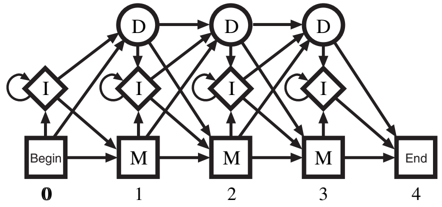
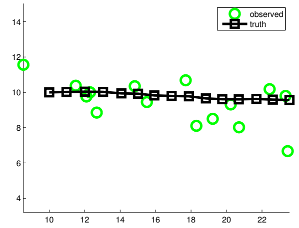
Royal Institute of Technology

STAT. METH. IN CS – PF, SMC

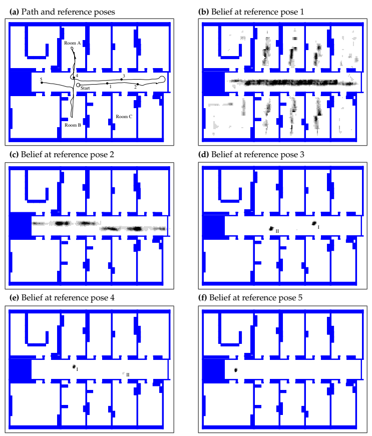
Lecture 10





CONTINUOUS STATES

PARTICLES FILTER - ROBOT POSITION FROM OBSERVATIONS



COMBINING INFO FROM VARIOUS SENSORS

- Inertial sensors
- Camera
- Barometer

- Inertial sensors
- Radar
- Barometer
- Map

- Inertial sensors
- Cameras
- Radars
- Wheel speed sensors
- Steering wheel sensor

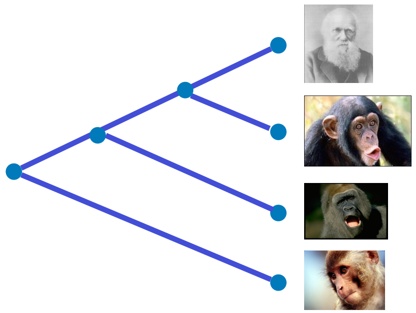
- Inertial sensors
- Ultra-wideband

Aim: Motion capture, find the motion (position, orientation, velocity and acceleration) of a person (or object) over time.

PHYLOGENY

Input: species

Output: tree where proximity correlates with similarity

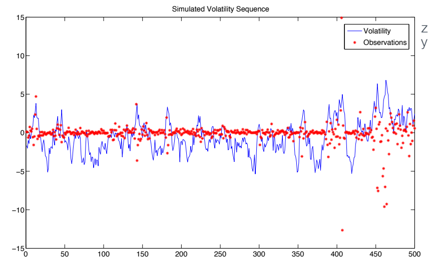


STOCHASTIC VOLATILITY

$$Z_1 \sim \mathcal{N}(0, \frac{\sigma^2}{1-\sigma^2})$$

$$Z_n = \alpha Z_{n-1} + \sigma V_n \quad \text{where} \quad V_n \sim \mathcal{N}(0,1)$$

$$Y_n = \beta e^{Z_n/2} W_n \qquad W_n \sim \mathcal{N}(0,1)$$



PARTICLE FILTERING & SEQUENTIAL MONTE CARLO

- Probabilistic (Monte Carlo) recursive inference
- SSM (or HMM)
- Applications
  - tracking
  - time series forecasting
  - on-line parameter learning

