8. Machine Learning and Real-world Data.

(a) $\mu = (A, B)$. A: the matrice of the transition probability $\Rightarrow \underset{\text{special states}}{3+2}$ 5×5

B: the matrice of the emission probability $\Rightarrow \underset{\text{special outputs}}{6+2}$ 5×8

states $S = \{S_0, F, L_1, L_2, S_e\}$. S_0: start state. S_e: end state

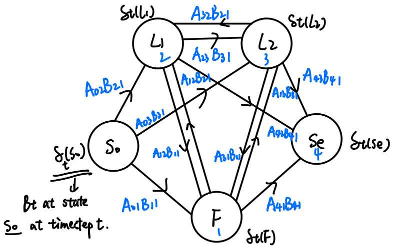
output $K = \{k_0, 1, 2, 3, 4, 5, 6, k_e\}$ k_0: start output. k_e: end output

$A_{ij} = P(X_t = S_j \mid X_{t-1} = S_i)$, $i, j \in [0, 5)$

$B_{ij} = P(O_t = k_j \mid X_t = k_i)$, $i \in [0,5)$, $j \in [0,8)$

(b) A state diagram : Actually, we haven't learnt about this

For example: the sample term is "1" (→ correspond to k_1). Here I list all the possible transition pairs.



I label some arcs with the possibility.

c) probabilies and training data counts

ci) The training data count: $Count_{trans}(X_t = \overset{\text{"L2"}}{3} \mid X_{t-1} = \overset{\text{"F"}}{1}) = 0$

And $Count_{trans}(X_t = 3 \mid X_{t-1} = \overset{\text{"L1"}}{2})$ may be quite large. Hence $A_{13} = P(X_t = 3 \mid X_{t-1} = 1) = 0$

This can be modelled by HMM, since the possibility including transition directly from F to S_2 will be zero. The most probable hidden states sequence won't include $(F \to S_2)$

cii) $Count_{tran}(X_t = \overset{Se}{4} \mid X_{t-1} = \overset{F}{1}) = 1$. $A_{14} = P(X_t = 4 \mid X_{t-1} = 1) = 1$

This can be modelled by HMM. The hidden path with F on the last roll is likely to yield higher probability

ciii). Training data counts: $Count_{trans}(X_t = 3 \mid X_{t-1} = m)$ + $Count_{emission}(O_t \mid X_t = 3)$ — They may both become small. (m can be any state aside from 3) will be small. Hence $P(X_t = 3 \mid X_{t-1} = m)$ will be small too. This behaviour reflects in HMM producing result sequences with fewer L_2. It may not be exactly less than 2 in a row, so to some extent, it can be modelled.

(iv) It's hard to tell the effect on probabilities and counts, since we're not sure which dice the croupier switches to and we only count the associations between output V and the dice at the same timestep. Therefore, it can't be modelled by HMM.