Entropy Computation in Partially Observed Markov Chains

F. Desbouvries

INT / CITI Dept. & CNRS UMR 5157

91011 Évry, France

1. Embedded Markovian models : HMC $\subset$ PMC $\subset$ TMC
   - Hidden Markov Chains (HMC)
   - Pairwise Markov Chains (PMC)
   - Triplet Markov Chains (TMC)
   - Examples
2. Efficient algorithms in Partially observed Markov Chains
   - Bayesian restoration (filtering, smoothing...)
   - Parameter estimation
   - Entropy Computation

Hidden Markov Chains (HMC)

$$\left\{ \begin{array}{lcl} \mathbf{x}_{n+1} & = & g_n(\mathbf{x}_n, \mathbf{u}_n) \\ \mathbf{y}_n & = & h_n(\mathbf{x}_n, \mathbf{v}_n) \end{array} \right. , \quad \mathbf{u} \text{ and } \mathbf{v} \text{ indep. and jointly indep.}$$

- $p(\mathbf{x}_{0:n}, \mathbf{y}_{0:n}) = \underbrace{p(\mathbf{x}_0)p(\mathbf{x}_1|\mathbf{x}_0)\cdots p(\mathbf{x}_n|\mathbf{x}_{n-1})}_{p(\mathbf{x}_{0:n})} \times \underbrace{p(\mathbf{y}_0|\mathbf{x}_0)\cdots p(\mathbf{y}_n|\mathbf{x}_n)}_{p(\mathbf{y}_{0:n}|\mathbf{x}_{0:n})}$

- **Filtering :**

$$p(\mathbf{x}_n|\mathbf{y}_{0:n}) = \frac{p(\mathbf{y}_n|\mathbf{x}_n)\int p(\mathbf{x}_n|\mathbf{x}_{n-1})p(\mathbf{x}_{n-1}|\mathbf{y}_{0:n-1})d\mathbf{x}_{n-1}}{p(\mathbf{y}_n|\mathbf{y}_{0:n-1})}$$

  - linear and Gaussian case : Kalman filtering
  - general case : extended Kalman , particle filtering $\cdots$

Extension HMC ⇒ Pairwise Markov Chains (PMC) ?

$$\begin{cases} \mathbf{x}_{n+1} & = & g_n(\mathbf{x}_n, \mathbf{u}_n) \\ \mathbf{y}_n & = & h_n(\mathbf{x}_n, \mathbf{v}_n) \end{cases}, \quad \mathbf{u} \text{ and } \mathbf{v} \text{ indep. and jointly indep.}$$

- $$\begin{cases} p(\mathbf{x}_{n+1}|\mathbf{x}_{0:n}) & = & p(\mathbf{x}_{n+1}|\mathbf{x}_n); \\ p(\mathbf{y}_{0:n}|\mathbf{x}_{0:n}) & = & \prod_{i=0}^{n} p(\mathbf{y}_i|\mathbf{x}_{0:n}); \\ p(\mathbf{y}_i|\mathbf{x}_{0:n}) & = & p(\mathbf{y}_i|\mathbf{x}_i) \ \forall i, \ 0 \leq i \leq n. \end{cases}$$

- moreover the pair $\mathbf{z}_n = \begin{bmatrix} \mathbf{x}_n \\ \mathbf{y}_{n-1} \end{bmatrix}$ is Markovian.

Extension HMC $\Rightarrow$ Pairwise Markov Chains (PMC) ?

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \text{ Markovian} \Rightarrow \begin{cases} p(\mathbf{x}|\mathbf{y}) \text{ and } p(\mathbf{y}|\mathbf{x}) \text{ Markovian} \\ p(\mathbf{x}) \text{ and } p(\mathbf{y}) \text{ not necessarily Markovian} \end{cases}$$

- **Modeling power :** $\begin{cases} \text{PMC} \not\Rightarrow \text{HMC} \\ p(\mathbf{y}|\mathbf{x}) \text{ Markovian} \end{cases}$

- **Bayesian restoration :** $p(\mathbf{x}|\mathbf{y})$ Markovian

Extension PMC $\Rightarrow$ Triplet Markov Chains (TMC) ?

