtain the distributions of the two sum of squares (be specific with regard to their family of distributions, degrees of freedom, and non-centrality parameters).