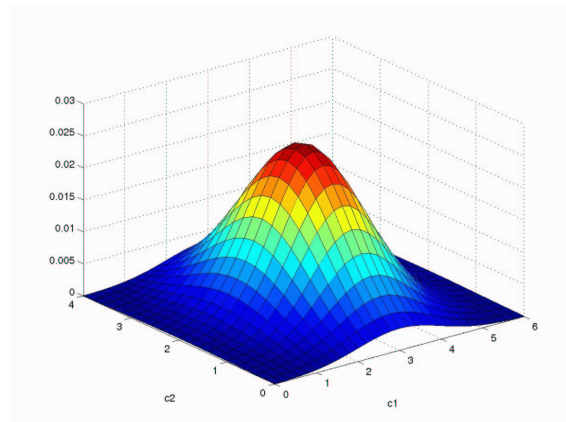
Centre for Vision Speech & Signal Processing
University of Surrey, Guildford GU2 7XH.

HMM part 4

Dr Philip Jackson

- Discrete HMMs
  - Quantised output probabilities
- Continuous HMMs
  - Gaussian output pdfs
- Maximum Likelihood estimation
- Revised B-W formula



Parameters of a discrete HMM, λ

State transition probabilities,

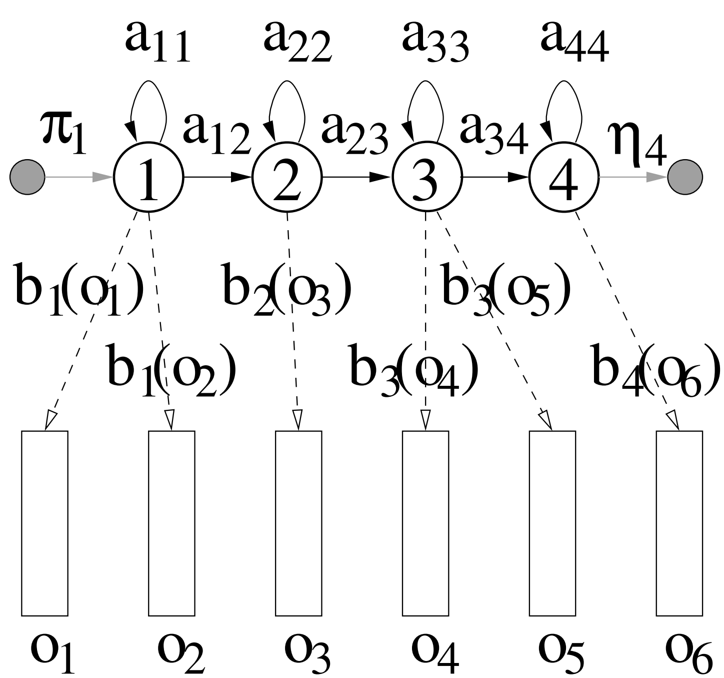
$$A = \{\pi_j, a_{ij}, \eta_i\} = \{P(x_t = j | x_{t-1} = i)\} \qquad \text{for } 1 \leq i, j \leq N$$

where N is the number of states

Discrete output probabilities,

$$B = \{b_i(k)\} = \{P(o_t = k | x_t = i)\} \qquad \text{for } 1 \leq i \leq N, \; 1 \leq k \leq K$$

where K is the number of observation types



generating
a state sequence
$X = \{1, 1, 2, 3, 3, 4\}$
and observations
$\mathcal{O} = \{o_1, o_2, \ldots, o_6\}$

Parameters of a continuous HMM, λ

State-transition probabilities,

$A = \{\pi_j, a_{ij}, \eta_i\} = \{P(x_t = j | x_{t-1} = i)\}$ for $1 \leq i, j \leq N$

where N is the number of states

Continuous output probability densities,

$B = \{b_i(o_t)\} = \{p(o_t | x_t = i)\}$ for $1 \leq i \leq N$

where the output pdf for each state i can be Gaussian

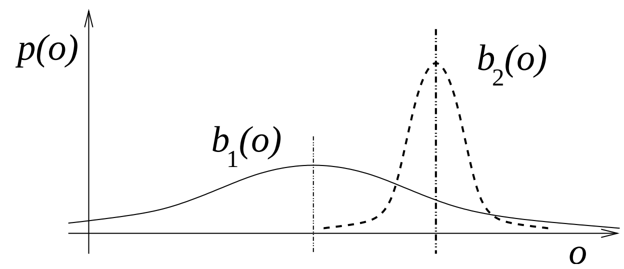
$$\begin{aligned} b_i(o_t) &= \mathcal{N}(o_t; \mu_i, \Sigma_i) \\ &= \frac{1}{\sqrt{2\pi\Sigma_i}} \exp\left(\frac{-(o - \mu_i)^2}{2\Sigma_i}\right) \end{aligned} \quad (1)$$

evaluated at o_t with mean μ_i and variance Σ_i

Univariate Gaussian (scalar observations)

