1. Here is a procedure written in C and a few questions about it:

```
int xxx (A, B, n, m) int *A, *B, n, m;
{
    int i = 0, j = 0;
    while (j < m) {
        while (A[i] < B[j] && i < n) i++;
        if (i == n) return(0);
        if (A[i] == B[j]) {i++; j++;}
        else return (0);
    }
    return(1);
}
```

(a) [3] What does this procedure do? When does it return 0? When should it return 1?

**Solution**

There were a few creative interpretations, but most noticed—I believe—that this was simply a detailed version of a procedure I described in our first meeting, and was meant to test whether the numbers in array $B$ are a subset of those in array $A$; and it does so, provided the arrays are sorted. It returns 1 when $B \subseteq A$ and 0 otherwise.

Let $X$ be the number of $A|B$ comparisons made in line 5 of $xxx$.
Let $Y$ be the number of $A|B$ comparisons made in line 7 of $xxx$.
All the answers below may be given in terms of the input parameters $m, n$.

(b) [6] What is the smallest possible value of $X$? How can this value occur?
In this and the following parts: list all the ways in which the extreme values can occur.

**Solution**

All the answers below assume that on entry, $m, n \geq 0$. 
Smallest value: \(\min(1,m)\). When \(m = 0\) the loop in lines 5–9 is not performed even once. The comparison is performed only once if either \(n = 0\) or \(A[0] > B[0]\).

(c) [6] What is the smallest possible value of \(Y\)? How can this value occur?

Solution
The smallest possible value, 0, occurs when \(m = 0\) or \(n = 0\) or \(B[0] > A[n - 1]\).

(d) [6] What is the largest possible value of \(X\)? How can this value occur?

Solution
The largest value is \(n\) or \(n + 1\), depending how the compiler arranges the two comparisons in line 5 – it is \(n + 1\) if they are done in the order shown there. This value is obtained if the first \(j(\geq 0)\) terms of \(B\) are found in \(A\), and then it is discovered that the \(j + 1\)st term \(B[j] > A[n - 1]\).

(e) [6] What is the largest possible value of \(Y\)? How can this value occur?

Solution
\(Y\) assumes the value \(\min(m,n)\) if all the \(B\) elements (except possibly the last) are found in \(A\).

(f) [4] Give any comments that you have—I do not ask here for calculations—about the expected values of \(X\) and \(Y\).

Solution
This was possibly an unfair question – and I did not give it a heavy weight: It is very hard to say anything meaningful about the expected values of \(X\) or \(Y\) without any additional information, beyond the fact that they must be between their maximal and minimal values. I do not think there is any good reason here to invent any elaborate imaginary scenarios...

2. (a) [10] Prove by induction that \(n^4 - 4n^2\) is divisible by 3 for any \(n \in \mathbb{N}\).

Note: This is a two-stage proof (you will find out what this means when you do it!).

Solution
Note: We use the notation \(a \mid b\) to denote “\(a\) divides \(b\),” which can also be rephrased as “\(b\) is a multiple of \(a\).”

Most did a a fair job of this. Not a particularly good one. In fact, there were numerous lapses in notation, or terminology, and I was not sure if they represent carelessness or a deeper lack of understanding of the principles underpinning mathematical induction. For example, a common error was to use \(n = 0\) as base, whereas the claim was to show for \(n \in \mathbb{N}\), a set which does not contain
the zero. It so happens the property under consideration holds at zero... But what if I happened to ask about another property that holds in \( n \neq 0 \)? would you have abandoned the problem? I suggest you review the notion of well-founded set and the exact definition of mathematical induction!

In the set \( \mathbb{N} \), \( n = 1 \) is a minimal element, and therefore a proper choice for a basis, and \( q(n) \equiv n^4 - 4n^2 \) is \(-3\) there -- hence the desired property holds. Then we assume an induction hypothesis, that the claim \( 3 \mid q(n) \) holds for all elements \( n \in \mathbb{N} \) up to some element \( t \) and want to prove the induction step, for \( t + 1 \):

\[(t + 1)^4 - 4(t + 1)^2 = t^4 + 4t^3 + 6t^2 + 4t + 1 - 4t^2 - 8t + 4 = (t^4 - 4t^2) + (6t^2 - 3) + 4t^2 - 4t.\] (*)

The two parenthesized terms in the right-hand side are divisible by 3. The first via the induction hypothesis, and the second, since the integer coefficients there are multiples of 3.

