INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

08th November 2011

Subject CT3 – Probability & Mathematical Statistics

Time allowed: Three Hours (15.00 – 18.00)

Total Marks: 100

INSTRUCTIONS TO THE CANDIDATES

1. *Please read the instructions on the front page of answer booklet and instructions to examinees sent along with hall ticket carefully and follow without exception*

2. *Mark allocations are shown in brackets.*

3. *Attempt all questions, beginning your answer to each question on a separate sheet. However, answers to objective type questions could be written on the same sheet.*

4. *In addition to this paper you will be provided with graph paper, if required.*

5. *Please check if you have received complete Question Paper and no page is missing. If so kindly get new set of Question Paper from the Invigilator*

AT THE END OF THE EXAMINATION

Please return your answer book and this question paper to the supervisor separately.

Q. 1) Let X be a discrete random variable taking values 0, 1 and 2 with probabilities p, 1-2p and p respectively with $0 \le p \le 0.5$.

Find the value of p for which Var (X) will be maximum.

[3]

Q. 2) Verify whether f(x), as defined below, is a probability density function:

$$f(x) = \begin{cases} \frac{1}{18}(3+2x) & for\ 2 \le x \le 4 \\ 0 & otherwise \end{cases}$$

(2)

If yes, find P[$2 \le X \le 3$]. (2)

[4]

Q. 3) The joint density function of X and Y is given by

$$f(x,y) = \frac{e^{-yx^2/2}}{\sqrt{2\pi/y}}.ye^{-y}, \quad -\infty < x < \infty,\ y > 0$$

a) State the probability density functions and hence identify the statistical distributions of

i) Y (1)

ii) X conditional on Y = y. (1)

b) Compute E(Y) and Var(Y). (2)

c) Compute E(X|Y = y) and Var(X|Y = y). (2)

d) Hence, compute E(X) and Var(X). (4)

[10]

Q. 4) You are the producer for the television show Actuarial Idol. Each year, 1000 actuarial clubs audition for the show. The probability of a club being accepted is 0.20.

The number of members of an accepted club has a distribution with mean 20 and variance 20. Club acceptances and the number of club members are mutually independent.

Your annual budget for persons appearing on the show equals 10 times the expected total number of persons plus 10 times the standard deviation of the total number of persons.

Calculate your annual budget for persons appearing on the show.

[5]

Q. 5) Food packages, after filling, are weighed on an automatic device which rejects all packages below 5.155 kg as they are considered "under-weight". All rejected packages are carefully weighed and their weights recorded. The table below represents the frequency distribution obtained from a sample of 200 such weights:

Weights of rejected packages (kg)	**Frequency**
5.075 - 5.085	1
5.085 - 5.095	3
5.095 - 5.105	3
5.105 - 5.115	7
5.115 - 5.125	10
5.125 - 5.135	23
5.135- 5.145	56
5.145 - 5.155	97
	200

a) Determine the mean and standard deviation of this sample (6)

Assume the following:

- The distribution of weights of the filled packets are normal
- The distribution of weights of the rejected packets are normal
- The standard deviation of the whole distribution may be estimated as 4 times the standard deviation of this sample

b) Using the fact that the overall probability to reject a package is 0.05:

i) Show that an estimate of the mean weight of the whole distribution is 5.246 kg. (3)

ii) Hence, estimate the mean weights of the packages which are not rejected. (1)

[10]

Q. 6) A lamp requires a particular type of light bulb whose lifetime has a distribution with mean 3 (months) and variance 1. As soon as a bulb burns out, it is replaced with a new one.

What is the smallest number of bulbs you need to purchase so that, with probability at least 0.9772, the lamp burns for at least 40 months?

[Hint: Use Central Limit Theorem; $\Phi(2) = 0.9772$]

[5]

Q. 7) Suppose you roll 10 four-faced dice, with faces labeled 1, 2, 3, 4, and each equally likely to appear on top.

Let X_i denote the number showing up on the i^{th} die when rolled for i = 1, 2 … 10. Assume X_i-s are independent. Let $S = X_1 + X_2 + \ldots + X_{10}$ denote the sum of the 10 numbers obtained.

