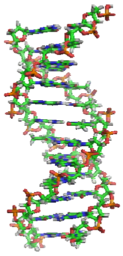
Hidden Markov Model approach for the Assignment of Genome-wide Copy Number Alterations

1. Genetic Background
2. Expectation – Maximization
3. Hidden Markov Models

Genome

- Contains entire heredity information of a cell
- Build up of the four bases A, C, G, T
- Human genome contains 3 billion bases.



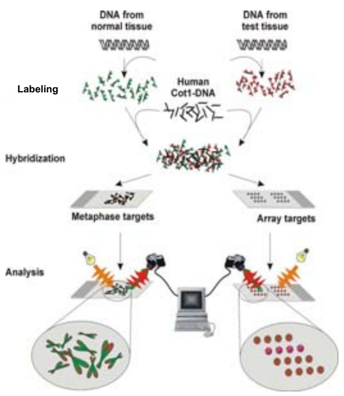
Copy Number Aberrations

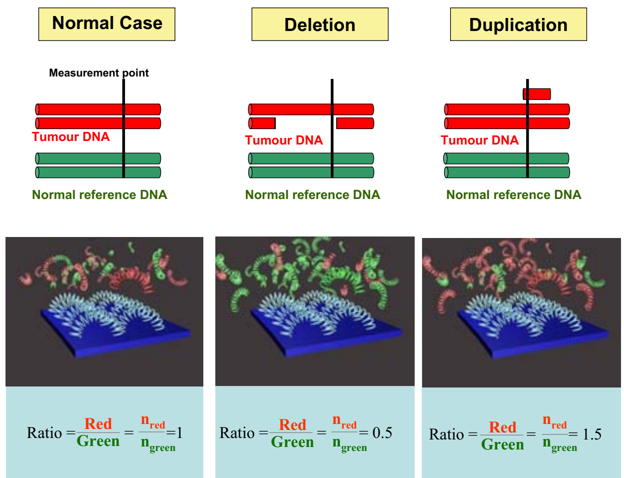
- Genome is present in two copies in each cell
- One or both copies can get lost (deletion)
- Additional copies can emerge (gain, amplification)
- Size of deletions/amplifications: 1 bp - few Mbp, even a whole chromosome can be aberrated

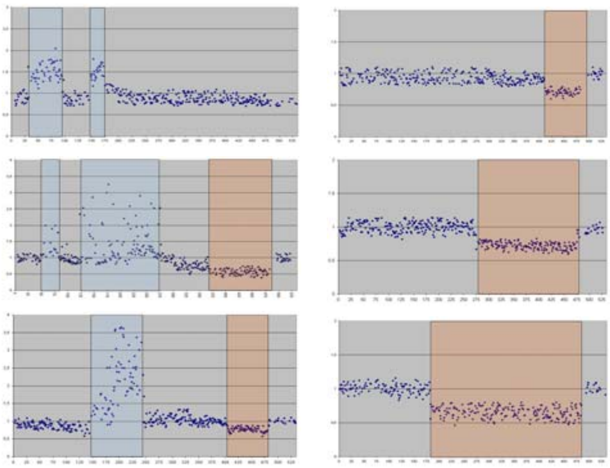
Array CGH

- Comparative Genomic Hybridization
- Comparison of two genomes
- Compare the genomes of two individuals or of two different tissues of the same individual

Metaphase-CGH and Microarray-CGH







Expectation-Maximization (EM) Algorithm

- Maximum likelihood estimators for parameters in models that contain unobserved (latent) variables.
  - incomplete-data problems (HMM, missing data problems)
  - models with "artificially" introduced latent variables (mixture model)
- EM emerged from a number of previous, "intuitive" algorithms
- generalized as EM by Dempster, Laird, and Rubin in 1977[1]
- Recursive algorithm (E-step; M-step)
- EM is often computationally easier than other methods

[1] Dempster, A.P., Laird. N.M., Rubin, D.B.: *Maximum-Likelihood from incomplete data via the EM algorithm.* J. Royal Statist. Society, 1977

Parameter estimation for a Mixture Model - Intuitive approach

Probability distribution as a mixture of K components:

$$p(x|\theta) = \sum_{k=1}^{K} \alpha_k f(x|\vartheta_k) \qquad \sum_k \alpha_k = 1 \qquad \theta = (\alpha_k, \vartheta_k)$$

Two Gaussians:

$$p(x|\theta) = \alpha \cdot N\left(x|\mu_1, \sigma_1^2\right) + (1-\alpha) \cdot N\left(x|\mu_2, \sigma_2^2\right)$$

Five parameters to estimate :$\theta = (\alpha, \mu_1, \sigma_1, \mu_2, \sigma_2)$