It remains to show that \( 3 \mid (4t^3 - 4t) \), and the factor 4 can be dropped since it cannot affect divisibility by 3. It is then needed to handle \( p(t) \equiv t^3 - t \) and you were asked to prove this by induction as well. The basis would be 1 again, and \( 3 \nmid p(1) = 0 \). Assume the claim for all \( t \) up to some \( m \in \mathbb{N} \) and the induction step requires to prove \( 3 \nmid p(m + 1) \). Now \( p(m + 1) = m^3 + 3m^2 + 3m + 1 - m - 1 = (m^3 - m) + 3(m^2 + m) \) which again is divisible by three for the same two reasons listed for (*).

(b) [5] Prove the claim of part (a) directly (no induction, simply from elementary properties of the integers).

**Solution**
The simplest way is probably to write

\[q(n) = n^4 - 4n^2 = n^2(n^2 - 4) = n^2(n - 2)(n + 2).\]

We consider \( n \) in terms of its divisibility by 3, and observe it can be of three different forms. For some \( k \in \mathbb{N} \) we can have

(i) \( n = 3k \),
(ii) \( n = 3k + 1 \), or
(iii) \( n = 3k + 2 \).

We now consider the above product representation for \( q(n) \), and observe that in case (i), the factor \( n^2 \) is divisible by three, in case (ii), so is \( n + 2 \), and the factor \( n - 2 \) provides 3-divisibility in the last case, (iii).

3. (a) We consider binary search trees. To keep the problem manageable we restrict it to trees of three nodes only. The search-tree property (I hope I am only reminding you) is that the keys in the
left-hand subtree are smaller than the root-key, and the keys in the right-hand subtree are larger than the root-key.

[4] Show the five possible trees.

Solution
I believe all found the correct five trees:

![Five possible trees](image)

Such a tree is created by inserting a sequence of three numbers into an empty tree. For example, inserting the sequence 4, 2, 1 results in a tree: whereas the sequence 7, 3, 12 produces:

![Trees with numbers](image)

Since the shape of the tree is determined by the relative size of the numbers, rather than by the actual values of the number, the following describes all the information needed to determine the probability of getting each of the possible trees:

Let the elements of the given triplet, in the order given, be $a, b, c$. They are always distinct – no repeated values.

The value of $a$ is arbitrary.

The value $b$ is larger than $a$ in probability $1/4$ (and smaller in probability $3/4$).

The value $c$ is the smallest of the three in probability $1/8$, is the largest in probability $1/2$.

[20] Determine the probabilities of all the trees shapes.

*Hint:* The probability of getting a tree like the one above on the left is $3/32$.

Solution
This seems to have been an easy one, to my delight. The key, as most realized, is to know that $a$ is always at the root, and except in the middle tree, $b$ is above $c$; then it remains to characterize the tree shapes by the inequalities that determine them, and use the data given above. The only preliminary calculation needed was that

$$\Pr(c \text{ is intermediate, between } a \text{ and } b) = 1 - \frac{1}{8} - \frac{1}{2} = \frac{3}{8}.$$
Then, for the above trees, from left to right, we evaluate

\[ Pr(a > b > c) = \frac{3}{4} \times \frac{1}{8} = \frac{3}{32}. \]

\[ Pr(a > c > b) = \frac{3}{4} \times \frac{3}{8} = \frac{9}{32}. \]

\[ Pr(c > a > b) + Pr(b > a > c) = \frac{1}{2} \times \frac{3}{4} + \frac{1}{4} \times \frac{1}{8} = \frac{13}{32}. \]

\[ Pr(b > c > a) = \frac{1}{4} \times \frac{3}{8} = \frac{3}{32}. \]

\[ Pr(c > b > a) = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}. \]

And as a minimal check reveals, the five values above sum to one.

4. (a) [6] What is the meaning of \([az^n]f(z)\), when \(a\) is a constant in \(z\)?

**Solution**

This, to my surprise, proved a real hard-to-figure-out question.

Sometimes *substitution* is the best technique:

\[ [x^k]g(x) = \text{the coefficient of } x^k \text{ in the power development of } g(x) \quad \implies \quad [az^n]f(z) = \text{the coefficient of } az^n \text{ in the power development of } f(z). \]

That was the answer I wanted to see. There is the other matter — how do we compute it: and that is dealt with in the second part, which nearly all of you found the answer for, either from the above definition, or by that most sensible of fall-back methods: looking it up.

(b) [8] Given that \(f(z) = \sum_{k \geq 0} f_k z^k\), what is the value of \([az^n]f(z)\)?

**Solution**

How do we find it? by looking for \(z^n\) in the power series development of \(f(z)\), and it clearly comes with an \(f_n\) factor, so if you multiply and divide by \(a\) you get the term \((f_n/a) \cdot (az^n)\), and \(f_n/a\) is your answer.

