Centre for Vision Speech & Signal Processing
University of Surrey, Guildford GU2 7XH.

HMM tutorial 4

by Dr Philip Jackson

- Discrete & continuous HMMs
  - Discrete output pdfs
  - Continuous output pdfs
- Revised B-W formulae
- Implementing B-W re-estimation
  - Forward-backward algorithm
  - Accumulation & update
- Summary

http://www.ee.surrey.ac.uk/Personal/P.Jackson/ce.adsp/

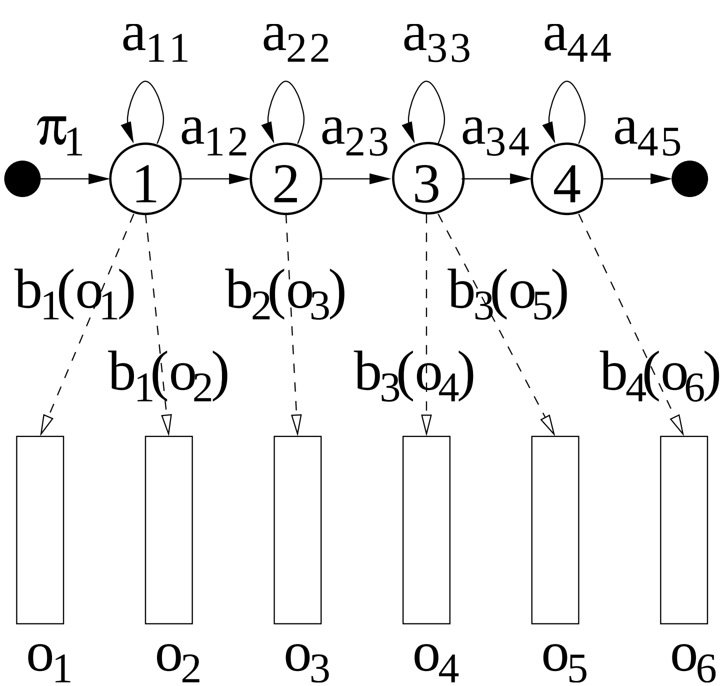
Discrete & continuous HMMs

Types of HMM: Discrete $\lambda = \{\pi, A, B\}$

(a) Initial-state probabilities,
$\pi = \{\pi_i\} = \{P(x_1 = i)\}$ for $1 \leq i \leq N$;

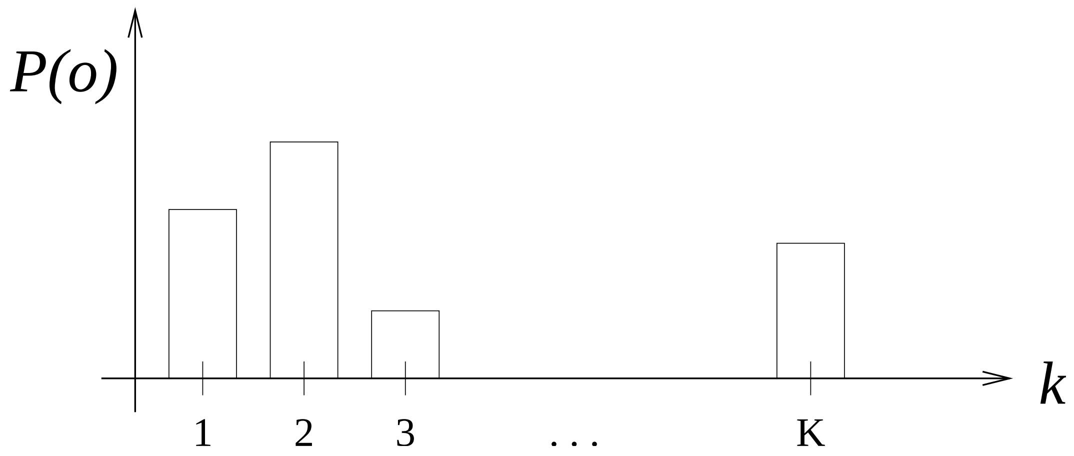
(b) State-transition probabilities,
$A = \{a_{ij}\} = \{P(x_t = j | x_{t-1} = i)\}$ for $1 \leq i, j \leq N$;

(c) Discrete output probabilities,
$B = \{b_i(k)\} = \{P(o_t = k | x_t = i)\}$ for $1 \leq i \leq N$ and $1 \leq k \leq K$.