No closed-form solution for the ML-estimator

$$L(\theta) = \prod_{i=1}^{N} \sum_{k=1}^{K} \alpha_k f(x_i|\vartheta_k) \qquad \text{N independent samples}$$

$$\log L(\theta) = \sum_{i=1}^{N} \log \left[\sum_{k=1}^{K} \alpha_k f(x_i|\vartheta_k) \right]$$

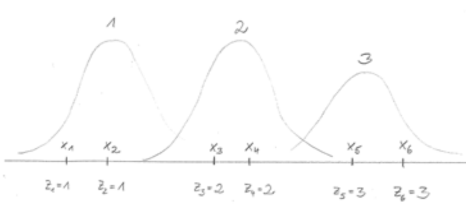
ML-equations are non-linear (even for a single observation)

Making it easier by complicating it first

- Introduction of new, latent variables z to the model
- x = incomplete (observed) data; (z, x) = complete data

Introduction of latent variables to the Mixture Model

Variable z_i tells from which component of the p.d.f. x_i is;
$1 \leq z_i \leq K, \quad 1 \leq i \leq N$



Knowing the z_i solves the problem!

Distribution of the latent variables, given the observations and the parameters

Initial guess of the parameters: $\theta_t = (\alpha_k, \vartheta_k)$

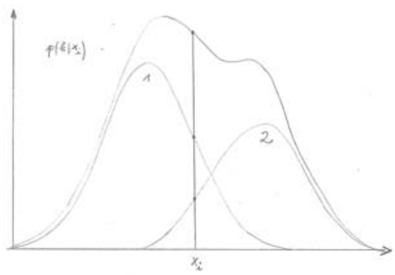
$$\begin{aligned} p\left(z_i = k | x_i, \theta_t\right) &= \frac{p(z_i = k|\theta_t) \cdot p(x_i|z_i = k, \theta_t)}{p(x_i|\theta_t)} \\ &= \frac{\alpha_k \cdot f\left(x_i|\vartheta_k\right)}{\sum_k \alpha_k \cdot f\left(x_i|\vartheta_k\right)} \\ &= \omega_{ik} \quad \text{("membership probability" of } x_i \text{ to component } k) \end{aligned}$$

Expected number of observations that come from component k:

$$c_k = \sum_i \omega_{ik} \qquad \sum_k c_k = N$$

"Membership probability" for two Gaussian components

$$\omega_{ik} = \frac{\alpha_k N\left(x_i|\mu_k, \sigma_k^2\right)}{\alpha_1 N\left(x_i|\mu_1, \sigma_1^2\right) + \alpha_2 N\left(x_i|\mu_2, \sigma_2^2\right)} \qquad k = (1, 2)$$



Re-estimation of the parameters for Gaussian components

Having the membership probability ω_{ik}:

$$\begin{aligned} \alpha_k &= \frac{c_k}{N} = \frac{\sum_i \omega_{ik}}{N} \\ \mu_k &= \frac{\sum_i \omega_{ik} x_i}{\sum_i \omega_{ik}} \\ \sigma_k^2 &= \frac{\sum_i \omega_{ik}(x_i - \mu_k)^2}{\sum_i \omega_{ik}} \end{aligned}$$

Alternate re-calculation of ω_{ik} and $(\mu_k,\ \sigma_k^2,\ \alpha_k)$ finds a maximum likelihood estimator.

General Approach to the EM

Observed data: x; latent data: z; parameters of the model: θ

$$L(\theta) = p(x|\theta) = \frac{p(z, x|\theta)}{p(z|x, \theta)}$$

$$\log p(x|\theta) = \log p(z, x|\theta) - \log p(z|x, \theta) \qquad (1)$$

calculate the cond. expectation with respect to $z|x, \theta_t$ for some known θ_t
(generate z according to the distribution $p(z|x, \theta_t)$ and average)

Expected values of log-likelihood

$$\log p(x|\theta) = \log p(z, x|\theta) - \log p(z|x, \theta)$$

$$\log p(x|\theta) = \underbrace{\int_z \log p(z, x|\theta) \cdot p(z|x, \theta_t)\, dz}_{Q(\theta, \theta_t)} - \underbrace{\int_z \log p(z|x, \theta) \cdot p(z|x, \theta_t)\, dz}_{H(\theta, \theta_t)}$$

$$\log p(x|\theta) = Q(\theta, \theta_t) - H(\theta, \theta_t)$$

Increasing the log-likelihood

Find a new value of θ so that $\Delta L = L(\theta) - L(\theta_t) \geq 0$

$$\begin{aligned} \Delta L &= \log p(x|\theta) - \ log p(x|\theta_t) \\ &= Q(\theta, \theta_t) - Q(\theta_t, \theta_t) + \underbrace{H(\theta_t, \theta_t) - H(\theta, \theta_t)}_{\text{always } \geq 0} \geq 0 \end{aligned}$$

$$H(\theta, \theta_t) \leq H(\theta_t, \theta_t) \qquad \textit{upper bound}$$

$$\theta_{t+1} = \underset{\theta}{\operatorname{argmax}}\, Q(\theta, \theta_t) \qquad \textit{increases likelihood}$$

