第三次作业：正则语言的性质

4.1.2

4.1.2 考察字符串 $s=0^p1^p0^p1^p \in L$，

(e)解：假设该语言L是正则的，且L的泵长度为p。考察字符串 $s=0^p1^p0^p1^p \in L$，因此该语言不是正则的

由泵引理可知 $|xy| \le p$，因此设 $y=0^k$。$0^{p+k}1^p0^p1^p \notin L$。与假设矛盾，因此该语言不是正则的。由泵引理

可知 $|xy| \le p$，因此设 $y=0^k$。$0^{p+k}1^p1^p0^p \notin L$，与假设矛盾，因此该语言不是正则

(f)解：假设该语言L是正则的，且L的泵长度为p。考察字符串 $s=0^p1^p1^p0^p \in L$（即 $w=0^p1^p$）。

(g)解：假设该语言L是正则的，且L的泵长度为p，考察字符串 $s=0^p1^p \in L$，即 $w=0^p$

由泵引理可知：$|xy| \le p$，因此设 $y=0^k$，由于 $0^{p+k}1^p \notin L$，与假设矛盾，因此不是正则

(h)解：假设该语言L是正则的，且L的泵长度为p，考察字符串 $s=0^p1^p \in L$，即 $w=0^p$

由泵引理可知：$|xy| \le p$，因此设 $y=0^k$，由于 $0^{p+k}1^p \notin L$ 与假设矛盾，故不为正则

4.1.3

4.1.3 且 $|y|=l$，$|z|=m$，其对应素数表示为 $q=2^{l+m}x+2^my+z$

(a)解：假设有字符串 $w=xyz$ 属于正则语言L，

对y打q圈后的字符串对应数字为 $p=2^{ql+m}x+2^m\sum_{j=0}^{q-1}2^{jl}\cdot y+z$

由费马小引理得 $2^{q-1}\equiv 1 \pmod q$，故 $2^{q-1}=kq+1$，$(2^{q-1})^l=(kq+1)^l$

易知 $(kq+1)^l$ 的展开式常数项为1，其余项模q后0

有 $2^{(q-1)l}\equiv 1 \pmod q$　$2^{ql-l+l}\equiv 2^l \pmod q \Rightarrow 2^{ql}-1\equiv 2^l-1 \pmod q$

即 $(2^{ql}-1)=k'q+(2^l-1)$

又 $\because \frac{2^{ql}-1}{2^l-1}=\frac{k'q}{2^l-1}+1=1+2^l+\cdots+2^{(q-1)l}$

由于等式左侧为整数，得 $\frac{k'q}{2^l-1}$ 为整数。$\because$ q为素数，因此q与 2^l-1 互质，得 $\frac{k'}{2^l-1}\in Z$

有 $\frac{2^{ql}-1}{2^l-1}\equiv 1 \pmod q$

下面考察P是否为素数：$P=2^{ql+m}x+2^m\sum_{j=0}^{q-1}2^{jl}\cdot y+z$

$=(2^{ql+m}-2^{l+m})x+2^m(\frac{2^{ql}-1}{2^l-1}-1)y+q$

$=2^{l+m}(2^{(q-1)l}-1)x+2^m(\frac{2^{ql}-1}{2^l-1}-1)y+q$

由上式得三项均为q的倍数，即 $q|P$，因此P是合数不属于正则语言L

综上：L不是正则语言

(b) 假设正则语言L的泵长度为P，设素数P_0是大于$P+1$的素数

考察字符串$w=0^{P_0}1^{(P_0-1)!}$，易得P_0与$(P_0-1)!$互质，从而$w\in L$

假设$|y|=k$，则w打P_0圈得到$w'=0^{P_0(k+1)}1^{(P_0-1)!}$

由于$k\in[1,P]$，又$P_0>P+1$，因此$k\in[1,P_0-1]$，即$k+1\in[2,P_0-1]$

故$P_0(k+1)$与$(P_0-1)!$有公因子$k+1$，不互质，故$w'\notin L$，L不是正则语言

4. 2. 2

4.2.2

解：对于正则语言L，存在一个DFA $M=(Q,\Sigma,\delta,q_0,F)$，使$L(M)=L$。由题意，对于L/a构造DFA

$M'=(Q,\Sigma,\delta,q_0,F')$，其中$F'=\{q\mid\delta(q,a)\in F, q\in Q\}$

对于L/a中的任意字符串w，$wa\in L\Leftrightarrow\delta(q_0,wa)\in F\Leftrightarrow\delta(\delta(q_0,w),a)\in F\Leftrightarrow\delta(q_0,w)\in F'$

因此$L(M')=L/a$，因此L/a是正则语言

4. 2. 7

4.2.7

解：设存在DFA $M_A=(Q_A,\Sigma_A,\delta_A,q_{0A},F_A)$，$M_B=(Q_B,\Sigma_B,\delta_B,q_{0B},F_B)$使对于语言L、M，有$L(M_A)=L$，$L(M_B)=M$。构造$M_{alt}=(Q,\Sigma,\delta,q_0,F')$以识别alt(L,M)

