How To Identify Hidden Markov Models

An Algebraic Statistical Answer

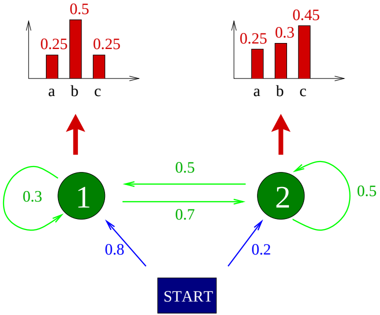
Alexander Schönhuth

Department of Mathematics
UC Berkeley

March 2010

Guideline

Hidden Markov Chains



- Initial probabilities $\pi = (0.8, 0.2)^T$
- Transition probabilities

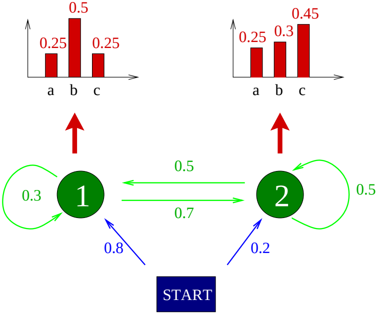
$$M = (m_{ij} := P(j \to i))_{i,j=1,2}$$
$$= \begin{pmatrix} 0.3 & 0.5 \\ 0.7 & 0.5 \end{pmatrix}$$

- Emission probabilities, e.g. $E_{1b} = 0.5, E_{2c} = 0.45$.

Random source (X_t) with values in $\Sigma = \{a, b, c\}$:

e.g.: $P_X(X_1 = a, X_2 = b) = \pi_1 e_{1a}(a_{11}e_{1b} + a_{12}e_{2b}) + \pi_2 e_{2a}(a_{21}e_{1b} + a_{22}e_{2b})$

Alternative Formulation



- Initial probabilities $\pi = (0.8, 0.2)^T$
- Transition probabilities

$$M = (m_{ij} := P(j \to i))_{i,j=1,2} = \begin{pmatrix} 0.3 & 0.5 \\ 0.7 & 0.5 \end{pmatrix}$$

- Emission probabilities, e.g. $E_{1b} = 0.5, E_{2c} = 0.45$.

Alternative formulation: Let

$$T_v := M \begin{pmatrix} E_{1v} & 0 \\ 0 & E_{2v} \end{pmatrix}, v = a, b, c$$

then

$$P_X(X_1 = v_1, ..., X_n = v_n) = \langle \begin{pmatrix} 1 \\ 1 \end{pmatrix} | T_{v_n} \cdot \ldots \cdot T_{v_1} | \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} \rangle.$$

Identifiability of HMMs

[Blackwell and Koopmans, 1957]

- Identifiability Problem: When do two HMMs give rise to equivalent processes? Solutions:
  - Ito, Amari, Kobayashi (1992): exponential runtime
  - AS (2008): linear runtime
- When is a process due to an HMM? Solution:
  - Heller (1965): characterization of no practical use
- Remark: Hidden Markov process on d hidden states is uniquely determined by its distribution on strings of length $2d - 1$.

Identifiability of HMMs

[Blackwell and Koopmans, 1957]

- Identifiability Problem: When do two HMMs give rise to equivalent processes? Solutions:
  - Ito, Amari, Kobayashi (1992): exponential runtime
  - AS (2008): linear runtime
- When is a process due to an HMM? Solution:
  - Heller (1965): characterization of no practical use
- Remark: Hidden Markov process on d hidden states is uniquely determined by its distribution on strings of length $2d - 1$.
- Question of practical relevance: when is a distribution on strings of length n induced by an HMM with at most $\lceil \frac{n}{2} \rceil$ hidden states?
- If so, how many are there?
- Algebraic statistics can give answers!

Algebraic Statistical Models

The Hidden Markov Model

Definition: Hidden Markov Model for d hidden states and strings of length n over the Σ:

$$\begin{array}{cccc} \mathbf{f}_{n,d}: & \mathbb{C}^{d(d-1)+d|\Sigma|} & \longrightarrow & \mathbb{C}^{|\Sigma|^n} \\ & ((T_a = MO_a)_{a\in\Sigma}), \pi) & \mapsto & (\langle \mathbf{1}|T_{v_n}\cdot \ldots \cdot T_{v_1}|\pi\rangle)_{v=v_1\ldots v_n\in\Sigma^n}. \end{array}$$

where $\pi, \mathbf{1} = (1, \ldots, 1)^T \in \mathbb{C}^d$, $M, O_a \in \mathbb{C}^{d^2}$ such that

$$\begin{array}{rcl} \mathbf{1}^T M & = & \mathbf{1}^T \\ (O_a)_{ij} & = & \begin{cases} E_{ia} & i = j \\ 0 & i \neq j \end{cases}, \quad \sum_a O_a = \mathrm{Id}. \end{array}$$

Wanted: Equations describing

$$\overline{\mathrm{im}\,(\mathbf{f}_{n,d})}$$

Algebraic Statistical Models

The Matrix Markov Model

Definition: Matrix Markov Model of rank d for strings of length n

$$\mathbf{g}_{n,d} : \begin{array}{ccc} \mathbb{C}^{|\Sigma| d^2} & \longrightarrow & \mathbb{C}^{|\Sigma|^n} \\ ((T_a)_{a\in\Sigma}, \pi) & \mapsto & (\langle \mathbf{1} | T_{v_n} \ldots T_{v_1} | \pi \rangle)_{v=v_1\ldots v_n \in \Sigma^n} . \end{array}$$

where $\mathbf{1}^T (\sum_a T_a) = \mathbf{1}$.

