EECS 391
Intro to AI

Probabilistic Reasoning over Time

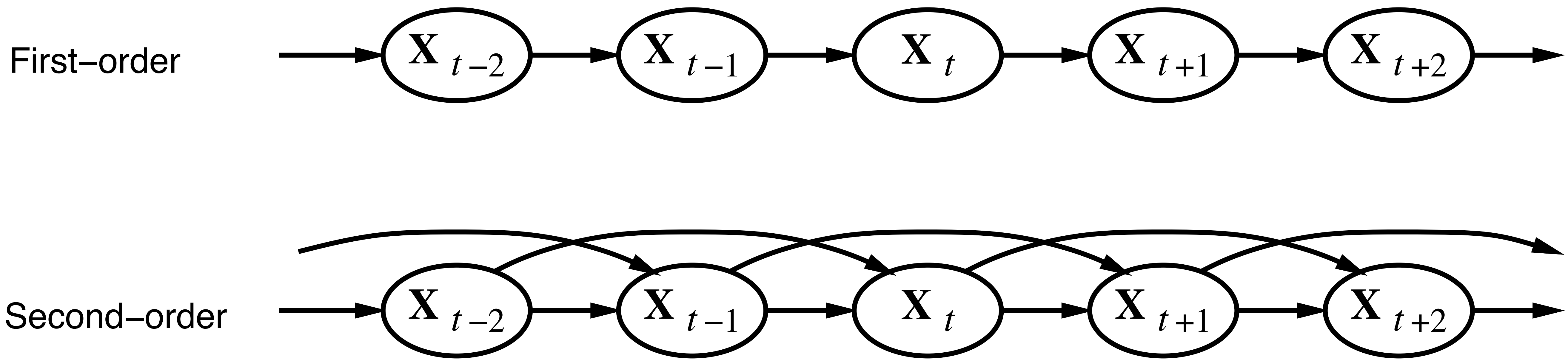
L23 Tue Dec 5

Markov processes (Markov chains)

Construct a Bayes net from these variables: parents?

Markov assumption: $\mathbf{X}_t$ depends on **bounded** subset of $\mathbf{X}_{0:t-1}$

First-order Markov process: $\mathbf{P}(\mathbf{X}_t|\mathbf{X}_{0:t-1}) = \mathbf{P}(\mathbf{X}_t|\mathbf{X}_{t-1})$
Second-order Markov process: $\mathbf{P}(\mathbf{X}_t|\mathbf{X}_{0:t-1}) = \mathbf{P}(\mathbf{X}_t|\mathbf{X}_{t-2}, \mathbf{X}_{t-1})$



Sensor Markov assumption: $\mathbf{P}(\mathbf{E}_t|\mathbf{X}_{0:t}, \mathbf{E}_{0:t-1}) = \mathbf{P}(\mathbf{E}_t|\mathbf{X}_t)$

Stationary process: transition model $\mathbf{P}(\mathbf{X}_t|\mathbf{X}_{t-1})$ and
sensor model $\mathbf{P}(\mathbf{E}_t|\mathbf{X}_t)$ fixed for all t

Inference tasks

Filtering: $\mathbf{P}(\mathbf{X}_t|\mathbf{e}_{1:t})$
belief state—input to the decision process of a rational agent

Prediction: $\mathbf{P}(\mathbf{X}_{t+k}|\mathbf{e}_{1:t})$ for $k > 0$
evaluation of possible action sequences;
like filtering without the evidence

Smoothing: $\mathbf{P}(\mathbf{X}_k|\mathbf{e}_{1:t})$ for $0 \leq k < t$
better estimate of past states, essential for learning

Most likely explanation: $\arg\max_{\mathbf{x}_{1:t}} P(\mathbf{x}_{1:t}|\mathbf{e}_{1:t})$
speech recognition, decoding with a noisy channel

HMMs in Barber's notation

The HMM defines a Markov chain on hidden (or 'latent') variables $h_{1:T}$. The observed (or 'visible') variables are dependent on the hidden variables through an emission $p(v_t|h_t)$. This defines a joint distribution

$$p(h_{1:T}, v_{1:T}) = p(v_1|h_1)p(h_1)\prod_{t=2}^{T} p(v_t|h_t)p(h_t|h_{t-1})$$

For a stationary HMM the transition $p(h_t|h_{t-1})$ and emission $p(v_t|h_t)$ distributions are constant through time.

