1. (20 points) Let $L_1$ and $L_2$ be languages and define

$$\text{shuffle}(L_1, L_2) = \{x_1 y_1 x_2 y_2 \ldots x_n y_n \mid x_1 \ldots x_n \in L_1, y_1 \ldots y_n \in L_2\}.$$ 

Show that if the language $L_1$ is not regular and $L_2$ is any language then the languages $\text{shuffle}(L_1, L_2)$ and $\text{shuffle}(\overline{L_1}, L_2)$ cannot both be regular. Recall for any language $L$, $\overline{L} = \Sigma^* \setminus L$ denotes the complement language of $L$. 

*Hint: Recall closure properties of regular languages.*

**Solution:** Let $L_1$ be an irregular language, $L_2$ be any language, and assume for contradiction that $\text{shuffle}(L_1, L_2)$ and $\text{shuffle}(\overline{L_1}, L_2)$ both are regular. Then since the regular languages are closed under union, we see that

$$\text{shuffle}(L_1, L_2) \cup \text{shuffle}(\overline{L_1}, L_2) = \{x_1 y_1 \ldots x_n y_n \mid x_1 \ldots x_n \in L_1, y_1 \ldots y_n \in L_2\}$$

$$\cup \{x_1 y_1 \ldots x_n y_n \mid x_1 \ldots x_n \in L_1, y_1 \ldots y_n \in \overline{L_2}\}$$

$$= \{x_1 y_1 \ldots x_n y_n \mid x_1 \ldots x_n \in L_1, y_1 \ldots y_n \in \Sigma^*\}$$

$$= \text{shuffle}(L_1, \Sigma^*)$$

since $L_2 \cup \overline{L_2} = \Sigma^*$. Now, recall the alternating language $L_{alt}$ shown in discussion. We showed that if $L$ is a regular language, then $L_{alt}$ is regular. It follows that since $\text{shuffle}(L_1, \Sigma^*)$ is regular, so is $(\text{shuffle}(L_1, \Sigma^*))_{alt}$. However, by inspection we see that $(\text{shuffle}(L_1, \Sigma^*))_{alt} = L_1$, which we assumed to be irregular. Thus, we have a contradiction and our conclusion follows.

2. (40 points) In this problem we investigate the limits of the Pumping Lemma as it was stated in class and look for an alternative that remedies one of these shortcomings.

(a) (10 points) Let $L_1$ be the language

$$L_1 = \{a^i b^p \mid i \geq 0 \text{ and } p \text{ is a prime}\}.$$ 

Prove that the language $L_2 = b^* \cup L_1$ satisfies the conditions of the Pumping Lemma. I.e. show that there exists a $p \in \mathbb{N}$ such that for every word $w \in L_2$
with $|w| \geq p$ we can write $w = xyz$ such that $|xy| \leq p$, $|y| > 0$, and for every $i \geq 0$, $xy^iz \in L_2$.

(b) (20 points) Prove the following generalization of the Pumping Lemma:

Let $L$ be a regular language. There exists a $p \in \mathbb{N}$ such that for every $w \in L$ and every partition of $w$ into $w = xyz$ with $|y| \geq p$ there are strings $a, b, c$ such that $y = abc$, $|b| > 0$, and for all $i \geq 0$, $xab^icy \in L$.

(c) (10 points) Prove that the language $L_2$ is not regular.

Solution:

(a) Set $p = 1$ and let $w \in L_2$ with $|w| \geq 1$. Then either $w = b^m$, $m \geq 1$ for some $m \in \mathbb{N}$ or $w = a^jb^p$ for $j > 0$ and $p$ a prime. In the first case, set $x = \epsilon$, $y = b$, and $z = b^{m-1}$. It is clear that $xy^iz = b^{i+m-1} \in L_2$ for every $i \geq 0$.

In the second case, set $x = \epsilon$, $y = a$, and $z = a^{j-1}b^p$. Then it is clear that $xy^iz = a^{i+j-1}b^p \in L_2$ for every $i \geq 0$. Thus, $L_2$ satisfies the conditions of the Pumping Lemma.

(b) Let $L$ be a regular language and $M$ a DFA for $L$ with $p = |Q|$ states. Let $w \in L$ and let $x, y, z$ be any strings such that $w = xyz$ and $|y|$. Denote by $q_0, q_1, \ldots, q_{|w|}$ the states (in order) that $M$ visits when processing the word $w$. After processing $x$, $M$ will be in state $q_{|x|}$ and when processing $y$, the series of states that get visited are

$$q_{|x|} \rightarrow q_{|x|+1} \rightarrow \cdots \rightarrow q_{|x|+|y|}.$$  

Since there are $|y| + 1$ states $q_i$ with $|x| \leq i \leq |y|$ and $|y| + 1 > p = |Q|$, the pigeonhole principle guarantees the existence of indices $|x| \leq i < j \leq |y|$ such that $q_i = q_j$. Now, writing $w = w_1w_2\ldots w_{|w|}$ for each $w_k \in \Sigma$, we set $a = w_{|x|+1}\ldots w_{i-1}$, $b = w_i\ldots w_{j-1}$, and $c = w_j\ldots w_{|x|+|y|-1}$. For every $k \geq 0$, $M$ will have the following computation path when processing $xab^kc$:

$$q_0 \rightarrow \cdots \rightarrow q_{i-1} \rightarrow \left(q_i \rightarrow \cdots \rightarrow q_{j-1}\right) \rightarrow \cdots \rightarrow \left(q_k \rightarrow \cdots \rightarrow q_{j-1}\right) \rightarrow q_j \rightarrow \cdots \rightarrow q_{|w|}$$  

$k$ times

Since $q_{|w|}$ is an accepting state as $w \in L$, it follows that $xab^kc \in L$ for every $k \geq 0$. The conclusion follows.