- Idea: approximate $p(z_{1:t}|y_{1:t})$ with

$$\sum_{s=1}^{S} \hat{w}_t^s I\left(z_{1:t} = z_{1:t}^s\right)$$

where $\hat{w}^s_t$ is normalized weight of sample s at time t

- New belief state obtained by importance sampling

NOTATION

Generic form

hidden state

$$z_t := g(u_t, z_{t-1}, \epsilon_t)$$

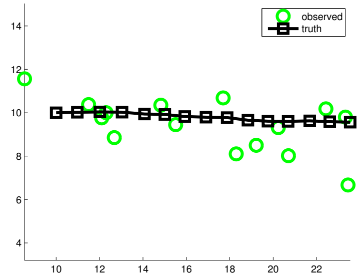
Denoted $g(z_t|z_{t-1})$

observation

$$y_t := h(z_t, u_t, \delta_t)$$

Denoted $h(y_t|z_t)$

Estimate belief state (filtering) $p(z_t|y_{1:t})$



NOTATION

Generic form

hidden state

$$\boldsymbol{z}_t := g(\boldsymbol{u}_t, \boldsymbol{z}_{t-1}, \boldsymbol{\epsilon}_t)$$

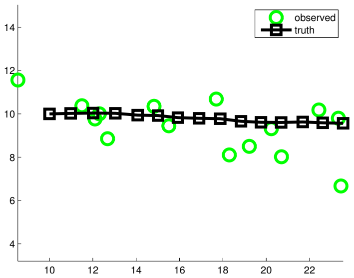
Denoted $g(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}) = p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1})$

observation

$$\boldsymbol{y}_t := h(\boldsymbol{z}_t, \boldsymbol{u}_t, \boldsymbol{\delta}_t)$$

Denoted $h(\boldsymbol{y}_t|\boldsymbol{z}_t) = p(\boldsymbol{y}_t|\boldsymbol{z}_t)$

Estimate belief state (filtering) $p(\boldsymbol{z}_t|\boldsymbol{y}_{1:t})$



COMPARISON SSM VS HMM

$$g(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}) = p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1})$$
$$h(\boldsymbol{y}_t|\boldsymbol{z}_t) = p(\boldsymbol{y}_t|\boldsymbol{z}_t)$$

- For HMMs DP is possible since ≤ constant #of values (states) for z_n
- Here typically infinite (R)
- Two options
  - Solve analytically (or possibly numerically)
  - Use a Particle Filter (PF) algorithm or generally, Sequential Monte Carlo (SMC)

SMC

- We want densities $\{\pi_n\}_{n\in[N]}$
- Typically $\pi_n(Z_{1:n}) = \gamma_n(Z_{1:n})/C_n$
  - where γ_n is a likelihood which we can evaluate pointwise
  - C_n is a normalizing constant
- So for a SSM

$$\pi_n(Z_{1:n}) = p(Z_{1:n}|Y_{1:n}) = p(Z_{1:n}, Y_{1:n})/p(Y_{1:n})$$

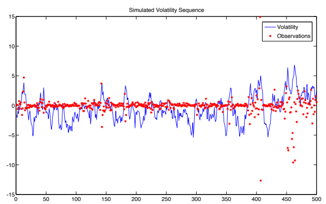
Y will be “data” so constant

pointwise computable

TWO PROBLEMS

- Problem 1:

  $\pi_n(Z_{1:n})$ is typically complex, high dimensional, and hard to sample from



- Problem 2:

  if we can sample, sampling $Z_{1:n}$ typically takes $\Omega(n)$ time

IMPORTANCE SAMPLING

- Assume we have $q_n(Z_{1:n})$ such that $\pi_n(Z_{1:n}) > 0 \Rightarrow q_n(Z_{1:n}) > 0$
- Let $w_n(Z_{1:n}) := \gamma_n(Z_{1:n})/q_n(Z_{1:n})$

Our proposal, we can sample from and compute it.