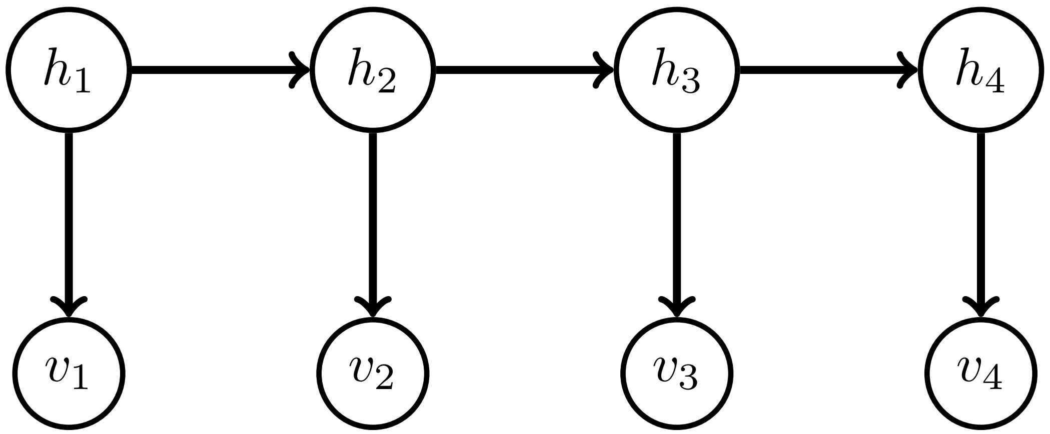


Figure : A first order hidden Markov model with 'hidden' variables $\mathrm{dom}(h_t) = \{1, \ldots, H\}$, $t = 1 : T$. The 'visible' variables v_t can be either discrete or continuous.

Barber's notation

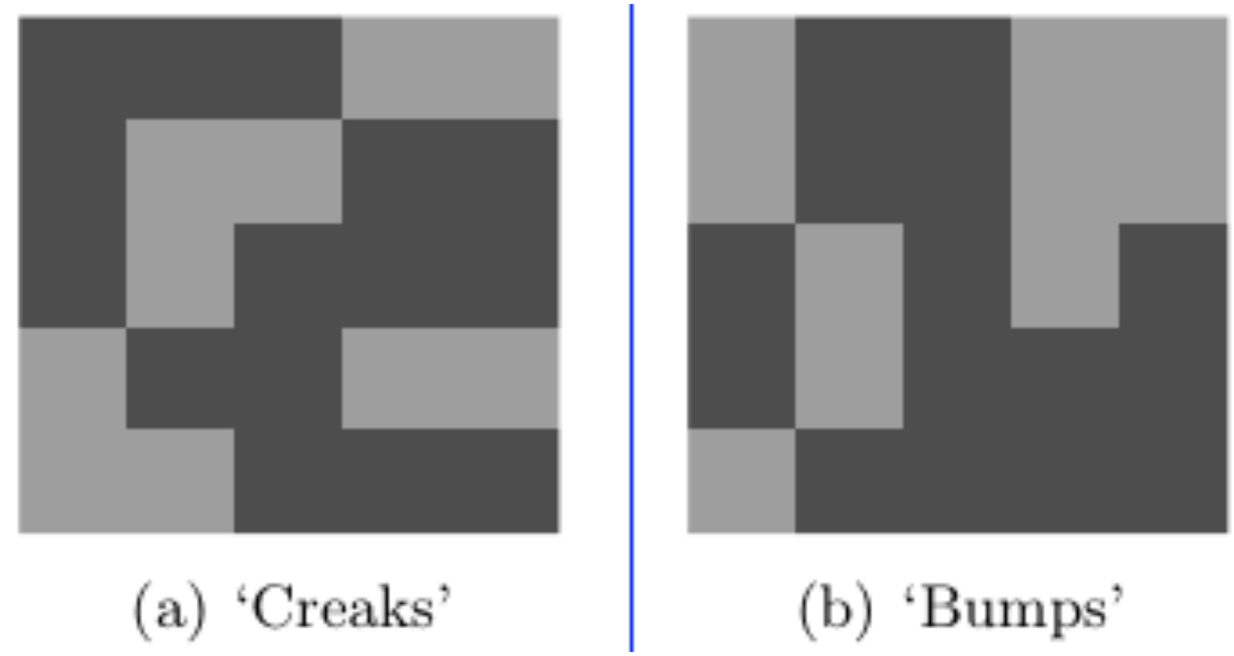
23.2.1 The classical inference problems

The common inference problems in HMMs are summarised below:

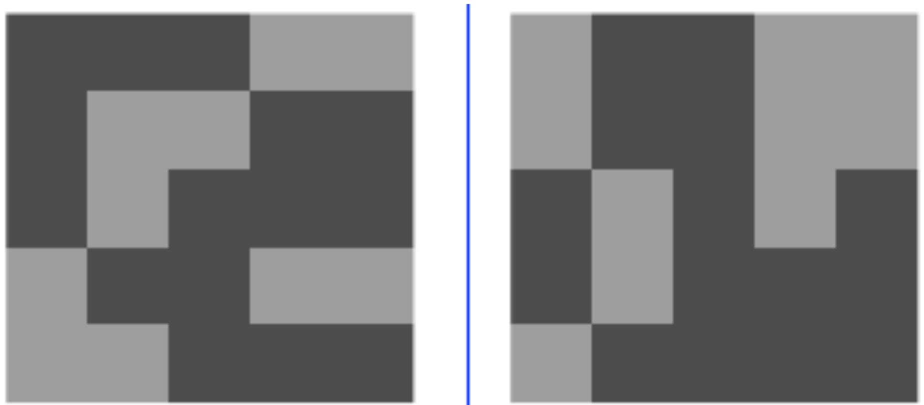
Filtering	(Inferring the present)	$p(h_t\|v_{1:t})$	
Prediction	(Inferring the future)	$p(h_t\|v_{1:s})$	$t > s$
Smoothing	(Inferring the past)	$p(h_t\|v_{1:u})$	$t < u$
Likelihood		$p(v_{1:T})$	
Most likely Hidden path	(Viterbi alignment)	$\underset{h_{1:T}}{\operatorname{argmax}}\, p(h_{1:T}\|v_{1:T})$	

Barber’s Creaks and Bumps example

- You’re in bed and hear a burglar downstairs.
- Want to infer the location of a burglar using knowledge of house
- Use a 5x5 grid of floor below
  - (a) prob. floor creaking at each location
  - (b) prob. that burglar will bump into something at each location
  - light square is prob. 0.9 dark 0.1
- Don’t know where there was a creak or a bump

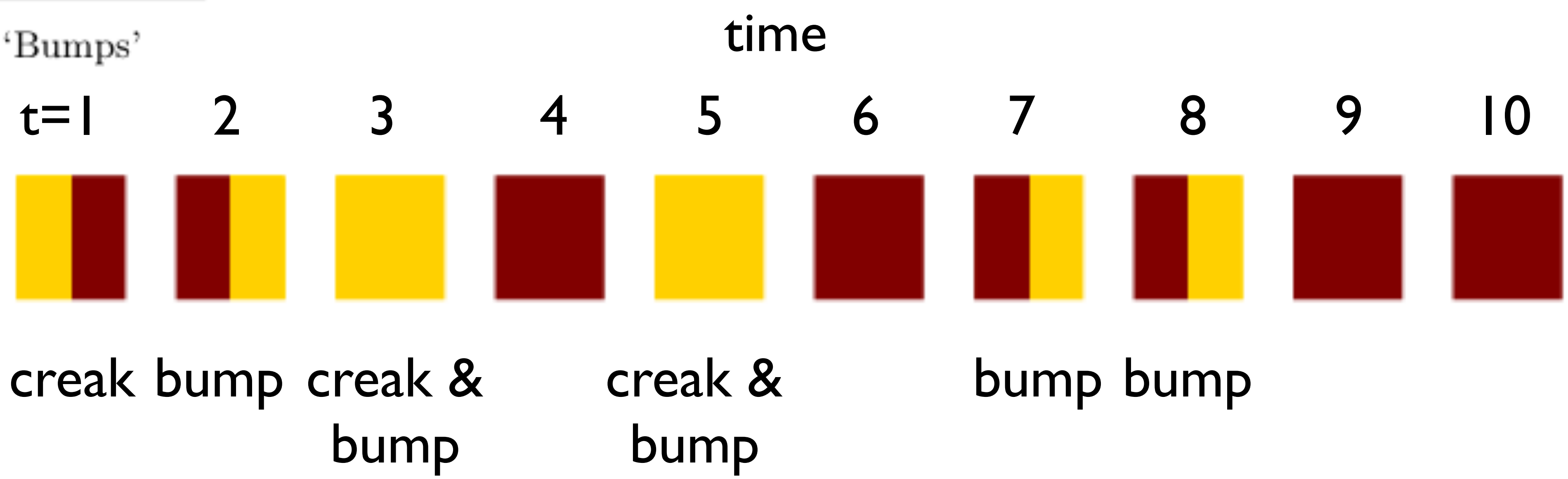


(a) ‘Creaks’ (b) ‘Bumps’

