1. 在变量a, b, c上的简单算术表达式由以下文法定义：

   (a) $S \to T$
   (b) $S \to S + T$
   (c) $S \to S - T$
   (d) $S \to S * T$
   (e) $S \to S/T$
   (f) $T \to a$
   (g) $T \to b$
   (h) $T \to c$
   (i) $T \to (S)$

   写出下列表达式的推导过程:

   (a) $a + (b * c)$
   (b) $a/(b - c)$
   (c) $(a + b) * a - c * b/(a + a)$

   答:

   (a) $S \Rightarrow T$
   $\Rightarrow T + T$
   $\Rightarrow a + T$
   $\Rightarrow a + (S)$
   $\Rightarrow a + (S * T)$
   $\Rightarrow a + (T * T)$
   $\Rightarrow a + (b * T)$
   $\Rightarrow a + (b * c)$

   (b) $S \Rightarrow S/T$
   $\Rightarrow T/T$
   $\Rightarrow a/T$
   $\Rightarrow a/(S)$
   $\Rightarrow a/(S - T)$
   $\Rightarrow a/(T - T)$
   $\Rightarrow a/(b - c)$

   (c) $S \Rightarrow S - S$
   $\Rightarrow S * T - S/T$
   $\Rightarrow (S) * T - S * T/T$
   $\Rightarrow (S + T) * T - T * T/T$
   $\Rightarrow (T + T) * T - c * T/T$
   $\Rightarrow (a + T) * T - c * b/T$
   $\Rightarrow (a + b) * T - c * b/(S)$
   $\Rightarrow (a + b) * a - c * b/(S + T)$
   $\Rightarrow (a + b) * a - c * b/(T + T)$
   $\Rightarrow (a + b) * a - c * b/(a + T)$
   $\Rightarrow (a + b) * a - c * b/(a + a)$

2. 证明【幻灯片】第2章 形式语言基础 第18页命题(3)。
   答：
   证明：
   <u>"仅当"：</u> "$B \overset{*}{\Rightarrow} w$" $\Longrightarrow$ "w中b的个数比a的个数多1"
   由$B \overset{*}{\Rightarrow} w$，第1步推导必然是$B \Rightarrow bS$或$B \Rightarrow aBB$

   - 若第1步推导是$B \Rightarrow bS$；必然有$w = bw_1$且$S \overset{*}{\Rightarrow} w_1$；
     因为$|w_1| = k - 1$，根据归纳假设（命题(1)），w_1中包含相等个数的a和b，因此，w中b的个数比a的个数多1
   - 若第1步推导是$B \Rightarrow aBB$；必然有$w = aw_1w_2$ 且$B \overset{*}{\Rightarrow} w_1$和$B \overset{*}{\Rightarrow} w_2$；因为$|w_1| < k - 1$且$|w_2| < k - 1$，根据归纳假设（命题(3)），$w_1$和$w_2$中$b$的个数比$a$的个数多1，因此，$w$中$b$的个数比$a$的个数多1

   <u>"当"：</u> "w中b的个数比a的个数多1" $\Longrightarrow$ "$B \overset{*}{\Rightarrow} w$"
   w的第1个符号只有a或b两种情况（$|w| = k$）

   - 若w的第1个符号是b；则$w = bw_1$且$|w_1| = k - 1$；
     因为w_1中包含相等个数的a和b，根据归纳假设（命题(1)），有$S \overset{*}{\Rightarrow} w_1$，因此，有$B \Rightarrow bS \overset{*}{\Rightarrow} bw_1 = w$，即$B \overset{*}{\Rightarrow} w$
   - 若w的第1个符号是a；则$w = aw_1$且$|w_1| = k - 1$；
     这时w_1中b的个数比a的个数多2，可将w_1分为两部分，即$w_1 = w_{11}w_{12}$，使w_{11}和w_{12}中b的个数都比a的个数多1
     因为$|w_{11}| < k - 1$且$|w_{12}| < k - 1$，根据归纳假设（命题(3)），有$B \overset{*}{\Rightarrow} w_{11}$和$B \overset{*}{\Rightarrow} w_{12}$，因此，有$B \Rightarrow aBB \overset{*}{\Rightarrow} aw_{11}w_{12} = aw_1 = w$，即$S \overset{*}{\Rightarrow} w$