(c) Solve the divide-and-conquer recurrence

\[ T_n = 5T_{\lfloor \frac{n}{2} \rfloor} + an, \quad n \geq 2, \quad T_1 = 0. \]

Here I missed by not providing you with the value of \(T_0\), which should by rights be zero as well. It is not needed for the first part, but is needed for the second one.
[12] Solve it for $n$-values which are powers of 3.

**Solution**

In this case the above equation simplifies to

$$T_{3^k+1} = 5T_{3^k} + an, \quad k \geq 0, \quad T_1 = 0.$$ 

and the obvious representation $b_k \equiv T_{3^k}$ shows it is a standard linear first order recursion,

$$b_{k+1} = 5b_k + a3^{k+1}, \quad k \geq 0, \quad b_0 = 0,$$

for the like of which you know a complete explicit solution: Some used it, some calculated from scratch (all using the same notation, which was suggestive – but I am not sure of what).

The general solution I gave would lead to $a_n = 5, b_n = a3^n$, hence

$$b_k = \sum_{j=0}^{k-1} \left( \frac{3}{5} \right)^j = \frac{3^{k+1} - 1}{2} \left( \frac{3}{5} \right)^k.$$ 

Since this is obtained from a formula we established, I do not need to verify this is a solution, but it is a good idea to check agreement with some particular value, such as $T_9 = 24a$ or $T_{27} = 147a$, to guard against arithmetic errors.


**Solution**

This is a rather different kettle of fish: no formulas to direct you, no established algorithms, beyond the suggestion I made in class: “you compute successive values from the recursion, looking for a pattern, try to make a guess, and verify by substitution that the guess satisfies the recurrence.” —or some other words to a similar effect.

With the added input $T_0 = 0$ we get $T_2 = 2a$. Successive values failed to show a useful suggestion, untill you notice the jumps at $n = 6, 18, 54$, and realize the jumps are related to the fact that as this values are reduced successively in the recurrence (by repeatedly dividing by 3 and then taking the floor), they are the values that start a sequence of $n$s which leave the final remainder 2, rather than 1.

So the result depends on the digits when the number is written to base 3! Following this observation it is all down hill.

Let $n = (d_kd_{k-1} \ldots d_0)_3$ be the base-3 representation of $n$, i.e.: $n = d_k3^k + d_{k-1}3^{k-1} + \ldots + d_0$.

This representation also lets us see that

$$\left\lfloor \frac{n}{3} \right\rfloor = d_k3^{k-1} + d_{k-1}3^{k-2} + \ldots + d_1,$$
and

\[ \lfloor \frac{n}{3} \rfloor = d_k 3^{k-r} + d_{k-r-1} 3^{k-2} + \ldots + d_r. \]

This also tells us that \( \lfloor \frac{n}{3^k} \rfloor = \lfloor \frac{n}{2^d} \rfloor \), and this holds recursively for higher powers as well. When we iterate the recurrence \( k - 1 \) times we get

\[
T_n = na + 5T_{\lfloor \frac{n}{3} \rfloor} \]
\[
= na + 5 \left( \lfloor \frac{n}{3} \rfloor a + 5T_{\lfloor \frac{n}{3^2} \rfloor} \right) \quad \text{and so on,}
\]
\[
= a \left( n + 5 \lfloor \frac{n}{3} \rfloor + 25 \lfloor \frac{n}{3^2} \rfloor + \ldots + 5^{k-1} \lfloor \frac{n}{3^{k-1}} \rfloor \right) + 5^k T_{d_k}
\]
\[
= a \sum_{r=0}^{k-1} 5^r \sum_{j=r}^{k} d_j 3^{j-r} + 5^k T_{d_k}
\]
\[
= \frac{3a}{2} \sum_{j=0}^{k} d_j 3^j \left( \left( \frac{5}{3} \right)^{j+1} - 1 \right) - ad_k 5^k + 5^k T_{d_k}
\]
\[
= \frac{5a}{2} \sum_{j=0}^{k} d_j 5^j - \frac{3a}{2} n + 5^k (T_{d_k} - ad_k).
\]

There does not see to be much more one can do with this... However, substitution into the original recurrence verifies this is the correct solution.

Let us see how we rederive from this the specialized solution we obtained above. When \( n = 3^k \), then \( d_k = 1 \) and all other \( d_j \) are zero, hence we find

\[
T_{3^k} = \frac{5a}{2} 5^k - \frac{3a}{2} 3^k + 5^k (0 - a) = \frac{3a}{2} \left( 5^k - 3^k \right).
\]

as we saw before.