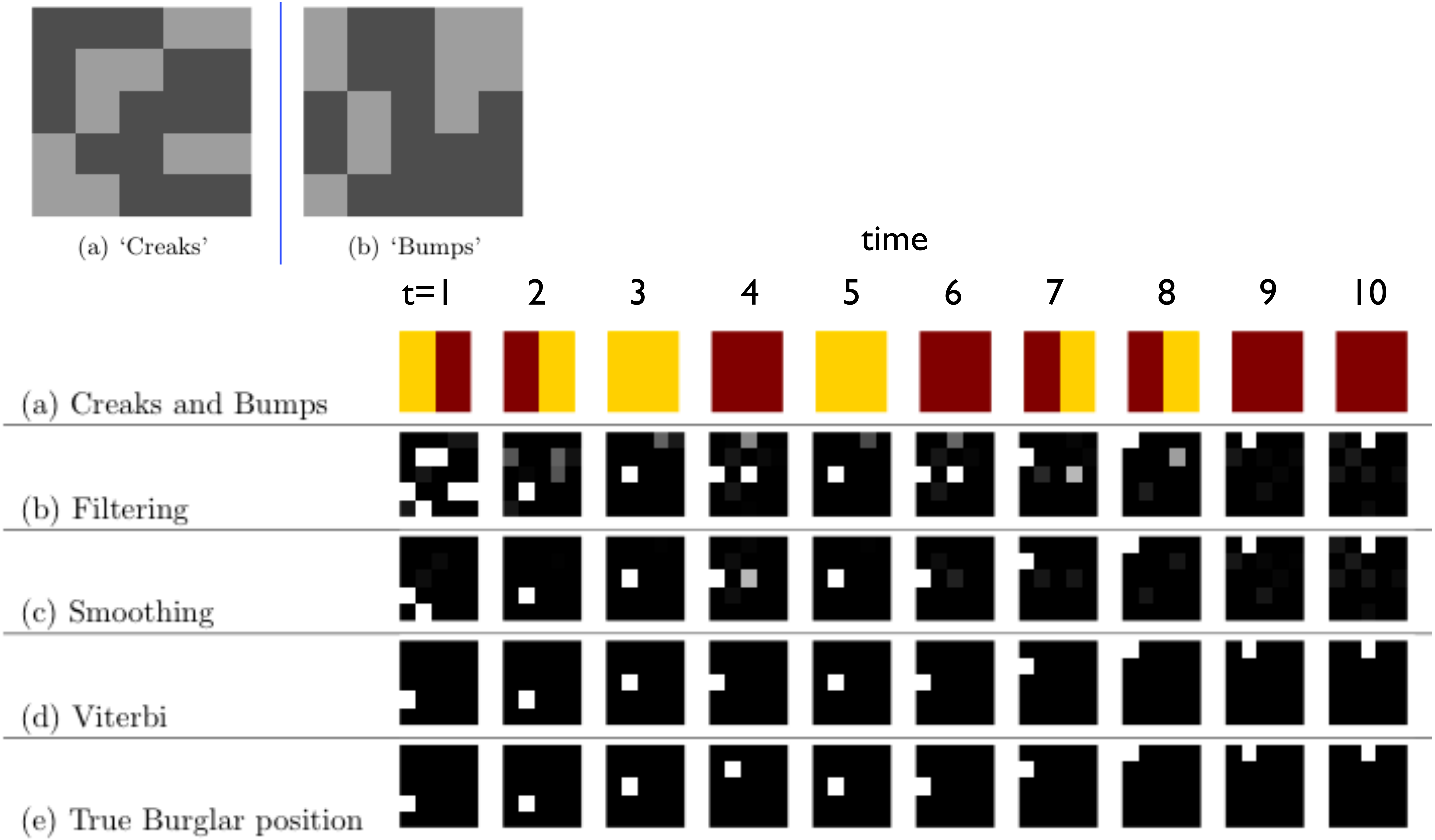


(a) 'Creaks' (b) 'Bumps'

