第4章　情報源

情報源　　無記憶情報源（独立情報源）

出現記号間に従属性が存在しない

マルコフ情報源

出現記号間に従属性が存在する

正規マルコフ情報源

エルゴードマルコフ情報源

4.1　情報源モデル

離散的情報源（ディジタル情報源）　・・・　本書で扱う
連続的情報源（アナログ情報源）

記号列を構成する記号集合

$S = \{s_1, s_2, \ldots,\ s_n\}$

s_k　連続情報源記号　or　情報源シンボル

S　　情報源アルファベット

英語　S={a,b, c, …, z, space}

日本語　S={あ，い，う，．．．，ん}

情報源モデル　$S = \{s_1, s_2, \ldots, s_n\}$ ＋確率情報

（記号が存在する割合）

4.2 情報源の種類

情報源から発生する記号列に含まれる記号相互の関係

次に発生する記号がこれまでに発生した記号にしない／する
$P(s_k|s_l) = P(s_k)$が成り立つ／成り立たない

記憶のない情報源（無記憶情報源，独立情報源）

$$P(s_k|s_l) = P(s_k)$$

記憶のある情報源（マルコフ情報源）

$$P(s_k|s_l) \neq P(s_k)$$

マルコフ連鎖という確率過程により特徴づけられる
（例）英語では，tの後にhが出やすい．zの後にxはでない．

4.3　無記憶情報源（独立情報源）モデル

確率事象で構成できる

$$S = \begin{Bmatrix} s_1, & s_2, & \dots, & s_n \\ P(s_1), & P(s_2), & \dots, & P(s_n) \end{Bmatrix}$$

記号当たりの発生平均情報量（発生エントロピー）

$$H(S) = -\sum_{k=1}^{n} P(s_k) log P(s_k)$$ ［bit／情報源記号］

情報源からの記号の発生速度　r^*

単位時間当たりの発生平均情報量　$H^*(S) = r^* H(S)$

［bit／単位時間］

4.4　通報

通報:　情報源から発生する記号系列

長さnの通報

$$\boldsymbol{x}_t{}^{(n)} = x_{t-(n-1)}, \ x_{t-(n-2)}, \ \ldots, \ x_{t-1}, \ x_t$$

x_t　時刻tで発生する記号
時刻tでs_kが発生→$x_t = s_k$

記号が発生した時刻が問題にならないとき

$$\boldsymbol{x}^{(n)} = x_1, x_2, \ \ldots \ , \ x_n$$

4.5　マルコフ情報源

定義：　通報による条件付き確率

$$\forall t, \qquad \forall n\ (\geq m) \quad P(x_t|\boldsymbol{x}_{t-1}^{(n)}) = P(x_t|\boldsymbol{x}_{t-1}^{(m)})$$

左辺：通報$\boldsymbol{x}_{t-1}^{(n)}$が発生した後，記号$x_t$が発生する確率

条件付き発生確率$P(x_t|x_{t-1}^{(n)})$がnより小さい長さmの通報発生後の記号x_tの発生確率で表される.

m重マルコフ情報源（$m > 1$）
単純マルコフ情報源（$m = 1$）

$$\forall t,\ \forall n \quad P(x_t|\boldsymbol{x}_{t-1}^{(n)}) = P(x_t|x_{t-1})$$

「情報源アルファベットS＋条件付き発生確率」により規定される.

