COMPUTER SCIENCE TRIPOS Part IA

Wednesday 2 June 1993 9 to 12

Paper 2

Answer **five** *questions.*
At least **one** *question from each section is to be answered.*
Submit the answers in five **separate** *bundles each with its own cover sheet.*
Write on **one** *side of the paper only.*

SECTION A

1 Catalan numbers may be characterised through the set β of well-formed bracketings. The following rules define membership of β:

(a) the null string $\lambda \in \beta$;

(b) $\mathrm{S} \in \beta \Rightarrow (\mathrm{S}) \in \beta$; (NESTING)

(c) $\mathrm{S}_1 \in \beta, \mathrm{S}_2 \in \beta \Rightarrow \mathrm{S}_1\mathrm{S}_2 \in \beta$. (CONCATENATION)

Show that the number of different well-formed bracketings that can be made with $2n$ brackets is

$$\frac{1}{n+1}\binom{2n}{n}.$$

Suppose that an extra rule

(b') $\mathrm{S} \in \beta \Rightarrow < \mathrm{S} > \in \beta$; (ANGLE-NESTING)

is introduced in addition to (a)–(c). How many bracketings of length $2n$ will there now be?

2 Two teams A and B play a match in which the winner is the first team to win n games. If A needs i games to win and B needs j games to win, denote by $\mathrm{P}(i, j)$ the probability that A will win. B is the better team, and in any particular game A's probability of winning is only 2/5. Write down a relation between $\mathrm{P}(i, j)$, $\mathrm{P}(i-1, j)$ and $\mathrm{P}(i, j-1)$.

Of what order is the computation of $\mathrm{P}(k, k)$ for given k? Show how to lay out the results for maximum re-use of computed values, and work out $\mathrm{P}(2, 2)$.

3 In a meteorological experiment the annual rainfall in millimetres is recorded for Aberdeen, Bangor, Canterbury and Dublin: readings are taken in each of the years 1981–83. The data are represented as a set of triples (r, t, y), where rainfall $r \in \mathbb{N}$ is a natural number of millimetres, the town $t \in T$ is identified by a letter A–D, and the year $y \in Y$ by a digit 1–3. Show how to identify the data with subsets of each of $(\mathbb{N} \times T) \times Y$, $(T \times Y) \times \mathbb{N}$ and $(Y \times \mathbb{N}) \times T$. Which of these relations **either** *must* **or** *may* define a (partial) function from $(\mathbb{N} \times T) \to Y$, $(T \times Y) \to \mathbb{N}$, $(Y \times \mathbb{N}) \to T$ respectively?

(a) Taking one which *must* be a function $(P \times Q) \to R$ say, show how to convert it to a function whose domain is P and whose range is the set of functions $Q \to R$.

(b) Assume that one of the other relations, which *may* define a partial function, does so. Show how to convert it to a partial function whose range is a set of partial functions.

Illustrate your answer with suitably chosen example data.

SECTION B

4 X and Y are independent random variables having Poisson distributions with parameters α and β respectively. By using probability generating functions, or otherwise, prove that $X + Y$ has a Poisson distribution and give its parameter.

Find the conditional distribution for X given that $X + Y = n$, and give its mean and variance. Explain your result in words.

5 A single die is repeatedly thrown, and accumulating counts of 1s, 2s, ..., 6s are recorded. Find the probability that the event *'The accumulated counts of* 1s, 2s, ..., 6s *are equal'* will ever occur.

[Hint: you will need Stirling's approximation $n! \approx (2\pi)^{\frac{1}{2}}\, n^{n+\frac{1}{2}}\, e^{-n}$]

6 In a chip factory three machines manufacture 25%, 35%, and 40% of the total production. Of their output, 5%, 4%, and 2% respectively are rejected as faulty. What is the probability that from a collection of chips drawn at random from the output two which are faulty were made on the same machine? What is the probability that three faulty chips were all made on different machines?

SECTION C

7 Discuss the requirements of an Automatic Teller Machine (ATM) service under the headings:

(*a*) security;

(*b*) availability of service;

(*c*) integrity of stored data;

(*d*) procedure for dispute resolution.

8 Consider these ML functions for performing arithmetic in radix k, where k is an ML variable whose value is a positive integer.

```
fun value []               = 0
  | value (x::xs)          = x + (k*value xs);

fun carry c []             = [c]
  | carry c (x::xs)        = ((c+x) mod k) ::
                                   carry ((c+x) div k) xs;

fun sum c [] ys            = carry c ys
  | sum c (x::xs) []       = carry c (x::xs)
  | sum c (x::xs) (y::ys) = ((c+x+y)mod k) ::
                                   sum ((c+x+y)div k) xs ys;
```

(*a*) State and justify the rule of structural induction for lists.

(*b*) Your client would like you to prove the correctness of `sum`, expressed by the property

`value(sum 0 xs ys) = value(xs)+value(ys).`

Generalize this formula so that it permits a useful structural induction proof, explaining your reasons.

(*c*) Prove the base case of the structural induction.

(*d*) Prove the inductive step of the structural induction.

(*e*) What does the correctness proof say about the case where k equals 1? Discuss whether other properties are necessary to ensure correctness.

Proofs may assume the analogous correctness property for `carry` and standard mathematical laws. State these assumptions clearly.

9 A *multiset* is an unordered collection of elements. An element may occur zero or more times. Two multisets are equal if they contain the same number of occurrences of each element. For example, the multisets $\{a, a, b\}_m$ and $\{a, b, a\}_m$ are equal; they differ from $\{a, b\}_m$.

Let M and N stand for multisets. Write $\texttt{occs}(x, M)$ for the number of occurrences of x in M. Write $\texttt{mplus}(M, N)$ for the sum of M and N, defined by

$$\texttt{occs}(x, \texttt{mplus}(M, N)) = \texttt{occs}(x, M) + \texttt{occs}(x, N)$$

for all x. Write $\texttt{mequal}(M, N)$ for the equality test for M and N.

The question concerns how to represent multisets of strings within ML. For each of the given data representations (a), (b) and (c) describe how you would implement `occs`, `mplus` and `mequal`. If possible, incorporate simple efficiency improvements. In each case state the approximate running time of $\texttt{occs}(x, M)$, $\texttt{mplus}(M, N)$ and $\texttt{mequal}(M, N)$ using O-notation.

(a) Represent multisets using lists of strings, for example $\{a, b, a\}_m$ by $[a, b, a]$.

(b) Represent multisets by lists of pairs of the form (x, k), for example $\{a, b, a\}_m$ by $[(a, 2), (b, 1)]$.

(c) Represent multisets using binary trees.

Compare the representations (a), (b) and (c), stating the advantages and drawbacks of each.

[Answers need not contain ML code. You may refer to algorithms and data structures from the notes *Problem Solving in ML*.]

SECTION D

10 What is an operating system *kernel*, and what runtime services does an executable binary program expect from a typical time-sharing kernel?

How may transfer of control (program flow) between the loaded application and the resident kernel be achieved?

11 Describe and compare the following two programming approaches:

(*a*) execution of a compiled and linked set of modules;

(*b*) using an interpreter.

Include examples of each, showing when one method is preferable to the other. Is there an absolute distinction between the two methods or is there a spectrum of approaches?

12 Outline the data protection principles set down in the Data Protection Act, and discuss the possible justifications for them.