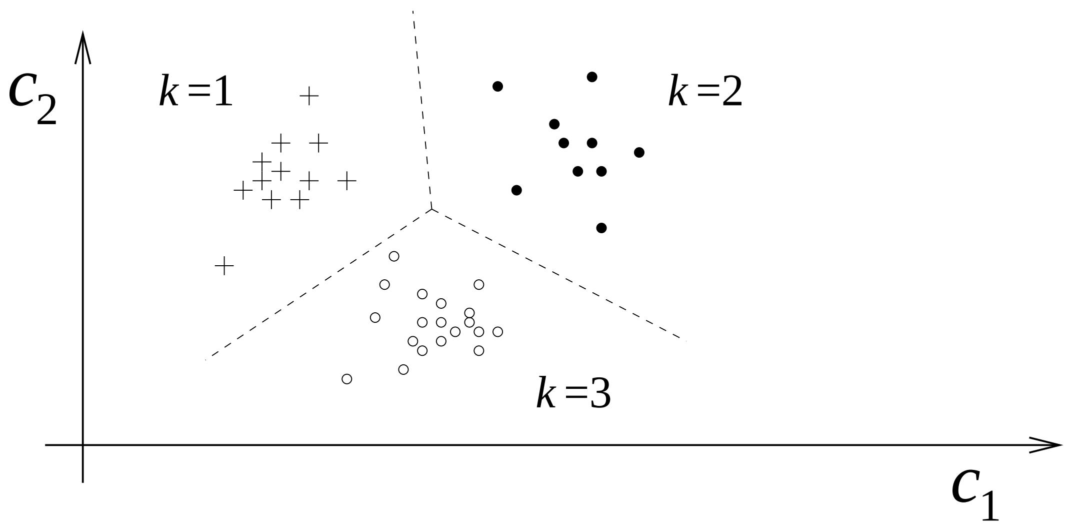


producing *discrete* observations with a state sequence $X = \{1, 1, 2, 3, 3, 4\}$.

Discrete output pdfs



Discretised observations



Types of HMM: Continuous $\lambda = \{\pi, A, B\}$

(a) Initial-state probabilities,
$\pi = \{\pi_i\} = \{P(x_1 = i)\}$ for $1 \leq i \leq N$;

(b) State-transition probabilities,
$A = \{a_{ij}\} = \{P(x_t = j | x_{t-1} = i)\}$ for $1 \leq i, j \leq N$;

(c) Continuous output probabilities,
$B = \{b_i(o_t)\} = \{P(o_t | x_t = i)\}$ for $1 \leq i \leq N$,

where the output probability for each state,

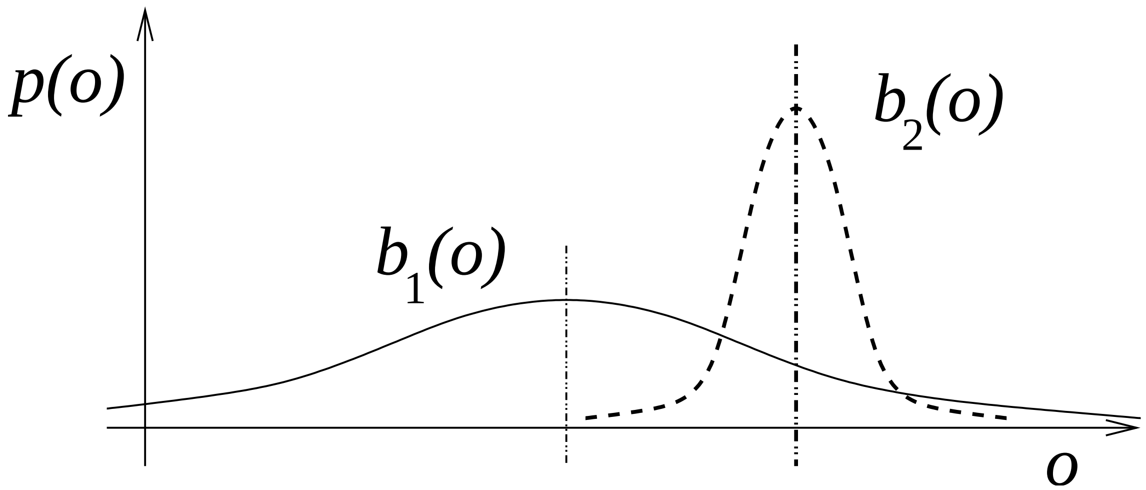
$$b_i(o_t) = f\left(o_t; \kappa_i\right), \tag{11}$$

is a function of the observations $f(o_t)$ that depends on some model parameters κ_i.