其中$Q=\{(q_i,q_j,s)\mid q_i\in q_A, q_j\in q_B, s\in\{0,1\}\}$，$s=0$表示当前读入偶数个字符，$s=1$表示读入奇数个字符

$\Sigma=\Sigma_A\cup\Sigma_B$，即$M_{alt}$的字母表为$M_A$与$M_B$的并

$$\delta((q_i,q_j,s),a)=\begin{cases}(\delta_A(q_i,a),q_j,1), & s=0\\ (q_i,\delta_B(q_j,a),0), & s=1\end{cases}$$

$F=\{(F_i,F_j,0)\mid F_i\in F_A, F_j\in F_B\}$，即$M_{alt}$的终止状态同时满足$F_A$与$F_B$，且$s=0$

易知上述M_{alt}满足$L(M_{alt})=alt(L,M)$

因此alt(L,M)是正则语言

4. 2. 8

4.2.8

解：设存在DFA $M_A=(Q_A,\Sigma_A,\delta_A,q_{0A},F_A)$，使对语言L，有$L(M_A)=L$

下面构造 $M=(Q,\Sigma,\delta,q_0,F)$.

Q所代表的M中的状态有如(q,S)的形式，S表示在M_A中接受当前已经读入长度的字符串后，能够到达接受状态的所有状态

转移函数满足$\delta((q,S),a)=(\delta_A(q,a),T)$，T表示接受字符后能到达S中的$M_A$的状态集合

此外，$q_0=(q_{0A},F_A)$对应在A中的起始状态，接受字符串长为0.

集合对应M_A中的F_A.

易得：M满足$L(M)=half(L)$

故 $half(L)$是正则语言

4.2.9

4.2.9

解：设存在DFA $M_A=(Q_A,\Sigma_A,\delta_A,q_{0A},F_A)$，使对于语言L，有$L(M_A)=L$.

下面构造 $M=(Q,\Sigma,\delta,q_0,F)$

Q的含义与4.2.8中相同. $S=\{p\in Q_A \mid \delta_A(p,w')\in F_A, w'\in\Sigma^{f(|w|)}\}$

转移函数满足$\delta((q,S),a)=(\delta_A(q,a),T)$

将T构造为$S=\{p\in Q_A \mid \delta_A(p,w')\in F_A, w'\in \Sigma^{f(|w|+1)}\}$

此外$q_0=(q_{0A},F_A)$，对应在A中的起始状态，且接受的字符串长为0

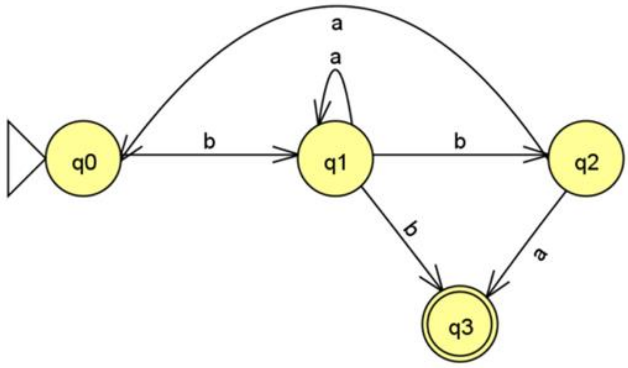
终止状态F对应所有满足$q_0'\in S$的状态(q_0',S)