(a) Creaks and Bumps

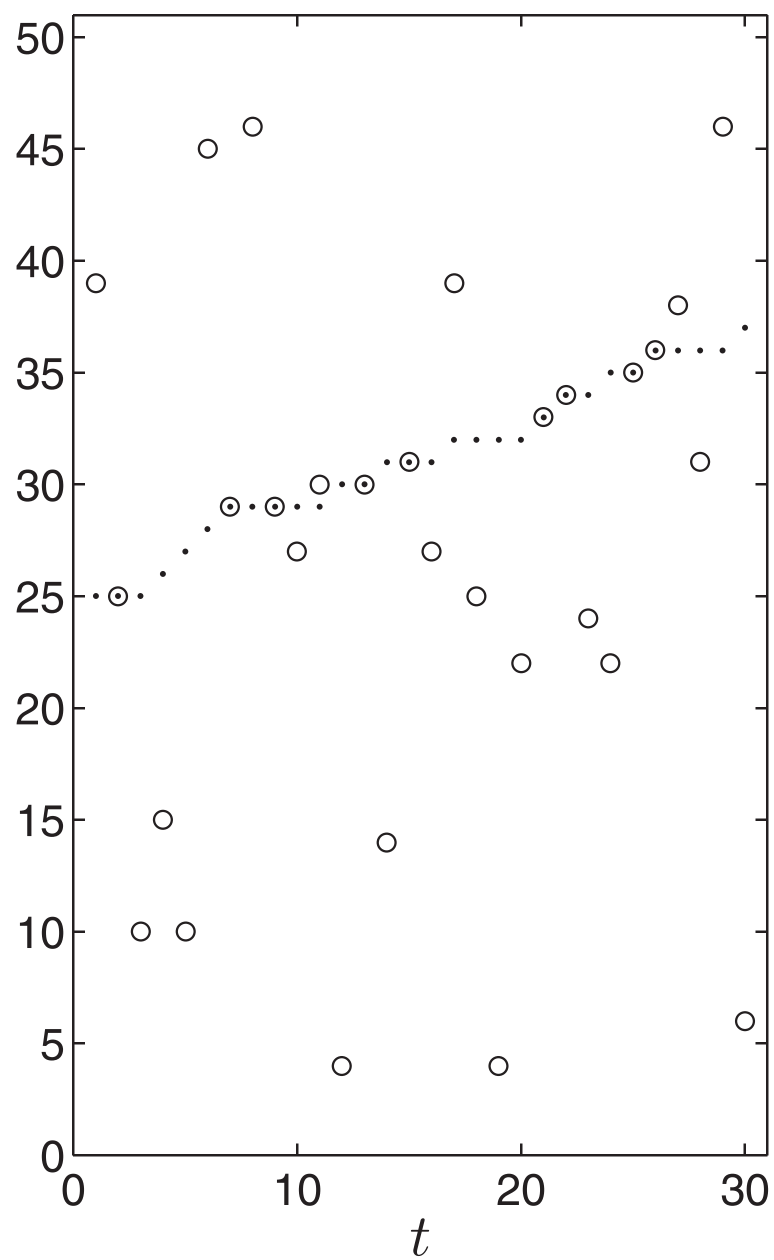


- each panel represents the visible information at each time step
  - don't know location just creak or bump at each time step
  - left half creak, right half: bump
  - lighter shade: presence of a creak or bump, darker shade: absence
- Where is the burglar?



