NATIONAL UNIVERSITY OF SINGAPORE

FINAL EXAMINATION

ST2334 Probability and Statistics

(Semester 1: AY 2005–2006)

November 2005 — Time Allowed : 2 Hours

INSTRUCTIONS TO CANDIDATES

1. This examination paper contains **FOUR(4)** questions and comprises **THREE(3)** printed pages.
2. Answer **ALL** the questions for TOTAL **60** marks.
3. Read the questions **CAREFULLY**, and label some important quantities clearly at your own convenience.
4. This is a **CLOSED** book examination. Students are allowed to bring in a piece of A4 help-sheet with both sides written.
5. Any textbooks or lecture notes are not allowed.

1. **[15 marks]** Let X and Y be continuous random variables with joint density function

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{x}{5} + \dfrac{y}{20}, & 0 < x < 1, 1 < y < 5 \\ 0, & \text{elsewhere} \end{cases}.$$

   (a) Find the marginal density of X.

   (b) Show that the conditional density of $Y|X = x$ is given by

$$f_{Y|X}(y|x) = \frac{4x + y}{4\,(4x + 3)}, \qquad 1 < y < 5.$$

   Argue why X and Y are dependent.

   (c) Suppose $X = 0.5$. Find the conditional cumulative distribution function of Y.

   (d) Find $P(X + Y > 3)$.

2. **[17 marks]** An Internet Service Provider believes that the daily usage rate (e.g., $=$ $1/3$ if a customer spends 8 hours online in a day) of their customers, X, follows a probability density function given by

$$f_X(x) = \begin{cases} \dfrac{9x}{4}, & 0 < x \le \dfrac{2}{3} \\ \dfrac{3}{2}, & \dfrac{2}{3} < x \le 1 \\ 0, & \text{elsewhere} \end{cases}.$$

The subscription fee is calculated on a daily basis. Denote the daily fee by Y. If a customer is online for more than $2/3$ of the time in a day, s/he needs to pay

$$\text{usage rate} \times \$2.$$

Otherwise s/he has to pay a minimum fee $\$1\frac{1}{3}$. Simple calculation shows that the average daily fee of a customer, $E(Y)$, is \$1.5.

   (a) Show that $E(Y^2) = \dfrac{62}{27}$.

Assume that daily usage of any day is independent of any other days. In a month of 30 days,

   (b) What is an approximate probability that the monthly subscription fee is lower than 47?

   (c) Compute the probability that there are 15 days with daily fees more than $\$1\frac{1}{3}$

      (i) exactly, and

      (ii) by normal approximation.

3. **[10 marks]** In the past, the gas mileage of a compact car has a normal distribution with mean 30 miles per gallon and standard deviation of 5 miles per gallon. Engineers redesigned the engine of this car and tested the gas mileage of a random sample of 100 new cars. They want to prove that the mean gas mileage of new cars, μ, is higher than that of the old cars. Assume that the mileage of new cars is normally distributed with standard deviation of 4 miles per gallon.

   (a) If the manufacturer always rejects the null hypothesis, H_0: $\mu = 30$, when the sample mean mileage is greater than 30.7. What is the level of significance of the test?

   (b) Suppose a sample mean of 30.8 has been observed. If you sample a large number of 100 new cars, what is the proportion of time that you will observe an even larger sample mean.

   (c) If we change the sample size to $n = 25$ and everything else remains the same, do you need to use a t distribution when determining the rejection region of the test or when constructing confidence intervals? Why?

4. **[18 marks]** In a study on the effect of coffee on performance of company workers, eleven workers were selected at random and were asked to perform the task without drinking coffee. Ten workers were selected at random to take coffee and then perform a task. Their performance was recorded on a scale from 0 to 20. (Higher scores indicate better performance.) The results are summarized as follows:

Without Coffee	With Coffee
$n_1 = 11$	$n_2 = 10$
$\overline{y}_1 = 16.2$	$\overline{y}_2 = 18.9$
$s_1 = 2.1$	$s_2 = 2.0$

   (a) Compute a point estimate for the pooled variance.

   (b) (i) Construct a 95% confidence interval for the difference between the average performance of the two groups.

   (ii) Give an interpretation of the interval in (i).

   (iii) Can you conclude that the average performance is different for the two groups of workers at $\alpha = 5\%$ based on the interval in (i)?

   (iv) What is the possible error that you have made with your decision in the previous test? What is the chance of making it?

   (v) State clearly what are the assumptions you have made in (i).

[END OF PAPER]