For a given state i,

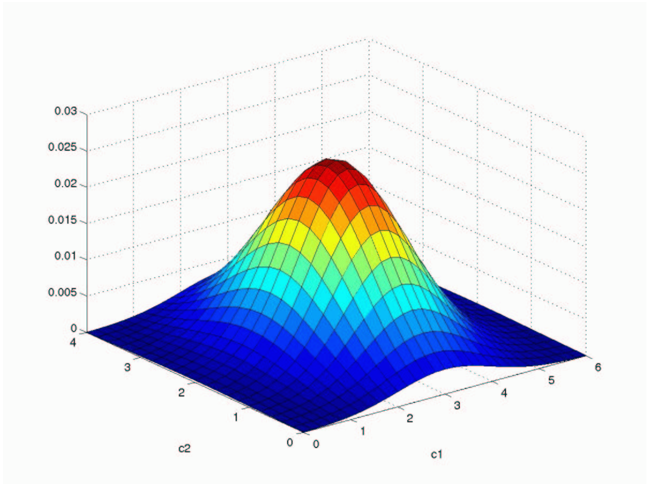
$$b_i(o_t) = \frac{1}{\sqrt{2\pi\Sigma_i}} \exp\left[\frac{-(o_t-\mu_i)^2}{2\Sigma_i}\right]$$



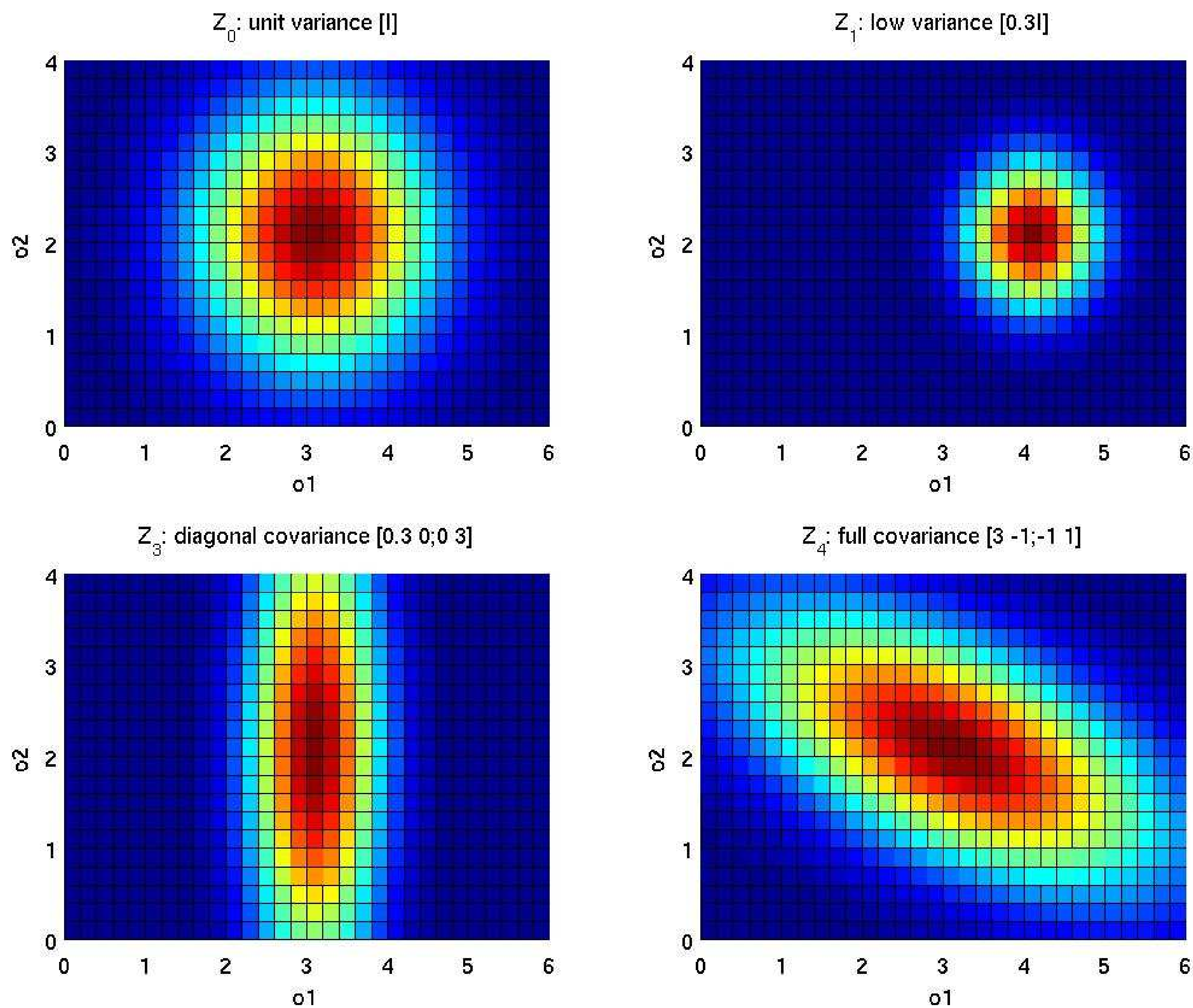
Multivariate Gaussian (vector observations)

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

where the dimensionality of the observation space is K.



