School of Mathematical Sciences
INTRODUCTION TO MATHEMATICAL STATISTICS II
Semester 1, 2004

OUTLINE FOR WEEKS 3 AND 4

PREPARATION:

Chapter 3 of WMS is about discrete distributions, where the outcomes in S are countable and the probabilities assigned to them add to one. We will be looking at examples of discrete distributions, their properties and some approximations.

We will not cover the negative binomial distribution (WMS optional Section 3.6, pp. 116–118). However in Section 3 of the lecture notes we include the normal distribution and its approximation to the binomial distribution, which WMS leave until later. Note too that the geometric distribution is covered in WMS Section 3.5, pp. 110–116.

You will need to revise summations, integrals and derivatives from your previous mathematics courses.

Lecture	Section	Overheads	Title	WMS	Pages
5	2.6	2–28:2–33	Sequences of events	2.9	59–65
	2.7	2–34:2–43	Bayes' Rule and Bayes' Odds	2.10	67–72
6	3.1	3–1:3–3	Discrete random variables	3.1	83–84
	3.2	3–4:3–5	Probability distributions	3.2	84–88
	3.3	3–6:3–7	Expectation	3.3	88–97
7	3.3	3–8:3–15	Expectation of functions of r.v.'s	3.3	88–97
	3.4	3–16:3–17	Expectation of linear functions	"	"
8	3.5	3–18:3–19	Random sampling	2.12	74–76
	3.6	3–20:3–28	Binomial distribution	3.4	97–109

TUTORIAL 2/ASSIGNMENT 1:

The questions below are for Tutorial 2. You should attempt these during Weeks 3 and 4, so that you come to the Tutorial well prepared.

Assignment 1 consists of questions marked (*) on the first **two** Tutorial sheets: these are Questions 9 and 10 from Tutorial sheet 1, and Questions 3 and 6 from Tutorial sheet 2.

Due date: by 10am on Friday 2 April (Week 5). Solutions to the assignments will be available when marked assignments are returned to students.

1. **Output of machines** (To be done in groups at the tutorial.)

   Output is produced from three machines, labelled B_1, B_2 and B_3, in the proportions $0.25, 0.5$ and 0.25, respectively. The overall error rate is 5%, so if A is the event that a part is produced in error, then $P(A) = 0.05$. We also know that the three machines have different error rates of 2%, 4% and 10% for B_1, B_2 and B_3, respectively.

(a) Complete the following table of *conditional probabilities* $P(A|B_i), P(\bar{A}|B_i)$ for $i = 1, 2, 3$:

Machine	B_1 B_2 B_3	Total
A		
$\bar{A}$		
Total		

Which of the row or column totals are meaningful here?

(b) Complete the table of *joint probabilities* $P(A \cap B_i), P(\bar{A} \cap B_i)$ for $i = 1, 2, 3$:

Machine	B_1 B_2 B_3	Total
A		
$\bar{A}$		
Total		1.000

(c) Now convert these to $P(B_i|A), P(B_i|\bar{A})$ for $i = 1, 2, 3$:

Machine	B_1 B_2 B_3	Total
A		
$\bar{A}$		
Total		

Which of the row or column totals are meaningful here?

Suppose we find a part that is defective. What is the probability that it came from machine B_3?

2. **Bayes' Rule and the Law** (To be done in groups at the tutorial.)

In 1990, a man was sentenced to 16 years in jail for raping three women. The main evidence linking the accused to the attacks was the allegedly close match between the DNA genetic fingerprint of samples found at the scene of the crime and the accused. At the trial, the forensic scientist said that the match was so good, that only 1 in 3 million people would match this. In his summing up, the judge told the jury that so large a figure 'approximates pretty well to certainty'.

But in 1993, the Court of Appeal quashed the conviction declaring that the verdict was unsafe. The forensic scientist and judge had fallen into a trap known as the 'prosecutor's fallacy'. The DNA evidence gives the probability of getting as good a DNA match from a randomly chosen (and therefore presumably innocent) person in the population, i.e. $P(\text{match}|I)$. However the judge and forensic scientist confused this with the probability of the accused being innocent given so good a match, i.e. $P(I|\text{match})$.

(a) There was little other corroborating evidence, so assume the prior probability of guilt was 1 in a million. What was the *prior odds* of guilt?

(b) Find the *likelihood ratio*.

What assumption are you making here?

(c) Hence find the *posterior odds* of guilt.

Is this 'guilty beyond reasonable doubt'? Justify your answer.

3. * **Diagnosis of a particular patient**

   Suppose a man is feeling sick and he makes an appointment to visit his doctor. After examining the patient, but not yet seeing the result of the blood test, the doctor's opinion is that there is a 30% chance the patient has a particular disease.

   It is known that 95% of men with this disease produce a positive blood test. But it is also known that 2% of men without the disease will also produce a postive result (this is called a *false positive*).

   How should the doctor revise her opinion after seeing a positive blood test result?

   To answer this question, find the posterior odds *that the patient has the disease given a positive blood test.*

4. (This question will be discussed in the tutorial.)

   The number of houses N that a fire company can serve depends on the distance r in city blocks that a fire engine can cover in a specified period of time. If we assume that N is proportional to the area of a circle R blocks from the fire station, then $N = c\pi R^2$, where c is a constant, and R is the number of blocks that a fire engine can cover in the specified time interval. For a particular fire company, $c = 8$, the probability distribution for R is shown in the following table, and $P(R = r) = 0$ for $r \le 20$ and $r \ge 27$:

r	21	22	23	24	25	26
$P(R = r)$	.05	.20	.30	.25	.15	.05

   Find the expected number of houses that the fire company can serve.

5. (This question will be discussed in the tutorial if there is time.)

   Let Y be a discrete random variable with mean μ and variance σ^2. If a and b are constants, prove that

   (a) $E(aY + b) = aE(Y) + b = a\mu + b$.

   (b) $\text{var}(aY + b) = a^2\text{var}(Y) = a^2\sigma^2$.

6. * Let Y be the discrete uniform random variable with values $y_i = i$ for $i = 1, \ldots, n$. Show that the mean of this distribution is $\mu = (n + 1)/2$, and the variance is $\sigma^2 = (n^2 - 1)/12$.

   [Note that you can assume the result $\sum_{i=1}^{n} i^2 = n(n + 1)(2n + 1)/6$.]

7. (This is an additional exercise to be attempted in your own time.)

   A rental company has a piece of equipment which is leased, on average, only one day in five.

   (a) If rental on one day is independent of rental on any other day, find the probability distribution of Y, the number of days between rentals (up to and including the day of rental).

(b) Find the expected value of Y and its variance.

8. (This is an additional exercise to be attempted in your own time.)

   A missile protection system consists of n radar sets operating independently, each with probability 0.9 of detecting an incoming missile.

   (a) If $n = 5$ and a missile enters the zone, what is the probability that at least one radar set detects the missile?

   (b) How large does n need to be if we require the probability of detecting a missile to be at least 0.999?

Associate Professor Patty Solomon
12 March 2004