Kyushu University 100th 2011
知の新世紀を拓く

形式言語理論

Formal Language Theory

5. 正規集合に対する反復補題

九州大学
KYUSHU UNIVERSITY

本日の内容

- 準備
  - 鳩の巣原理
- 正規言語
- 正規言語に対する反復補題
- 反復補題を使う

準備

鳩の巣原理

【鳩の巣原理】$n > m$ とする. n 羽の鳩が m 個の巣にいるならば，少なくとも一つの巣には 2 羽以上の鳩がいる.

ココ

$$n = 10$$
$$m = 9$$

正規言語

正規言語(regular languages)

【定義】アルファベット Σ 上の言語 L が正規であるとは,
L を受理する決定性有限オートマトンが存在するときをいう.

上の定義と定理2.1, 2.2, 2.3から次の定理が導かれる.

【定理3.1】Lをアルファベット Σ 上の言語 とするとき, 次の(1)〜(4)は等価である.

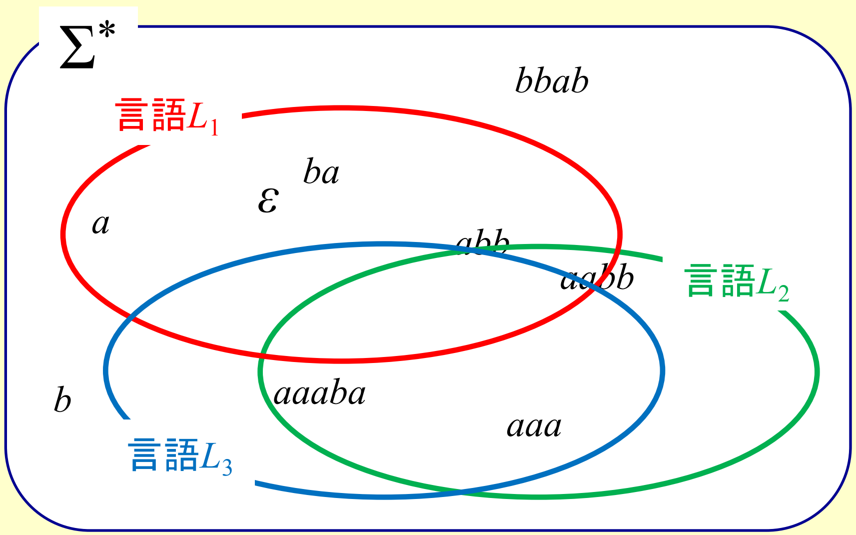
(1) Lは正規である (Lを受理する決定性有限オートマトンが存在する).

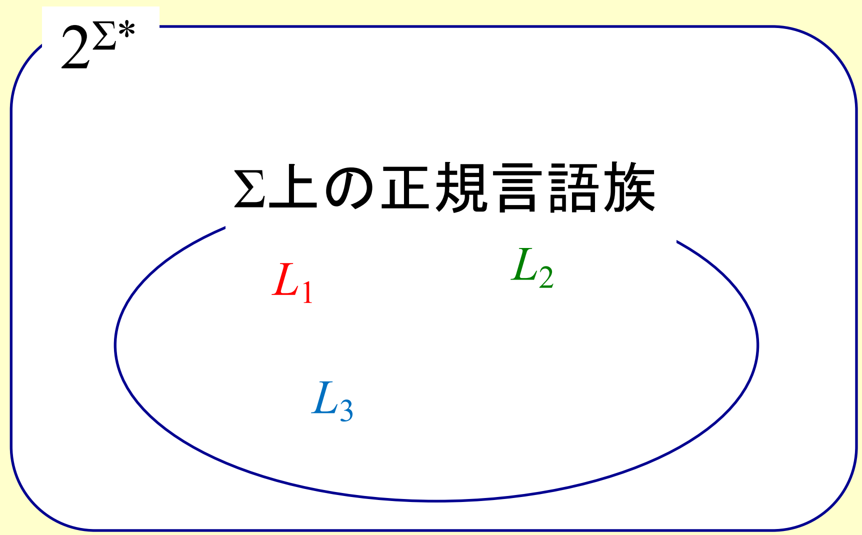
(2) Lを受理する非決定性有限オートマトンが存在する.

