

Fall 2011 BIOSTAT 615/815 Problem Set #1

Due is Tuesday September 27th, 2011 08:40AM (before the class starts)

Problem 0. Introductory Questionnaire

1. Have you had any prior programming language experience? If so, in which languages (e.g. C, C++, Java, R, SAS, Matlab...)?
2. How much are you familiar with the following algorithms or concepts? (Scale 0 to 10)
 - () Quicksort algorithm
 - () Dynamic Programming
 - () Dijkstra's algorithm
 - () Hidden Markov Model
 - () E-M algorithm
 - () Markov Chain Monte-Carlo (MCMC)
 - () Gibbs Sampling
3. How would you rate your current C++ programming skills and experience? (0 to 10).
4. Have you ever enrolled in 615 or 815 before?

Problem 1. Extension of Fisher's Exact Test

Implement a program `fullFastFishersExactTest`, as an extended version from the `fastFisherExactTest` presented in the class. The following two additional features are required for the `fullFastFishersExactTest`.

1. When more than or less than 4 input arguments were specified, report an error message with an adequate guide on how to use the program.
2. In addition to the 2-sided p-values, calculate and output one sided p-values (in both directions). The two-sided and one-sided p-values can be calculated by

- $p_{2sided}(a, b, c, d) = \sum_x \Pr(x) I[\Pr(x) \leq \Pr(a)]$
- $p_{greater}(a, b, c, d) = \sum_{x \geq a} \Pr(x)$
- $p_{less}(a, b, c, d) = \sum_{x \leq a} \Pr(x)$

where $\Pr(x)$ is the hypergeometric probability of $[\#row1, col1] = x$ with row sums of $a + b$ and $c + d$, and column sums of $a + c$ and $b + d$.

Below is an example of expected output of the program.

```
user@host:~/> ./fullFastFishersExactTest 2 7 8 2
Two-sided log10(p) = -1.63801, p-value = 0.0230141
One-sided (less) log10(p) = -1.73232, p-value = 0.0185217
One-sided (greater) log10(p) = -0.000428027, p-value = 0.999015

user@host:~/> ./fullFastFishersExactTest 20 70 80 20
Two-sided log10(p) = -15.2289, p-value = 5.90393e-16
One-sided (less) log10(p) = -15.3764, p-value = 4.20368e-16
One-sided (greater) log10(p) = 8.0232e-14, p-value = 1

user@host:~/> ./fullFastFishersExactTest
Usage: fullFastFishersExactTest [#row1col1] [#row1col2] [#row2col1] [#row2col2]
```

Turn in a hard copy of your full source code, and example output. Also, e-mail your source code (.cpp only as an attached file) to hmkang@umich.edu with title 'BIOSTAT 615/815 Homework #1 - [Your Full Name]'

Problem 2 - Pointers and Arrays

Consider the following program ps-1-2.cpp.

```
#include <iostream>

int main(int argc, char** argv) {
    int nv = argc;
    int& nr = argc;
    int* pr = &argc;

    char** ppc = argv;
    char* pc = *argv;
    char c1 = **argv;
    char c2 = argv[1][2];

    std::cout << "argc = " << argc << std::endl;
    std::cout << "nv = " << nv << std::endl;
    std::cout << "nr = " << nr << std::endl;
    std::cout << "pr[0] = " << pr[0] << std::endl << std::endl;

    std::cout << "pc = " << pc << std::endl;
    std::cout << "ppc[0] = " << ppc[0] << std::endl;
    std::cout << "argv[0] = " << argv[0] << std::endl;
    std::cout << "c1 = " << c1 << std::endl;
    std::cout << "c2 = " << c2 << std::endl << std::endl;

    nr = 10;
    ++argv;

    std::cout << "argc = " << argc << std::endl;
    std::cout << "nv = " << nv << std::endl;
    std::cout << "nr = " << nr << std::endl;
    std::cout << "pr[0] = " << pr[0] << std::endl << std::endl;

    std::cout << "pc = " << pc << std::endl;
    std::cout << "ppc[0] = " << ppc[0] << std::endl;
    std::cout << "argv[0] = " << argv[0] << std::endl;
    std::cout << "c1 = " << c1 << std::endl;
    std::cout << "c2 = " << c2 << std::endl << std::endl;

    return 0;
}
```

What is the output of the following program ps-1-2.cpp when you run the following command?

```
user@host:~/> ./ps-1-2 Hello World
```

Briefly explain why each output line changes or does not change between the first and second half of the outputs.

Problem 3 - Revisiting towerOfHanoi

Consider the following version of modified towerOfHanoi program

```
#include <iostream>

void towerOfHanoi(int n, int s, int i, int d, int& ra) {
    ++ra;
    if ( n > 0 ) {
        towerOfHanoi(n-1,s,d,i,ra);
        std::cout << "Disk " << n << " : " << s << " -> " << d << std::endl;
        towerOfHanoi(n-1,i,s,d,ra);
    }
}

int main(int argc, char** argv) {
    int n = atoi(argv[1]);
    int numCalls = 0;
    towerOfHanoi(n, 1, 2, 3, numCalls);
    std::cout << "Total of " << numCalls << " function calls have been made" << std::endl;
}
```

1. What is the total number of function calls reported when $n = 3$?
2. What would be the total number of function calls reported for an arbitrary n ? Explain why.
3. What would happen if the last argument of towerOfHanoi() function was int ra instead of int& ra?
4. (BIOSTAT815 only) Modify towerOfHanoi function to reduce the total number of function calls by factor of 2 or more.