(c) Let $p$ be the pumping length of part (b). Assume for contradiction that $L_2$ is regular and consider the word $w = ab^q \in L_2$ for a prime $q$ such that $q > p$. Now, write $w = xyz$ where $x = ab^{q-p}$, $y = b^p$, and $z = \epsilon$. Then by part (b) we can find strings $\alpha, \beta, \gamma$ (symbols chosen as to not clash with the $a$s and $b$s of $L_2$) such that $y = \alpha\beta\gamma$ with $|\beta| > 0$ such that $x\alpha\beta^iz \in L_2$ for every $i > 0$. Since $y$ only consists of $b$s we can write $\beta = b^j$ for some $j > 0$. Then for every $i \geq 0$

$$x\alpha\beta^iz = ab^{q-p}b^{p-j+i+j} = ab^{q+(i-1)j} \in L_2,$$  

(1)
so in particular for $i = q + 1$, $ab^{i+q}j = ab^{(j+1)}$ $\in L_2$. However, $q(j+1) > q$ since $j > 0$ and $q \mid q(j+1)$ so $q(j+1)$ cannot be a prime. This is a contradiction, and thus, $L_2$ is not regular.

3. **(40 points)** For a language $L$ over alphabet $\Sigma$, we define

$$L_{\frac{1}{3} - \frac{1}{3}} = \{xz \in \Sigma^* \mid \exists y \in \Sigma^* \text{ with } |x| = |y| = |z| \text{ such that } xyz \in L\}.$$ 

Prove that if $L$ is regular, then $L_{\frac{1}{3} - \frac{1}{3}}$ need not be regular.

**Intuition:** We will prove this using contradiction. First assume that $L_{\frac{1}{3} - \frac{1}{3}}$ is regular. Consider the regular language $L = 0^*21^*$. First, we'll show that for this $L$, $L_{\frac{1}{3} - \frac{1}{3}} \cap \{0,1\}^* = \{0^n1^n \mid n > 0\}$. We know that $\{0^n1^n \mid n > 0\}$ is not regular which is a contradiction because by the closure properties of regular languages, the intersection of two regular languages must be regular. Hence our assumption is false.

**Solution:** Let’s assume that $L_{\frac{1}{3} - \frac{1}{3}}$ is regular. Consider the regular language $L = 0^*21^*$. Let $L_1 = L_{\frac{1}{3} - \frac{1}{3}} \cap \{0,1\}^*$. Let’s call $L_{eq} = \{0^n1^n \mid n > 0\}$. First, we’ll show that $L_1 = L_{eq}$. This can be proved by showing that $L_1 \subseteq L_{eq}$ and $L_{eq} \subseteq L_1$.

- $L_{eq} \subseteq L_1$
  Consider any word $w = 0^n1^n \in L_{eq}$ such that $n > 0$. Consider the substrings $x = 0^n$ and $z = 1^n$. Let’s set $y = 21^{n-1}$. Now, the string $xyz = 0^n21^{2n-1} \in L$ and $|x| = |y| = |z|$. Therefore, $w \in L_{\frac{1}{3} - \frac{1}{3}}$. Also, $w \in \{0,1\}^*$ as it only consists of 0s and 1s. Thus, $w \in L_1$.

- $L_1 \subseteq L_{eq}$
  Consider any word $w \in L_1$. Since $w \in \{0,1\}^*$, $w$ must contain only 0s and 1s. Let $w = 0^a1^b$ for some $a, b$. Since $w \in L_{\frac{1}{3} - \frac{1}{3}}$, we can write $w = xz$ such that $\exists y, xyz \in L$ and $|x| = |y| = |z|$. Since $xyz \in L$ and $x, z$ have only 0s and 1s, $y$ must contain a 2. Let $y = 0^b21^j$. This means that no matter what $i$ and $j$ are, $x$ must contain only 0s and $z$ must contain only 1s since $xyz \in L$. Therefore, $x = 0^a$ and $z = 1^b$. However, $|x| = |z|$ and so $a = b$. That is, $w = 0^a1^a$ and hence $w \in L_{eq}$.

Thus, $L_1 = L_{eq}$. We know that $L_{eq}$ is not a regular language (shown in class). However, $L_1$ is the intersection of two regular languages which by the closure properties of regular languages must be regular. Therefore, there is a contradiction and our assumption that $L$ is regular was wrong. We’ve now given an example where $L$ is regular, but $L_{\frac{1}{3} - \frac{1}{3}}$ is not regular.