EM - Iteration

$$\theta_0 \rightarrow \theta_t$$

$$Q(\theta, \theta_t) = \int_z \log p(z, x|\theta) \cdot p(z|x, \theta_t)\, dz \qquad \textit{E-step}$$

$$\theta_{t+1} = \underset{\theta}{\operatorname{argmax}}\, Q(\theta, \theta_t) \qquad \textit{M-step}$$

$$\theta_{t+1} \rightarrow \theta_t$$

EM - discrete latent variables, multiple independent observations

$$Q(\theta, \theta_t) = \sum_z \log p(z, x|\theta) \cdot p(z|x, \theta_t) \qquad \textit{summation}$$

$$L(\theta) = \prod_{i=1}^{N} p(x_i|\theta) \qquad \text{i.i.d.}$$

$$Q(\theta, \theta_t) = \sum_i \sum_z \log p(z, x_i|\theta) \cdot p(z|x_i, \theta_t)$$

Upper bound of H

Definition of the function H:

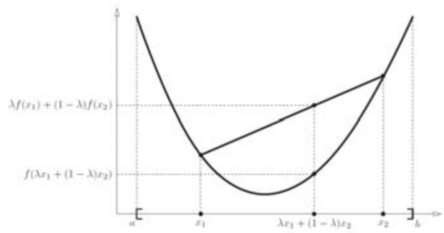
$$H(\theta, \theta_t) = \int_z \log p(z|x, \theta) \cdot p(z|x, \theta_t)\, dz$$

It follows that $H(\theta, \theta_t)$ is bounded:

$$H(\theta, \theta_t) \leq H(\theta_t, \theta_t)$$

Jensens Inequality

$$\Phi\left[\sum p_i x_i\right] \leq \sum p_i \Phi(x_i) \qquad \sum p_i = 1$$



Figur: Convex function

The upper bound of $H(\theta, \theta_t)$

$$\begin{aligned}
& H(\theta_t, \theta_t) - H(\theta, \theta_t) \\
= \; & \int_z \log p(z|x, \theta_t) \cdot p(z|x, \theta_t)\, dz - \int_z \log p(z|x, \theta) \cdot p(z|x, \theta_t)\, dz \\
= \; & \int_z -\log\left[\frac{p(z|x, \theta)}{p(z|x, \theta_t)}\right] \cdot p(z|x, \theta_t)\, dz \qquad \textit{Jensens inequality} \\
\geq \; & -\log\left\{\int_z \frac{p(z|x, \theta)}{\cancel{p(z|x, \theta_t)}} \cdot \cancel{p(z|x, \theta_t)}\, dz\right\} \\
= \; & 0
\end{aligned}$$

EM - What have we won?

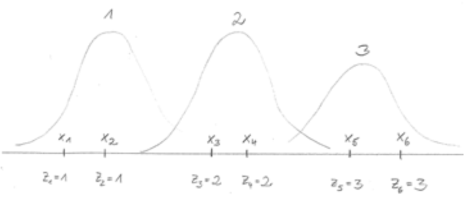
$$L(\theta) = p(x|\theta) \qquad \textit{maximize w.r.t. } \theta$$

$$Q(\theta, \theta_t) = \int_z \log p(z, x|\theta) \cdot p(z|x, \theta_t)\, dz \qquad \textit{... as well}$$

It *does* make sense (see mixture model).

Parameter estimation for a Mixture Model - Formal approach

Latent variables z_i have the same meaning as before.
$1 \leq z_i \leq K, \quad 1 \leq i \leq N$



E-step - Calculation of Q

The probability distribution was:

$$p(x_i|\theta) = \sum_{k=1}^{K} \alpha_k \cdot f(x_i|\vartheta_k) \qquad \theta = (\alpha_k; \vartheta_k) \qquad \vartheta_k = (\mu_k; \sigma_k^2)$$

In the E-step, we have to calculate:

$$Q(\theta, \theta_t) = \sum_{i=1}^{N} \sum_{k=1}^{K} \log p(z_i = k, x_i|\theta) \cdot p(z_i = k|x_i, \theta_t)$$

OBS!! Sums over z_k with $k \neq i$ disappear.