IMPORTANCE SAMPLING

- Assume we have $q_n(Z_{1:n})$ such that $\pi_n(Z_{1:n}) > 0 \Rightarrow q_n(Z_{1:n}) > 0$
- Let $w_n(Z_{1:n}) := \gamma_n(Z_{1:n})/q_n(Z_{1:n})$

Computable pointwise

IMPORTANCE SAMPLING

- Assume we have $q_n(Z_{1:n})$ such that $\pi_n(Z_{1:n}) > 0 \Rightarrow q_n(Z_{1:n}) > 0$
- Let $w_n(Z_{1:n}) := \gamma_n(Z_{1:n})/q_n(Z_{1:n})$
- Sample

$$Z_{1:n} \sim q_n(Z_{1:n})$$

- let

$$W_n^s := w_n(Z_{1:n})/\sum_s w_n(Z_{1:n})$$

- we get estimates

$$\hat{\pi}_n(Z_{1:n}) := \sum_s W_n^s \delta_{Z_{1:n}^s}(Z_{1:n})$$

$$\hat{C}_n := \frac{1}{S}\sum_s w_n(Z_{1:n}^s)$$

NOTICE

$$E[\hat{C}_n] = E_{q_n}[\sum_s w_n(Z_{1:n}^s)/S] = \frac{1}{S}\sum_s E[w_n(Z_{1:n}^s)]$$

$$= \frac{1}{S}\sum_s E\left[\frac{\gamma_n(Z_{1:n}^s)}{q_n(Z_{1:n}^s)}\right] = \frac{1}{S}SC_n = C_n$$

since

$$E_{q_n}\left[\frac{\gamma_n(Z_{1:n}^s)}{q_n(Z_{1:n}^s)}\right] = \int q_n(Z_{1:n})\frac{\gamma_n(Z_{1:n})}{q_n(Z_{1:n})}\,dZ_{1:n}$$

$$= \int \gamma_n(Z_{1:n})\,dZ_{1:n} = C_n$$

recall

$$1 = \int \pi_n(Z_{1:n})\,dZ_{1:n} = \int \frac{\gamma_n(Z_{1:n})}{C_n}\,dZ_{1:n} = \frac{1}{C_n}\int \gamma_n(Z_{1:n})\,dZ_{1:n}$$

BIASED ESTIMATE

The following estimate is biased:

$$I_n^{\mathrm{IS}}(\varphi_n) = \sum_{s \in [S]} W_n^s \varphi(Z_{1:n}^s)$$

Why?

PROPOSAL DESIGN

- Base proposal on previous approximation

=3

$$q_n(Z_{1:n}) := q'_n(\hat{\pi}_{n-1}(Z_{1:n-1}), Z_n)$$

A DECOMPOSABLE PROPOSAL MITIGATES PROBLEM 2

A decomposable proposal satisfies

$$q_n(Z_{1:n}) = q_{n-1}(Z_{1:n-1}) q_n(Z_n | Z_{1:n-1})$$

which gives

$$q_n(Z_{1:n}) = q_1(Z_1) \prod_{k \in [n]} q_n(Z_k | Z_{1:k-1})$$

So to get $Z_{1:n}^s \sim q_n(Z_{1:n})$

we can sample

Will be sampled from approximation of π_{k-1}

$$Z_1^s \sim q_1(Z_1)$$

$$Z_{1:k}^s \sim q_n(Z_k | Z_{1:k-1}), \quad \forall\ 2 \le k \le n$$

OBTAINING UNNORMALIZED WEIGHTS

=3

$$w_n(Z_{1:n}) = \frac{\gamma_n(Z_{1:n})}{q_n(Z_{1:n})} = \frac{\gamma_{n-1}(Z_{1:n-1})}{q_{n-1}(Z_{1:n-1})} \frac{\gamma_n(Z_{1:n})}{\gamma_{n-1}(Z_{1:n-1}) q_n(Z_n | Z_{1:n-1})}$$

$$= w_{n-1}(Z_{1:n-1}) \alpha_n(Z_{1:n}) = w_1(Z_1) \prod_{k=2}^{n} \alpha_k(Z_{1:k})$$

where

$$\alpha_n(Z_{1:n}) := \frac{\gamma_n(Z_{1:n})}{\gamma_{n-1}(Z_{1:n-1}) q_n(Z_n | Z_{1:n-1})}$$

SEQUENTIAL IMPORTANCE SAMPLING

- n =1
  - Sample $Z_1^s \sim q_1(Z_1)$
  - Compute weights $w_1(Z_1^s)$ and $W_1(Z_1^s)$