$$\left\{\begin{array}{ll} \mathbf{r} & \text{auxiliary} \\ \mathbf{x} & \text{hidden} \\ \mathbf{y} & \text{observed} \end{array}\right. \quad : \quad \text{the triplet } \underbrace{\begin{bmatrix} \mathbf{r}_n \\ \mathbf{x}_n \\ \mathbf{y}_{n-1} \end{bmatrix}}_{\mathbf{t}_n} \text{ is Markovian}$$

- **Modeling power :** $\left\{ \begin{array}{l} \text{TMC} \not\Rightarrow \text{PMC} \\ \\ \end{array} \right.$

Extension PMC $\Rightarrow$ Triplet Markov Chains (TMC) ?

$$\left\{ \begin{array}{ll} \mathbf{r} & \text{auxiliary} \\ \mathbf{x} & \text{hidden} \\ \mathbf{y} & \text{observed} \end{array} \right. : \quad \text{the triplet } \underbrace{\begin{bmatrix} \mathbf{r}_n \\ \mathbf{x}_n \\ \mathbf{y}_{n-1} \end{bmatrix}}_{\mathbf{t}_n} \text{ is Markovian}$$

- **Modeling power :** $\left\{ \begin{array}{l} \text{TMC} \quad \not\Rightarrow \text{PMC} \\ p(\mathbf{y}|\mathbf{x}) \ \text{ marginal of the MC } \ p(\mathbf{y}, \mathbf{r}|\mathbf{x}) \end{array} \right.$

Extension PMC $\Rightarrow$ Triplet Markov Chains (TMC) ?

$$\left\{\begin{array}{ll}\mathbf{r} & \text{auxiliary} \\ \mathbf{x} & \text{hidden} \\ \mathbf{y} & \text{observed}\end{array}\right. \quad : \quad \text{the triplet } \underbrace{\begin{bmatrix}\mathbf{r}_n \\ \mathbf{x}_n \\ \mathbf{y}_{n-1}\end{bmatrix}}_{\mathbf{t}_n} \text{ is Markovian}$$

- **Modeling power :** $\left\{\begin{array}{l}\text{TMC} \quad \not\Rightarrow \text{PMC} \\ p(\mathbf{y}|\mathbf{x}) \;\; \text{marginal of the MC} \;\; p(\mathbf{y}, \mathbf{r}|\mathbf{x})\end{array}\right.$

- **Bayesian restoration :** $\left\{\begin{array}{l}\text{we restore} \;\; \mathbf{x}_n^* = \begin{bmatrix}\mathbf{x}_n \\ \mathbf{r}_n\end{bmatrix}; \\ \widehat{\mathbf{x}}_{n|n} \;\; \text{is obtained by marginalization .}\end{array}\right.$

1. Embedded Markovian models : HMC $\subset$ PMC $\subset$ TMC

   - Hidden Markov Chains (HMC)

   - Pairwise Markov Chains (PMC)

   - Triplet Markov Chains (TMC)

   - Examples

2. Efficient algorithms in Partially observed Markov Chains

   - Bayesian restoration (filtering, smoothing...)

   - Parameter estimation

   - Entropy Computation

Example 1 : $\mathbf{x}$ discrete, $\mathbf{r}$ discrete

- Hidden semi-Markov Chains (e.g.[Yu and Kobayashi, 2003]) :
  - r_n the time during which x_n remains in the same state
  - q : probability distribution (on $\mathbb{N}^*$) of state duration

$$\begin{cases} p(x_{n+1}|x_n, r_n) & = \begin{cases} \delta_{x_n}(x_{n+1}) & \text{if} \quad r_n > 1 \\ p(x_{n+1}|x_n) & \text{if} \quad r_n = 1 \end{cases} \\ p(r_{n+1}|r_n) & = \begin{cases} \delta_{r_n - 1}(r_{n+1}) & \text{if} \quad r_n > 1 \\ q(r_{n+1}) & \text{if} \quad r_n = 1 \end{cases} \\ p(y_{n+1}|x_{n+1}, r_{n+1}, \mathbf{t}_n) & = p(y_{n+1}|x_{n+1}) \end{cases}$$

