アルゴリズムとデータ構造　第３回演習　　年　　組　　番 氏名

1. 要素があらかじめ小さい順に並んでいる配列 data に対して，次の手順で key を探索する関数を作成せよ。(1) まず配列の中央の値と key を比較し，もし等しければ探索終了。(2) そうでなければ， key が配列の前半と後半のどちらに含まれるか分かるので，それを新しい探索範囲とする。(3) 新しい探索範囲の中央の値と key を比較する。(4) 以上の手順を繰り返し，範囲を次々に半分に絞り込んで探索していく。

```
int binary_search(int key, int data[], int n)
{
    int left, right, mid;

    left = 0;       /* 探索範囲の左端 */
    right = n - 1;  /* 探索範囲の右端 */

    while (left <= right) {


        if (key == data[mid]) {


        } else if (key < data[mid]) {


        } else {


        }
    }
    return -1;
}
```

2. この **2 分探索**の最大計算量（最悪計算量）を考えてみる。
   (a) 最初の探索範囲には n 個の要素がある。 (1)～(3)のループを 1 回行うごとに探索範囲（分割された広いほう）は α 分の 1 に狭まるので，k 回目では n の α^k 分の 1 個になる。α の値を示せ。

   (b) 最悪の場合のループ回数を k 回とすると，k-1 回目に探索範囲が最後の 1 個になって k 回目に最後の比較をするので，およそ $n \cdot \left(\frac{1}{\alpha}\right)^{k-1}=1$ という式が成り立つ。この式から最大計算量 k を求め，n を横軸としたグラフを描いて線形探索と比較せよ。

3. この **2 分探索**の平均計算量を考えてみる。ただし，簡単のため $n = 2^k - 1$ とする。
   (a) key が配列の中央の値と同じなら 1 回目で見つかるので，1 回目で見つかる値は 1 個である。以下，2 回目で見つかる値は 2 個，3 回目は 4 個と続く。最後の k 回目で見つかる値は何個か示せ。

   (b) 平均計算量が以下の式で求められる理由を考えよ。

$$C = \frac{1}{n}\{1 \cdot 1 + 2 \cdot 2 + 3 \cdot 4 + \cdots + (k-1) \cdot 2^{k-2} + k \cdot 2^{k-1}\}$$

   【発展】この値は，$2C - C$を計算することによって求めることができる。難しいので説明は省略…。

$$2C - C = \frac{1}{n}\left\{k \cdot 2^k - \left(1 + 2 + 2^2 + \cdots + 2^{k-1}\right)\right\} = \frac{1}{n}\left(k \cdot 2^k - \frac{1-2^k}{1-2}\right) = \cdots$$

4. 【発展】下記のプログラムは，2 分探索を**再帰**を用いて記述したものである。1 の関数とは引数が異なり，配列の中の data[first]～data[last]の範囲から値 key を探索する。この関数の動作を考えてみよ。

```
int binary_search2(int key, int data[], int first, int last)
{
    int mid;

    if (first > last)
        return -1;

    mid = (first + last) / 2;
    if (key == data[mid])
        return mid;
    if (key < data[mid])
        return binary_search2(key, data, first, mid - 1);
    /* else */
        return binary_search2(key, data, mid + 1, last);
}
```

5. 【発展】下記のプログラムは，2 分探索を改良した**内挿探索（補間探索）**と呼ばれるものである。これは，配列内で添字が left から right まで増えるにつれて，値が data[left]から data[right]までほぼ一定の増加率で増えると仮定すると，探索値 key は次の比例式の mid の位置付近にあるだろうという推測に基づく。

$$right - left : data[right] - data[left] = mid - left : key - data[left]$$

内挿探索は非常に高速だが，データによっては非常に遅くなってしまう。どのような場合か考えてみよ。

```
int interpolation_search(int key, int data[], int n)
{
    int left, right, mid;
    int ldata, rdata;

    left = 0;       /* 探索範囲の左端 */
    right = n - 1;  /* 探索範囲の右端 */

    while (left <= right) {
        ldata = data[left];
        rdata = data[right];

        /* mid の計算式で分母が 0 にならないようにする（重複値がある場合） */
        if (ldata == rdata) {
            if (key == ldata) return left;
            /* else */ break;
        }
        mid = left + (right - left) * (key - ldata) / (rdata - ldata);

        if (key == data[mid]) {
            return mid;
        } else if (key < data[mid]) {
            right = mid - 1;
        } else {
            left = mid + 1;
        }
    }
    return -1;
}
```