   综上所述，命题（3）成立。

3. 证明【幻灯片】第2章 形式语言基础 第51页$L(G) = \{a^{2^k} \mid k \geqslant 1\}$。

   答:

   证明:

步骤 1：建立推导长度与生成字符串长度之间的关系

对推导的步骤数进行数学归纳，证明在经过特定的推导步数后，生成的字符串长度为 2^k。

基础情况（$k = 1$）

当 $k = 1$ 时，需要证明 G 可以生成长度为 $2^1 = 2$ 的字符串 a^2。

推导过程： $S \Rightarrow ACaB$
$\Rightarrow AaaCB$
$\Rightarrow AaaE$
$\Rightarrow AaEa$
$\Rightarrow AEaa$
$\Rightarrow aa$
因此，$k = 1$ 时，G 可以生成长度为 $2^1 = 2$ 的字符串 aa。

基础情况成立。

归纳假设

假设对于某个 $k \geq 1$，命题 $P(k)$ 成立，即文法 G 可以生成长度为 2^k 的字符串 a^{2^k}。

需要证明： 命题 $P(k+1)$ 成立，即 G 可以生成长度为 2^{k+1} 的字符串 $a^{2^{k+1}}$。

证明 $P(k+1)$

(a) **初始状态：** S
(b) **应用产生式 (1)：** $S \to ACaB$
(c) **对 Ca 部分连续应用产生式 (2)：** C由A右侧（即第2个位置）移到B左侧（即倒数第2个位置）k趟
(d) 已获得 $Aa2^kCB$
(e) 此时如使用产生式 (4)，$CB \to E$，根据归纳假设，则会推导出 $a2^k$
(f) 此时如使用产生式 (3)，$CB \to DB$，最终推导出 $a2^{k+1}$

因此，G 可以生成长度为 2^{k+1} 的字符串 $a^{2^{k+1}}$，命题 $P(k+1)$ 得证。

通过数学归纳法，证明了对于所有 $k \geq 1$，文法 G 可以生成长度为 2^k 的字符串 a^{2^k}。

附加说明

为什么只能生成长度为 2^k 的字符串？

- 产生式 (2)：$Ca \to aaC$，每次应用都会使 a 的数量翻倍，并保留 C。
- 最终通过产生式 (4) 和 (8) 消除 A、B、C 和 E，得到纯 a 串。
- 因此，文法 G 只能生成长度为 2^k 的字符串 a^{2^k}。

通过对推导步数的归纳，我们严格证明了文法 G 生成的语言为：

$$L(G) = \{a^{2^k} \mid k \geq 1\}$$

4. 给定文法G：

$$S \to aSa \mid bSb \mid a \mid b \mid \varepsilon$$

求$L(G)$？并加以证明。

答：

给定的文法 G 生成的语言 $L(G)$ 是由字母表 $\{a, b\}$ 构成的所有**回文串**，即：

$$L(G) = \{w \in \{a, b\}^* \mid w = w^R\}$$

其中，w^R 表示字符串 w 的反转。

证明：

我们需要证明以下两个部分：

(a) $L(G) \subseteq P$，即文法 G 生成的所有字符串都是回文串；
(b) $P \subseteq L(G)$，即所有的回文串都可以由文法 G 生成。