Algebraic Statistical Models

The Matrix Markov Model

Definition: Matrix Markov Model of rank d for strings of length n

$$\mathbf{g}_{n,d} : \begin{array}{ccc} \mathbb{C}^{|\Sigma| d^2} & \longrightarrow & \mathbb{C}^{|\Sigma|^n} \\ ((T_a)_{a\in\Sigma}, \pi) & \mapsto & (\langle \mathbf{1} | T_{v_n} \ldots T_{v_1} | \pi \rangle)_{v=v_1\ldots v_n \in \Sigma^n} . \end{array}$$

where $\mathbf{1}^T (\sum_a T_a) = \mathbf{1}$.

Reminder: Hidden Markov Model on d hidden states for strings of length n

$$\mathbf{f}_{n,d} : \begin{array}{ccc} \mathbb{C}^{d^2 + (|\Sigma|-1)d} & \longrightarrow & \mathbb{C}^{|\Sigma|^n} \\ ((T_a = MO_a)_{a\in\Sigma}, \pi) & \mapsto & (\langle \mathbf{1} | T_{v_n} \ldots T_{v_1} | \pi \rangle)_{v=v_1\ldots v_n \in \Sigma^n} . \end{array}$$

Hence

$$\mathrm{im}\,(\mathbf{f}_{n,d}) \subset \mathrm{im}\,(\mathbf{g}_{n,d}).$$

Matrix Markov Model

Dimension

Lemma: Let $S : \mathbb{C}^d \to \mathbb{C}^d$ be an isomorphism such that $\mathbf{1}^T S = \mathbf{1}^T$ and

$$\begin{aligned} T'_a &:= S^{-1} T_a S \\ x' &:= S^{-1} x. \end{aligned}$$

Then

$$\mathbf{g}_{n,d}((T_a)_{a\in\Sigma}, x) = \mathbf{g}_{n,d}((T'_a)_{a\in\Sigma}, x')$$

Theorem: Let $n \geq 2d - 1$.

$$\dim \overline{\mathrm{im}\, \mathbf{g}_{n,d}} = (|\Sigma| - 1)d^2 + d.$$

Differences

The Strassen Condition

Lemma: Equivalent statements:

(i)

$$\mathbf{g}_{n,d}((T_a)_{a\in\Sigma}, x) \in \operatorname{im} \mathbf{f}_{n,d}$$

(ii) [Strassen Condition]

$$\forall a,b,c \in \Sigma : \quad (T_a)^{-1} T_c (T_b)^{-1} = (T_b)^{-1} T_c (T_a)^{-1}.$$

Remark: The Strassen Condition reflects that

$$T_c (T_a)^{-1}, T_c (T_b)^{-1}$$

are simultaneously diagonalizable.

Differences

The Strassen Condition

Lemma: Equivalent statements:

(i)

$$\mathbf{g}_{n,d}((T_a)_{a\in\Sigma}, x) \in \mathrm{im}\, \mathbf{f}_{n,d}$$

(ii) [Strassen Condition]

$$\forall a,b,c \in \Sigma : \quad (T_a)^{-1}T_c(T_b)^{-1} = (T_b)^{-1}T_c(T_a)^{-1}.$$

Remark: The Strassen Condition reflects that

$$T_c(T_a)^{-1}, T_c(T_b)^{-1}$$

are simultaneously diagonalizable.

Corollary:

$$\Sigma = \{a,b\} : \quad \overline{\mathrm{im}\, \mathbf{f}_{n,d}} = \overline{\mathrm{im}\, \mathbf{g}_{n,d}}$$

since (ii) trivially satisfied for binary alphabet $\Sigma = \{a,b\}$.

Main Theorem
The Hankel Matrix

- Let $v \in \Sigma^n$. Write

$$p(v = v_1 \ldots v_n) := \mathbb{P}(X_1 = v_1, \ldots, X_n = v_n).$$

  for stochastic process (X_t).

- Write

$$uv = u_1 \ldots u_m v_1 \ldots v_n \in \Sigma^{m+n}$$

  for concatenation of
  $u = u_1 \ldots u_m \in \Sigma^m$ and
  $v = v_1 \ldots v_n \in \Sigma^n$.

Main Theorem

The Hankel Matrix

- Let $v \in \Sigma^n$. Write

$$p(v = v_1 \ldots v_n) := \mathbb{P}(X_1 = v_1, \ldots, X_n = v_n).$$

for stochastic process (X_t).

- Write

$$uv = u_1 \ldots u_m v_1 \ldots v_n \in \Sigma^{m+n}$$

for concatenation of $u = u_1 \ldots u_m \in \Sigma^m$ and $v = v_1 \ldots v_n \in \Sigma^n$.