Examples of bivariate (2D) Gaussians



Continuous pdf parameter estimation

Example 0: Least-squares estimate of the mean

For a set of samples $\mathcal{O} = \{o_1, o_2, \ldots, o_T\}$ from a single class, we form their total squared distance from the mean μ:

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

and set its derivative w.r.t. μ to zero

$$\frac{\partial E}{\partial \mu} = -2 \sum_{t=1}^{T} (o_t) + 2T\hat{\mu} = 0$$

$$\Rightarrow \qquad \hat{\mu}_{\text{LS}} = \frac{1}{T} \sum_{t=1}^{T} (o_t) \qquad (3)$$

giving the formula to evaluate the sample mean.

Example 1: Maximum likelihood estimate

Assuming the observations $\mathcal{O}$ are independent univariate Gaussian random variables $o_t \sim \mathcal{N}(\mu, \Sigma)$, let us now find the ML estimate the value of μ.

The likelihood function is the product of each $p(o_t|\mu, \Sigma)$

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

Taking the logarithm gives a more convenient expression,

$$\begin{aligned} L(\mu, \Sigma) &= \ln\left[\frac{1}{(2\pi\Sigma)^{T/2}} \prod_{t=1}^{T} \exp\left(-\frac{(o_t - \mu)^2}{2\Sigma}\right)\right] \\ &= \left[-\frac{T}{2}\ln(2\pi\Sigma) + \sum_{t=1}^{T}\left(-\frac{(o_t - \mu)^2}{2\Sigma}\right)\right] \qquad (4) \end{aligned}$$

Example 1: ML estimate of the mean

Differentiating the likelihood function w.r.t. the mean,

$$\begin{aligned}
\frac{\partial L}{\partial \mu} &= \frac{\partial}{\partial \mu}\left[-\frac{T}{2}\ln(2\pi\Sigma) + \sum_{t=1}^{T}\left(-\frac{(o_t-\mu)^2}{2\Sigma}\right)\right] \\
&= \frac{-1}{2\Sigma}\,\frac{\partial}{\partial \mu}\sum_{t=1}^{T}(o_t-\mu)^2 \\
&= \frac{1}{\Sigma}\sum_{t=1}^{T}(o_t-\mu) \qquad (5)
\end{aligned}$$

At the maximum, $\partial L/\partial \mu = 0$, and so

$$\hat{\mu}_{\text{ML}} = \frac{1}{T}\sum_{t=1}^{T} o_t \qquad (6)$$

Example 2: ML estimate of the variance

Similarly to estimate the variance Σ (assuming μ to be known), we differentiate the likelihood function w.r.t. Σ:

$$\begin{aligned} \frac{\partial L}{\partial \Sigma} &= \frac{\partial}{\partial \Sigma}\left[-\frac{T}{2}\ln(2\pi\Sigma) + \sum_{t=1}^{T}\left(-\frac{(o_t-\mu)^2}{2\Sigma}\right)\right] \\ &= \frac{-T}{2}\frac{1}{\Sigma} + \frac{1}{2\Sigma^2}\sum_{t=1}^{T}(o_t-\mu)^2 \qquad (7) \end{aligned}$$

At the maximum $\partial L/\partial\Sigma = 0$, and re-arranging gives

$$\hat{\Sigma}_{\mathrm{ML}} = \frac{1}{T}\sum_{t=1}^{T}(o_t-\mu)^2 \qquad (8)$$

Baum-Welch training of Gaussian state parameters

For observations produced by an HMM with a continuous multivariate Gaussian distribution, i.e.:

$$b_i(\mathbf{o}_t) = \mathcal{N}(\mathbf{o}_t; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \tag{9}$$