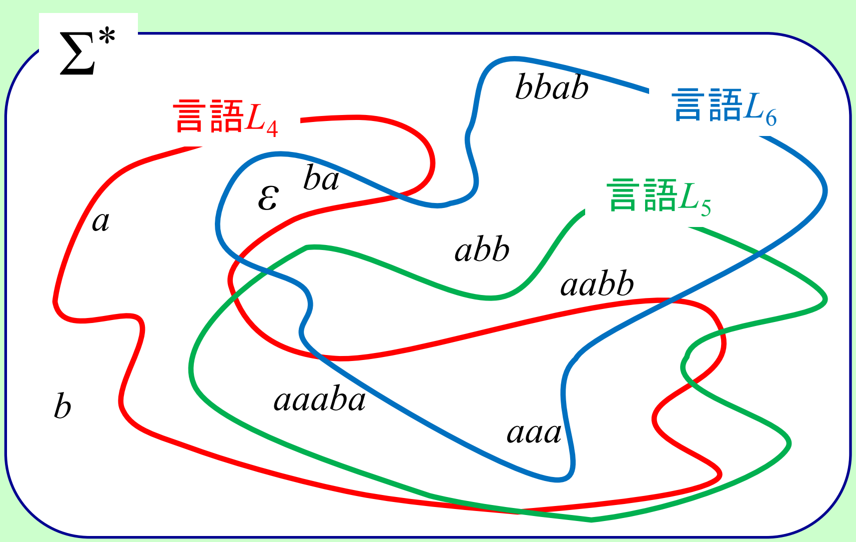
(3) Lを受理するε動作付き非決定性有限オートマトンが存在する.

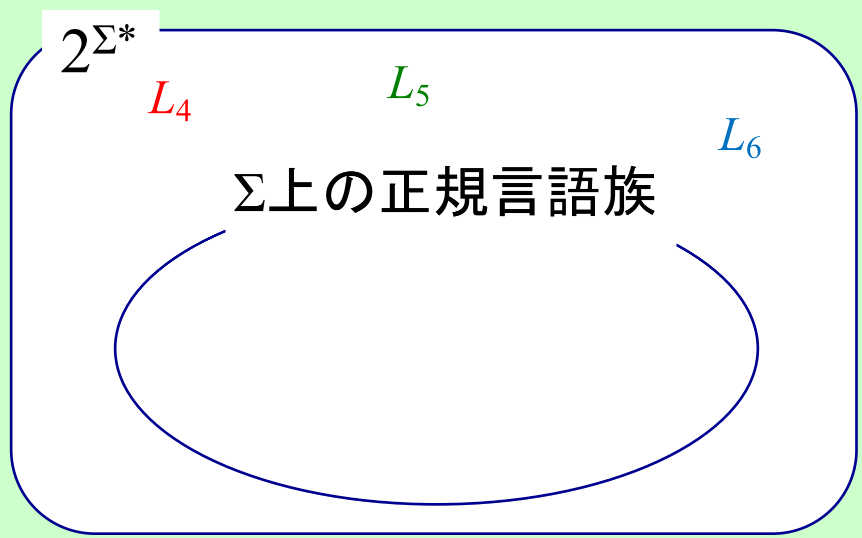
(4) Lを表す正規表現が存在する.

直観的な理解：正規言語と非正規言語









正規言語に関する例題

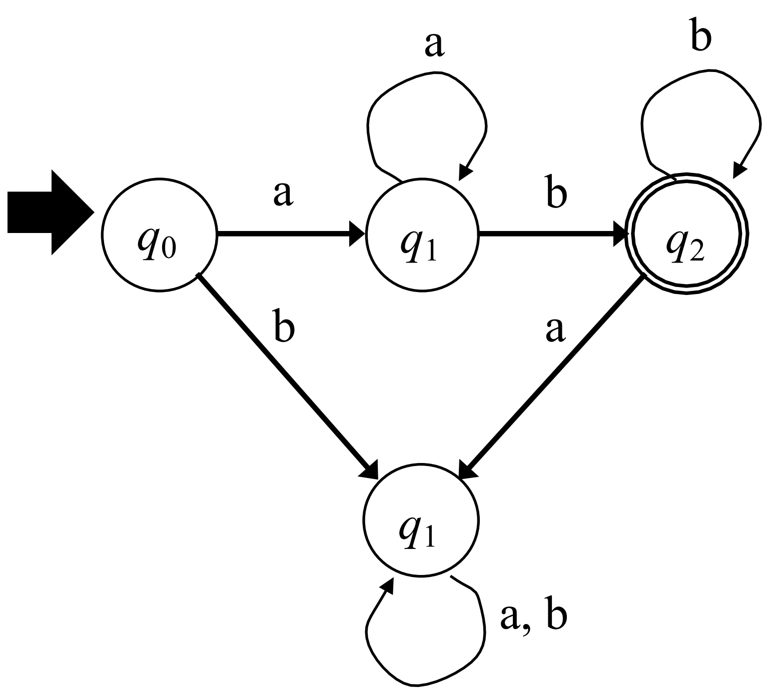
【例題3.1】言語$L_1=\{a^p b^q \mid p>0$ かつ $q>0\}$は正規言語であるか？証明を添えて答えよ.

【例題3.2】言語$L_2=\{a^p b^p \mid p>0\}$は正規言語であるか？証明を添えて答えよ.

- Lが正規言語であると主張したい.
  → Lを受理する有限オートマトンを示す.
- Lが正規言語でないと主張したい.
  → Lを受理する有限オートマトンが存在しないことを証明.