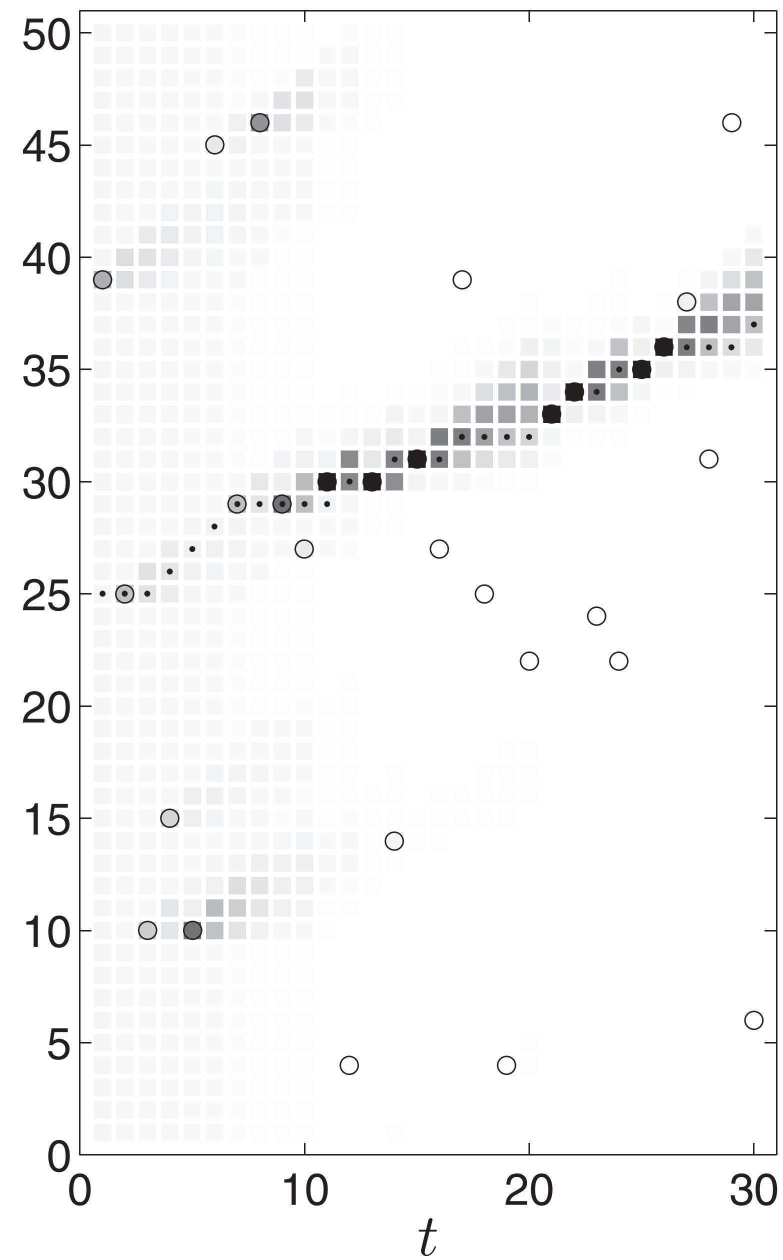
Filtering	(Inferring the present)	$p(h_t\|v_{1:t})$	
Prediction	(Inferring the future)	$p(h_t\|v_{1:s})$	$t > s$
Smoothing	(Inferring the past)	$p(h_t\|v_{1:u})$	$t < u$
Likelihood		$p(v_{1:T})$	
Most likely Hidden path	(Viterbi alignment)	$\underset{h_{1:T}}{\operatorname{argmax}}\, p(h_{1:T}\|v_{1:T})$	