E-step - Calculation of Q

Calculate joint-probability term in Q:

$$Q(\theta, \theta_t) = \sum_{i=1}^{N} \sum_{k=1}^{K} \log p(z_i = k, x_i|\theta) \cdot p(z_i = k|x_i, \theta_t)$$

The joint probability is an easy expression:

$$p(z_i = k, x_i|\theta) = \alpha_k \cdot f(x_i|\vartheta_k)$$

... what is confirmed by the expression:

$$\sum_{k=1}^{K} p(z_i = k, x_i|\theta) = p(x_i|\theta) = \sum_{k=1}^{K} \alpha_k \cdot f(x_i|\vartheta_k)$$

E-step - Calculation of Q

Calculate conditional-probability term in Q:

$$Q(\theta, \theta_t) = \sum_{i=1}^{N} \sum_{k=1}^{K} \log p(z_i = k, x_i|\theta) \cdot p(z_i = k|x_i, \theta_t)$$

Bayes rule:

$$p(z_i = k|x_i, \theta_t) = \frac{p(x_i, z_i = k|\theta_t)}{p(x_i|\theta_t)} = \underbrace{\frac{\alpha_k \cdot f(x_i|\vartheta_k)}{f(x_i|\theta_t)}}_{\text{at step t}} = \omega_{ik}$$

E-step - Calculation of Q

$$Q(\theta, \theta_t) = \sum_{i=1}^{N} \sum_{k=1}^{K} \log p(z_i = k, x_i|\theta) \cdot p(z_i = k|x_i, \theta_t)$$

$$\begin{aligned} Q(\theta, \theta_t) &= \sum_{i=1}^{N} \sum_{k=1}^{K} \log \left[\alpha_k \cdot f(x_i|\vartheta_k)\right] \cdot \omega_{ik} \\ &= \underbrace{\sum_{i=1}^{N} \sum_{k=1}^{K} \omega_{ik} \log \alpha_k}_{Q_A} + \underbrace{\sum_{i=1}^{N} \sum_{k=1}^{K} \omega_{ik} \log f(x_i|\vartheta_k)}_{Q_B} \end{aligned}$$

Both parts can be maximized separately

M-Step - Maximization of Q_A

$$Q_A = \sum_{i=1}^{N} \sum_{k=1}^{K} \omega_{ik} \log \alpha_k = \sum_{k} c_k log \alpha_k \quad \text{with} \quad \sum_{k} \alpha_k = 1$$

$$\alpha_k = \frac{c_k}{N} = \frac{\sum_i \omega_{ik}}{N} \qquad \text{Lagrange}$$

M-Step - Maximization of Q_B

$$
\begin{aligned}
Q_B &= \sum_{i=1}^{N} \sum_{k=1}^{K} \omega_{ik} \log f(x_i|\vartheta_k) \\
&= \sum_{i=1}^{N} \omega_{i1} \log f(x_i|\vartheta_1) + \sum_{i=1}^{N} \omega_{i2} \log f(x_i|\vartheta_2) + \ldots
\end{aligned}
$$

$\sum_{i=1}^{N} \omega_{ik} \log f(x_i|\vartheta_k)$ maximize K of these terms w. r. t. ϑ_k

M-Step for a mixture of Gaussians

$$
\begin{aligned}
\mu_k &= \frac{\sum_i \omega_{ik} x_i}{\sum_i \omega_{ik}} \\
\sigma_k^2 &= \frac{\sum_i \omega_{ik} (x_i - \mu_k)^2}{\sum_i \omega_{ik}} \\
\alpha_k &= \frac{\sum_i \omega_{ik}}{N} = \frac{c_k}{N}
\end{aligned}
$$

These are the results we obtained before in an intuitive way.

EM - Convergence

- Each iteration increases or holds constant the likelihood
- converges to a local maximum or saddle point
- multiple initial values, simulated annealing

MAP - EM

Maximize the posterior probability:

$$p(\theta|x) = \frac{p(x|\theta) \cdot p(\theta)}{p(x)}$$

$$
\begin{aligned}
L(\theta) &= \log p(x|\theta) + \log p(\theta) && \textit{introduce latent variables ...} \\
&= \log p(z, x|\theta) - \log p(z|x, \theta) + \log p(\theta) && \textit{expectation ...} \\
&= Q(\theta, \theta_t) - H(\theta, \theta_t) + \log p(\theta) && \textit{additional term: } \log p(\theta)
\end{aligned}
$$

MAP-EM: Iteration

$$\theta_0 \to \theta_t$$

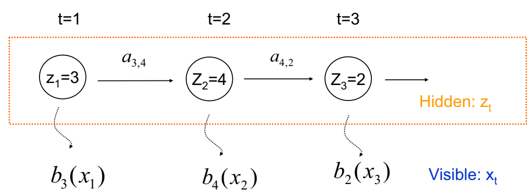
$$Q(\theta, \theta_t) = \int_z \log p(z, x|\theta) \cdot p(z|x, \theta_t)\, dz \qquad \textit{E-step}$$

$$\theta_{t+1} = \underset{\theta}{\operatorname{argmax}} \left[Q(\theta, \theta_t) + \log p(\theta)\right] \qquad \textit{M-step}$$

$$\theta_{t+1} \to \theta_t$$

Choosing suitable priors makes MAP-EM not harder than MLE.