例題3.1への解答

- $L_1 = \{a^p b^q \mid p>0$ かつ $q>0\}$ は正規言語である.
- 証明:
  - L_1 は以下のDFAによって受理される.

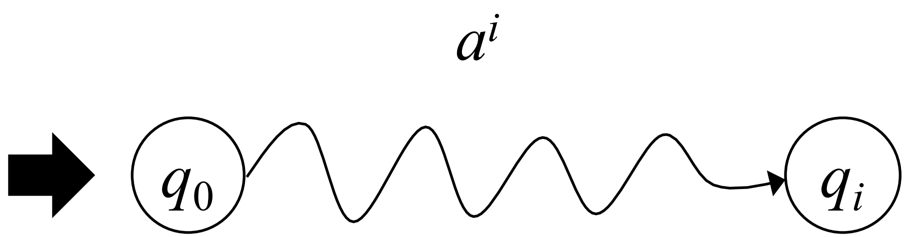


例題3.2への解答…？

- $L_2=\{a^p b^p \mid p>0\}$は正規言語ではない.
- 証明:
  - DFAは無数にある. その中にL_2を受理するDFAが存在しないことをどうやって示せばいいか？
  - 「色々考えたけど思いつかない」ではダメ.
- 「正規言語への反復補題」を使って証明する.

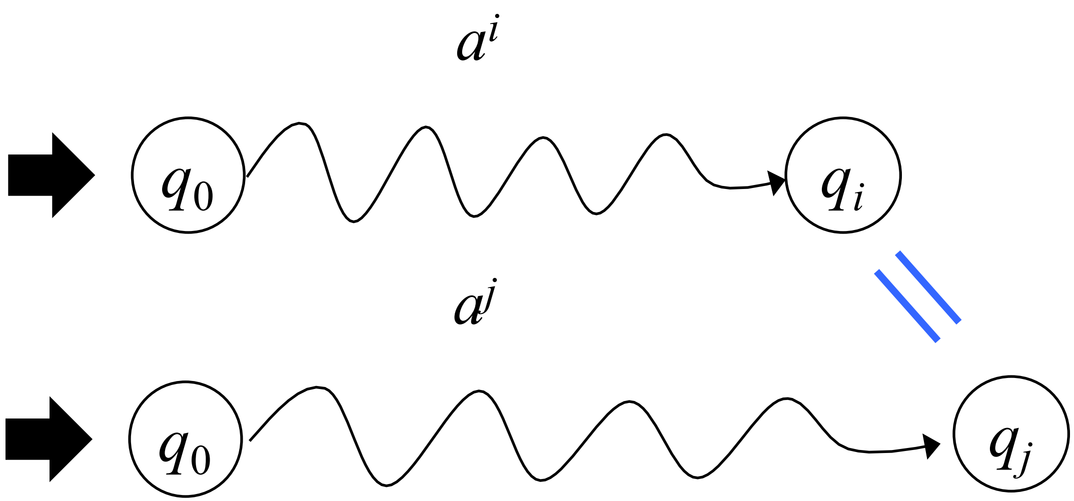
反復補題のアイデア（1／4）

- $L_2=\{a^p b^p \mid p>0\}$を受理する有限オートマトン M が存在すると仮定する. M の状態数を n とする.
- 各$i=1,2,\ldots$に対して，初期状態から文字列 a^i を読んで到達する状態を q_i とする.



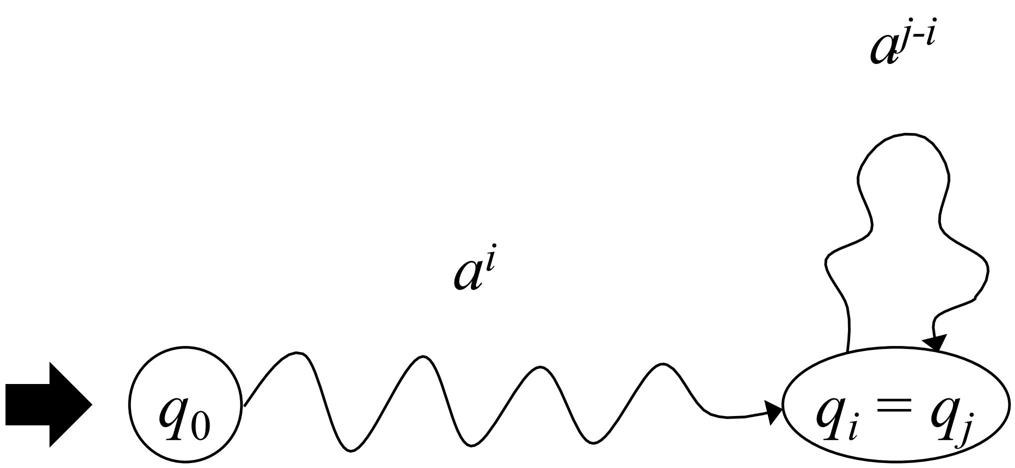
反復補題のアイデア（2／4）

- 状態 $q_0, q_1, \ldots, q_n$ について考える．鳩の巣原理より，$q_i = q_j$ となる整数 i, j $(0 \leq i < j \leq n)$ が存在する．