Robot localization



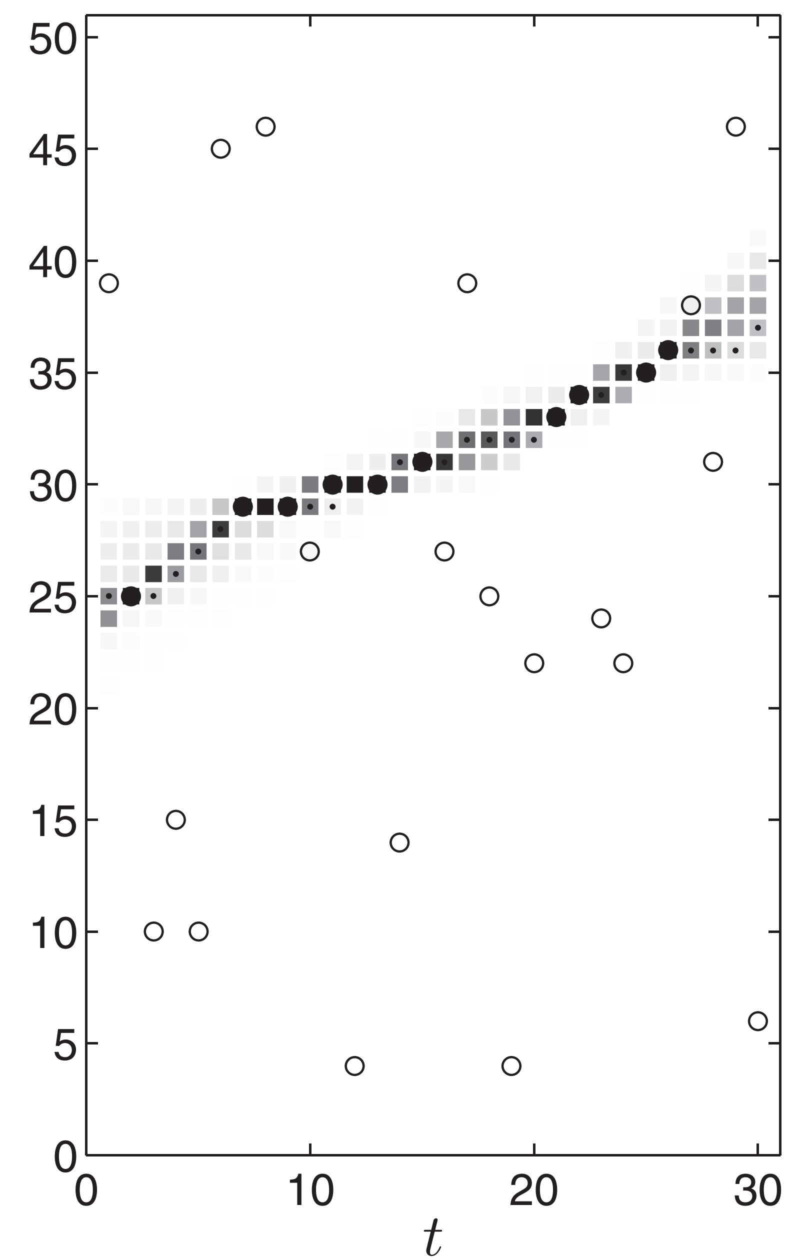
(a)

dots: true location
circles: measured locations
(noisy)



(b)

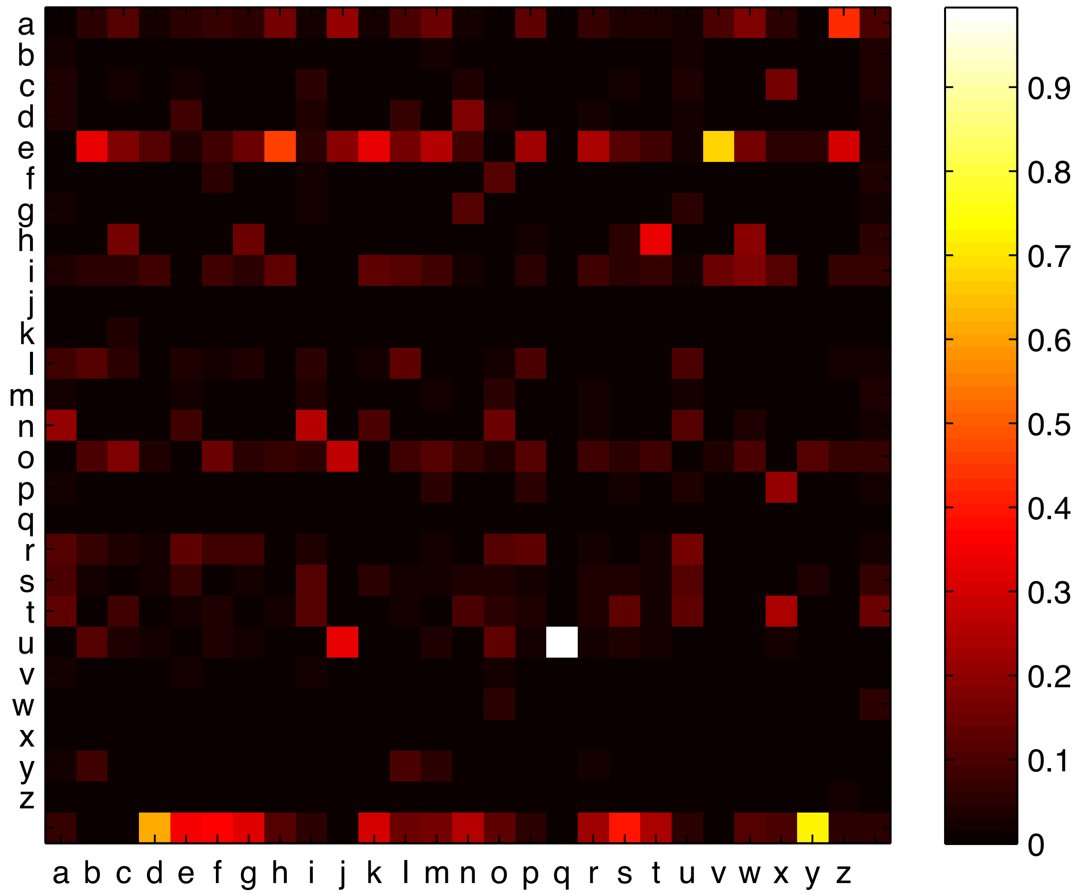
filtering distribution
at each time step
$p(h_t|v_{1:t})$