Hidden Markov Model



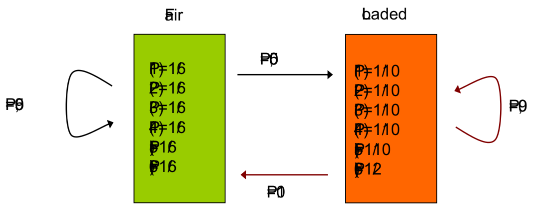
$$P(x_1, x_2, \ldots, z_1, z_2, \ldots | \theta) = p(z_1) \cdot \quad b_{z_1}(x_1) \cdot a_{z_1, z_2} \cdot \quad b_{z_2}(x_2) \cdot a_{z_2, z_3} \cdot \quad \ldots$$

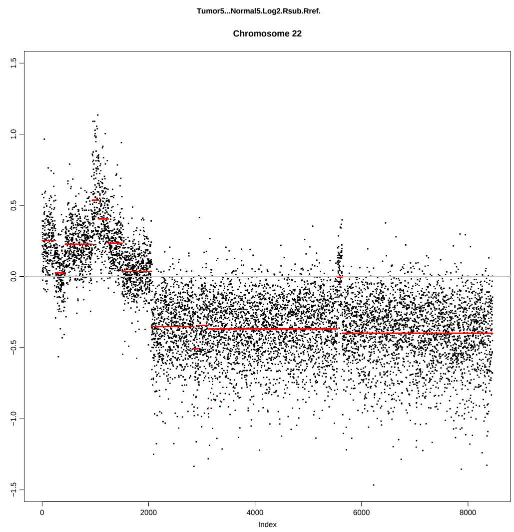
$$P(x, z | \theta) = p(z_1) \cdot \prod_{i=1}^{T} b_{z_i}(x_i) \cdot a_{z_i, z_{i+1}}$$

Oasionellydisbnest asino

Oservations x are visible: 3 1 2 4 54 63 4 63 63 4 61 4 63 6

Sate z is hdden:





Parameters of the Markov chain and their properties

$\theta = (a_{ij}, b_i, \pi_i)$

$\pi_i = p(z_1 = i)$	initial states	$1 \leq i \leq N$
$a_{ij} = p(z_{t+1} = j \mid z_t = i)$	transitions	$1 \leq i, j \leq T$
$b_i(x_t) = p(x_t \mid z_t = i)$	emissions	$1 \leq i \leq N$

HMM is homogeneous, z_t discrete, x_t discrete or continuous

Find the state path given the observations and the model

$$p(x, z|\theta) = p(z|x, \theta) \cdot p(x|\theta)$$

$$z^\star = \underset{z_1 \ldots z_T}{\operatorname{argmax}}\, p(z_1 \ldots z_T | x_1 \ldots x_T, \theta)$$

$$z^\star = \underset{z_1 \ldots z_T}{\operatorname{argmax}}\, p(z_1 \ldots z_T, x_1 \ldots x_T | \theta)$$

- Finds the copy numbers that underlie our data (or the occasions when the loaded dice is used)
- N^T pathes $\rightarrow 2 \cdot T \cdot N^T$ computations - brute force fails

Find the single best path for given x and θ - Viterbi

$$\boxed{p^\star = \max_z p(z|x, \theta) = \max_z p(z, x|\theta)}$$

A variable $\delta_t(i)$ defined in such a way:

$$\delta_t(i) = \max_{z_1, \ldots, z_{t-1}} p(z_1, \ldots, z_{t-1}, z_t = i, x_1, \ldots, x_t | \theta)$$

... can be recursively calculated like that:

$$\delta_{t+1}(j) = \max_i [\delta_t(i) a_{ij}] \cdot b_j(x_{t+1})$$

The Viterbi algorithm

$$\boxed{p^\star = \max_z p(z|x, \theta) = \max_z p(z, x|\theta)}$$

$\delta_1(i) = \pi_i \cdot b_i(x_1); \quad \psi_1(i) = 0$	$1 \leq i \leq N$
$\delta_t(j) = \max_i [\delta_{t-1}(i) a_{ij}] \cdot b_j(x_t)$	$2 \leq t \leq T;\ 1 \leq j \leq N$
$\psi_t(j) = \operatorname{argmax}_i [\delta_{t-1}(i) a_{ij}]$	$2 \leq t \leq T;\ 1 \leq j \leq N$

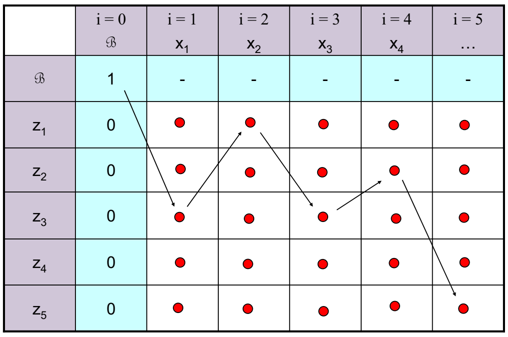
$p^\star = \max_i \delta_T(i)$ (max. prob.)

