Midterm 1

5:00-7:00pm, 7 October

Notes: There are three questions on this midterm. Answer each question in the space below it, using the back of the sheet to continue your answer if necessary. If you need more space, use the blank sheet at the end. Please write clear, concise *answers. The questions vary quite a bit in difficulty, so if you are having problems with part of a question, leave it and try the next one. All three questions carry approximately equal credit.*

Your Name:

---

1. In this problem, the notation $\#a(w)$ denotes the number of occurrences of the symbol a in a string w.

   Classify each of the following languages as either *regular*, *context-free but not regular*, or *not context-free*. In each case, prove that the language has the properties you claim, by describing a machine or a grammar for the language and/or by using an appropriate Pumping Lemma. [NOTE: When specifying a machine, it is not necessary to give the transition function in full: a clear operational description is sufficient.]

   (i) The language $\{ w \in \{0,1\}^* : \#0(w) > \#1(w) - 17\}$.
   (ii) The language $\{ w \in \{0,1\}^* : \#0(w) > \#1(w) \text{ and } \#1(w) < 17\}$.
   (iii) The language $\{ w \in \{0,1,2\}^* : \#0(w) > \#1(w) > \#2(w)\}$.

---

2. Which of the following statements are true? If the statement is true, give a *short* proof. If it is false, give a *simple* counterexample. You may assume without proof any result that was proved in class or on a homework, provided you state it clearly.

(i) If the language L^* is regular, then L is regular. [Recall that L^* denotes all finite concatenations of strings in L.]

(ii) If the language L is regular, then so is the language L^{odd} consisting of all concatenations of an odd number of strings from L. [Thus, e.g., if $L = \{a,bb\}$ then $L^{odd} = \{a,bb,aaa,aabb,abba,\ldots,bbbbbb,aaaaa,\ldots\}$.]

(iii) If L is regular, then L^R denotes the "reversal" of L, i.e., the language consisting of all strings in L written in reverse order.

(iv) If L is context-free, then L^R is also context-free, where L^R is also context-free, where L^R is as defined in part (iii).

(v) Let G be a context-free grammar with the following property: starting from any variable A, it is not possible to generate any string of the form uAv, where u, v are sequences of variables or terminals. Then the language $L(G)$ generated by G is regular.

3. Let L be a regular language. This problem concerns the following two languages related to L:

$$\text{Unary}(L) = \{1^t \colon \exists x \in L \text{ with } |x| = t\ \};$$

$$\text{Binary}(L) = \{y\colon\ \exists x \in L \text{ with } \text{bin}(|x|) = y\ \};$$

Here bin(n) is the standard binary encoding of integer n, without leading zeros. (So, e.g., bin(0) = 0, bin(1) = 1, bin(2) = 10 and so on.) Thus Unary(L) consists of the lengths of all strings in L encoded in unary, and Binary(L) consists of all these lengths encoded in binary.

(a) Let L be an arbitrary regular language over the alphabet $\{a,b\}$. Show that Unary(L) is a regular language over the alphabet $\{1\}$.

(b) In preparation for part (c), let L be an arbitrary regular language over the alphabet $\{1\}$. Show that there exist integers N and $T > 0$ such that, for all $t > N$, $1^t \in L$ if and only if $1^{t+T} \in L$.

(c) Let L be an arbitrary regular language over the alphabet $\{a,b\}$. Show that Binary(L) is a regular language over the alphabet $\{0,1\}$. [HINT: Use parts (a) and (b).]