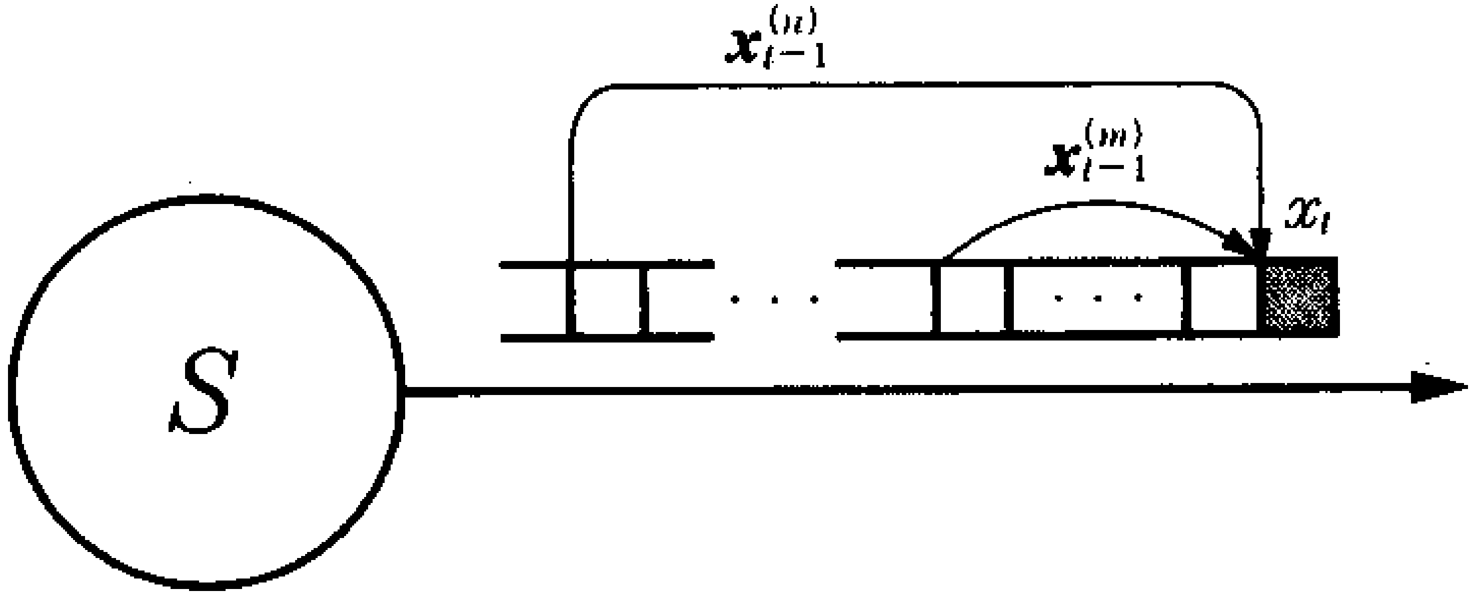


図 4.2　マルコフ情報源

4.6 マルコフ連鎖（マルコフチェーン）

時刻tでの状態を表す確率変数　$x(t)$
状態集合　$Q = \{q_1, q_2, \dots, q_N\}$

状態確率ベクトル　$\boldsymbol{u}(t) = (u_1(t),\ u_2(t), \dots,\ u_N(t))$

$$u_k(t) = P(x(t) = q_k)$$

$$0 \le u_k(t) \le 1\ (k = 1, 2, \dots, N),\ \sum_{k=1}^{N} u_k(t) = 1$$

状態遷移確率行列（状態遷移行列）

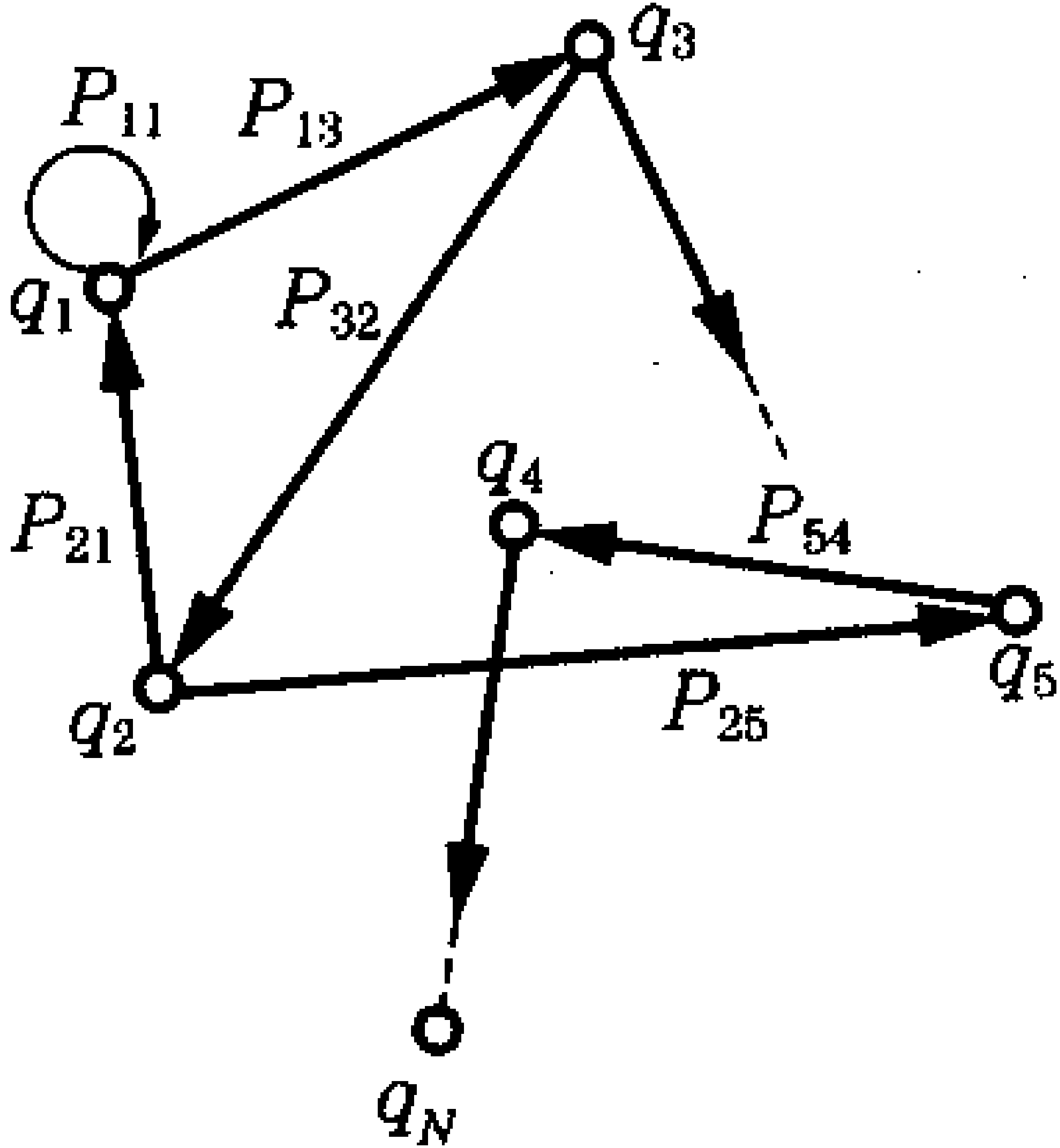
$$\boldsymbol{P} = \begin{bmatrix} P_{11} & \cdots & P_{1N} \\ \vdots & \ddots & \vdots \\ P_{N1} & \cdots & P_{NN} \end{bmatrix}$$

P_{kl}　状態遷移確率

$$P_{kl} = P(q_l | q_k) = P(q_k \to q_l)$$

$$0 \le P_{kl} \le 1\ \ (k, l = 1, 2, \dots, N), \qquad \sum_{l=1}^{N} P_{kl} = 1$$

$$
\text{状態遷移確率行列}：\boldsymbol{P} = \begin{array}{cc} & \begin{array}{ccccc} q_1 & \cdots & q_\ell & \cdots & q_N \end{array} \\ \begin{array}{c} q_1 \\ \vdots \\ q_k \\ \vdots \\ q_N \end{array} & \begin{bmatrix} P_{11} & \cdots & \uparrow & \cdots & P_{1N} \\ \vdots & \ddots & \uparrow & \cdot & \cdot \\ \rightarrow & \rightarrow & P_{k\ell} & \cdot & \cdot \\ \vdots & \cdot & \cdot & \cdot & \cdot \\ P_{N1} & \cdot & \cdot & \cdot & P_{NN} \end{bmatrix} \end{array}
$$