SEQUENTIAL IMPORTANCE SAMPLING

Here and later s means for all s in [S]

- n =1
  - Sample $Z_1^s \sim q_1(Z_1)$
  - Compute weights $w_1(Z_1^s)$ and $W_1(Z_1^s)$

SEQUENTIAL IMPORTANCE SAMPLING

- n =1
  - Sample $Z_1^s \sim q_1(Z_1)$
  - Compute weights $w_1(Z_1^s)$ and $W_1(Z_1^s)$
- n ≥ 2
  - Sample $Z_{1:n}^s \sim q_n(z_n|Z_{1:n-1})$
  - Compute weight

$$w_n(Z_{1:n}^s) = w_{n-1}(Z_{1:n-1}^s)\alpha(Z_{1:n}^s)$$

$$W_n(Z_{1:n}^s) \propto w_n(Z_{1:n}^s)$$

SEQUENTIAL IMPORTANCE SAMPLING

What's the best proposal?

- n =1
  - Sample $Z_1^s \sim q_1(Z_1)$
  - Compute weights $w_1(Z_1^s)$ and $W_1(Z_1^s)$
- n ≥ 2
  - Sample $Z_{1:n}^s \sim q_n(z_n|Z_{1:n-1})$
  - Compute weight

$$w_n(Z_{1:n}^s) = w_{n-1}(Z_{1:n-1}^s)\alpha(Z_{1:n}^s)$$

$$W_n(Z_{1:n}^s) \propto w_n(Z_{1:n}^s)$$

SEQUENTIAL IMPORTANCE SAMPLING

- n =1
  - Sample $Z_1^s \sim q_1(Z_1)$
  - Compute weights $w_1(Z_1^s)$ and $W_1(Z_1^s)$
- n ≥ 2
  - Sample $Z_{1:n}^s \sim q_n(z_n|Z_{1:n-1})$
  - Compute weight

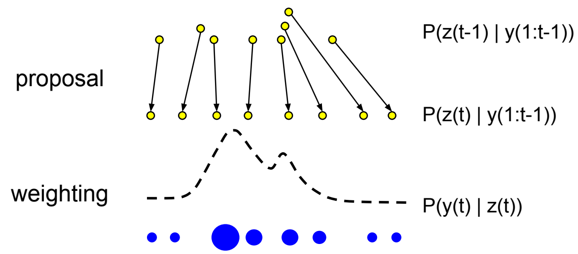
$$w_n(Z_{1:n}^s) = w_{n-1}(Z_{1:n-1}^s)\alpha(Z_{1:n}^s)$$

$$W_n(Z_{1:n}^s) \propto w_n(Z_{1:n}^s)$$

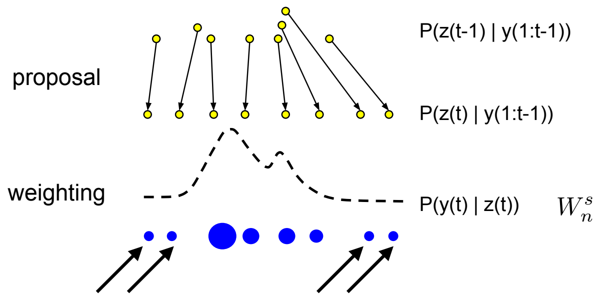
What's the best proposal?

$$q_n^{\text{opt}}(Z_{1:n}|Z_{1:n-1}) = \pi_n(Z_{1:n}|Z_{1:n-1})$$

PF FIGURE



PF FIGURE



Particles with small weights seem likely to never improve

RESAMPLING

- Intuitively appealing, practical, theoretically beneficial (lower variance)
- In the n:th step ample S new particles

$$P(s) = W_n^s$$

- So $Z_{1:n}^s$ gets N_n^s offsprings and

$$N_n^{1:S} = (N_n^1, \ldots, N_n^S)$$

follows a multinomial with parameters $(N, W_n^{1:S})$

- New unbiased approximation

$$\overline{\pi}(Z_{1:n}) := \sum_s \frac{N_n^s}{S} \delta_{Z_{1:n}}^s(Z_{1:n})$$

