Institute of Theoretical Computer Science October 2, 2013
Thomas Holenstein, Ueli Maurer, Angelika Steger, Emo Welzl, Peter Widmayer

Algorithms, Probability, and Computing Fall 2013
Honors Assignment Set 1

- The solution is due on **Wednesday, October 23**. Please bring a print-out of your solution with you to the honors class or send your solution as a PDF to `timon.hertli@inf.ethz.ch`.
- Please solve the exercises carefully and then write a nice and complete exposition of your solution using a computer, where we strongly recommend to use LaTeX. A tutorial can be found at `http://www.cadmo.ethz.ch/education/thesis/latex`.
- You are welcome to discuss the tasks with your colleagues, but we expect each of you to hand in your own, individual write-up.
- There will be three Honors assignments. All of them will be graded and the average grade will contribute 75% to your final grade (for the honors class). The remaining 25% consist of an oral presentation.
- This is a theory course, which means: if an exercise does not explicitly say "you do not need to prove your answer" or "justify intuitively", then a formal proof is **always** required.

Exercise 1 (Roll and Roll and Roll) *(50 Points)*

Alice plays the following game: In the first round, she throws a six-sided die d6 to get a number $X_1 \in \{1, \dots, 6\}$. In the next rounds, she throws X_1 many d6 and adds up the results to get a number $X_2 \in \{1, \dots, 36\}$. In the next round, she throws X_2 many d6, and so on. In general, she throws in the i-th round X_{i-1} many d6, and X_i is the sum of the results.

We want to find an efficient way to compute expectation and variance of X_i.

(a) Compute the expected value of X_5.

(b) Think for five minutes about an elementary way to compute the variance. How long would a computer program take to evaluate your formula? You do not need to write down your solution of this point.

(c) From now on, let $p_i(n) := \mathbf{Pr}[X_i = n]$, and let $f_i(X) := f_i := \sum_{n=0}^{\infty} p_i(n)X^n$ be the generating function of the p_i. Note that f_i has only finitely many terms, so we can evaluate the function without worrying about convergence.

Find a connection between $f_2(X)$ and $f_1(f_1(X))$.

(d) Express $f_i(X)$ in terms of f_1.

(e) Now consider any random variable Y with values in $\mathbb{N}_0$. Let $q(n) := \mathbf{Pr}[Y = n]$, and let $g := \sum_{n=0}^{\infty} q(n)X^n$ be the generating function of q. Find a connection between $M_1 := g'(1)$ and $\mathbb{E}[Y]$. (You may assume that all series converge.)

(f) Find a connection between $M_2 := g''(1)$ and the variance $\mathbf{Var}[Y]$.

(g) Use the last two points and your formula in (d) to derive a formula for the expectation and variance of X_i in terms of f_1 and its derivatives.

(h) If you wrote a computer program, how long would it take to compute the variance of X_i? (You may just count the number of algebraic operations, and $O(...)$ is fine.) *Hint:* Dynamic Programming.

(i) Compute the variance of X_3.

Exercise 2 (Tree Height) *(50 Points)*

Remember that in Section 1.3 of the APC lecture notes we have shown that the expected height of a random search tree for n keys is upper bounded by $c \ln n$ (for $c \approx 4.311$). In this exercise, we want to use generating functions to show parts of a proof of a *lower bound* for the expected height.

Let $p_{n,h}$ be the probability that a random serach tree for n keys has height *at most* h. For $h \in \mathbb{N}_0$, let $f_h \overset{\text{gf}}{\longleftrightarrow} (p_{n,h})_{n \geq 0}$ be the corresponding generating function. Note that the height is defined as the maximum depth, and that the depth of the root is 0. We define the height of an empty tree to be 0.

(a) What is f_0? What is $f_h(0)$?

(b) Prove that for $h \geq 1$, $D(f_h) = f_{h-1}^2$.

(c) Give f_1, f_2 explicitly as formal power series.

(d) Show that f_h is a polynomial. What is its degree? What is the coefficient of the monomial corresponding to the degree? You don't need to give a closed form, but you should try to simplify as much as possible.

(e) Let H_n be the height of a random search tree for n keys. Define $h_n := \mathbb{E}[H_n]$. Prove that $\sum_{h=0}^{\infty}(\frac{1}{1-X} - f_h) \overset{\text{gf}}{\longleftrightarrow} (h_n)_{n \geq 0}$.
NOTE: You have here an infinite sum of generating functions. This translates to an infinite sum for each coefficient of the resulting generating function. You can assume that these sums converge.

(f) From the properties above, one can show using elaborate methods that there exists $L, C > 0$ so that for all $1 \leq h \leq c \ln n - L \ln \ln n$, we have

$$p_{n,h} \leq C \exp\left(-ne^{-(h+L\ln h)/c}\right).$$

Let $h_0 := c \ln n - L \ln \ln n$. Prove that

$$\sum_{h=0}^{\lfloor h_0 \rfloor} p_{n,h} = O(1).$$

HINT: Prove first that there exists a constant D such that $n \geq De^{(h_0 + L \ln h_0)/c}$. You may want to change the summation index to $l := \lfloor h_0 \rfloor - h$ at some point. Try to uppper bound each summand by $C \exp(-D \exp(l))$ and then show convergence.

(g) Prove that $\mathbb{E}[H_n] \geq c \ln n - L \ln \ln n - O(1)$ where the L is from (f).