P_{11}
P_{13}
q_3
q_1
P_{32}
P_{21}
q_4
P_{54}
q_5
q_2
P_{25}
q_N

時刻tにおける状態確率ベクトル

$\boldsymbol{u}(0) \rightarrow \boldsymbol{u}(t)$について考える

$\boldsymbol{u}(t) = \boldsymbol{u}(t-1)\boldsymbol{P}$であるから$\boldsymbol{u}(t) = \boldsymbol{u}(0)\boldsymbol{P}^t$

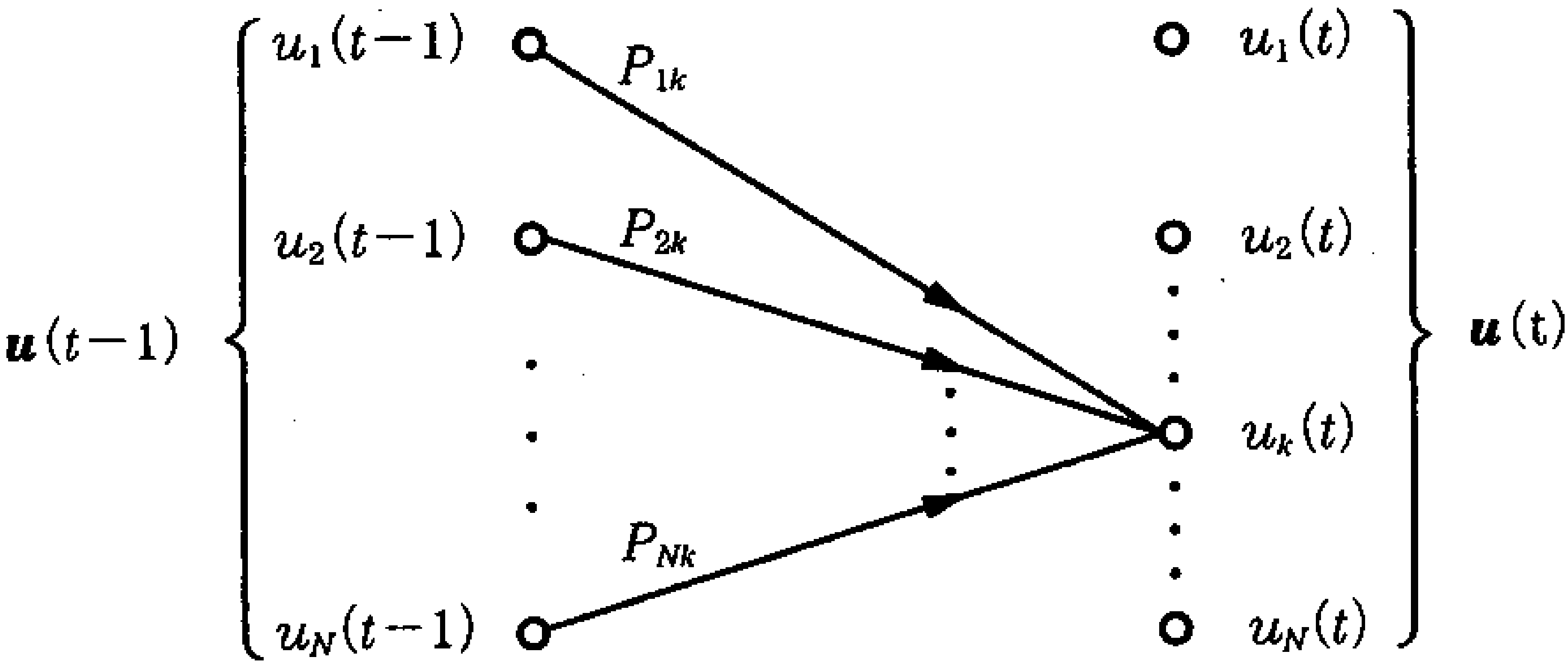


図 4.5　$\boldsymbol{u}(t)$ と $\boldsymbol{u}(t-1)$ の関係

4.7　マルコフ情報源モデル

単純マルコフ情報源モデル

　情報源アルファベット　$S=\{s_1,\ \ s_2,\ \ \ldots\ ,\ \ s_n\}$

　条件付き発生確率　　$P(s_k|s_l)\quad k,l=1,2,\ldots,n$

　（状態遷移確率，状態遷移行列）

状態遷移行列：

$$\boldsymbol{P}(1)=\boldsymbol{P}=\begin{array}{cc} & \begin{array}{ccccc} s_1 & \cdots & s_\ell & \cdots & s_n \end{array} \\ \begin{array}{c} s_1 \\ \vdots \\ s_k \\ \vdots \\ s_n \end{array} & \begin{bmatrix} P_{11} & \cdots & \uparrow & \cdots & P_{1n} \\ \vdots & \ddots & \uparrow & \cdot & \cdot \\ \rightarrow & \rightarrow & P_{k\ell} & \cdot & \cdot \\ \vdots & \cdot & \cdot & \cdot & \cdot \\ P_{n1} & \cdot & \cdot & \cdot & P_{nn} \end{bmatrix} \end{array}$$

発生平均情報量（発生エントロピー）

$$H(S|S)$$
$$= \sum_{l=1}^{n} H(S|s_l)P(s_l) = -\sum_{k=1}^{n}\sum_{l=1}^{n} P(s_k, s_l) log P(s_k|s_l)$$

［bit／情報源記号］

4.8 シャノン線図

2元マルコフ情報源$S = \{0, 1\}$を考える.

単純マルコフ情報源モデルを考える.