SYSTEMATIC RESAMPLING

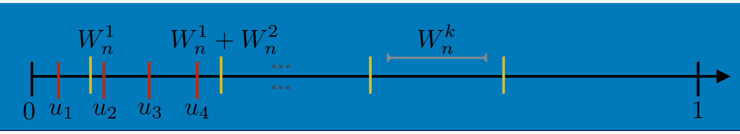
- Systematic resampling aims at lower the variance of N_n^s
- Procedure
  - sample $u_1 \sim U[0, 1/S]$

↑ Uniform distribution

SYSTEMATIC RESAMPLING

- Systematic resampling aims at lower the variance of N_n^s
- Procedure
  - sample $u_1 \sim U[0, 1/S]$
  - for s ≥ 2, $u_s := u_1 + (s-1)/S$
  - for all s,

$$N_n^s := |\{\, u_t \,:\, \sum_{k=1}^{s-1} W_n^k \le u_t \le \sum_{k=1}^{s} W_n^k \,\}|$$



PROPERTIES OF SYSTEMATIC RESAMPLING

- Systematic resampling
  - is one of of several resampling strategies
  - is unbiased
  - easy to implement
  - outperforms other alternatives in most cases
- Consequently, widely used in practice

ESTIMATE OF RATIO BETWEEN NORMALIZING CONSTANTS

- C_n/C_{n-1} can be estimated (consistently) by

$$\widehat{C_n/C_{n-1}} := \sum_{s \in S} W_{n-1}^s \alpha(Z_{1:n}^s)$$

- Motivation

$$\begin{aligned}
&\int \alpha_n(Z_{1:n})\pi_{n-1}(Z_{1:n-1})q_n(Z_n|Z_{1:n-1})\, d\, Z_{1:n} \\
&\quad = \int \frac{\gamma_n(Z_{1:n})\pi_{n-1}(Z_{1:n-1})q_n(Z_n|Z_{1:n-1})}{\gamma_{n-1}(Z_{1:n-1})q_n(Z_n|Z_{1:n-1})}\, d\, Z_{1:n} \\
&\quad = \frac{1}{C_{n-1}} \int \gamma_n(Z_{1:n})\, d\, Z_{1:n} \\
&\quad = C_n/C_{n-1}
\end{aligned}$$

$$\alpha_n(Z_{1:n}) := \frac{\gamma_n(Z_{1:n})}{\gamma_{n-1}(Z_{1:n-1})q_n(Z_n|Z_{1:n-1})}$$

SEQUENTIAL IMPORTANCE SAMPLING

- n =1…
- n ≥ 2
  - Sample $Z^s_{1:n} \sim q_n(z_n | Z_{1:n-1})$
  - Compute weight
    $w_n(Z^s_{1:n}) = w_{n-1}(Z^s_{1:n-1})\alpha(Z^s_{1:n})$ and $W_n(Z^s_{1:n}) \propto w_n(Z^s_{1:n})$
  - Resampling gives $\overline{Z}^1_{1:n}, \ldots, \overline{Z}^S_{1:n}$

SEQUENTIAL IMPORTANCE SAMPLING

- n =1…
- n ≥ 2
  - Sample $Z^s_{1:n} \sim q_n(z_n | Z_{1:n-1})$
  - Compute weight
    $w_n(Z^s_{1:n}) = w_{n-1}(Z^s_{1:n-1})\alpha(Z^s_{1:n})$ and $W_n(Z^s_{1:n}) \propto w_n(Z^s_{1:n})$
  - Resampling gives $\overline{Z}^1_{1:n}, \ldots, \overline{Z}^S_{1:n}$