Show that the moment generating function, $M_S(t)$, of S is given by:

$$M_S(t) = \left[\frac{e^t - e^{5t}}{4\,(1-e^t)}\right]^{10}$$

[3]

Q. 8) Ayush and Shriya play a game as follows:

- Ayush selects a positive integer θ ε { 1, 2, 3 … } at random and writes it down on a piece of paper
- Shriya has two chances to ask Ayush what the number θ is.
- Each time Ayush tosses a biased coin secretly, and reports to Shriya the number θ - 1 if a head comes up and the number θ + 1 if a tail comes up.
- Shriya has to guess the true value of θ based on the two numbers reported by Ayush after two independent tosses of the coin.

It is known to Shriya that the coin has a probability 1/3 of showing a head. Let X and Y be the two numbers which Ayush reported.

Shriya considers the following two estimators of θ:

- $T = \frac{1}{2}(X+Y) - \frac{1}{3}$
- $S = \begin{cases} \frac{1}{2}(X+Y) & \text{if } X \neq Y \\ X - \frac{1}{3} & \text{if } X = Y \end{cases}$

a) Show that:

$$P(S=\theta) = \frac{4}{9}, \quad P\left(S = \theta - \frac{4}{3}\right) = \frac{1}{9}, \quad P\left(S = \theta + \frac{2}{3}\right) = \frac{4}{9}.$$ (4)

b) Calculate the bias of S as an estimator of θ. Is S unbiased? (2)

c) Calculate the mean squared error of S as an estimator of θ. (2)

It can be shown that T is an unbiased estimator of θ with variance $\frac{4}{9}$.

d) Which of the estimators T or S should Shriya use then if she wants to minimize the error of her estimation? (2)

[10]

Q. 9) The table below contains information about the annual inflation rates (averaged over an past decade) of 22 countries, where for i = 1, 2, … ,11

- x_i = log$_e$(inflation rate) for the i^{th} developing country,
- y_i = log$_e$(inflation rate) for the i^{th} developed country.

Developing countries			Developed countries		
Nation	x_i	$(x_i - \bar{x})^2$	**Nation**	y_i	$(y_i - \bar{y})^2$
COSTA RICA	-1.4697	0.0223	GERMANY	-3.5066	0.7195
BAHAMAS	-2.8134	2.2288	FINLAND	-2.6593	0.0000
ZAIRE	-0.7985	0.2725	AUSTRALIA	-2.5257	0.0176
BARBADOS	-2.6593	1.7924	ITALY	-2.2073	0.2035
UGANDA	-0.3285	0.9840	DENMARK	-2.6593	0.0000
URUGUAY	-0.7985	0.2725	LUXEMBERG	-2.9957	0.1138
TURKEY	-0.8916	0.1839	FRANCE	-2.6593	0.0000
TANZANIA	-1.3093	0.0001	U. KINGDOM	-2.6593	0.0000
PERU	0.0770	1.9530	S. AFRICA	-1.9661	0.4793
EGYPT	-0.3147	1.0116	IRELAND	-2.4079	0.0627
ETHIOPIA	-3.2189	3.6040	BELGIUM	-2.9957	0.1138
Total	**-14.5254**	**12.3251**	**Total**	**-29.2422**	**1.7101**

Assume that $(x_1, \ldots, x_{11})$ and $(y_1, \ldots, y_{11})$ are realizations of two independent normal random samples, each of size 11, drawn from $N(\mu_x, \sigma_x^2)$ and $N(\mu_y, \sigma_y^2)$ respectively.

a) Calculate an unbiased estimate of σ_x^2 and an unbiased estimate of σ_y^2. (2)

b) Calculate a 95% equal-tailed confidence interval for σ_x^2/σ_y^2.

β ~ upper quantiles of $F_{a,b}$						
β	0.025	0.050	0.100	0.900	0.950	0.975
a=10, b=10	3.717	2.978	2.323	0.431	0.336	0.269
a=10, b=11	3.526	2.854	2.248	0.434	0.340	0.273
a=11, b=10	3.665	2.943	2.302	0.445	0.350	0.284
a=11, b=11	3.474	2.818	2.227	0.449	0.355	0.288

(5)

c) Describe how you would make use of the confidence interval found in (b) to test

$$H_0: \sigma_x^2/\sigma_y^2 = c \text{ against } H_1: \sigma_x^2/\sigma_y^2 \neq c,$$

for some specified constant 'c' at the 5% level. (2)

What is your conclusion for the test if c = 30? (1)

[10]

Q. 10) A model for the spread of a measles epidemic initiated by one infective in a household of size 3 predicts that the total number of people infected in the household has the following distribution:

Total number infected:	1	2	3
Probability :	$(1-\theta)^2$	$2\theta(1-\theta)^2$	$\theta^2(3-2\theta)$

Here θ, the probability of adequate contact, is an unknown parameter.