其中，P 表示由字母表 $\{a, b\}$ 构成的所有回文串。

证明 $L(G) \subseteq P$

我们需要证明所有由文法 G 生成的字符串都是回文串。

证明方法：结构归纳法

基础情况：

文法的基本产生式：

- $S \to a$
- $S \to b$

- $S \to \varepsilon$

生成的字符串分别为 a、b 和 ε，它们都是回文串。

归纳步骤：

假设非终结符 S 可以推导出回文串 w，即 $S \Rightarrow w$ 且 $w = w^R$。

那么，文法的递归产生式为：

- $S \to aSa$
- $S \to bSb$

我们需要证明通过这些产生式得到的字符串也是回文串。

推导：

(a) 对于产生式 $S \to aSa$：
如果 $S \Rightarrow w$，且 w 是回文串，那么：

$$S \Rightarrow aSa \Rightarrow awa$$

由于 w 是回文串，因此 $w = w^R$。
那么 awa 的反转为：

$$(awa)^R = a(w^R)a = awa$$

因此，awa 也是回文串。

(b) 对于产生式 $S \to bSb$：
类似地，如果 $S \Rightarrow w$，且 w 是回文串，那么：

$$S \Rightarrow bSb \Rightarrow bwb$$

由于 w 是回文串，因此 $w = w^R$。
那么 bwb 的反转为：

$$(bwb)^R = b(w^R)b = bwb$$

因此，bwb 也是回文串。

通过结构归纳法，我们证明了所有由文法 G 生成的字符串都是回文串，即 $L(G) \subseteq P$。

证明 $P \subseteq L(G)$

我们需要证明所有的回文串都可以由文法 G 生成。

证明方法：数学归纳法

基础情况：

长度为 0 或 1 的回文串：

- 当 $n = 0$ 时，字符串为 ε。
  由产生式 $S \to \varepsilon$ 得到。
- 当 $n = 1$ 时，字符串为 a 或 b。
  由产生式 $S \to a$ 或 $S \to b$ 得到。

归纳假设：

假设长度小于 n 的任意回文串都可以由 S 推导得到。

归纳步骤：

对于长度为 $n \geq 2$ 的回文串 w，我们需要证明 w 可以由 S 推导得到。

推导：

- 由于 w 是回文串，令 w 的第一个字符和最后一个字符相同，即：

  $$w = cxc$$

  其中 $c \in \{a, b\}$，x 是一个字符串，且 $x = x^R$（即 x 也是回文串），长度为 $n-2$。
- 根据归纳假设，x 可以由 S 推导得到，即 $S \Rightarrow x$。
- 根据 c 的取值，应用相应的产生式：
  - 如果 $c = a$，则：

    $$S \Rightarrow aSa \Rightarrow axa$$

  - 如果 $c = b$，则：

    $$S \Rightarrow bSb \Rightarrow bxb$$

- 因此，$w = cxc$ 可以由 S 推导得到。

通过数学归纳法，我们证明了所有的回文串都可以由文法 G 生成，即 $P \subseteq L(G)$。

综合以上两部分的证明，我们得出结论：

$$L(G) = P = \{w \in \{a, b\}^* \mid w = w^R\}$$

即文法 G 生成了所有由字母表 $\{a, b\}$ 构成的回文串。

文法 G 生成了所有由 $\{a, b\}$ 构成的回文串，即：

$$L(G) = \{w \in \{a, b\}^* \mid w = w^R\}$$

5. 写出能生成下列各个语言的文法：

(a) $\{a^n b^m \mid n, m \geq 1\}$

$$S \to A\,B$$
$$A \to a \mid a\,A$$
$$B \to b \mid b\,B$$

- $S \to A\,B$：开始符号 S 推导为 A 和 B 的串联，即首先生成若干个 a，然后生成若干个 b。
- A 的产生式：
  - $A \to a$：生成一个 a。
  - $A \to a\,A$：在已有的 A 前再添加一个 a，实现递归，生成多个 a。
  - 这样 A 至少生成一个 a，保证 $n \geq 1$。
- B 的产生式：
  - $B \to b$：生成一个 b。
  - $B \to b\,B$：在已有的 B 前再添加一个 b，实现递归，生成多个 b。
  - 这样 B 至少生成一个 b，保证 $m \geq 1$。

上述推导表明，该文法能够生成所有满足 $n, m \geq 1$ 的 $a^n b^m$ 形式的字符串。

(b) $\{w \mid w \in \{a, b\}^+ 且 w 中不含有两个相邻的 a\}$

需要确保生成的字符串中没有两个连续的 a。也就是说，在生成一个 a 之后，必须生成至少一个 b 才能再次生成 a。

$$S \to b\,S \mid a\,A \mid b$$
$$A \to b\,S \mid b \mid \varepsilon$$

- S 是开始符号，表示可以开始生成字符串。
- 产生式 $S \to b\,S$ 表示可以连续生成任意多个 b。
- 产生式 $S \to a\,A$ 表示生成一个 a 后，进入状态 A。
- 产生式 $S \to b$ 表示可以以 b 结束。
- 非终结符 A 表示在生成了一个 a 之后的状态，此时不能再直接生成 a。
  - 产生式 $A \to bS$ 表示在 a 后必须生成一个 b，然后继续生成 S。
  - 产生式 $A \to b$ 表示在 a 后生成一个 b 并结束。
  - 产生式 $A \to \varepsilon$ 表示直接结束，这时生成以一个 a 结尾的串。
- 在状态 A 下，不能再生成 a，只能生成 b，这样就避免了连续生成两个 a 的情况。