Two estimates

$$\hat{\pi}_n(Z_{1:n}) = \sum_{s=1}^{S} W^s_n \delta_{Z^s_{1:n}}(Z_{1:n}) \text{ and } \overline{\pi}_n(Z_{1:n}) = \frac{1}{S}\sum_{s=1}^{S} \delta_{\overline{Z}^s_{1:n}}(Z_{1:n})$$

RESAMPLING IN PRACTICE

- Resampling introduce variance and weight may be approx. same
- Solution
  - resample only effective sample size
    $$\text{ESS} := \frac{1}{\sum_{s=1}^{S}(W^s_n)^2}$$
  - is below a threshold, typically,
    $$1/S$$

SIS EX.

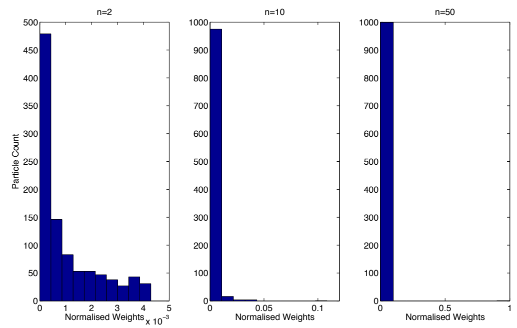


Figure 3: Empirical distributions of the particle weights obtained with the SIS algorithm for the stochastic volatility model at iterations 2, 10 and 50. Although the algorithm is reasonably initialised, by iteration 10 only a few tens of particles have significant weight and by iteration 50 a single particle is dominant.

SIS EX.

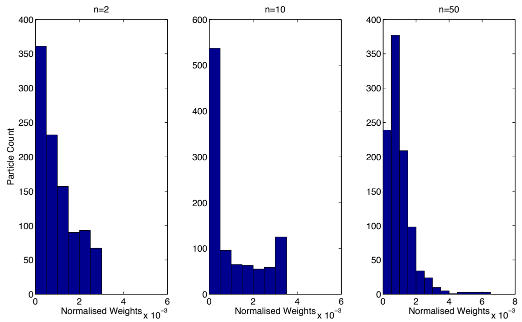
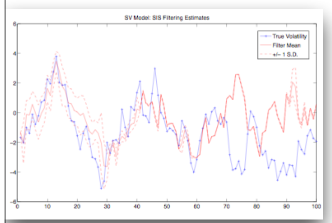
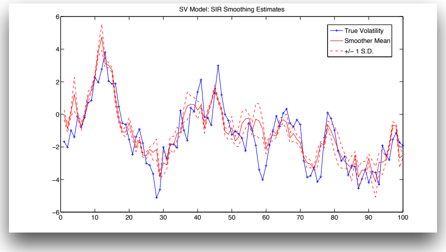


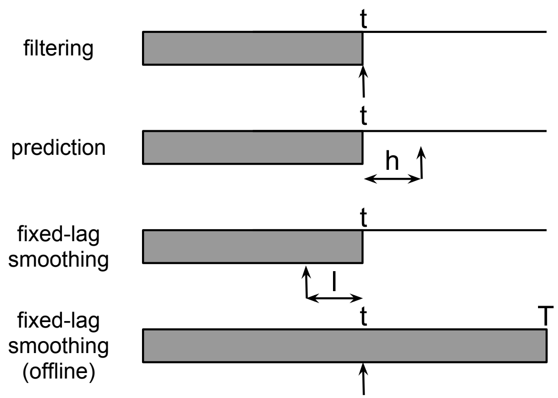
Figure 4: Empirical distribution of particle weights for an SIR algorithm applied to the stochastic volatility model. Notice that there is no evidence of weight degeneracy in contrast to the SIS case. This comes at the cost of reducing the quality of the path-space representation.