- Then

$$p(\underbrace{x_{n+1}, r_{n+1}, y_{n+1}}_{\mathbf{t}_{n+1}}|\mathbf{t}_{1:n}) = p(\mathbf{t}_{n+1}|\mathbf{t}_n) =$$

$$p(r_{n+1}|x_n, r_n, y_n)p(x_{n+1}|r_{n+1}, x_n, r_n, y_n)p(y_{n+1}|x_{n+1}, r_{n+1}, x_n, r_n, y_n)$$

Example 2 : **x** continuous, **r** discrete

- "Switching" or "Jumping" models :

$$\begin{cases} p(r_{n+1}|r_{0:n}) & = & p(r_{n+1}|r_n) \\ \mathbf{x}_{n+1} & = & \mathbf{F}(r_{n+1})\mathbf{x}_n + \mathbf{G}(r_{n+1})\mathbf{u}_n \\ \mathbf{y}_n & = & \mathbf{H}(r_n)\mathbf{x}_n + \mathbf{J}(r_n)\mathbf{v}_n \end{cases}$$

- Then

$$p(\underbrace{\mathbf{x}_{n+1}, r_{n+1}, \mathbf{y}_{n+1}}_{\mathbf{t}_{n+1}}|\mathbf{t}_{0:n}) = p(\mathbf{t}_{n+1}|\mathbf{t}_n) =$$

$$p(r_{n+1}|x_n, r_n, y_n)p(x_{n+1}|r_{n+1}, x_n, r_n, y_n)p(y_{n+1}|x_{n+1}, r_{n+1}, x_n, r_n, y_n)$$

Exemple 3 : $\mathbf{x}$ continuous, $\mathbf{r}$ continuous

- $$\begin{cases} \mathbf{x}_{n+1} & = & \mathbf{F}_n \mathbf{x}_n + \mathbf{G}_n \mathbf{u}_n \\ \mathbf{y}_n & = & \mathbf{H}_n \mathbf{x}_n + \mathbf{J}_n \mathbf{v}_n \end{cases},$$

- $$\underbrace{\begin{bmatrix} \mathbf{u}_{n+1} \\ \mathbf{v}_{n+1} \end{bmatrix}}_{\mathbf{r}_{n+1}} = \underbrace{\begin{bmatrix} \mathbf{A}_n^{\mathbf{u},\mathbf{u}} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_n^{\mathbf{v},\mathbf{v}} \end{bmatrix}}_{\mathbf{A}_n} \begin{bmatrix} \mathbf{u}_n \\ \mathbf{v}_n \end{bmatrix} + \underbrace{\begin{bmatrix} \xi_n^{\mathbf{u}} \\ \xi_n^{\mathbf{v}} \end{bmatrix}}_{\xi_n}$$ [Sorenson 1966]

- $$\begin{bmatrix} \mathbf{x}_{n+1} \\ \mathbf{r}_{n+1} \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{F}_n & \overline{\mathbf{G}}_n & \mathbf{0}_{n_\mathbf{x} \times n_\mathbf{y}} \\ \mathbf{0}_{n_\mathbf{r} \times n_\mathbf{x}} & \mathbf{A}_n & \mathbf{0}_{n_\mathbf{r} \times n_\mathbf{y}} \\ \mathbf{H}_n & \overline{\mathbf{J}}_n & \mathbf{0}_{n_\mathbf{y} \times n_\mathbf{y}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_n \\ \mathbf{r}_n \\ \mathbf{y}_{n-1} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{n_\mathbf{x} \times 1} \\ \xi_n \\ \mathbf{0}_{n_\mathbf{y} \times 1} \end{bmatrix}$$

1. Embedded Markovian models : HMC $\subset$ PMC $\subset$ TMC
   - Hidden Markov Chains (HMC)
   - Pairwise Markov Chains (PMC)
   - Triplet Markov Chains (TMC)
   - Examples
2. Efficient algorithms in Partially observed Markov Chains
   - Bayesian restoration (filtering, smoothing...)
   - Parameter estimation
   - Entropy Computation

Efficient algorithms for Partially Observed Markov Chains

- Bayesian restoration

  filtering : $p(x_n|y_{0:n})$

  - Kalman Filter *(Ait-El-Fquih & Desbouvries, IEEE tr. SP Aug. 2006)*
  - Particle Filtering *(Desbouvries & Pieczynski , NSIP'03)*

  smoothing : $p(x_n|y_{0:N})$

  - Forward Backward, Viterbi *(Pieczynski 2000)*
  - RTS, Two-Filter, Frazer-Potter, Bryson-Frazier ... *(Ait-El-Fquih & Desbouvries, SSP'05, MaxEnt'06)*
  - Particle filtering *(Ait-El-Fquih & Desbouvries , NSSPW'06)*