Gaussian output pdfs

Univariate Gaussian (scalar observations)

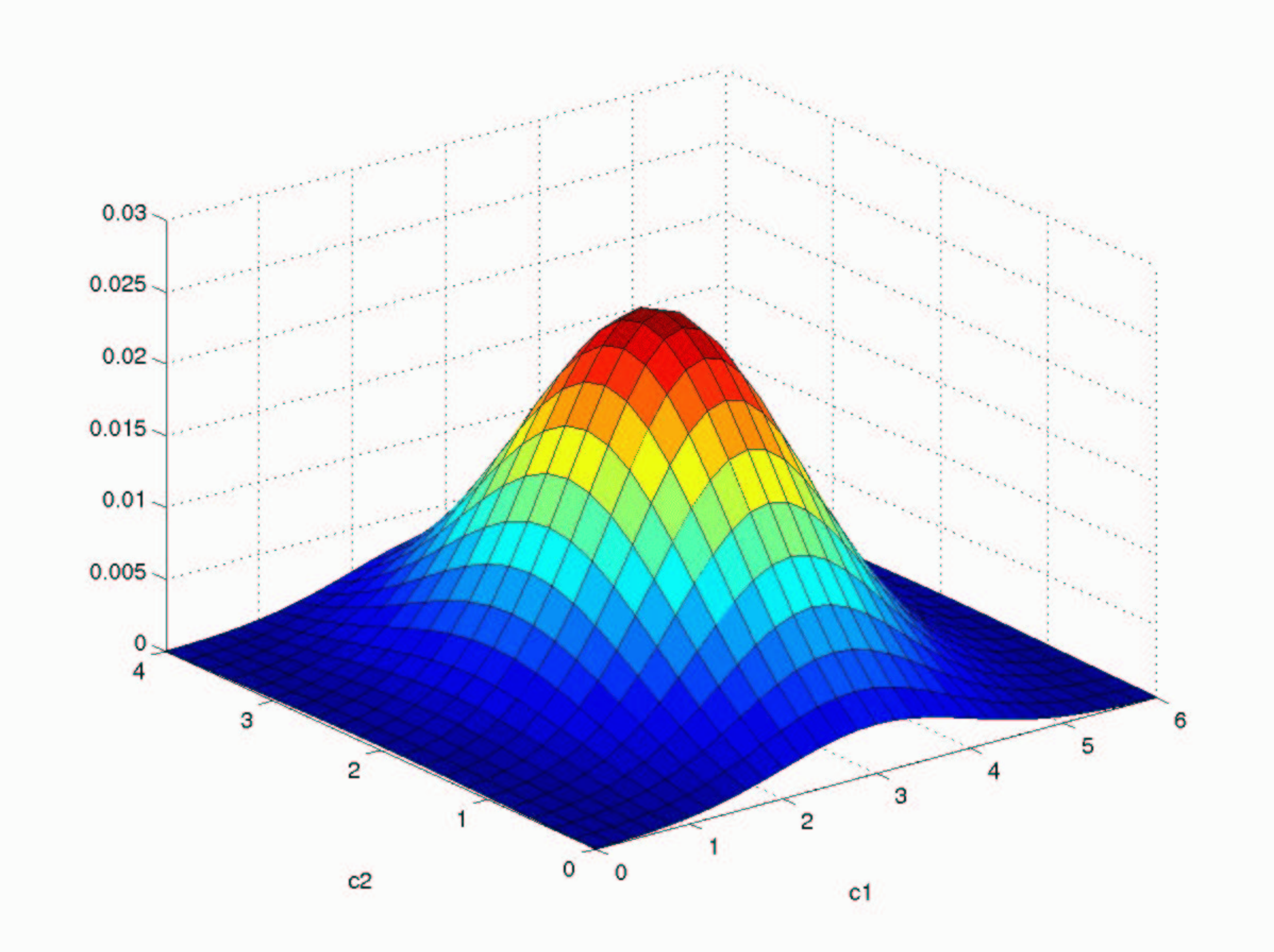
$$b_i(o_t) = \frac{1}{\sqrt{2\pi\Sigma_i}} \exp\left[-\frac{(o_t-\mu_i)^2}{2\Sigma_i}\right].$$



Multivariate Gaussian (vector observations)

$$b_i(\mathbf{o}_t) = \frac{1}{\sqrt{(2\pi)^K |\boldsymbol{\Sigma}_i|}} \exp\left[-\frac{1}{2}(\mathbf{o}_t - \boldsymbol{\mu}_i)\boldsymbol{\Sigma}_i^{-1}(\mathbf{o}_t - \boldsymbol{\mu}_i)'\right],$$

where K is the dimensionality of the observation space.



