1. (a) Justify the identity
\[ \sum_{x=0}^{n} \binom{D}{x} \binom{N-D}{n-x} = \binom{N}{n} \]
using a ‘combinatorial’ argument - namely one based on the interpretation of \( \binom{n}{x} \) as the number of ways of choosing \( x \) objects from \( n \). Note this identity verifies that the hypergeometric pmf given on lecture slide 181 correctly sums to 1.
(b) Write down the formula equating the coefficients of \( s^n \) on both sides of the equation \((1+s)^D(1+s)^{N-D} = (1+s)^N \) (as stated on lecture slide 183) as an alternative verification that the hypergeometric pmf sums to 1.
(c) Consider multiplying two polynomials together to obtain a third as follows:
\[ \sum_{i=0}^{n} a_i x^i \times \sum_{j=0}^{n} b_j x^j = \sum_{k=0}^{2n} c_k x^k \]
Write down a simple formula for \( c_k \) in terms of the \( \{a_i\} \) and \( \{b_j\} \) assuming we define \( a_i = 0 \) if \( i < 0 \) or \( i > n \) and \( b_j = 0 \) if \( j < 0 \) and \( j > n \).

2. The number of cars passing an intersection on a country road in any time interval of duration \( t \) hours has a Poisson distribution with mean \( 5t \). A man notes the successive waiting times between cars, but only records completed minutes eg if the time is 2 minutes 48 seconds he records 2 minutes. Times less than 1 minute are recorded as 0.
(a) What is the probability that four cars actually do arrive in a half hour period?
(b) What is the distribution of the actual waiting time \( T \) between successive cars?
(c) Let the recorded waiting time be \( R \). By calculating \( P(i \leq T < i+1) \) for \( i = 0, 1, \ldots \) find the pmf of \( R \). Name the distribution of \( R \) and give the value of its parameter(s).
(d) What percentage of recorded times will exceed 15 minutes?

3. Let \( X \) be the number of zeros in \( n = 50 \) independent random decimal digits (each digit from 0, 1, \ldots, 9 has equal probability of occurring). Find
(a) the exact probability \( P(X = 2) \);
(b) the Poisson approximation to this probability;

4. The temperature in a restaurant is maintained within the range 19.5°C to 22°C. The digital thermometer on the control panel of the plant shows the value of \( T \) rounded to the nearest integer. Assuming that \( T \) is uniformly (continuously) distributed in the indicated range,
(a) find and plot the pdf and cdf of \( T \) and
(b) find and plot the pmf and cdf of the thermometer reading \( U \).
**Additional problems:** Ghahramani: Sections 5.2, 7.3 and 7.4 - skip remainder for now until later in the course when we study Poisson processes ; Section 7.1 - skip problems 8-17.

---

**Homework set 5**

**NB:** This week one of the homework questions is based on the work you do in the computer lab. Solutions to the homework problems are to be given to your tutor at the **start** of your tutorial in the period from **Monday 16 April to Friday 20 April**.

All four problems should be attempted. Recall that only two (randomly chosen) of them will be marked. The form and neatness of work can be considered in marking. Working and/or reasoning must be given to obtain full credit.

1. Tim was given $n$ one-dollar scratch tickets, each winning:
   - $2 \text{ with probability } 1/8,$
   - $10 \text{ with probability } 1/100,$
   - $1,000 \text{ with probability } 1/10,000,$
   - $20,000 \text{ with probability } 1/1,000,000$

   (a) Let $A_n$ be the event that Tim wins on at least one of the $n$ tickets he was given. Compute $\Pr(A_n)$.

   (b) What is the minimum number $n$ of tickets Tim should have to get $\Pr(A_n) > 1/2$?

   (c) Compute the expected value of Tim’s winnings when $n = 8$.

   (d) What is Tim’s expected net gain if he had to pay for the tickets and $n = 8$?

2. Calculate the ratio $r(x) = \frac{p_X(x)}{p_X(x-1)}$ for a Poisson random variable $X$ with parameter $\lambda$. What can you say about the shape of the Poisson distribution from this result? Under what circumstances is the ratio $r(x) = 1$?

3. A telephone company employs five information operators who receive requests for information independently of one another, each according to a Poisson distribution with parameter $\lambda = 0.5t$ for any given $t$-minute period.

   (a) What is the probability that during a given 2-minute period, the first operator receives no requests?

   (b) What is the probability that during a given 2-minute period, exactly four of the five operators receive no requests?

   (c) Write an expression for the probability that during a given 2-minute period, all of the operators receive exactly the same number of requests.

4. Based on your work in Computer lab:
   (a) reproduce your manual calculations from Question 1 parts (b) and (c)
   (b) reproduce your answer to Question 1 part (d).
620-201 Probability — 2007

Computer Lab 5

In this lab you

- investigate the shape of the pmf for a negative binomial random variable for different values of its parameters.
- study the spatial distribution of various facilities in the city of Coventry in England.

Exercise A - Negative binomial distribution pmf

1. The m-file Lab5ExA calculates and plots the pmf for a negative binomial random variable. When you run the program in the command window you are prompted to enter the values for the parameters \( r \) and \( p \). This program handles the generalised form of the negative binomial so it works for any \( r > 0 \), not just integer \( r \) (see lecture slide 173).

   (a) Copy the program Lab5ExA from the server and study it. Using lecture slide 175 check that the formulae used for the mean and standard deviation are accurate. Using lecture slide 170 make sure you understand how the code calculating the negative binomial coefficient \( \binom{r}{z} \) works.

   (b) Run the program for \( r = 3 \) and \( p = 0.5 \). Verify using the pmf directly that the probabilities plotted at 1 and 2 are actually identical. Referring to \( r(z) \) as defined on lecture slide 177, would you expect each of the values \( r(1) \), \( r(2) \) and \( r(3) \) to be greater than, equal to or less than 1? Check your expectations by hand calculations.

   (c) Run the program for \( r = 1.5 \) and \( p = 0.5 \). Verify the probability at 3 directly using the pmf on slide 173.

   (d) Run the program for a variety of \( r \) and \( p \) values including non-integer \( r \). Summarise the effect of the parameters on the symmetry (or lack of it) of the distribution. For what values does the distribution appear similar to a normal bell shaped curve?

Exercise B - Coventry spatial data

1. As introduced in lectures the Poisson distribution counts the number of discrete events occurring in continuous time or space, where those events occur at a uniform rate. In this exercise we look at the distribution of various facilities across the city of Coventry in England -namely fish and chip shops, churches and post offices - to see if they fit the Poisson model.

   Copy the Excel spreadsheet Coventry.xls from the 620-201 folder on the Maths and Stats server to Student Data and look at the Coventry map (try zooming the view to make the map larger - this will help you later). The basic idea is to divide the total map area into a number of squares and count the number of ‘events’ (here facilities) in each square. If the events are uniform then this frequency distribution should be Poisson. Also we need to estimate the Poisson parameter first before comparing the fit.
Work through all the spreadsheet questions. The spreadsheet instructions are self explanatory and answers are available once you have completed the exercises so you can check your results. This spreadsheet (as well as many other interesting exercises) is available from the web site http://home.ched.coventry.ac.uk/Volume/ maintained by Neville Hunt and Sidney Tyrrell of Coventry University. This link has been added to ‘Interesting links’ on the subject home page.