- Consider the (infinite-dimensional) Hankel matrix

$$\mathcal{P}_p := [p(uv)]_{u,v \in \Sigma^*} \in \mathbb{R}^{\Sigma^* \times \Sigma^*}.$$

- **Example** ($\Sigma = \{0, 1\}$):

$$\mathcal{P}_p = \begin{pmatrix} p(\Box) & p(0) & p(1) & \ldots \\ p(0) & p(00) & p(10) & \ldots \\ p(1) & p(01) & p(11) & \ldots \\ p(00) & p(000) & p(100) & \ldots \\ p(01) & p(001) & p(101) & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Main Theorem

The Hankel Matrix

- Let $v \in \Sigma^n$. Write

$$p(v = v_1 \ldots v_n) := \mathbb{P}(X_1 = v_1, \ldots, X_n = v_n).$$

for stochastic process (X_t).

- Write

$$uv = u_1 \ldots u_m v_1 \ldots v_n \in \Sigma^{m+n}$$

for concatenation of
$u = u_1 \ldots u_m \in \Sigma^m$ and
$v = v_1 \ldots v_n \in \Sigma^n$.

- Consider the (infinite-dimensional) Hankel matrix

$$\mathcal{P}_p := [p(uv)]_{u,v \in \Sigma^*} \in \mathbb{R}^{\Sigma^* \times \Sigma^*}.$$

- **Example** ($\Sigma = \{0,1\}$):

$$\mathcal{P}_p = \begin{pmatrix} p(\square) & p(0) & p(1) & \ldots \\ p(0) & p(00) & p(10) & \ldots \\ p(1) & p(01) & p(11) & \ldots \\ p(00) & p(000) & p(100) & \ldots \\ p(01) & p(001) & p(101) & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Remark: For p hidden Markov process on d hidden states:

$$\operatorname{rk} \mathcal{P}_p \leq d.$$

Main Theorem

Set Theoretic Lemma

Lemma: Let $n \geq 2d - 1$. Then the following statements are equivalent:

(i)

$$(p(v))_{v \in \Sigma^n} \in (\operatorname{im} \mathbf{g}_{n,d} \setminus \operatorname{im} \mathbf{g}_{n,d-1}) \tag{1}$$

(ii)

$$\operatorname{rk} \mathcal{P}_{p,d-1,d-1} = \operatorname{rk} \mathcal{P}_{p,\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil} = \mathcal{P}_{p,\lceil \frac{n}{2} \rceil, \lfloor \frac{n}{2} \rfloor} = d \tag{2}$$

Main Theorem

Set Theoretic Lemma

Lemma: Let $n \geq 2d - 1$. Then the following statements are equivalent:

(i)

$$(p(v))_{v\in\Sigma^n} \in (\mathrm{im}\,\mathbf{g}_{n,d} \setminus \mathrm{im}\,\mathbf{g}_{n,d-1}) \tag{1}$$

(ii)

$$\mathrm{rk}\,\mathcal{P}_{p,d-1,d-1} = \mathrm{rk}\,\mathcal{P}_{p,\lfloor\frac{n}{2}\rfloor,\lceil\frac{n}{2}\rceil} = \mathcal{P}_{p,\lceil\frac{n}{2}\rceil,\lfloor\frac{n}{2}\rfloor} = d \tag{2}$$

Example: $n = 4, d = 2$

$$\mathcal{P}_{p,4,2} = \begin{pmatrix} p(\square) & p(0) & p(1) & p(00) & p(01) & p(10) & p(11) \\ p(0) & p(00) & p(10) & p(000) & p(010) & p(100) & p(110) \\ p(1) & p(01) & p(11) & p(001) & p(011) & p(101) & p(111) \\ p(00) & p(000) & p(100) & p(0000) & p(0100) & p(1000) & p(1100) \\ p(01) & p(001) & p(101) & p(0001) & p(0101) & p(1001) & p(1101) \\ p(10) & p(010) & p(110) & p(0010) & p(0110) & p(1010) & p(1110) \\ p(11) & p(011) & p(111) & p(0011) & p(0111) & p(1011) & p(1111) \end{pmatrix}$$

Main Theorem
Invariants

Let

$$
\begin{aligned}
I &:= \langle \det\left(p(u_i v_j)\right)_{1\le i,j\le d+1} \mid 0 \le |u_i|, |v_j| \le \lceil \tfrac{n}{2} \rceil, |u_i v_j| \le n \rangle \\
J &:= \langle \det\left(p(u_i v_j)\right)_{1\le i,j\le d} \mid 0 \le |u_i|, |v_j| \le d-1 \rangle.
\end{aligned}
$$

Note: To obtain equations for strings of length n, replace

$$p(uv) = \sum_{w,|w|=n-|uv|} p(uvw) \quad \text{for} \quad |uv| < n.$$

Theorem:

$$\overline{\operatorname{im} \mathbf{g}_{n,d}} = \operatorname{rad} I : \operatorname{rad} J.$$

Outlook

- Larger alphabets
- Pathological cases
- Transporter theorem

Thanks for the attention!