Generative Models and Stochastic Algorithms for Population Average Estimation and Image Analysis

Stéphanie Allassonnière

CIS, JHU

July, 15th 2008

Context and motivations :

* Describing shapes

* Shape matching : many elaborated registrations theories

* **Defining and infering population average (atlas)**
  - **Statistical estimation in presence of unobserved variables**
  - **Proper statistical model = generative**
  - **Consistency issues addressed**

* Discrimination/Classification

Outline

Three generative statistical models and stochastic algorithms

1. Bayesian Mixed Effect (BME) gray level Template
    - 1.1 Mathematical framework for deformable models
    - 1.2 Past approaches to compute a population average
    - 1.3 Generative statistical models
    - 1.4 Statistical estimation of the model parameters
    - 1.5 Experiments on USPS database and 2D medical images
2. Bayesian Mixed Effect DTI Template
3. Noisy ICA

Outline

Three generative statistical models and stochastic algorithms

Linearized deformations :

- Non rigid deformations
- Let ψ be the deformation from I_0 to I_1 and v a vector field :

$$\psi = Id + v$$

- Resulting deformed image given by :

$$I_1(x) \simeq I_0(x - v(x))$$

Advantages :

- Easy to use in computations
- Relevant for certain classes of shapes

Drawbacks :

- Invertibility not guaranteed : Overlap $\rightarrow$ Shape topology may change !

Some solutions previously given :

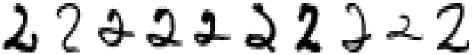
Only for the template !

- Using one of the data images y_1^n.

Some solutions previously given :

Only for the template !

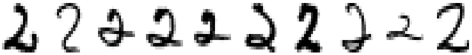
- Using one of the data images y_1^n.

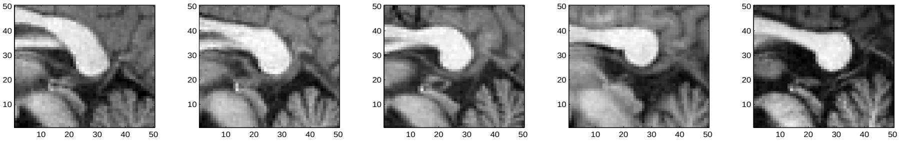


Some solutions previously given :

Only for the template !

- Using one of the data images y_1^n.





Some solutions previously given :

Only for the template !

- Procrustes' mean :

$$(\hat{I}_0, \hat{v}_1, \dots, \hat{v}_n) = \arg \min_{I_0, v_1, \dots, v_n} \sum_{i=1}^{n} \left(\frac{1}{2} \|v_i\|_V^2 + \frac{1}{2\sigma^2} |y_i \circ \phi^{v_i} - I_0|^2 \right)$$

Some solutions previously given :

Only for the template !

- Procrustes' mean :

$$(\hat{I}_0, \hat{v}_1, \dots, \hat{v}_n) = \arg \min_{I_0, v_1, \dots, v_n} \sum_{i=1}^{n} \left(\frac{1}{2} \|v_i\|_V^2 + \frac{1}{2\sigma^2} |y_i \circ \phi^{v_i} - I_0|^2 \right)$$

- A statistical interpretation : (Glasbey & Mardia)

$$y_i(x + v_i(x)) = I_0(x) + \epsilon_x, \;\; \epsilon_x \sim \mathcal{N}(0, \sigma^2), \;\; x \in \Lambda$$

Problems :

- Needs interpolation
- **Unobserved warping variables**
- Not a generative statistical model

Issues :

Describing databases as some i.i.d. sample of some parametric **generative** statistical model.

- Model parameters = template + deformation law (define the space V) + noise

Learning the parameters to avoid the previous problems :

- An intrinsic template I_0,
- A weighting term σ^2,
- A global geometric behavior in the class.

Generative Statistical Model :

→ Deformable template framework :

$$y^i_{j,k} = I_0(r_{j,k} - v_i(r_{j,k})) + \sigma\varepsilon_{j,k}$$

Conditions chosen for the template and the deformation fields :

Parametric Model of Splines : let $(p)_1^{k_p}$ and $(g)_1^{k_g}$ be two sets of control points (fixed, uniformly distributed) :

$$I_0(x) = K_p\alpha(x) = \sum_{k=1}^{k_p} K_p(x, p_k)\alpha(k) \ ,$$

$$v_\beta(x) = (K_g\beta)(x) = \sum_{k=1}^{k_g} K_g(x, g_k)\beta(k).$$

What to learn ?

$$\text{Photometry}: \begin{cases} \alpha & : \text{ to code the template} \\ \sigma^2 & : \text{ the noise variance} \end{cases}$$

$$\text{Geometry}: \quad \beta_1^n : \text{ to code each deformation}$$

BUT : does not give the geometrical behavior in the training set.

$\Longrightarrow$ Introduce a prior on β :

$$\beta_i \sim \nu(d\beta)$$

Parameters of ν = parameters to learn.