(c) $\{a^n b^m \mid 1 \leq n \leq m \leq 2n\}$

需要生成所有形如 $a^n b^m$ 的字符串，其中：

- $n \geq 1$
- m 是介于 n 和 $2n$ 之间的整数，即 $n \leq m \leq 2n$

$$S \to a\,S\,b$$
$$S \to a\,S\,b\,b$$
$$S \to a\,b$$
$$S \to a\,b\,b$$

- **生成字符串的形式：** 该文法生成的字符串都是以若干个 a 开头，接若干个 b 结尾，形式为 $a^n b^m$。
- **控制 n：** 每次应用产生式都会生成至少一个 a，因此 $n \geq 1$。
- **控制 m：** 对于每个 a，可以选择生成一个 b 或两个 b。这可以通过以下产生式实现：
  - $S \to a\,S\,b$：为当前的 a 生成一个 b。
  - $S \to a\,S\,b\,b$：为当前的 a 生成两个 b。
- **保证条件 $n \leq m \leq 2n$：**
  - **最小值 $m = n$：** 当每个 a 都只对应一个 b 时，$m = n$。
  - **最大值 $m = 2n$：** 当每个 a 都对应两个 b 时，$m = 2n$。
- **终止条件：** 当不再递归时，使用 $S \to a\,b$ 或 $S \to a\,b\,b$ 来结束生成过程。

(d) $\{a^i b^j c^k \mid i \neq j$ 或 $j \neq k\}$

$L = L_1 \cup L_2$
$L_1 = \{a^i b^j c^k \mid i \neq j\}$
$L_2 = \{a^i b^j c^k \mid j \neq k\}$

$L_1 \to EX$
$E \to aEb \mid A \mid B$
$A \to aA \mid a$
$B \to bB \mid b$

$X \to C \mid \varepsilon$
$C \to cC \mid c$

$L_2 \to YF$
$F \to bFc \mid B \mid C$
$B \to bB \mid b$
$C \to cC \mid c$
$Y \to A \mid \varepsilon$
$A \to aA \mid a$
$S \to L_1 \mid L_2$

L_1 对c无约束
只有条件 $\imath \neq j$，即a的个数不等于b的个数（$\#a \neq \#b$）
两种情况：$\#a > \#b$ 或 $\#b > \#a$
E 可生成 $a^n E b^n$，此时 $\#a = \#b$
将 E 替换为 B 或 A
生成 $a^i b^j$， $i \neq j$
最后附加任意多余的 C

(e) $\{w \mid w \in \{a,b\}^+$ 且w中a的个数是b的个数的两倍$\}$

$S \to SaSaSbS \mid SaSbSaS \mid SbSaSaS \mid \varepsilon$
或
$S \to aab \mid aba \mid baa$
$S \to Saab \mid aSab \mid aaSb \mid aabS$
$S \to Saba \mid aSba \mid abSa \mid abaS$
$S \to Sbaa \mid bSaa \mid baSa \mid baaS$

(f) $\{x \mid x \in \{a,b\}^+$且x不是ww的形式$\}$
L 包含所有非空字符串，这些字符串**不能**写成某个字符串 w 与自身的连接，即 $x \neq ww$。

- **产生式规则**：

$$S \to A \mid B \mid AB \mid BA$$
$$A \to a \mid aAa \mid aAb \mid bAa \mid bAb$$
$$B \to b \mid aBa \mid aBb \mid bBa \mid bBb$$

- **开始符号** S：
  - $S \to A$ 或 $S \to B$：生成奇数长度的字符串。
  - $S \to AB$ 或 $S \to BA$：生成偶数长度的字符串，且在连接时保证两个部分的中心字符不同。
- **非终结符** A **和** B：
  - A 生成所有**以奇数长度**、**以某个字符**（a **或** b）**为中心**的字符串，其中中心字符可以是 a 或 b，并通过组合生成不同的字符串。
  - B 类似，但用于生成另一类奇数长度的字符串。
- **避免生成形如** ww **的字符串**：
  - 对于奇数长度的字符串，由于长度不是偶数，无法写成两个相同字符串的连接，因此符合要求。
  - 对于偶数长度的字符串，通过 AB 或 BA 生成，其中 A 和 B 生成的部分中心字符不同，因此即使长度相等，也无法将字符串分成两个相同的子串。

https://cs.stackexchange.com/questions/19151/the-complement-of-ww-context-free

6. 文法G的产生式如下：

   (a) $S \to aBc$
   (b) $S \to aSBc$
   (c) $cB \to Bc$
   (d) $aB \to ab$
   (e) $bB \to bb$