Parameter estimation examples

Example 0: LS estimate of the mean

We have a set of measurements $\mathcal{O} = \{o_1, o_2, \ldots, o_T\}$, from which we would like to estimate the mean, μ.

Starting with a least-squares approach, we can write an expression for the squared distance of the samples from the mean:

$$
\begin{aligned}
E &= \sum_{t=1}^{T} (o_t - \mu)^2 \\
&= \sum_{t=1}^{T} \left(o_t^2\right) - 2\mu \sum_{t=1}^{T} (o_t) + T\mu^2. \qquad (12)
\end{aligned}
$$

Example 0 (continued)

To find the minimum, the derivative is set to zero:

$$\frac{\partial E}{\partial \mu} = 2T\hat{\mu} - 2\sum_{t=1}^{T}(o_t)$$

$$\Rightarrow \quad 2T\hat{\mu}_{\mathsf{LS}} - 2\sum_{t=1}^{T}(o_t) = 0$$

$$\Rightarrow \quad \hat{\mu}_{\mathsf{LS}} = \frac{1}{T}\sum_{t=1}^{T}(o_t)\,, \tag{13}$$

giving the usual formula for evaluating the sample mean.

Example 1: ML estimate of the mean

Assuming that observations are *continuous* and *normal*,

$$o_t = \mu + n_t \quad \text{for } t \in \{1, 2, \ldots, T\},$$

where $n_t \sim \mathcal{N}(0, \Sigma)$ are independent Gaussian random variables with zero mean and variance Σ, estimate the value of μ from a set of T observations.

The likelihood function is of univariate form (i.e., scalar):

$$p(\mathcal{O}|\lambda) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\Sigma}} \exp\left[-\frac{(o_t - \mu)^2}{2\Sigma}\right].$$

Taking the logarithm and solving the ML equation, gives

$$\widehat{\mu}_{\text{ML}} = \frac{1}{T} \sum_{t=1}^{T} o_t.$$

Example 2: ML estimate of the variance

Estimate the variance Σ, assuming μ to be known.

It can be shown that the ML estimate of variance is

$$\widehat{\Sigma}_{\text{ML}} = \frac{1}{T} \sum_{t=1}^{T} (o_t - \mu)^2.$$

ML estimates for a univariate Gaussian

Thus, we have derived the maximum likelihood estimates of the mean μ and the variance Σ from their moments:

$$\widehat{\mu}_{\mathsf{ML}} = \frac{1}{T} \sum_{t=1}^{T} o_t \tag{14}$$

and

$$\widehat{\Sigma}_{\mathsf{ML}} = \frac{1}{T} \sum_{t=1}^{T} (o_t - \mu)^2. \tag{15}$$

ML estimates for a multivariate Gaussian

Similarly, we can derive maximum likelihood estimates of the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ using their respective moments, for the multivariate case:

$$\hat{\boldsymbol{\mu}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{o}_t \tag{16}$$

and

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{T} \sum_{t=1}^{T} (\mathbf{o}_t - \boldsymbol{\mu})(\mathbf{o}_t - \boldsymbol{\mu})', \tag{17}$$

where $'$ denotes the vector transpose.

B-W re-estimation of Gaussian state parameters

Assuming that the observations come from an HMM with a *continuous* multivariate Gaussian distribution, i.e.:

$$b_j(\mathbf{o}_t) = \mathcal{N}\left(\mathbf{o}_t; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right), \tag{18}$$

we can make a soft (i.e., probabilitistic) allocation of the observations to the states. Thus, if $\gamma_t(j)$ denotes the likelihood of being in state j at time t then eqs. 16 and 17 become weighted averages,

$$\hat{\boldsymbol{\mu}}_j = \frac{\sum_{t=1}^{T} \gamma_t(j)\mathbf{o}_t}{\sum_{t=1}^{T} \gamma_t(j)} \tag{19}$$

and

$$\hat{\boldsymbol{\Sigma}}_j = \frac{\sum_{t=1}^{T} \gamma_t(j)(\mathbf{o}_t - \boldsymbol{\mu}_j)(\mathbf{o}_t - \boldsymbol{\mu}_j)'}{\sum_{t=1}^{T} \gamma_t(j)}, \tag{20}$$

normalised by a denominator which is the total likelihood of all paths passing through state j.

Tutorial 4 summary

- Recap. of likelihoods α_t and β_t
- Recap. of Viterbi algorithm
- Re-estimating models, $\Lambda = \{\lambda\}$
  - Occupation and transition
  - Baum-Welch formulae
- Gaussian pdf examples, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
  - Univariate
  - Multivariate
  - B-W re-estimation

Next: More on output pdfs, grammars. . .