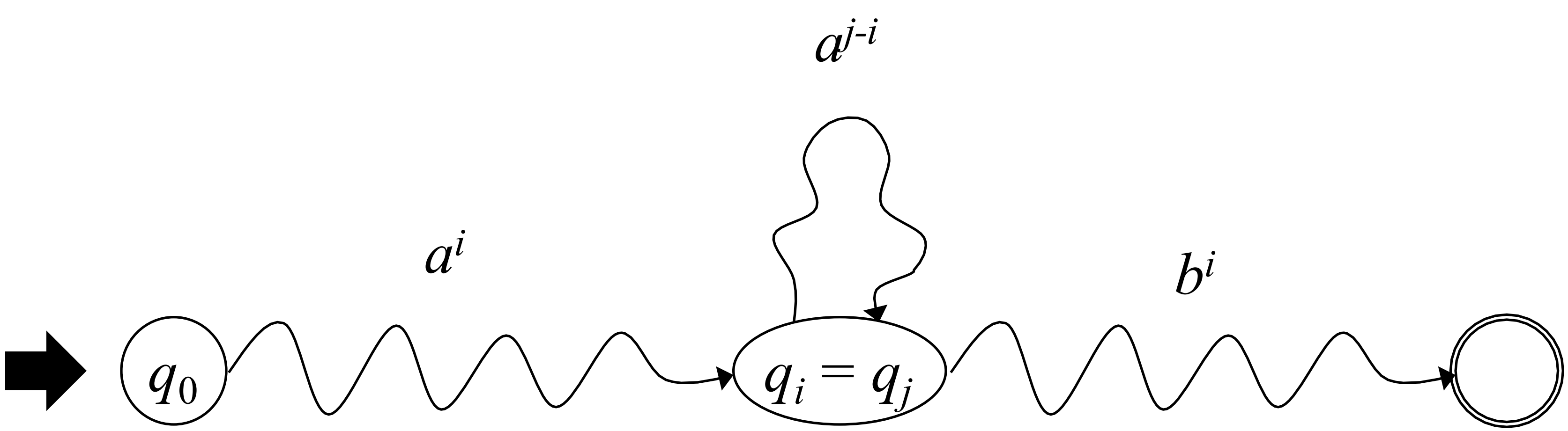
反復補題のアイデア（3／4）

- 状態 $q_0, q_1, \ldots, q_n$ について考える．鳩の巣原理より，$q_i = q_j$ となる整数 i, j $(0 \leq i < j \leq n)$ が存在する．



反復補題のアイデア（4／4）

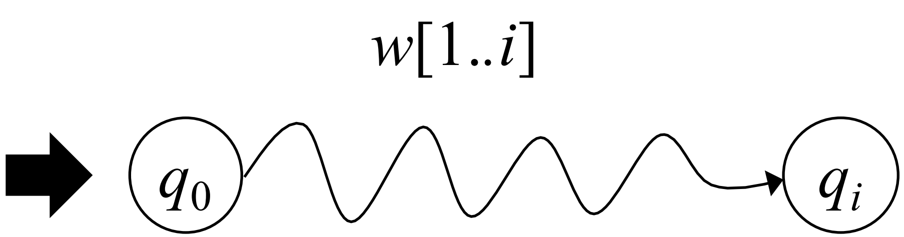
- $a^i b^i \in L_2$ であるため，初期状態から $a^i b^i$ を読んで到達する状態は受理状態.
- このとき，初期状態から文字列 $a^j b^i$ を読むと同じ受理状態へ到達する. これは矛盾.
- よって，背理法により，L_2 は正規言語ではない.



