

## 4664 Spam Detection

It is well-known that the number of occurrences of the term “free” can distinguish spam and nonspam emails. Your task is to build a spam detection module, based on the number of term “free” in an email.

The core of this detection module is a spam classifier, which is represented by two variables: *Low* and *High*. An email that contains  $X$  “free” words is classified by this module as a spam if  $Low \leq X \leq High$ , otherwise it is not.

To measure the goodness of a classifier, we introduce several information-retrieval terminologies:

		Actual	
		Spam	Non-Spam
Predicted	Spam	$TP$	$FP$
	Non-Spam	$FN$	$TN$

$TP$  (true positive) is the number of emails which are truly predicted as spam;  $FN$  (false negative) is the number of emails which are wrongly predicted as non-spam, and so on.

The portion of emails that are correctly identified as spam is denoted as precision ( $P$ ), which is formulated as  $P = TP/(TP + FP)$ . The portion of spam emails that are successfully identified is denoted as recall ( $R$ ), which is formulated as  $R = TP/(TP + FN)$ . To balance between precision and recall, we use the  $F$ -measure which is formulated as  $F = 2 \times P \times R/(P + R)$ . For example, when  $TP = 5$ ,  $FP = 3$ ,  $FN = 2$ ,  $TN = 4$ , we have  $R = 5/7$ ,  $P = 5/8$ , and  $F = 2/3$ .

When there is no spam, we can report all emails as non-spam with  $F = 1.0$  (perfect classifier).

Our data mining team has manually analyzed several emails and labeled them as spam or non-spam. Your task is to find the values of *Low* and *High* that yield the best classifier, i.e., the one that maximizes the  $F$ -measure.

### Input

The input consists of several test cases, where each case contains of two lines:

- $N$  : The maximum number of term “free” in any emails ( $1 \leq N \leq 2 \times 10^6$ )
- $a_0 A B M$  : parameters of random number generator. ( $2 \leq M \leq 10$ ;  $0 \leq a_0, A, B < M$ )

This random number generator generates a sequence of number:

$$a_i = (A * a_{i-1} + B) \bmod M \text{ for } i \geq 1$$

Specifying:

$pos_i = a_{2i}$  ( $0 \leq i \leq N$ ): the number of spam emails with  $i$  number of term “free”.

$neg_i = a_{2i+1}$  ( $0 \leq i \leq N$ ): the number of non-spam emails with  $i$  number of term “free”.

The input is terminated by EOF.

### Output

For each simulation print the  $F$ -measure of the best classifier (accurate to 6 decimal places).

**Explanation for the 1st case:**

This random number generator generates a sequence of 1, 2, 0, 1, 2, ... The number of spam emails is:  $pos_i = \{1, 0, 2, 1\}$ , and the number of non-spam emails is  $neg_i = \{2, 1, 0, 2\}$ .

The optimal classifier treats emails with number of term “free” between 2 and 3 as spam, with  $R = 3/4$  and  $P = 3/5$ , resulting  $F = 2/3$ . Another way of producing optimal classifier is to consider emails with number of term “free” equals to 2 as spam.

### Sample Input

```
3
1 1 1 3
5
2 3 4 5
```

### Sample Output

```
0.666667
0.923077
```