$z_T^\star = \operatorname{argmax}_i \delta_T(i)$

$z_t^\star = \psi_{t+1}(z_{t+1}^\star)$ (backtracking)

$T \cdot N^2$ computations!

Viterbi - Trellis



Forward-Variable

Allows to calculate $p(x|\theta)$ (sum over all state sequences)

Let $\alpha_t(i) = p(x_1, x_2, \ldots, x_t, z_t = i|\theta)$

then $\alpha_{t+1}(j) = \left[\sum_{j=1}^{N} \alpha_t(i) a_{ij}\right] b_j(x_{t+1})$

$\alpha_1(i) = \pi_i \cdot b_i(x_1) \qquad 1 \leq i \leq N$

$\alpha_{t+1}(j) = \left[\sum_{j=1}^{N} \alpha_t(i) a_{ij}\right] b_j(x_{t+1}) \qquad 1 \leq t \leq T-1; \quad 1 \leq j \leq N$

$p(x|\theta) = \sum_{i=1}^{N} \alpha_T(i)$

Backward-Variable

$\beta_t(i) = p(x_{t+1}, x_{t+2}, \ldots, x_T|z_t = i, \theta)$ Definition

---

$$\beta_T(i) = 1 \qquad 1 \leq i \leq N \qquad \text{Initialization}$$

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(x_{t+1}) \beta_{t+1}(j) \qquad t = T-1, T-2, \ldots 1$$

Other auxiliary variables: $\gamma_t(i)$

Probability of being in state i at position t, given the observations and the model:

$\gamma_t(i) = p(z_t = i|x, \theta)$ Definition

Can be expressed in terms of the forward-backward variables:

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i}{p(x|\theta)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)} \qquad \sum_{i=1}^{N} \gamma_t(i) = 1$$

Individually most likely state at position t:

$$z_t = \operatorname*{argmax}_i \gamma_t(i) = \operatorname*{argmax}_i p(z_t = i|x, \theta)$$

Other auxiliary variables: $\xi_t(i,j)$

Probability of being in state i at position t and in state j at position $t+1$, given the observations and the model:

$\xi_t(i,j) = p(z_t = i, z_{t+1} = j|x, \theta)$ Definition

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(x_{t+1}) \beta_{t+1}(j)}{p(x|\theta)} = \frac{\alpha_t(i) a_{ij} b_j(x_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(x_{t+1}) \beta_{t+1}(j)}$$

Problem

- Up to now, the "model", i.e. the parameters θ were supposed to be given
- For arrayCGH (and many other problems): How can we possibly know these parameters?

Solution

$$\boxed{\theta^\star = \operatorname*{argmax}_\theta p(x|\theta)}$$

- Adjust θ so that the probability of the observed sequence becomes maximal
- ML estimate of the HMM
- EM can solve this (Baum-Welch algorithm)

E-step: calculation of the auxiliary function Q

$$Q(\theta, \theta_t) = \sum_z \log p(z, x|\theta) \cdot p(z|x, \theta_t)$$

Probability of the "complete data" given the model θ:

$$p(x, z|\theta) = \pi(z_1) \cdot b_{z_1}(x_1) \cdot a_{z_1,z_2} \cdot b_{z_2}(x_2) \cdot a_{z_2,z_3} \cdot \ldots \cdot b_{z_T}(x_T)$$

$$p(x, z|\theta) = p(z_1) \prod_{t=2}^{T} p(z_t|z_{t-1}) \prod_{t=1}^{T} p(x_t|z_t)$$

$$\log p(x, z|\theta) = \log p(z_1) + \sum_{t=2}^{T} \log p(z_t|z_{t-1}) + \sum_{t=1}^{T} \log p(x_t|z_t)$$

E-step: Q separates into 3 terms

$$\begin{aligned}
Q(\theta, \theta_t) &= \sum_z \log p(z_1) p(z|x, \theta_t) \\
&+ \sum_z \sum_{t=2}^{T} \log p(z_t|z_{t-1}) p(z|x, \theta_t) \\
&+ \sum_z \sum_{t=1}^{T} \log p(x_t|z_t) p(z|x, \theta_t) \\
&= Q_A + Q_B + Q_C
\end{aligned}$$