状態集合 $Q = S = \{0,\ 1\}$

状態遷移確率

$$P(0|0) = a,\ P(1|0) = 1 - a$$
$$P(0|1) = 1 - b,\ P(1|1) = b$$

状態遷移行列

$$P = \begin{bmatrix} a & 1 - a \\ 1 - b & b \end{bmatrix}$$

状態遷移行列

$$P = \begin{bmatrix} a & 1-a \\ 1-b & b \end{bmatrix}$$

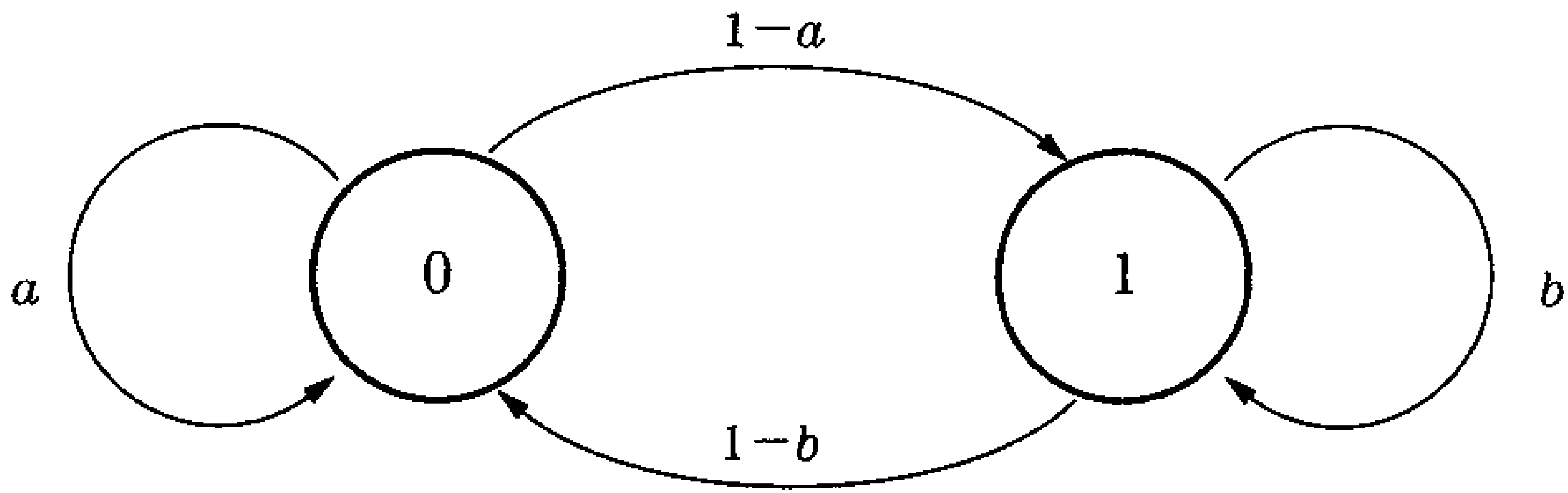


図 4.7　2 元単純マルコフ情報源のシャノン線図（$Q=S$ の場合）

2重マルコフ情報源

情報源アルファベット $S = \{0, 1\}$

状態集合 $Q = S^2 = \{q_1, q_2, q_3, q_4\} = \{00,\ 01,\ 10,\ 11\}$

条件付き発生確率