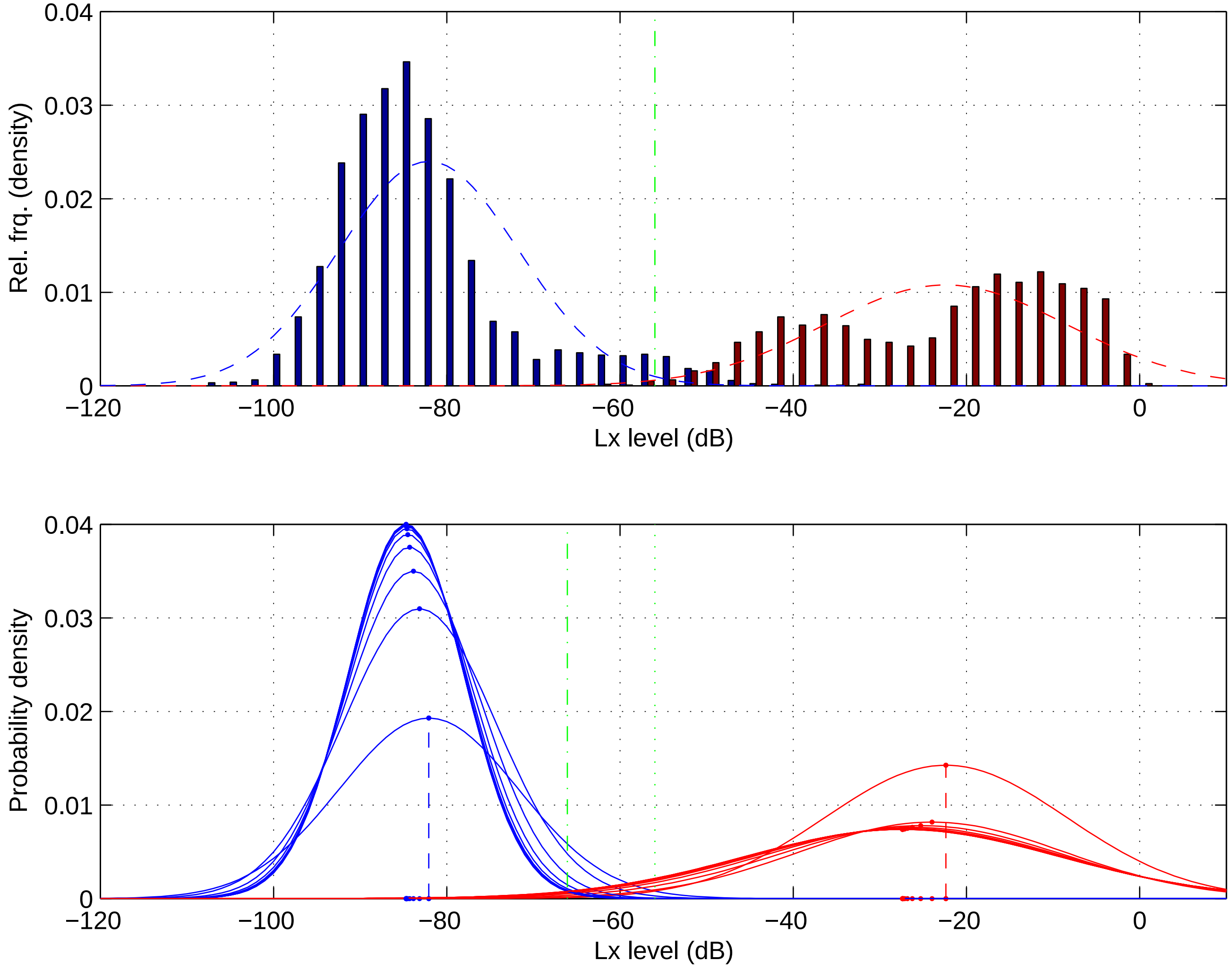
we can again make a soft (i.e., probabilitistic) allocation of the observations to the states. Thus, if $\gamma_t(i)$ denotes the likelihood of occupying state i at time t then the ML estimates of the Gaussian output pdf parameters become weighted averages,

$$\hat{\boldsymbol{\mu}}_i = \frac{\sum_{t=1}^{T} \gamma_t(i) \mathbf{o}_t}{\sum_{t=1}^{T} \gamma_t(i)} \tag{10}$$

$$\hat{\boldsymbol{\Sigma}}_i = \frac{\sum_{t=1}^{T} \gamma_t(i)(\mathbf{o}_t - \boldsymbol{\mu}_i)(\mathbf{o}_t - \boldsymbol{\mu}_i)^{\mathsf{T}}}{\sum_{t=1}^{T} \gamma_t(i)} \tag{11}$$

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

Baum-Welch training of two univariate Gaussians



Part 4 summary

- Discrete HMM
  - Quantised output probabilities, $P(o_t)$
- Continuous HMM
  - Gaussian output pdfs, $p(o_t)$
- Gaussian pdf examples, $b_i(\mathbf{o}_t) = \mathcal{N}(\mathbf{o}_t; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$
  - Univariate
  - Multivariate
  - ML solution
  - B-W re-estimation