The 3 parts can be maximized separately

E-step: Calculation of Q_A

$$\begin{aligned}
Q_A &= \sum_z \log p(z_1|\theta) p(z|x, \theta_t) \\
&= \sum_{z_1,\ldots,z_T} \log p(z_1|\theta) p(z_1, \ldots, z_T|x_1, \ldots, x_T, \theta_t) \\
&= \ldots \\
&= \sum_{i=1}^{N} \log p(z_1 = i|\theta) p(z_1 = i|x, \theta_t) \\
&= \sum_{i=1}^{N} \log \pi_i \cdot \gamma_1(i)
\end{aligned}$$

M-step: Re-estimation of the initial probabilities π_i

Maximize $\quad Q_A = \sum_{i=1}^{N} \log \pi_i \cdot \gamma_1(i) \quad$ w.r.t. π_i $\quad$ considering $\quad \sum_i \pi_i = 1$

Introduce Lagrange multiplicator:

$$\frac{\partial}{\partial \pi_i} \left\{ \sum_{i=1}^{N} \log \pi_i \cdot \gamma_1(i) + \lambda \left(\sum_i \pi_i - 1 \right) \right\} = 0$$

Re-estimation formula for the initial prob.: (using $\sum_i \gamma_1(i) = 1$)

$$\pi_i = \gamma_1(i)$$

E-step: Calculation of Q_B

$$\begin{aligned}
Q_B &= \sum_z \sum_{t=2}^{T} \log p(z_t|z_{t-1}) p(z|x, \theta_t) \\
&= \ldots \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=2}^{T} \log p(z_t = j|z_{t-1} = i) p(z_t = j, z_{t-1} = i|x, \theta_t) \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} \log a_{ij} \left(\sum_t \xi_t(i, j) \right)
\end{aligned}$$

M-step: Re-estimation of the transition probabilities a_{ij}

To maximize with respect to a_{ij}:

$$Q_B = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\sum_t \xi_t(i,j) \right) \log a_{ij}$$

Contraints:

$$\sum_j a_{ij} = 1$$

Method of Lagrange multipliers:

$$a_{ij} = \frac{\sum_t \xi_t(i,j)}{\sum_t \gamma_t(i)}$$

Re-estimation formula for the transition probabilities: Interpretation

$$a_{ij} = \frac{\sum_t \xi_t(i,j)}{\sum_t \gamma_t(i)}$$

$$a_{ij} = \frac{\sum_t p(z_t = i, z_{t+1} = j | x, \theta)}{\sum_t p(z_t = i | x, \theta)}$$

$$a_{ij} = \frac{\text{expected number of } i \to j \text{ transitions}}{\text{expected number of transitions away from i}}$$

Having a training set with known states, one would exactly use these expressions to re-estimate the transition probabilities

E-step: Calculation of Q_C

$$Q_C = \sum_z \sum_{t=1}^{T} \log p(x_t|z_t) p(z|x, \theta_t)$$

$$p(x_t|z_t, \theta) = p(x_t|z_t, \mu_{z_t}, \sigma_{z_t}) = \frac{1}{\sqrt{2\pi\sigma_{z_t}^2}} exp \left[\frac{-(x_t - \mu_{z_t})^2}{2\sigma_{z_t}^2} \right]$$

$$\begin{aligned} Q_C &= \sum_z \sum_{t=1}^{T} \left[\frac{-(x_t - \mu_{z_t})^2}{2\sigma_{z_t}^2} - \log \sigma_{z_t} \right] p(z|x, \theta_t) \\ &= \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\frac{-(x_t - \mu_i)^2}{2\sigma_i^2} - \log \sigma_i \right] p(z_t = i | x, \theta_t) \\ &= \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\frac{-(x_t - \mu_i)^2}{2\sigma_i^2} - \log \sigma_i \right] \gamma_t(i) \end{aligned}$$

M-step: Re-estimation of the parameters for the emission probability

Maximize $\quad Q_C = \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\frac{-(x_t - \mu_i)^2}{2\sigma_i^2} - \log \sigma_i \right] \gamma_t(i) \quad$ w. r. t. $\quad \mu_i, \sigma_i$

Weighted maximum likelihood:

$$\frac{\partial}{\partial \mu_i} = 0 \qquad \frac{\partial}{\partial \sigma_i} = 0$$

$$\mu_i = \frac{\sum_t \gamma_t(i) x_t}{\sum_t \gamma_t(i)} \qquad \sigma_i^2 = \frac{\gamma_t(i)(x_t - \mu_i)^2}{\sum_t \gamma_t(i)}$$

Re-estimation of the CDHMM model parameters for Gauss-shaped emission probability

$\pi_i = \gamma_1(i)$	initial state probabilities
$a_{ij} = \frac{\sum_t \xi_t(i,j)}{\sum_t \gamma_t(i)}$	transition probabilities
$\mu_i = \frac{\sum_t \gamma_t(i) x_t}{\sum_t \gamma_t(i)}$	mean of the emission distribution
$\sigma_i^2 = \frac{\sum_t \gamma_t(i)(x_t - \mu_i)^2}{\sum_t \gamma_t(i)}$	variance of the emission distr.