正規言語に対する反復補題

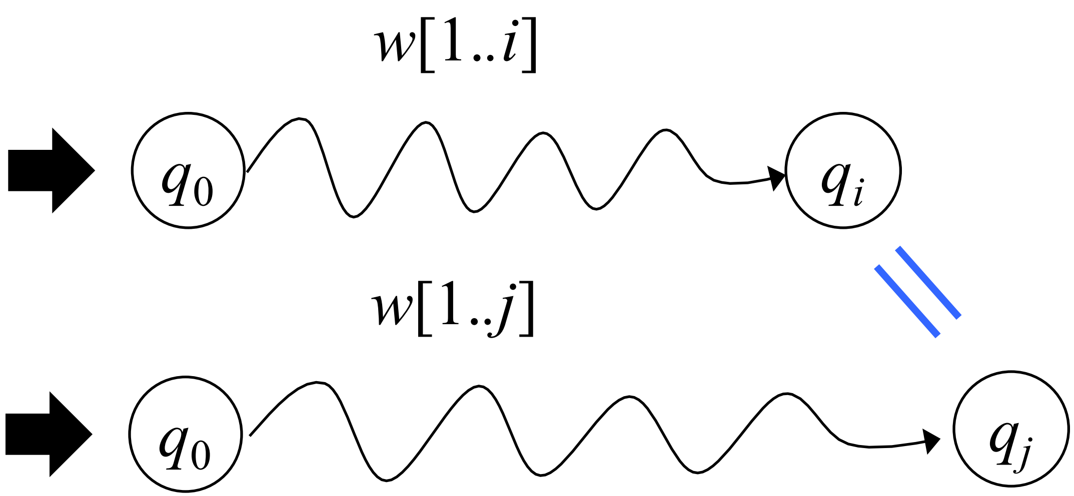
先ほどの議論の一般化（1／7）

- M を任意の有限オートマトンとし，その状態数を n とする．w を $w \in L(M)$ かつ $n \leq |w|$ を満たす任意の文字列とする．
- 各 $i = 0,1,\ldots,n$ に対して，初期状態から文字列 $w[1..i]$ を読んで到達する状態を q_i とする．



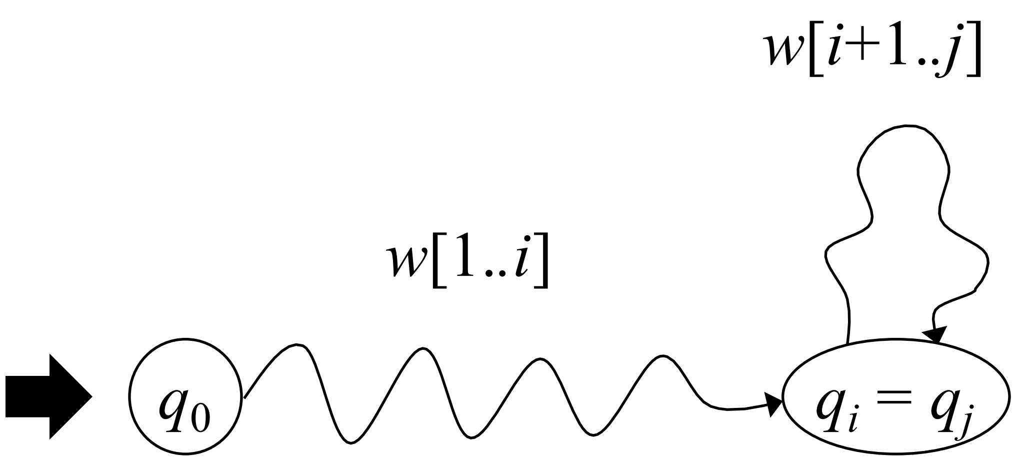
先ほどの議論の一般化(2／7)

- 状態 $q_0, q_1, \ldots, q_n$ について考える．鳩の巣原理により，$q_i = q_j$ となる整数 i, j $(0 \le i < j \le n)$が存在する．



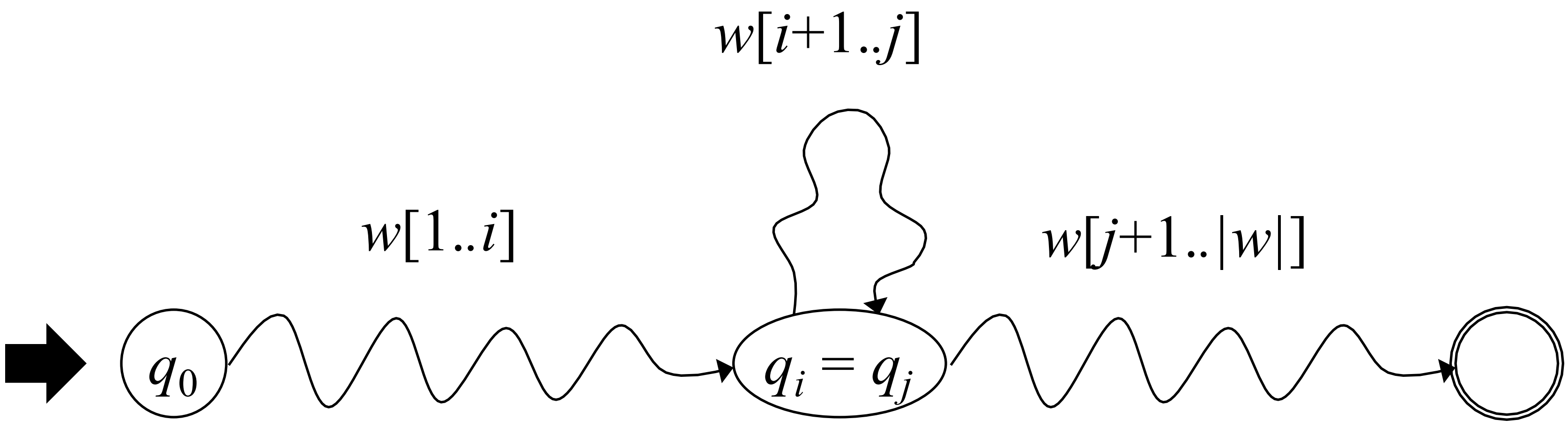
先ほどの議論の一般化(3／7)