$$P(0|00) = a,\ P(1|00) = 1 - a$$
$$P(0|01) = b,\ P(1|01) = 1 - \mathrm{b}$$
$$P(0|10) = c,\ P(1|10) = 1 - \mathrm{c}$$
$$P(0|11) = d,\ P(1|11) = 1 - d$$

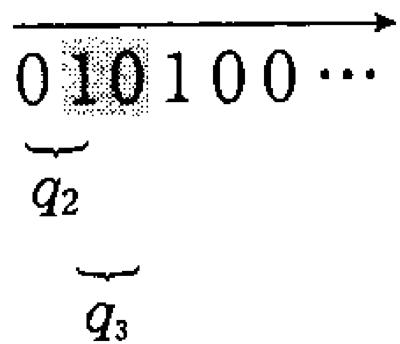


図 4.9　2 重マルコフ情報源からの通報

$$
\boldsymbol{P}(2)=\begin{bmatrix} a & 1-a & 0 & 0 \\ 0 & 0 & b & 1-b \\ c & 1-c & 0 & 0 \\ 0 & 0 & d & 1-d \end{bmatrix}
$$

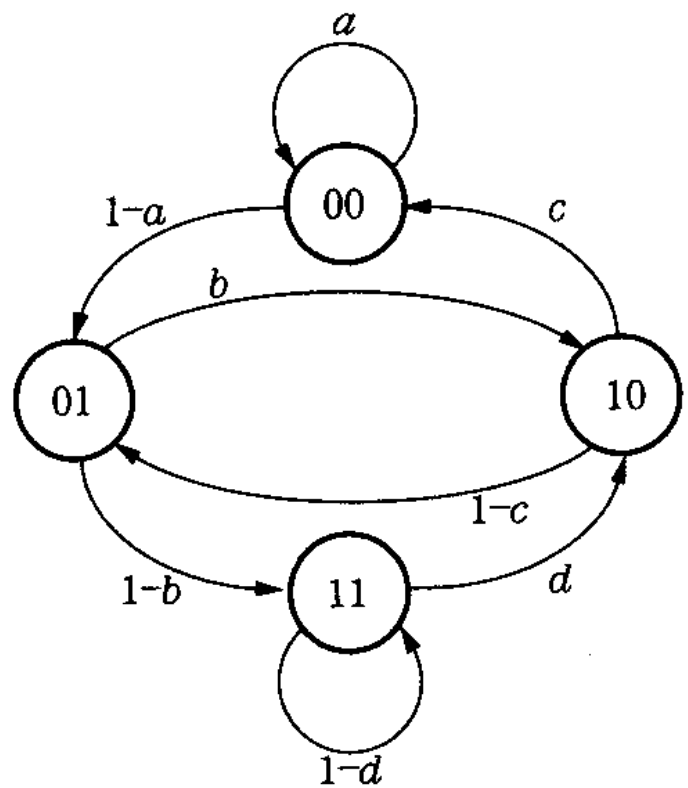


図 4.8　2 元 2 重マルコフ情報源のシャノン線図（$Q=S^2$ の場合）