- Parameter Estimation

  Gaussian case, EM algorithm *(Ait-El-Fquih & Desbouvries, Icassp'06)*

- Entropy Computation

Entropy Computation in Partially Observed Markov Chains

$$H(X_{0:N}|y_{0:N}) = -\sum_{x_{0:N}} p(x_{0:N}|y_{0:N}) \log p(x_{0:N}|y_{0:N}).$$

- Brute force : $O(K^N)$ operations
- HMC : *Hernando, Crespi & Cybenko, IEEE Tr. IT July 2005.* $O(K^2 N)$

Entropy Computation in Partially Observed Markov Chains

$$H(X_{0:N}|y_{0:N}) = -\sum_{x_{0:N}} p(x_{0:N}|y_{0:N}) \log p(x_{0:N}|y_{0:N}).$$

- Brute force : $O(K^N)$ operations
- HMC : *Hernando, Crespi & Cybenko, IEEE Tr. IT July 2005.* $O(K^2 N)$
- PMC :

$$H(X_{0:n}|y_{0:n}) = H(X_n|y_{0:n}) + \sum_{x_n} \underbrace{H(X_{0:n-1}|x_n, y_{0:n})} p(x_n|y_{0:n}).$$

$$\overbrace{H(X_{0:n-1}|x_n, y_{0:n})} = H(X_{n-1}|x_n, y_{0:n}) + \sum_{x_{n-1}} H(X_{0:n-2}|x_{n-1}, y_{0:n-1}) p(x_{n-1}|x_n, y_{0:n}).$$

Entropy Computation in Partially Observed Markov Chains

$$H(X_{0:N}|y_{0:N}) = -\sum_{x_{0:N}} p(x_{0:N}|y_{0:N}) \log p(x_{0:N}|y_{0:N}).$$

- Brute force : $O(K^N)$ operations
- HMC : *Hernando, Crespi & Cybenko, IEEE Tr. IT July 2005.* $O(K^2N)$
- PMC : recursive alg. + FB $(p(x_n|y_{0:n}), p(x_{n-1}|x_n, y_{0:n}.))$ $O(K^2N)$

$$H(X_{0:n}|y_{0:n}) \quad = \quad H(X_n|y_{0:n}) + \sum_{x_n} \underbrace{H(X_{0:n-1}|x_n, y_{0:n})} p(x_n|y_{0:n}).$$

$$\overbrace{H(X_{0:n-1}|x_n, y_{0:n})} = H(X_{n-1}|x_n, y_{0:n}) + \sum_{x_{n-1}} H(X_{0:n-2}|x_{n-1}, y_{0:n-1}) p(x_{n-1}|x_n, y_{0:n}).$$

Entropy Computation in Partially Observed Markov Chains

$$H(X_{0:N}|y_{0:N}) = -\sum_{x_{0:N}} p(x_{0:N}|y_{0:N}) \log p(x_{0:N}|y_{0:N}).$$

- Brute force : $O(K^N)$ operations
- HMC : *Hernando, Crespi & Cybenko, IEEE Tr. IT July 2005.* $O(K^2N)$
- PMC : recursive alg. + FB $(p(x_n|y_{0:n}), p(x_{n-1}|x_n, y_{0:n}.))$ $O(K^2N)$

$$H(X_{0:n}|y_{0:n}) = H(X_n|y_{0:n}) + \sum_{x_n} \underbrace{H(X_{0:n-1}|x_n, y_{0:n})} p(x_n|y_{0:n}).$$

$$\overbrace{H(X_{0:n-1}|x_n, y_{0:n})} = H(X_{n-1}|x_n, y_{0:n}) + \sum_{x_{n-1}} H(X_{0:n-2}|x_{n-1}, y_{0:n-1}) p(x_{n-1}|x_n, y_{0:n}).$$

- TMC : $H(X_{0:N}, R_{0:N}|y_{0:N})$ OK ; $H(X_{0:N}|y_{0:N})$ not available.