Deformations $(\beta_i)_{1 \leq i \leq n}$ = random hidden variables

Generative Model :

One component per class :

$$\begin{cases} (\Gamma_g, \theta_p) \sim \nu_g \otimes \nu_p \text{ with } \theta_p = (\alpha, \sigma^2) \\ \beta_1^n \sim \otimes_{i=1}^n \mathcal{N}(0, \Gamma_g) \mid \Gamma_g \\ y_1^n \sim \otimes_{i=1}^n \mathcal{N}(v_{\beta_i} I_\alpha, \sigma^2 \mathrm{Id}_\Lambda) \mid \beta_1^n, \theta_p \end{cases}$$

where $\nu_g(d\Gamma_g), \nu_p(d\sigma^2, d\alpha)$ are prior laws on the parameters.

Remark : big structures to learn even in the case of small size training set

Generative Model (2) :

General case : mixtures of deformable templates (τ_m components per class)

Hidden random variables : $(\beta_i)_{1 \leq i \leq n}$ and the image labels $(\tau_i)_{1 \leq i \leq n}$.

$$\begin{cases} \rho \sim \nu_\rho \\ \theta = (\theta_g^\tau, \theta_p^\tau)_{1 \leq \tau \leq \tau_m} \sim \otimes_{\tau=1}^{\tau_m} (\nu_g \otimes \nu_p) \\ \tau_1^n \sim \otimes_{i=1}^n \sum_{\tau=1}^{\tau_m} \rho_\tau \delta_\tau \mid \rho \\ \beta_1^n \sim \otimes_{i=1}^n \mathcal{N}(0, \Gamma_g^{\tau_i}) \mid \theta, \ \tau_1^n \\ y_1^n \sim \otimes_{i=1}^n \mathcal{N}(v_{\beta_i} I_{\alpha_{\tau_i}}, \sigma_{\tau_i}^2 Id_\Lambda) \mid \beta_1^n, \ \theta, \ \tau_1^n \end{cases}$$

Generative Statistical Model :

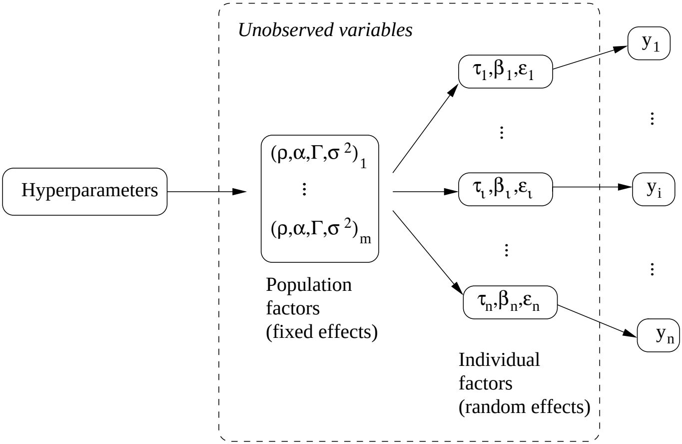


FIG.: Mixed effect structure for our BME-template

How to learn the parameters ? the MAP Estimator :

Parameters θ are estimated by maximum posterior likelihood :

$$\hat{\theta} = \arg\max P(\theta|y)$$

where
$\theta \in \Theta = \{ (\alpha, \sigma^2, \Gamma_g) | \alpha \in \mathbb{R}^{k_p},\ \sigma^2 > 0,\ \Gamma_g \in \mathcal{S}ym^+_{2k_g,*}(\mathbb{R}) \}$.
$\mathcal{S}ym^+_{2k_g,*}(\mathbb{R})$ is the set of positive definite symmetric matrices.

Let $\Theta_* = \{ \theta_* \in \Theta \mid E_P(\log q(y|\theta_*)) = \sup_{\theta\in\Theta} E_P(\log q(y|\theta))\}$
where P denotes the distribution governing the observations.

How to do in practice ?

Since β_1^n are unobserved variables, a natural approach to reach the MAP estimator is the **EM algorithm**.

Iteration l of the algorithm :
E Step : Compute the posterior law on $\beta_i, i = 1, \ldots, n$.

M Step : Parameter update :

$$\theta_{l+1} = \arg\max_{\theta} E\left[\log q(\theta, \beta_1^n, y_1^n) | y_1^n, \theta_l\right].$$

BUT : the E step is not tractable !

Details of the maximization step :

Geometry :

$$\theta_{g,l+1} = \Gamma_{g,l+1} = \frac{1}{n + a_g}(n[\beta\beta^t]_l + a_g \Sigma_g).$$

where

$$[\beta\beta^t]_l = \frac{1}{n}\sum_{i=1}^{n} \int \beta\beta^t \nu_{l,i}(\beta) d\beta,$$

is the empirical covariance matrix with respect to the posterior density function.

$\longrightarrow$ Importance of the prior !

Solution proposed : Stochastic version of the EM algorithm :

Idea : Couple SAEM with MCMC procedure (Delyon, Lavielle, