Hidden Markov models and dynamic programming

Matthew Macauley

Department of Mathematical Sciences
Clemson University
http://www.math.clemson.edu/~macaule/

Math 4500, Spring 2017

The occasionally dishonest casino

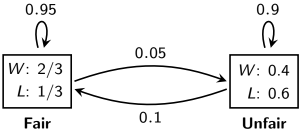
3 canonical questions

Given a sequence of roles by the casino:

`WWWLWLWLWLWWLWWLLLWWWWLWWLWWLWLWLLLWLWWLLWWLWLWLLWWLLLWLWWWWLWLWWWWL`

one may ask:

1. Evaluation: How likely is this sequence given our model?
2. Decoding: When was the casino rolling the fair vs. the unfair die?
3. Learning: Can we deduce the probability parameters if we didn't know them? (e.g., "*how loaded are the die?*", and "*how often does the casino switch?*")



Problem #1: Evaluation

For CpG identification, we need the *posterior probabilities* $P(\pi_t = k \mid x)$, for each $k \in Q$ and $t = 1, 2, \ldots, \ell$. By Bayes' theorem,

$$P(\pi_t = k \mid x) = \frac{P(x,\, \pi_t = k)}{P(x)}$$

We can compute $P(x,\, \pi_t = k)$ recursively:

$$\begin{aligned} P(x,\, \pi_t = k) &= P(x_1 x_2 \cdots x_t,\, \pi = k) \cdot P(x_{t+1} x_{t+2} \cdots x_\ell \mid x_1 x_2 \cdots x_t,\, \pi_t = k) \\ &= P(x_1 x_2 \cdots x_t,\, \pi = k) \cdot P(x_{t+1} x_{t+2} \cdots x_\ell \mid \pi_t = k) \\ &= f_k(t) \cdot b_k(t). \end{aligned}$$

The forward-backward algorithm

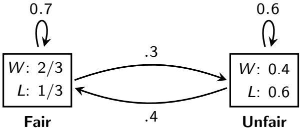
Given an emitted sequence $x = x_1 x_2 x_3 \cdots x_\ell$, we will use the

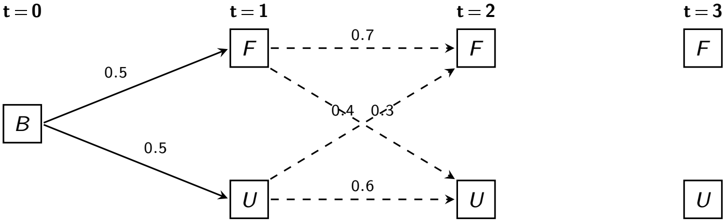
- **forward algorithm** to compute $f_k(t)$: *the probability of getting* $x = x_1 x_2 x_3 \cdots x_t$ *and ending up in state* k.
- **backward algorithm** to compute $b_j(t)$: *the probability of observing* $x_{t+1} \cdots x_\ell$ *from state* k.

It is also straightforwrd to compute $P(x)$ using either of these algorithms.

The forward algorithm

Example. Compute $P(x)$, for $x = \text{LWW}$.





Forward algorithm

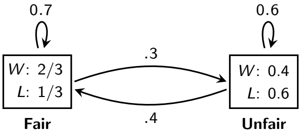
1. Initialize ($t = 0$): Set $f_B(0) = 1$, and $f_j(0)$, for all $j \in Q$.
2. Recursion: do for $t = 1, 2, \dots, \ell$:

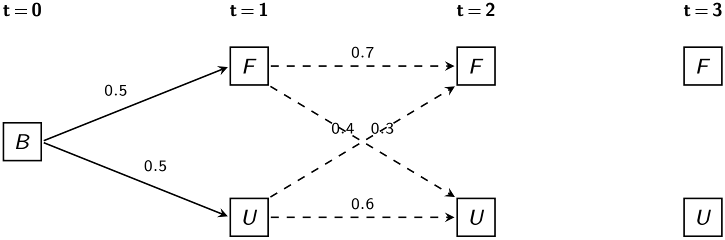
   for each $k \in Q$, define $$f_k(t) := e_k(x_t) \sum_{j \in Q} f_j(t-1) a_{jk}$$

3. Termination: Set $P(x) = \sum_{k \in Q} f_k(\ell)$.

The forward algorithm

Example. Compute $P(x)$, for $x =$ LWW.



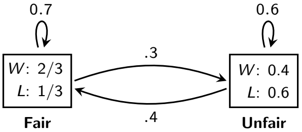


t=0. $f_B(0) = 1, \quad f_F(0) = 0, \quad f_U(0) = 0.$

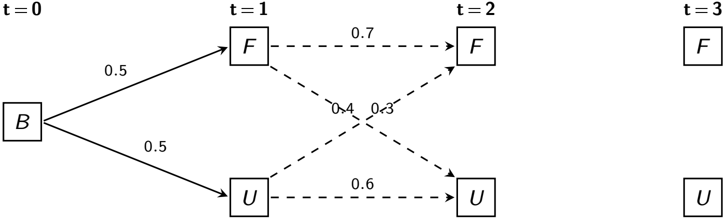
t=1. $f_F(1) = P(x_1 = L,\ \pi_1 = F) = f_B(0) \cdot a_{BF} \cdot e_F(L) = 1 \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}.$

$f_U(1) = P(x_1 = L,\ \pi_1 = U) = f_B(0) \cdot a_{BU} \cdot e_U(L) = 1 \cdot \frac{1}{2} \cdot \frac{6}{10} = 0.3.$

The forward algorithm



Example. Compute $P(x)$, for $x = \text{LWW}$.



$$\underline{t=2}: \quad f_F(2) = P(x_1x_2 = LW,\ \pi_2 = F) = f_F(1)\cdot a_{FF}\cdot e_F(W) + f_U(1)\cdot a_{UF}\cdot e_F(W) = \tfrac{1}{6}(.7)\tfrac{2}{3} + (.3)(.4)\tfrac{2}{3} \approx 0.1578.$$

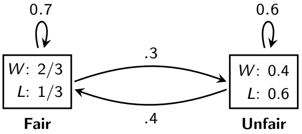
$$f_U(2) = P(x_1x_2 = LW,\ \pi_2 = U) = f_F(1)\cdot a_{FU}\cdot e_U(W) + f_U(1)\cdot a_{UU}\cdot e_U(W) = \tfrac{1}{6}(.3)(.4) + (.3)(.6)(.4) = 0.092.$$

$$\underline{t=2}: \quad f_F(3) = P(x_1x_2x_3 = LWW,\ \pi_3 = F) = f_F(2)\cdot a_{FF}\cdot e_F(W) + f_U(2)\cdot a_{UF}\cdot e_F(W) = (.1578)(.7)\tfrac{2}{3} + (.092)(.4)\tfrac{2}{3} \approx .0982.$$

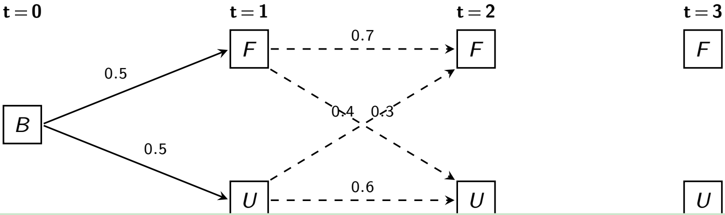
$$f_U(3) = P(x_1x_2 = LWW,\ \pi_3 = U) = f_F(2)\cdot a_{FU}\cdot e_U(W) + f_U(2)\cdot a_{UU}\cdot e_U(W) = (.1578)(.3)(.4) + (.092)(.6)(.4) \approx 0.0410.$$

Now, $P(x) = P(x = LWW) = f_F(3) + f_U(3) \approx .0982 + .0410 = .1392.$

The backward algorithm



Example. Compute $P(x)$, for $x =$ LWW.



Backward algorithm and $b_j(\ell)$

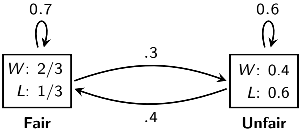
1. Initialize ($t = \ell$): Set $b_k(\ell) = 1$ for all $j \in Q$.
2. Recursion: do for $t = \ell - 1, \dots, 2, 1$:

$$\text{for each } j \in Q, \quad b_j(t) := P(x_{t+1}x_{t+2}\cdots x_\ell \mid \pi_t = j)$$
$$= \sum_{k \in Q} P(\pi_{t+1} = k \mid \pi_t = j) \cdot e_k(x_{t+1}) \cdot P(x_{t+2}\cdots x_\ell \mid \pi_{t+1} = k)$$
$$= \sum_{k \in Q} a_{jk} e_k(x_{t+1}) b_k(t+1).$$

3. Termination: Set $P(x) = \sum a_{Bk} e_k(x_1) b_k(1)$.

The backward algorithm

Example. Compute $P(x)$, for $x = \text{LWW}$.



t=3. $b_F(3) = 1, \quad b_U(3) = 1.$

t=2. $b_F(2) = a_{FF}e_F(W)b_F(3) + a_{FU}e_U(W)b_U(3) = (.7)\frac{2}{3} + (.3)(.4) = \frac{44}{75} \approx 0.5866.$

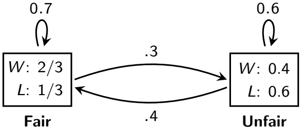
$b_U(2) = a_{UF}e_F(W)b_F(3) + a_{UU}e_U(W)b_U(3) = (.4)\frac{2}{3} + (.6)(.4) = \frac{38}{75} \approx 0.5067.$

t=1. $b_F(1) = a_{FF}e_F(W)b_F(2) + a_{FU}e_U(W)b_U(2) = (.7)\frac{2}{3} \cdot \frac{44}{75} + (.3)(.4)\frac{38}{75} \approx 0.3346.$

$b_U(2) = a_{UF}e_F(W)b_F(2) + a_{UU}e_U(W)b_U(2) = (.4)\frac{2}{3} \cdot \frac{44}{75} + (.6)(.4)\frac{38}{75} \approx 0.2780.$

The forward-backward algorithm

Example. Compute $P(x)$, for x = LWW.



- $P(\pi_1 = F \mid x_1x_2x_3 = LWW) = \dfrac{f_F(1)b_F(1)}{P(x)} \approx \dfrac{(1/6)(.3346)}{.1392} \approx 0.4006$
- $P(\pi_1 = U \mid x_1x_2x_3 = LWW) = \dfrac{f_F(1)b_F(1)}{P(x)} \approx \dfrac{(.3)(.2780)}{.1392} \approx 0.5991$
- $P(\pi_2 = F \mid x_1x_2x_3 = LWW) = \dfrac{f_F(1)b_F(1)}{P(x)} \approx \dfrac{(.1578)(.5866)}{.1392} \approx 0.6650$
- $P(\pi_1 = U \mid x_1x_2x_3 = LWW) = \dfrac{f_F(1)b_F(1)}{P(x)} \approx \dfrac{(.092)(.5067)}{.1392} \approx 0.3349$
- $P(\pi_3 = F \mid x_1x_2x_3 = LWW) = \dfrac{f_F(1)b_F(1)}{P(x)} \approx \dfrac{(.0982)(1)}{.1392} \approx 0.7055$
- $P(\pi_3 = U \mid x_1x_2x_3 = LWW) = \dfrac{f_F(1)b_F(1)}{P(x)} \approx \dfrac{(.041)(1)}{.1392} \approx 0.2945.$

Decoding and the Viterbi algorithm

Problem #2: Decoding

Given an observed path $x = x_1x_2x_3\cdots x_\ell$, what is the most likely hidden path $\pi = \pi_1\pi_2\pi_3\cdots\pi_\ell$ to emit x? That is, compute

$$\pi_{\max} = \arg\max_{\pi} P(\pi|x) = \arg\max_{\pi} P(x,\pi)$$

Assume that for each $j \in Q$, we've computed $\pi_1\pi_2\cdots\pi_{t-2}\pi_{t-1}$ of highest probability among those emitting $x_1x_2\cdots x_{t-1}$.

Denote the probability of this path by

$$v_j(t-1) = \max_{\pi=\pi_1\cdots\pi_{t-1}} P(\pi_{t-1}=j, x_{t-1}).$$

Then, for each $k \in Q$, say emitting $x_1x_2\cdots x_t$:

$$v_k(t) = \max_{\pi_1\cdots\pi_{t-1}} P(\pi_{t-1}=k, x_t) = \max_{j\in Q}\big\{v_j(t-1)a_{jk}e_k(x_t)\big\} = e_k(x_t)\max_{j\in Q}\big\{v_j(t-1)a_{jk}\big\}.$$

Decoding and the Viterbi algorithm

Viterbi algorithm

1. Initialize ($t = 0$): Set $v_B(0) = 1$, and $v_j(0)$, for all $j \in Q$.
2. Recursion: do for $t = 1, 2, \dots, \ell$:

$$\text{for each } k \in Q, \text{ define} \quad v_k(t) := e_k(x_t) \max_{j \in Q} \{v_j(t-1) a_{jk}\}$$

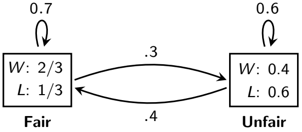
   Also, set $\text{ptr}_k(t) = r = \arg\max_j \{v_j(t-1) a_{jk}\}$.

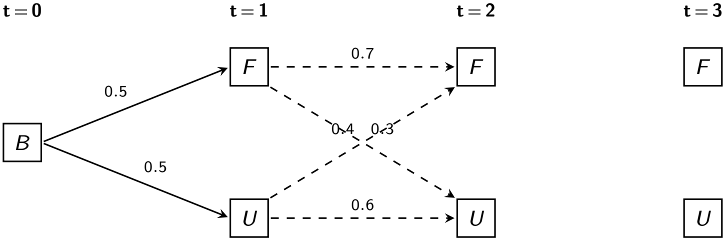
3. Termination: Set $P(x, \pi^*) = \max_{\pi} P(x, \pi) = \max_{j \in Q} \{v_j(\ell)\}$, and $\text{ptr}_k(\ell) = \pi^*_\ell$.

The maximum probability path can be found by tracing back through the pointers.

Decoding and the Viterbi algorithm

Example. Given $x =$ LWW, what is the most likely path $\pi = \pi_1\pi_2\pi_3$?





t=1. $\quad v_F(1) = \max_{\pi_1} P(\pi_1 = F,\ x_1 = L) = \max\{v_B(0) \cdot a_{BF} \cdot e_F(L)\} = 1 \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}.$

$$v_U(1) = \max_{\pi_1} P(\pi_1 = U,\ x_1 = L) = \max\{v_B(0) \cdot a_{BU} \cdot e_U(L)\} = 1 \cdot \frac{1}{2}(.6) = 0.3.$$