易得：上述M满足$L(M)=f(L)$

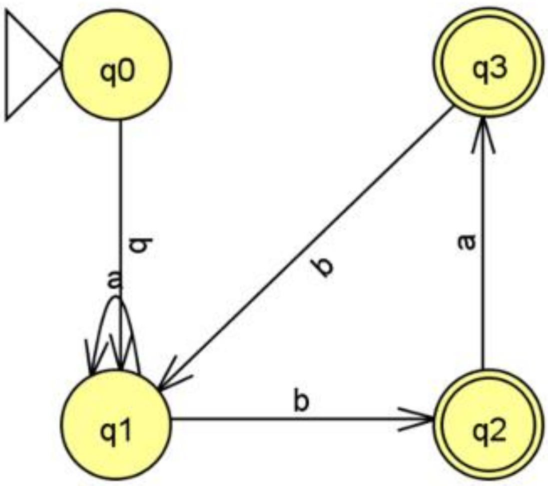
故a).b).c)中的$f(L)$均为正则语言

补充1：

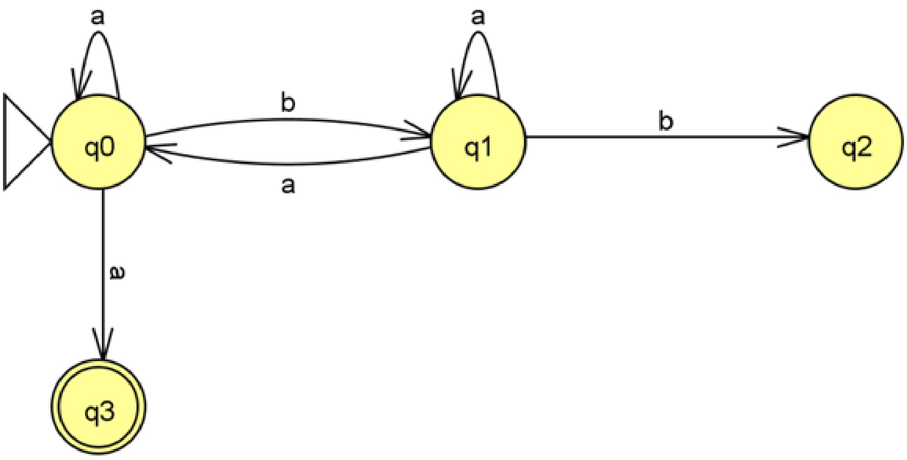
（1）构造 NFA 如下图：



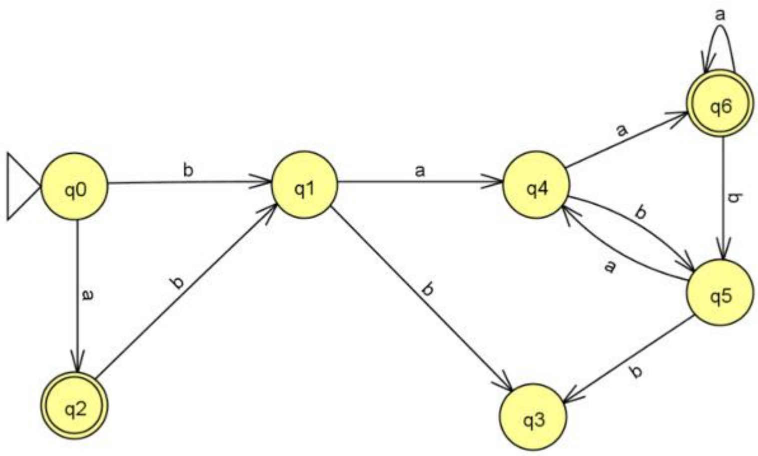
将其转换为 DFA 如下：



（2）构造 NFA 如下图：



将其转换为 DFA 如下：



补充 2：

解：a) $G_1 = (V, T, P_1, S)$

P_1: $S \rightarrow 0S \mid 1A \mid \varepsilon$

$A \rightarrow 1A \mid 0B \mid \varepsilon$

$B \rightarrow 0B \mid 1B$

b) $G_2 = (V, T, P_2, S)$

P_2: $S \rightarrow 0A \mid 1C \mid \varepsilon$

$A \rightarrow 0S \mid 1C \mid \varepsilon$

$B \rightarrow 1A \mid 0C$

$C \rightarrow 1A \mid 0B \mid \varepsilon$

补充 3：

L_2 是否是正则语言取决于 L_1。以下是原因：

(1) 如果 L_1 是一个正则语言，并且其字符串的长度是有限的，那么 L_2 也是正则语言。显然，在这种情况下，L_2 可以被一个确定性有限状态自动机 (DFA) 识别，就像问题中的示例所说明的那样。

(2) 考虑 L_1 表示基于 3 的整数集合，其中每个整数是 3 的倍数（即，其正则表达式为 10^+）。因此，L_2 表示在基 2 下的 3 的倍数集合。假设 L_2 是正则语言，然后考虑一个字符串 w_1，其长度大于 L_2 的 pumping 长度。根据抽水引理 (Pumping Lemma)，w_1 可以被分解为 $w_1 = xyz$，满足：

$$2^{|y|+|z|} \cdot |x| + 2^{|z|} \cdot |y| + |z| = 3^i$$

现在对 y 抽一次，得到 $w_2 = xyxyz$，它表示另一个 3 的倍数。w_2 表示的值为：

$$2^{|y|+|z|} \cdot |x| + 2^{|z|} \cdot |y| + 2^{|y|+|z|} \cdot |y| + |z|$$

将 w_2 和 w_1 表示的整数相减，得到：

$$[w_2] - [w_1] = 2^{|y|+|z|} \cdot |x| + 2^{|z|} \cdot |y| + 2^{|y|+|z|} \cdot |y| + |z| - (2^{|y|+|z|} \cdot |x| + 2^{|z|} \cdot |y| + |z|)$$

化简后得到：

$$[w_2] - [w_1] = 2^{|y|+|z|} \cdot |y|$$

由于 w_2 表示的值大于 w_1，所以 $[w_2] - [w_1]$ 必须是 3 的倍数。然而：

$$2^{|y|+|z|} \cdot |y|$$

显然无法整除 3（除非特定情况，且违反正则语言的封闭性）。因此，w_2 不属于 L_2，这与假设 L_2 是正则语言矛盾。

结论：

L_1 是正则语言的情况下，不能得出 L_2 也是正则语言的结论。