   请使用数学归纳法证明$L(G) = \{a^n b^n c^n \mid n \geq 1\}$。
   答：

   基本情况($n = 1$)：
   首先，证明基础情况，即当$n = 1$时，$L(G)$包含字符串abc。从文法产生式(a) 中，可以得到$S \to aBc$。 然后，使用产生式(d)。 可以重写aB为ab。因此，$L(G)$包含字符串abc它是一个a, b, c数量相等且数量为1的字符串。

   归纳假设
   假设对于某个正整数k。$L(G)$ 包含的字符串都是形如$a^k b^k c^k$的字符串。

   归纳步骤：
   要证明：对于$n = k + 1$，$L(G)$ 包含的字符串都是形如$a^{k+1} b^{k+1} c^{k+1}$

   (a) 首先执行k次产生式(b) $S \to aSBc$，产生$a^k S (Bc)^k$
   (b) 执行1次产生式(a) $S \to aBc$，产生$a^{(k+1)} (Bc)^{(k+1)}$

(c) 执行产生式(c) $cB \rightarrow Bc$ 和产生式(d) $aB \rightarrow ab$ 若干次可得 $a^{(k+1)}bB^kc^{(k+1)}$

(d) 执行k次产生式(e) $bB \rightarrow bb$ 获得 $a^{(k+1)}b^{(k+1)}c^{(k+1)}$

综上所述：$a^{(k+1)}b^{(k+1)}c^{(k+1)}$ 可由上述产生式产生，所以$L(G) = \{a^nb^nc^n \mid n \geq 1\}$成立

7. 根据乔姆斯基体系，判断下列各文法的类型：

(a) $G_1 = (\{S, A, B\}, \{a, b, c\}, P, S)$，其中产生式$P$为：
$S \rightarrow Aa$
$A \rightarrow Ba \mid c$
$B \rightarrow abc$

(b) $G_2 = (\{S, A, B\}, \{a, b, c, d\}, P, S)$，其中产生式$P$为：
$S \rightarrow aSB \mid d$
$aaA \rightarrow aaBc$
$A \rightarrow aA \mid b$
$B \rightarrow dcb$

(c) $G_3 = (\{S, X, Y\}, \{a, b, c\}, P, S)$，其中产生式$P$为：
$S \rightarrow aS \mid cY$
$X \rightarrow bS \mid a$
$Y \rightarrow bY \mid c$

(d) $G_4 = (\{S, X\}, \{a, b\}, P, S)$，其中产生式$P$为：
$S \rightarrow aX$
$X \rightarrow aX \mid \varepsilon$

(e) $G_5 = (\{S\}, \{a, b\}, P, S)$，其中产生式$P$为：
$S \rightarrow aS \mid b$
$aaS \rightarrow bb$
$aS \rightarrow \varepsilon$

答：

(a) 2型文法、上下文无关文法、CFG
(b) 1型文法、上下文有关文法、CSG
(c) 3型文法、正则文法、RG
(d) 3型文法、正则文法、RG
(e) 0型文法、无限制文法、PSG

8. 给出如下1型文法G

(a) $S \rightarrow Bc \mid CBC$
(b) $BC \rightarrow CB$
(c) $cC \rightarrow bc$
(d) $C \rightarrow c$
(e) $B \rightarrow b$

将G变换为等价的1^0文法G'。

答：不做要求

9. 给出如下右线性文法G

(a) $S \rightarrow abS \mid aaA$
(b) $A \rightarrow bbB \mid a$
(c) $B \rightarrow bab$

将G变换为等价的右正则文法G'。

答：不做要求

10. 给出如下右正则文法G且$L(G) = L$

(a) $S \rightarrow aS \mid bA$
(b) $A \rightarrow cA \mid \varepsilon$

将G变换为左正则文法G'，使得$L(G') = L^{\mathcal{R}}$。

答：不做要求

11. 仿照定理2.2的证明过程，证明定理2.3。

答：此题忽略