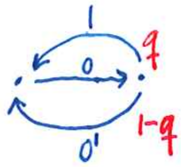
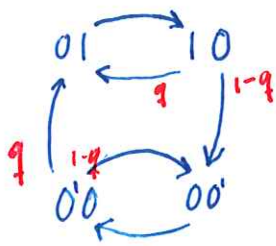
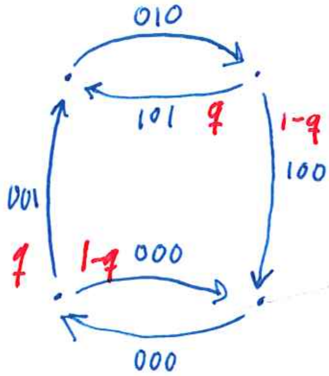
4.6



$\longleftrightarrow$

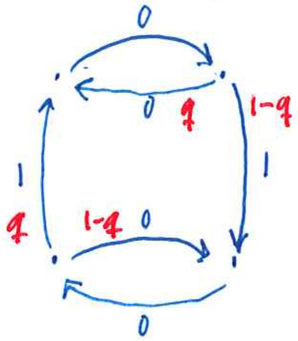


$\longleftrightarrow$



$S(0) = \frac{1}{2}\left(q \log \frac{1}{q} + (1-q) \log \frac{1}{1-q}\right)$

R18 $\longmapsto$

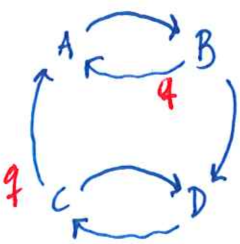


If $q = \frac{1}{2}$ it is impossible to know which node we are in. (Each infinite sequence could be generated in two ways.) But in that case the automaton can be rewritten as

i.e. identical to $t=0$.

So for $q=\frac{1}{2}$ the entropy does not change and $S(t) = \frac{1}{2} \log 2$.

If $q \neq \frac{1}{2}$ it is almost surely possible to know the node, since one solution branch will be increasingly likely and the other decreasingly likely as the length of the preceding sequence increases.



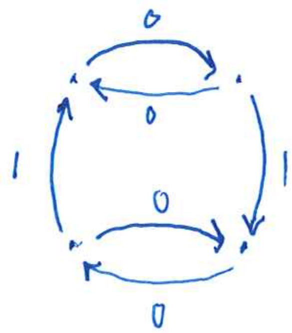
$$\left.\begin{aligned} p(A) &= q\,(p(C) + p(B)) \\ p(D) &= (1-q)(p(B) + p(C)) \\ p(C) &= p(D) \\ p(B) &= p(A) \end{aligned}\right\} \Rightarrow \begin{cases} p(A) = p(B) = \frac{q}{2} \\ p(C) = p(D) = \frac{1-q}{2} \end{cases}$$

$$\Rightarrow S(1) = \frac{1}{2}\left(q \log \frac{1}{q} + (1-q) \log \frac{1}{1-q}\right)$$

$$= S(0)$$

4.6 cont.

To find $s(2)$, look again at the automaton for $t=1$:



This automaton can generate the three-symbol sequences

$\{000, 001, 010, 100, 101\}$.

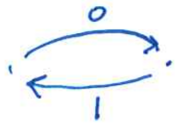
Check R18:

~~111~~	~~110~~	101	100	~~011~~	010	001	000
~~0~~	~~0~~	0	1	~~0~~	0	1	0

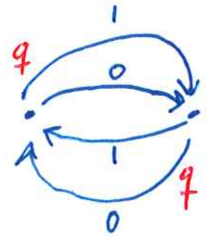
Given that the sequence at $t=1$ only contains the above-mentioned triplets (trigrams), R18 is almost reversible and $s(2) = s(1)$.

4.7

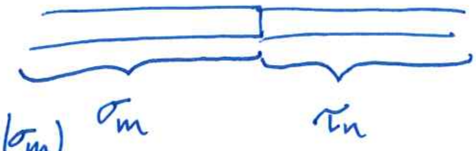
The periodic sequence is generated by



and the noisy sequence by



For the pre-sequences σ_m and post-sequences τ_n the corr. complexity is



$$(3.38) \quad \eta = \lim_{m\to\infty} \lim_{n\to\infty} \sum_{\sigma_m} p(\sigma_m) \sum_{\tau_n} p(\tau_n|\sigma_m) \log \frac{p(\tau_n|\sigma_m)}{p(\tau_n)}$$

If $q < \frac{1}{2}$ we will almost surely know the current node as $m \to \infty$, so

$$\eta = \lim_{n\to\infty} \sum_{z} p(z) \sum_{\tau_n} p(\tau_n|z) \log \frac{p(\tau_n|z)}{p(\tau_n)}.$$

Use Bayes' theorem:

