## School of Mathematical Sciences

**INTRODUCTION TO MATHEMATICAL STATISTICS II**

**Semester 1, 2004**

**OUTLINE FOR WEEKS 5 AND 6**

---

**PREPARATION**

Note that because of the mid-semester/Easter break, lecture 12 is the first lecture in Week 7.

| Lecture | Section | Overheads | Title | WMS | Pages |
|---|---|---|---|---|---|
| 9 | 3.4 | 3–16:3–17 | Expectation of linear functions | 3.3 | 88–97 |
| | 3.5 | 3–18:3–19 | Random sampling | 2.12 | 74–76 |
| | 3.6 | 3–20:3–29 | Binomial distribution | 3.4 | 97–109 |
| 10 | 3.6 | 3–20:3–29 | Binomial distribution (cont.) | 3.4 | 97–109 |
| | 3.7 | 3–30:3–35 | Hypergeometric distribution | 3.7 | 119–124 |
| 11 | 3.8 | 3–36:3–41 | Normal distribution (revision) | 4.5 | 170–175 |
| | 3.9 | 3–42:3–46 | Normal approx. to binomial | 7.5 | 354–360 |
| | 3.10 | 3–47:3–55 | Poisson distribution | 3.8 | 124–131 |

---

**TUTORIAL 3/ASSIGNMENT 2**

Tutorial 3 will be held in Week 6. Try to attempt the questions beforehand so that you can come to the tutorial well prepared; the questions provide you with a check of whether you are following the material presented so far. It can be hard at first to judge whether the question is easy or difficult, short or long! The questions in the tutorials and assignments, and the examples and exercises provided in the lectures, are your guide to the problems I want you to become familiar with learning to solve.

**The starred questions (*) are part of Assignment 2, which is due by 10am on Friday 7 May (Week 8).**

Questions 6, 7 and 8 involve the Poisson distribution, and questions 9 and 10 are about moment generating functions. If you are reading ahead, you may like to attempt the moment generating questions over the mid-semester break.

1. Question 5 from Tutorial sheet 2 will be discussed first in Tutorial 3.

2. **Case-control studies** (To be done in groups in the tutorial.)

   In the vicinity of a nuclear re-processing plant in the UK, four cases of childhood leukaemia were observed over a certain period. A retrospective *case-control* study was set up to investigate whether this incidence of childhood leukaemia was excessive, and possibly related to occupational radiation exposure in their fathers.

   For each leukaemia *case*, five children living in the local area who had similar characteristics but who did not have leukaemia (i.e. *controls*) were randomly chosen. The occupational radiation exposure (in mSv) in the fathers of the four cases and in the fathers of the 20 controls were then measured.

The results are set out in the following 2×2 table, in which each child is classified as a case or a control, and also according to whether their fathers were *exposed* or *unexposed*:

| Paternal exposure | Leukaemia cases | Local controls | Total |
|---|---|---|---|
| $\geq$ 100 mSv (exposed) | 3 | 1 | 4 |
| $<$ 100 mSv (unexposed) | 1 | 19 | 20 |
| Total | 4 | 20 | 24 |

(a) Use the hypergeometric distribution to find the probability of observing three exposed cases (i.e., $Y = 3$ cases whose fathers had been exposed to $\geq$ 100 mSv).

(b) Suppose it was observed that there were no exposed cases, i.e., $Y = 0$, but that the population of 24 children still contained four leukaemia cases, and that four fathers were exposed and 20 unexposed as described above. Write out the $2 \times 2$ table for this scenario, and find the probability that $Y = 0$.

3. **New South Wales Board of Studies** (To be done in groups in the tutorial.)

   Students' mathematics examination scripts are packed in boxes of 550. The examiner marks all the scripts in the box. The senior examiner then selects and re-marks a random sample of 50 scripts. If more than 3 scripts are found to have mistakes in the marking, the box of scripts is sent back to the examiner for thorough checking.

   (a) What distribution would be appropriate for the number of scripts with mistakes in the marking? Can you justify any assumptions you need to make?

   (b) What is the probability of accepting a box if it actually contains 4% mistakes in the marking?

   (c) An alternative procedure is for the senior examiner to randomly select 50 scripts and mark them *before* the examiner marks the full box. No marks are made on the scripts, rather, the senior examiner's marks are recorded on a sheet of paper, and the results of the examiner's marking are checked against this sheet. Again, if more than 3 scripts are found with mistakes in the marking, the box of scripts is sent back to the examiner for checking.

   Is this method equivalent to the standard method? What are the advantages and disadvantages of this alternative method?

*The following two quiz questions are to be done in groups in the tutorial.*

4. In a large population of university students, 20% of the students have experienced feelings of 'maths anxiety'. If you take a random sample of 10 students from this population, the probability that exactly 2 students have experienced maths anxiety is

(a) 0.3020
(b) 0.2634
(c) 0.2013
(d) 0.5
(e) 1

5. An opinion poll asks a random sample of voters 'Do you think politicians are underpaid?' Suppose that 25% of the population would respond 'yes'. If the sample size is 400, the probability that at least 90 respond 'yes' is approximately

   (a) 0.875
   (b) 0.125
   (c) 0.750
   (d) 0.225
   (e) None of the above.

6. (To be done in groups in the tutorial if there is time.)

   The university administration assures a mathematician that she has only 1 chance in 10,000 of being trapped in a lift (if only!) in the mathematics building. She goes to work 5 days a week, 52 weeks a year, for 10 years and always rides the lift up to her office when she first arrives.

   (a) What is the probability that she will never be trapped?
   (b) What is the probability that she will be trapped once?
   (c) What is the probability that she will be trapped twice?

   *Assume that the outcome on every day is independent. Is this a reasonable assumption?*

7. Customers arrive at a checkout counter in a department store according to a Poisson distribution at an average of seven per hour. During a given hour, what are the probabilities that:

   (a) No more than three customers arrive?
   (b) A least two customers arrive?
   (c) Exactly five customers arrive?

   If it takes approximately 10 minutes to serve each customer, find the mean, variance and standard deviation of the total service time for customers arriving during a one hour period. (Assume, probably unreasonably, that a sufficient number of servers are available so that no customer must wait for service.)

8. * The number of bacteria colonies of a certain type in samples of polluted water has a Poisson distribution with a mean of 2 per cubic centimetre.

   (a) If four 1-cubic-centimetre samples are independently selected from this water, find the probability that at least one sample will contain one or more bacteria colonies.

(b) How many 1-cubic-centimetre samples should be selected in order to have a probability of approximately 0.95 of seeing at least one bacteria colony?

9. Show that the *moment generating function* (mgf) of a geometric random variable $Y$ with probability of success $p$ is

$$m(t) = \frac{pe^t}{1 - (1 - p)e^t}.$$

Using this result, find $E(Y)$, $E(Y^2)$ and hence, $\mathrm{Var}(Y)$.

10. * If $Y$ has a binomial distribution with $n$ trials and probability of success $p$ show that the mgf for $Y$ is

$$m(t) = (pe^t + 1 - p)^n.$$

Using this result, find $E(Y)$, $E(Y^2)$ and hence, $\mathrm{Var}(Y)$.

Associate Professor Patty Solomon
29 March 2004