- 状態 $q_0, q_1, \ldots, q_n$ について考える．鳩の巣原理により，$q_i = q_j$ となる整数 i, j $(0 \leq i < j \leq n)$が存在する．



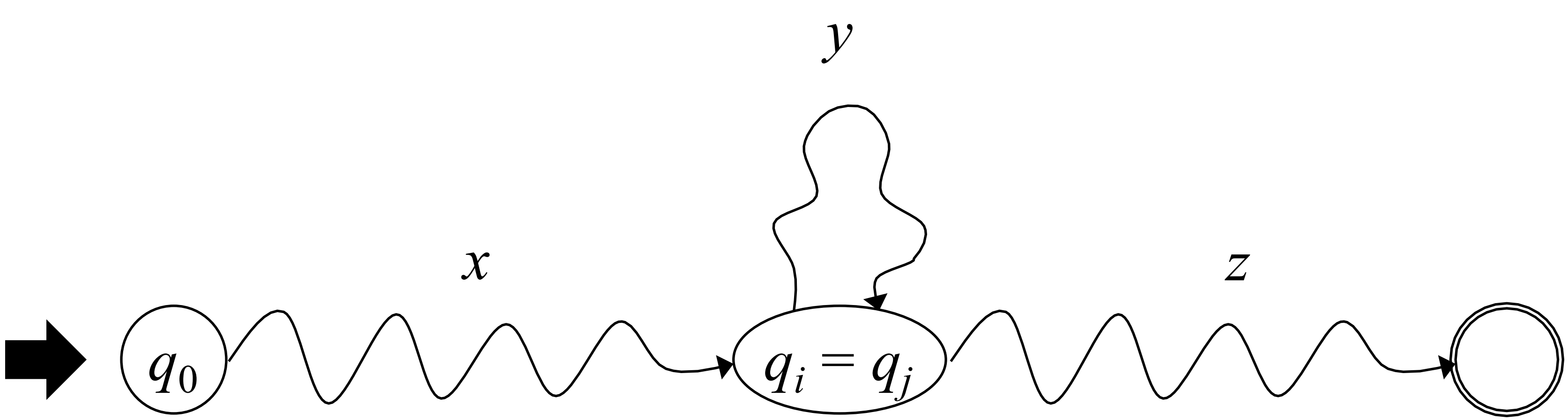
先ほどの議論の一般化（4／7）

- $w \in L(M)$ であるから，初期状態から w を読んで到達する状態は受理状態．



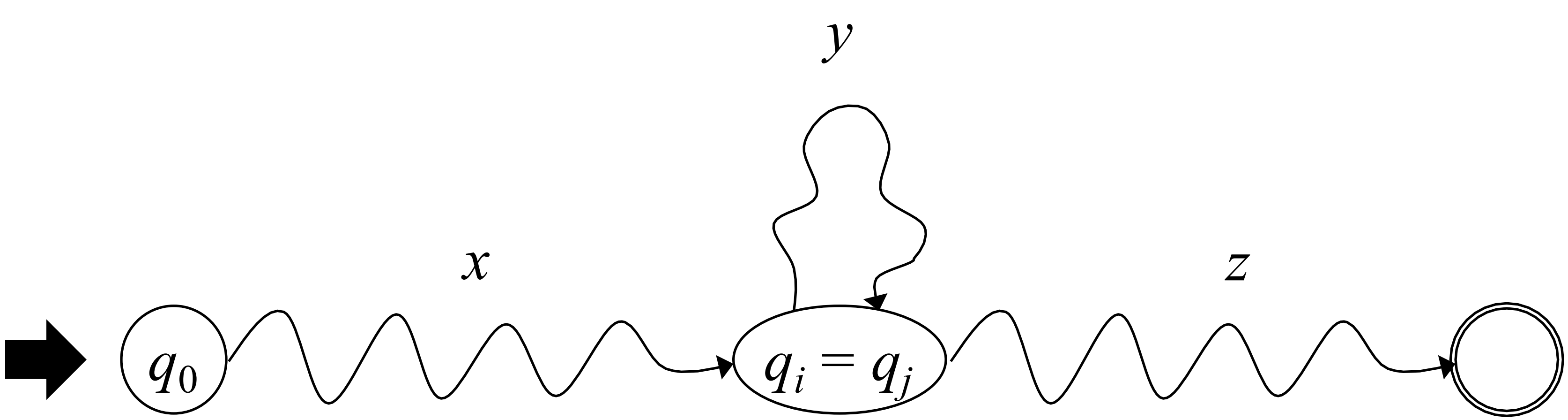
先ほどの議論の一般化（5／7）

- $x = w[1..i], y = w[i+1..j], z = w[j+1..|w|]$ とおく.