(c)

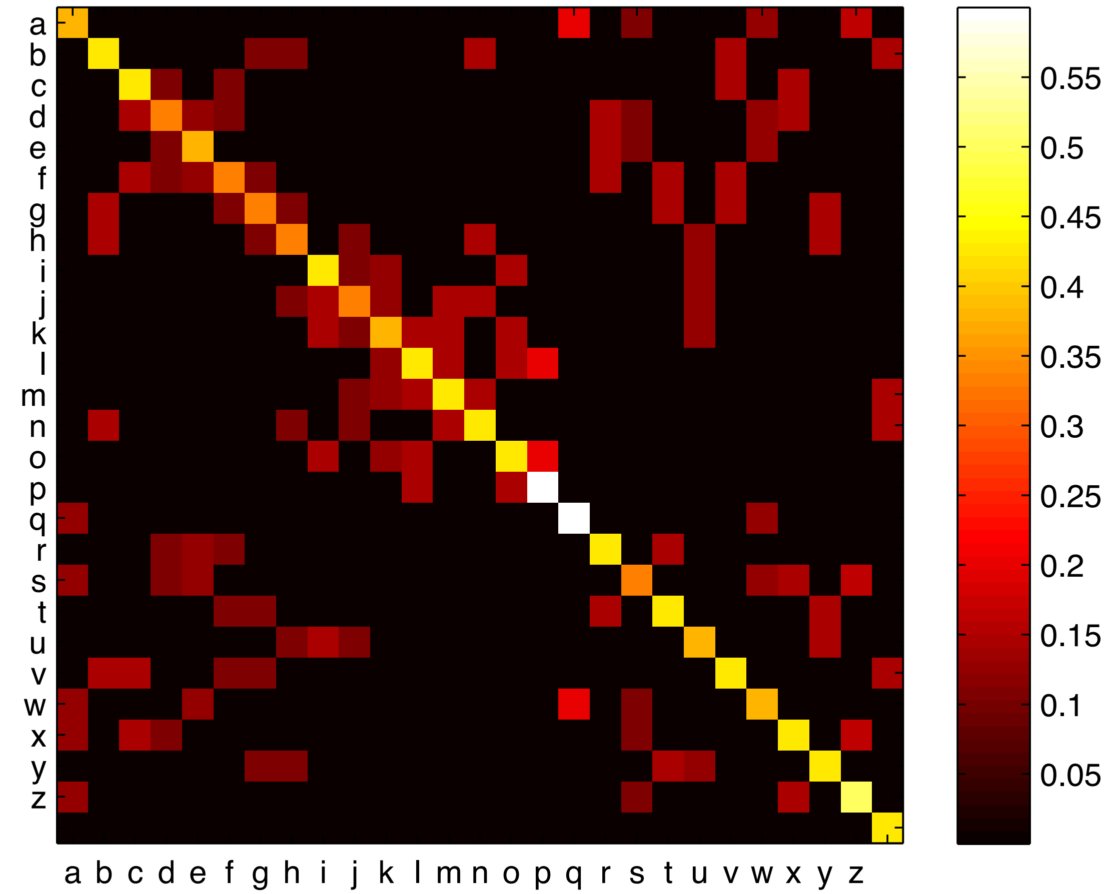
smoothing distribution
at each time step
$p(h_t|v_{1:T})$

A typing model



(a)

letter-to-letter transition matrix for English text



(b)

letter-to-letter transition matrix for "stubby fingers"