COMP 2804 — Assignment 4

Due: April 4, 2016 before 4:30pm, in the course drop box in Herzberg 3115. Note that 3115 is open from 8:30am until 4:30pm.

Assignment Policy: Late assignments will not be accepted. Students are encouraged to collaborate on assignments, but at the level of discussion only. When writing the solutions, they should do so in their own words. Past experience has shown conclusively that those who do not put adequate effort into the assignments do not learn the material and have a probability near 1 of doing poorly on the exams.

Important note: When writing your solutions, you must follow the guidelines below.

- The answers should be concise, clear and neat.
- When presenting proofs, every step should be justified.
- Assignments should be stapled or placed in an unsealed envelope.

Substantial departures from the above guidelines will not be graded. I will be providing solutions for some of these problems in the class on April 6th.

Question 1: (60 Marks - Each Question is worth 60/24= 2.5 Marks) Do Question 1-24 from the Winter 2015 COMP 2804 Final Exam. For each multiple choice question give a short reasoning for your choice of the answer. This exam paper can be found either from the course web-page and/or at the following address:

`http://people.scs.carleton.ca/~maheshwa/courses/2804/winter15exam.pdf`

Question 2: (10 Marks) It is know that the probability that a random integer consisting of 1000 digits is prime is approximately 1/2300. Estimate the expected number of integers with 1000-digits that need to be selected at random to find a prime number. Justify your answer.

Question 3: (10 Marks) If two teams A and B play a best-of-five series, and if team A has a $\frac{1}{4}$ chance of winning any game (and team B has $\frac{3}{4}$ chance of winning any game), then what is the expected number of games played in the series. (Note that in a best-of-five series, the teams play games until one team has won three games.)

Question 4: (10 Marks) As a promotion, the NewAge Cereal has placed a toy car in each of its cereal boxes. You can determine the color of the toy car, only by buying and then opening the cereal box. Each toy car is of a monochromatic color among possible $n \geq 1$ colors. Once you collect cars of all possible colors, then you win a real car. The company officials have ensured that a cereal box is equally likely to contain a car of any of the possible n-colors. Let X be the random variable equal to the number of cereal boxes that need to be purchased to obtain at least one toy car of each of the colors. Let X_j be the random variable equal to the number of additional cereal boxes that must be purchased after cars of j different colors have been collected until a car of new color is obtained, for $j = 0, 1, 2, \ldots, n-1$. Answer the following questions:

1. Show that $X = \sum_{j=0}^{n-1} X_j$.
2. Show that after cars of j distinct colors have been obtained, the probability that the color of the car in the next cereal box that is purchased is new (i.e. different from any of the j colors) is $\frac{n-j}{n}$.
3. Show that X_j has a geometric distribution with parameter $\frac{n-j}{n}$.
4. Show that $E(X) = n \sum_{j=1}^{n} \frac{1}{j}$.
5. Suppose that $n = 100$. Use the approximation $\sum_{j=1}^{n} \frac{1}{j} \approx \ln n + 0.5772$ to determine the expected number of cereal boxes that needs to be bought to collect cars of all different colors.

Question 5: (10 Marks)

1. Let X and Y be two independent random variables on a sample space S. Show that $E(XY) = E(X)E(Y)$.
2. Suppose a fair coin is flipped twice. Define two random variables X and Y, where X counts the number of heads, and Y counts the number of tails, in the two flips. Evaluate $E(X)$, $E(Y)$, and $E(XY)$. From these values, can you decide whether X and Y are dependent or independent?

Question 6: (Bonus Problem) Let $F(x)$ and $G(x)$ be two polynomials of degree d, where d is a positive integer. The polynomial F is given as a product of d monomials, and the polynomial G is given in the standard form. For example, $F(x) = (2x+1)(x-1)(x+2)(3x-1)$ and $G(x) = 6x^4 + 7x^3 - 12x^2 - 3x + 2$. To check whether $F(x) \stackrel{?}{=} G(x)$, we can convert $F(x)$ to the standard form, and then verify whether $F(x)$ and $G(x)$ are identical. Unfortunately, converting $F(x)$ to standard form is cumbersome and an expensive operation.

A simple randomized algorithm to check whether $F(x) = G(x)$ is as follows. Choose an integer r uniformly at random from the range $[1, ..., 100d]$. Evaluate $F(r)$ and $G(r)$. If $F(r) \neq G(r)$ report $F(x) \neq G(x)$, otherwise report $F(x) = G(x)$.

Since evaluating a degree d polynomial takes time proportional to $O(d)$, the above algorithm is very fast and simple. Observe the following. If $F(r) \neq G(r)$, then clearly $F(x) \neq G(x)$ and algorithm reports the correct answer. If $F(x) = G(x)$, then no matter what values of r we choose, $F(r) = G(r)$ and the algorithm reports the correct answer. But, if $F(r) = G(r)$, we cannot conclusively say that $F(x) = G(x)$, as we may land up choosing

r to be the root of the equation $F(x) - G(x) = 0$. First show that the above algorithm reports the wrong answer with probability at most $1/100$. Suggest some method(s) so that the probability of error can be further reduced, say for example to $1/10000$.