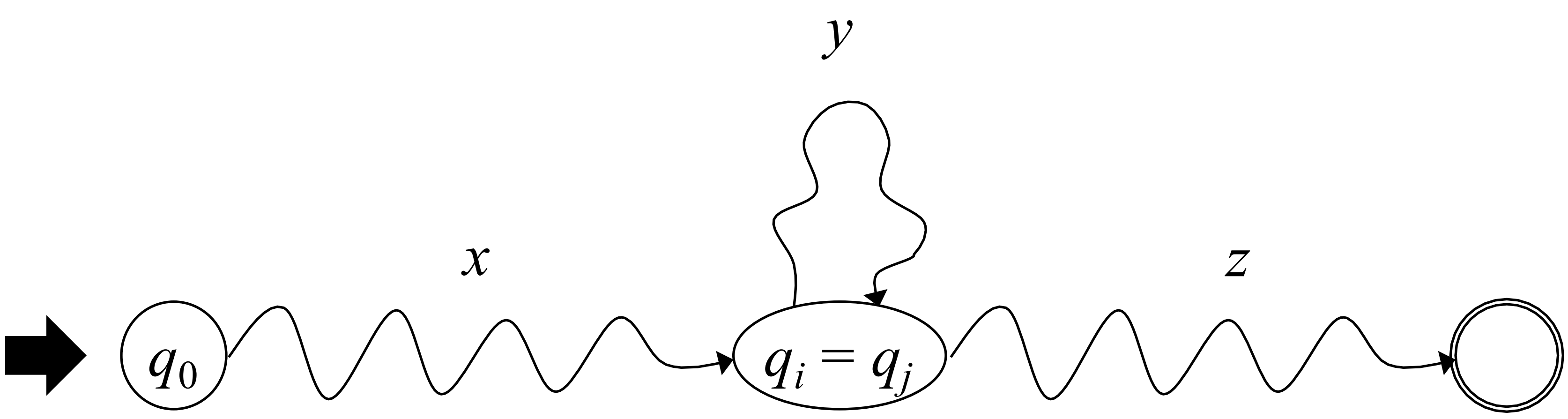
先ほどの議論の一般化（6／ 7）

- $x = w[1..i]$, $y = w[i+1..j]$, $z = w[j+1..|w|]$ とおく.
- すべての $k=0,1,2,\ldots$ に対して $xy^kz \in L(M)$ となる.



先ほどの議論の一般化(7／7)

- $x = w[1..i]$, $y = w[i+1..j]$, $z = w[j+1..|w|]$ とおく.
- すべての k=0,1,2,… に対して $xy^kz \in L(M)$ となる.
- $0 \leq i < j \leq n$ より, $|xy| \leq n$ かつ $|y| > 0$ である.



まとめると. . .

M を任意の有限オートマトンとする. M の状態数を n とするとき, 以下が成り立つ:

$n \leq |w|$ となる任意の文字列 $w \in L(M)$ に対して, 以下を満たす文字列 x, y, z が存在する.

1. $w = xyz$.
2. $|xy| \leq n$ かつ $|y| > 0$.
3. すべての k=0,1,2,… に対して $xy^k z \in L(M)$.

反復補題

【反復補題】言語 L が正規ならば，以下を満たす正整数 n が存在する:
$n \leq |w|$となる任意の文字列 $w \in L$ に対して，以下を満たす文字列 x, y, zが存在する.

1. $w = xyz$.
2. $|xy| \leq n$ かつ $|y| > 0$.
3. すべての k=0,1,2,… に対して $xy^k z \in L$.

反復補題を使う

例題（再掲）

【例題3.2】言語$L_2 = \{a^p b^p \mid p > 0\}$は正規言語であるか？証明を添えて答えよ.

- L_2が正規言語でないことを，反復補題を用いて証明.

例題3.2の証明

- $L_2=\{a^p b^p \mid p>0\}$が正規言語であると仮定する.
- 反復補題の n に対して $w = a^n b^n$ とおく.

$w \in L_2$かつ $n \leq |w|$を満たすwを<u>うまく</u>選ぶ

- $w \in L_2$かつ $n \leq |w|$ であるから
  $w = xyz,\ |xy| \leq n,\ |y| > 0$ となる x, y, z が存在して
  $xy^k z \in L_2 \ \ (k=0,1,\ldots)$ となる.
- $|xy| \leq n$より $xy \in \{a\}^+$である. $y=a^t$ とおくと,
  $xy^k z = a^{n+(k-1)t}\, b^n \in L_2\ (k=0,1,\ldots)$となる.
- $k=0$とおくと, $xy^k z=a^{n-t}\, b^n \notin L_2$となり矛盾.
- よって，背理法により，L_2は正規言語ではない.