The following data gives the frequency distribution of 334 such measles epidemics.

Total number infected:	1	2	3
Frequency :	34	25	275

a) Show that the maximum likelihood estimate of θ is 0.728. (6)

b) Carry out a chi-square goodness of fit test at 5% level of significance to examine if the model provides a good fit to these data. (6)

[12]

Q. 11) Six insurance companies were being compared with regards to premium being charged for house contents insurance for houses in a particular city. Independent random samples of five policies from each company are examined and the premiums were recorded.

Company	Premium Amounts (y)					Σy	Σy^2
A	151	168	128	167	134	748	113,254
B	152	141	129	120	115	657	87,251
C	175	155	162	186	148	826	137,394
D	149	148	137	138	169	741	110,479
E	123	132	142	161	152	710	101,742
F	145	131	155	172	141	744	111,676
						4,426	**661,796**

a) Compute an ANOVA table for these data, and show that there are no significant differences, at the 5% level, between mean premiums being charged by each company. (5)

b) A colleague points out that Company C has the largest mean premium of 165.2 and that Company B has the smallest mean premium of 131.4 and suggests performing t-test to compare these two companies.

i) Perform this t-test, using the estimate of variance from the ANOVA table, and in particular show that there is a significant difference at the 1% level. (3)

ii) Your colleague states that there is therefore a significant difference between the six companies. Discuss the apparent contradiction with your conclusion in part (a). (2)

[10]

Q. 12) An experiment is being conducted to check the efficacy of a newly invented medicine on mice. Let y_i be the value of the response variable for i^{th} mouse receiving a dose of x_i for i = 1, 2, …, 12.

A linear regression model is proposed to examine the impact of this medicine and is assumed to take the following form:

$$y_i = \alpha + \beta(x_i - \bar{x}) + e_i, \quad i = 1, 2, \ldots, 12$$

Here: $\bar{x} = \frac{1}{12}\Sigma x$ and e_i-s are independent normal variables with mean zero and unknown variance σ^2.

You are given the following summarized values for the data:

- Σy = 716
- Σy^2 = 43,827
- Σx = 453
- Σx^2 = 17,372
- Σxy = 27,503

a) Using the results given in the actuarial tables:

i) Show that the least square estimates for α and β are given by:

$$\hat{\alpha} = \bar{y} \quad \text{where} \quad \bar{y} = \frac{1}{12}\sum y$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} \quad \text{where} \quad \left.\begin{aligned} S_{xy} &= \sum xy - 12\bar{x}\bar{y} \\ S_{xx} &= \sum x^2 - 12\bar{x}^2 \end{aligned}\right\}$$ (3)

ii) Show that the estimate of σ^2 is given by:

$$\hat{\sigma}^2 = \frac{1}{10}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) \text{ where } S_{yy} = \sum y^2 - 12\bar{y}^2$$ (1)

b) Hence, deduce the equation for the regression line using the summarized data given above. (3)

c) Devise a t-test for testing the following hypotheses: H_o: β = 0 against H_1: β > 0. State your conclusion at 5% significance level. (4)

d) It has now been discovered that there was a mistake in recording the values of the response variable. The correct value should have been $10y_i$

i) Show that the least square estimate of β is $\hat{\beta}^* = 10\hat{\beta}$ where $\hat{\beta}$ is as defined in part (a) earlier. (2)

ii) Show that the estimate of σ^2 is $\hat{\sigma}^{*2} = 100\,\hat{\sigma}^2$ where $\hat{\sigma}^2$ is as defined in part (a) earlier (2)

iii) Hence establish how your t-test for the problem given in part (c) will change as a result? Restate your conclusion at the same 5% significance level. (3)

[18]

****