MAP estimate for the CDHMM

Maximize the posterior probability:

$$p(\theta|x) = \frac{p(x|\theta) \cdot p(\theta)}{p(x)} \qquad \textit{Maximum a posteriori}$$

$$\theta_{t+1} = \underset{\theta}{\operatorname{argmax}} \left[Q(\theta, \theta_t) + \log p(\theta) \right] \qquad \textit{M-step}$$

Choose prior so that the 3 parts of Q can be maximized separately:

$$p(\theta) \propto \prod_{i=1}^{N} \left[\pi_i^{\eta_i - 1} g(\tau_i, \nu_i, \alpha_i, \beta_i) \prod_{j=1}^{N} a_{ij}^{\eta_{ij} - 1} \right]$$

MAP re-estimation of the CDHMM model parameters for Gauss-shaped emission probability

$$
\begin{aligned}
\pi_i &= \frac{(\eta_i - 1) + \gamma_1(i)}{\sum_{i=1}^{N}(\eta_i - 1) + 1} \\
a_{ij} &= \frac{(\eta_{ij} - 1) + \sum_t \xi_t(i,j)}{\sum_{i=1}^{N}(\eta_{ij} - 1) + \sum_t \gamma_t(i)} \\
\mu_i &= \frac{\tau_i \nu_i + \sum_t \gamma_t(i) x_t}{\tau_i + \sum_t \gamma_t(i)} \\
\sigma_i^2 &= \frac{2\beta_i + \tau_i(\nu_i - \mu_i)^2 + \sum_t \gamma_t(i)(x_t - \mu_i)^2}{(2\alpha_i - 1) + \sum_t \gamma_t(i)}
\end{aligned}
$$

Non-informative prior: $\eta_i = 1, \tau_i = 0,\ \nu_i = 0,\ \alpha_i = 1/2,\ \beta_i = 0$

Segmental MAP

$p(\theta, z|x) = \frac{p(z,\theta,x)}{p(x)} = \frac{p(z,x|\theta)\cdot p(\theta)}{p(x)}$

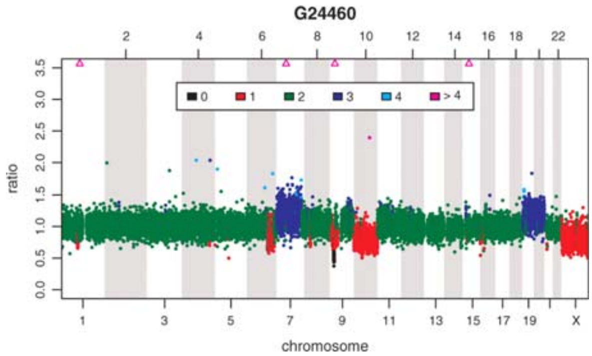
Find a θ that maximizes $p(\theta, z|x)$:

$$\theta = \underset{\theta}{\operatorname{argmax}} \max_z p(\theta, z|x) = \underset{\theta}{\operatorname{argmax}} \max_z p(x, z|\theta) \cdot p(\theta)$$

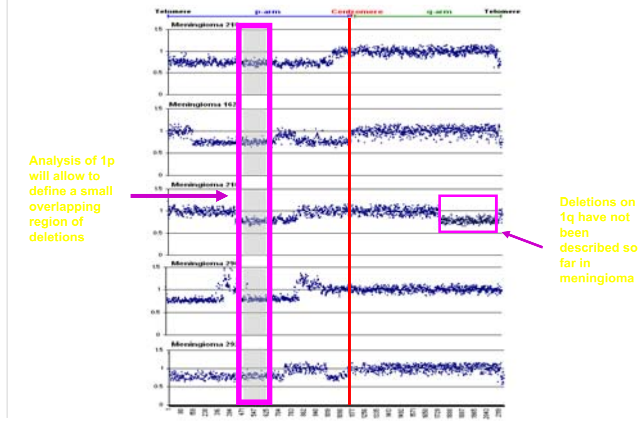
Alternate maximization over z and θ yields a sequence of non-decreasing $p(\theta, z|x)$:

$$
\begin{aligned}
z_{t+1} &= \underset{z}{\operatorname{argmax}}\, p(x, z|\theta_t) \qquad \text{Viterbi} \\
\theta_{t+1} &= \underset{\theta}{\operatorname{argmax}}\, p(x, z_{t+1}|\theta) \cdot p(\theta)
\end{aligned}
$$

SMAP - Result



Six meningiomas analyzed on chr. 1 array



Thanks