QED

例題

【例題3.3】 言語$L_3=\{w \mid w \in \{a, b\}^*, \#_a(w)=\#_b(w)\}$は正規言語でないことを証明せよ.

$\#_c(w)$ は文字cの文字列 w における出現回数を表す.
例 w=abbabのとき， $\#_\mathrm{a}(w)=2$, $\#_\mathrm{b}(w)=3$.

例題3.3の証明

- $L_3=\{w \mid w\in\{a,b\}^*, \#_a(w)=\#_b(w)\}$ が正規言語であると仮定する.
- 反復補題の n に対して $w=a^n b^n$ とおく.

例題3.2と同じ作戦

- $w\in L_3$ かつ $n\leq|w|$ であるから
  $w=xyz$, $|xy|\leq n$, $|y|>0$ となる x, y, z が存在して
  $xy^kz\in L_3$ $(k=0,1,\ldots)$ となる.
- $|xy|\leq n$ より $xy\in\{a\}^+$ である. $y=a^t$ とおくと,
  $xy^kz=a^{n+(k-1)t}b^n\in L_3$ $(k=0,1,\ldots)$ となる.
- $k=0$ とおくと, $xy^kz=a^{n-t}b^n\notin L_3$ となり矛盾.
- よって, 背理法により, L_3 は正規言語ではない.

QED

例題

【例題3.4】 言語$L_4 = \{uu \mid u \in \{a,b\}^*\}$は正規言語でないことを証明せよ.

例題3.4の証明

- $L_4 = \{uu \mid u \in \{a,b\}^*\}$が正規言語であると仮定する．
- 反復補題の n に対して $w = a^n b a^n b$ とおく．
- $w \in L_4$かつ $n < |w|$ であるから
  $w = xyz$, $|xy| \leq n$, $|y| > 0$ となる x, y, z が存在して
  $xy^k z \in L_4$ $(k=0,1,\ldots)$ となる．
- $|xy| \leq n$より $xy \in \{a\}^+$である． $y = a^m$ $(m>0)$とおくと，
  $xy^k z = a^{n+(k-1)m} b a^n b \in L_4$ $(k=0,1,\ldots)$となる．
- $k = 0$ とおくと $a^{n-m} b a^n b \in L_4$ となるが，
  $a^{n-m} b a^n b = uu$ となる $u \in \{a,b\}^*$は存在しない．矛盾．
- よって，背理法により，L_4は正規言語ではない．

QED

例題

【例題3.5】 言語$L_5=\{a^p \mid p$ は素数$\}$は正規言語でないことを証明せよ.

例題3.5の証明

- $L_5 = \{a^p \mid p$ は素数$\}$が正規言語であると仮定する.
- 反復補題の n に対して $n \leq p$ となる素数 p を選び$w = a^p$ とおく.
- $w \in L_5$ かつ $n \leq |w|$ であるから
  $w = xyz,\ |xy| \leq n,\ |y| > 0$ となる x, y, z が存在して
  $xy^kz \in L_5 \quad (k = 0,1,\ldots)$ となる.
- y の長さを m とおくと, $xy^kz = a^{p+(k-1)m} \in L_5$.
- $k = p+1$ とおくと $a^{p(m+1)} \in L_5$ となるが,
  $p(m+1)$は素数ではないので L_5 の定義に矛盾.
- よって，背理法により, L_5 は正規言語ではない.

QED

演習問題

【演習問題3.1】言語$L=\{uu^R \mid u \in \{a,b\}^*\}$は正規言語でないことを証明せよ.

w^R は文字列 w を反転したものを表す.
例）$w=$ ababbのとき，$w^R=$bbaba.

演習問題3.1の証明(解答例)

- $L = \{uu^R \mid u \in \{a,b\}^*\}$が正規言語であると仮定する.
- 反復補題の n に対して $w = a^n bba^n$とおく.
- $w \in L$ かつ $n < |w|$ であるから
  $w = xyz,\ |xy| \leq n,\ |y| > 0$ となる x, y, z が存在して
  $xy^k z \in L \ \ (k=0,1,\ldots)$.
- $|xy| \leq n$より $xy \in \{a\}^+$である. $y=a^m\,(m>0)$とおくと,
  $xy^k z = a^{n+(k-1)m} bba^n \in L \ (k=0,1,\ldots)$.
- $k = 0$ とおくと $a^{n-m} bba^n \in L$ となるが,
  $a^{n-m} bba^n = uu^R$ となる $u \in \{a,b\}^*$ は存在しない. 矛盾.
- よって, 背理法により, Lは正規言語ではない.

QED