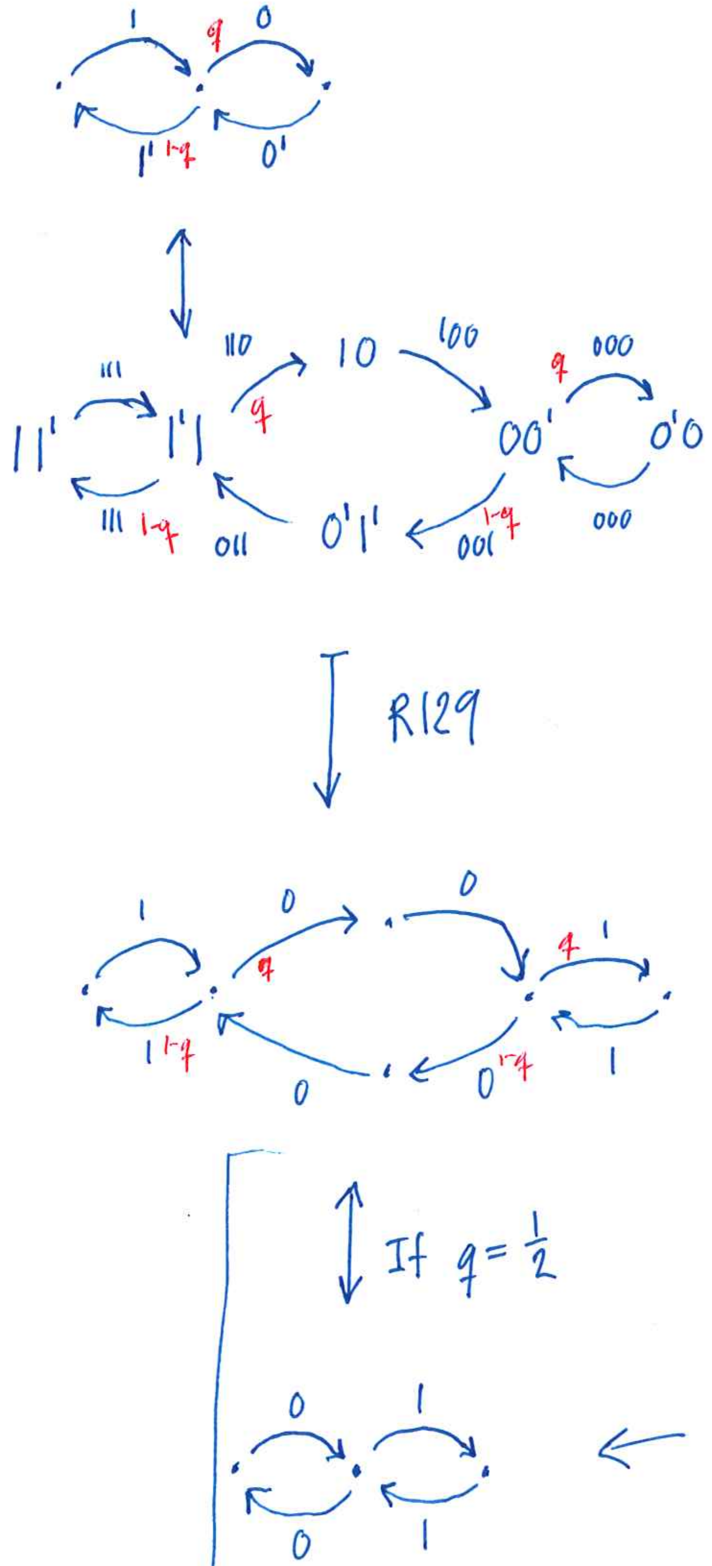
$$\eta = \lim_{n\to\infty} \sum_{z} p(z) \sum_{\tau_n} p(\tau_n|z) \log \frac{p(z|\tau_n)}{p(z)}$$

$$= \underbrace{\sum_{z} p(z) \log \frac{1}{p(z)}}_{= \log 2} + \underbrace{\lim_{n\to\infty} \sum_{z} p(z) \sum_{\tau_n} p(\tau_n|z) \log p(z|\tau_n)}$$

$= 0$, because as $n \to \infty$ we will be almost sure of $z|\tau_n$, i.e. $p(z|\tau_n) \xrightarrow[n\to\infty]{} 1$ if $p(\tau_n|z) > 0$.
(In other words, either $p(\tau_n|z) = 0$ or $p(z|\tau_n) \to 1$.)

$\Rightarrow \eta = \log 2$ for all $q < \frac{1}{2}$.

4.8



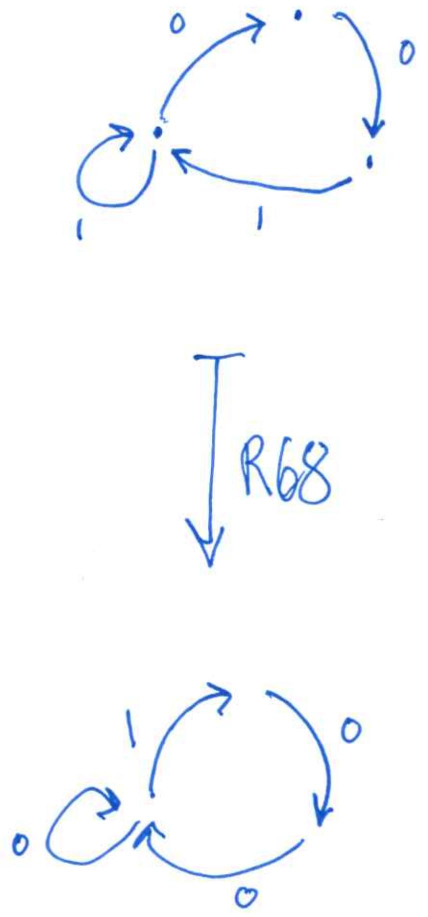
It looks the same as at $t=0$, so the entropy is $s(0) = s(1) = s(2) = \ldots = \frac{1}{2}\log 2$ when $q = \frac{1}{2}$.

If $q \neq \frac{1}{2}$, correlation information should increase and entropy decrease (compared to the situation $q = \frac{1}{2}$).

4.9

R68:

111	110	101	100	011	010	001	000
0	1	0	0	0	1	0	0



Look at the sequences:

t=0 00100111100110010010...
t=1 00100000100010010010...

at t=0, pairs of zeros, separated by blocks of ones. At t=1, only the last one in each block is preserved.

And at t=2, the same thinking applies.
So the sequence at t=2 is identical to t=1.

$s(0) = \frac{1}{2} \log 2$

$s(1) = \frac{1}{2} \log 2$

$s(2) = s(1)$

$s(3) = s(2)$

$\vdots$