Statistics Qualifying Exam

June 14, 2007

Name :

1. Let the random variable X have the Poisson pmf, $p(x) = \frac{m^x e^{-m}}{x!}$, $x = 0, 1, 2, 3, \ldots$ and 0, elsewhere. Derive the moment generating function for X.

2. Let X_1, X_2, X_3, X_4 be four iid random variables with the same pdf, $f(x) = 2x$, $0 < x < 1$, and 0, elsewhere. Find the pdf of $Y = min\{X_1, X_2, X_3, X_4\}$.

3. A fair die is cast and let $X = 0$ if $1, 2$, or 3 spots appear, let $X = 1$ if 4 or 5 spots appear, let $X = 2$ if 6 spots appear. Do this two independent times, obtaining X_1 and X_2. Compute $P(|X_1 - X_2| = 1)$.

4. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables from the uniform distribution on the interval $(0, 1)$.

   (a) Derive the cumulative distribution function of $X_{(i)}$, the i^{th} order statistic, for $i = 1, 2, \ldots, n$.

   (b) Derive the joint density function of $X_{(1)}$ and $X_{(n)}$.

   (c) Derive the density function of the range $R = X_{(n)} - X_{(1)}$.

5. Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$, $\mu \in (-\infty, \infty)$ and $\sigma^2 \in (0, \infty)$. Assume that σ^2 is known. We wish to estimate the parametric function $\tau(\mu) = \mu^2$ unbiasedly.

   (a) Based on X_1, produce an unbiased estimator of $\tau(\mu) = \mu^2$.

   (b) Obtain the uniformly minimum variance unbiased estimator (UMVUE) for $\tau(\mu) = \mu^2$.

6. Suppose that $X_1, \ldots, X_n$ are i.i.d. $Uniform(0, 4)$ random variables. Denote $U_n = \bar{X}_n$ for $n > 1$.

   (a) Show that $\sqrt{n}(U_n - 2) \xrightarrow{D} N(0, \frac{4}{3})$.

   (b) What is the limiting distribution of $\sqrt{n}(U_n^2 - 4)$?

7. It's claimed that many commercially manufactured dice are not fair. We roll such a dice 6000 times, 921 times we obtain "6". Perform the testing at alpha = 0.05 level. Find out the P-value and critical region.

8. It is known that a sample of 12; 11:2; 13:5; 12:3; 13:8; 11:9 comes from a population with the density function

$$f(x;\theta) = \begin{cases} \theta/x^{\theta+1}, & x > 1 \\ 0 & otherwise \end{cases}$$

where $\theta > 0$.

(a) Find the maximum likelihood estimate for θ.

(b) If the parameter is constrained in $\theta > 10$, find the MLE (provide step to show that it indeed achieves the maximum likelihood in the constrained parameter space.)

9. For ANOVA models with unequal sample size, we know that their type I and type III sum of squares are not the same. This is due to the fact that Type I SS weights each observation equally, while Type III SS weights each treatment equally.

Consider the bone data set where factor A is gender, $a = 2$ levels: male, female; and factor B is bone development, $b = 3$ levels: severely, moderately, or mildly depressed. The sample sizes are $3, 2, 2$ for male and $1, 3, 3$ for female. We use contrast to see what is being calculated for type I and type III SS. For type I contrast for gender effect, the hypothesis is $H_0 : \mu_{11} + \mu_{12} + \mu_{13} = \mu_{21} + \mu_{22} + \mu_{23}$, where $\mu_{11} = \mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11} \ldots$.
The contrast statement is

contrast 'gender Type III'
gender 3 - 3
gender*bone 1 1 1 -1 -1 -1;

Do the same for type III contrast for <u>bone</u> effect.

10. We are interested in examining the effect of Age, x, (in years) on triglycerides, y. We have 8 subjects: 4 of the subjects are exactly 20 years old, 1 is exactly 25 years old, 2 are exactly 30 years old, and 1 subject is exactly 35 years old. The least squares regression line is computed to be: $\hat{y} = 30 + 5x$.

Complete the ANOVA table below (fill in the blanks).

Use the MSE for the Model F-ratio:

Source of variation	Sum of Squares	Degree of Freedom	Mean Square	F-ratio
Model				
Error				
Lack of Fit				5.0
Pure Error				
Total (corrected)	8000			

11. Consider the following sas glm output from the "means" statement for 2 way ANOVA.

(a) Give the degree of freedoms for the main effects, interaction and error.

(b) Write down the factor effect model with model assumptions.

(c) Assume zero sum constraint. Compute the estimates for the parameters μ, α_i, β_j, $(\alpha\beta)_{ij}$.

The GLM Procedure

Level of height	N	sales Mean	sales Std Dev
1	4	44.0000000	3.16227766
2	4	67.0000000	3.74165739
3	4	42.0000000	2.94392029

Level of width	N	sales Mean	sales Std Dev
1	6	50.0000000	12.0664825
2	6	52.0000000	13.4313067

Level of height	Level of width	N	sales Mean	sales Std Dev
1	1	2	45.0000000	2.82842712
1	2	2	43.0000000	4.24264069
2	1	2	65.0000000	4.24264069
2	2	2	69.0000000	2.82842712
3	1	2	40.0000000	1.41421356
3	2	2	44.0000000	2.82842712