WITH AND WITHOUT RESAMPLING





OTHER ANALYSES ALSO POSSIBLE



A Tutorial on Particle Filtering and Smoothing:
Fifteen years later

Arnaud Doucet
The Institute of Statistical Mathematics,
4-6-7 Minami-Azabu, Minato-ku,
Tokyo 106-8569, Japan.
Email: Arnaud@ism.ac.jp

Adam M. Johansen
Department of Statistics,
University of Warwick,
Coventry, CV4 7AL, UK
Email: A.M.Johansen@warwick.ac.uk

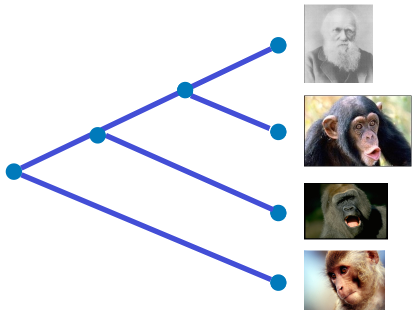
Read chapter 1-3 (not filtering and smoothing)

READ

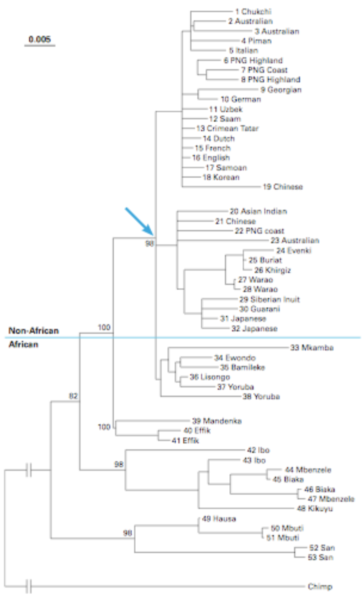
PHYLOGENY

Input: species

Output: tree where proximity correlates with similarity

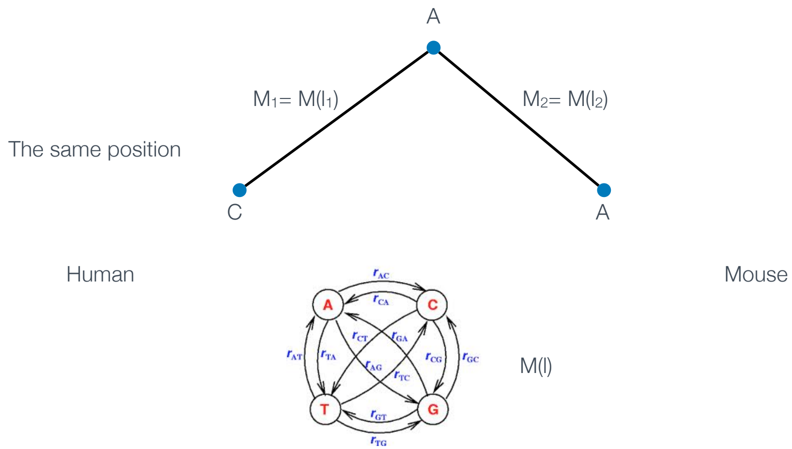


IS THE CHIMP OUR CLOSEST RELATIVE?



OUR ORIGIN

MARKOV MODEL OF SEQUENCE EVOLUTION



MARKOV MODEL OF SEQUENCE EVOLUTION

A C A C G T A C T G C G

$M_1 = M(l_1)$ $M_2 = M(l_2)$

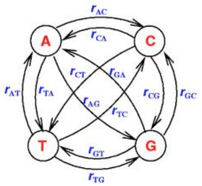
Same gene, same positions

A C A C G T A G T G C C

A A A C G T A C T G C A

Human

Mouse



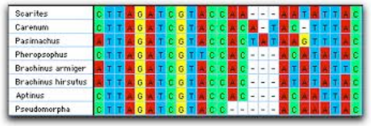
M(l)

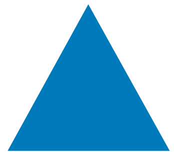
MARKOV MODEL OF SEQUENCE EVOLUTION

Human	A C A C G T A G T G C C
Mouse	A A A C G T A C T G C A

Uniform

In general,





A A C ....T A

THE END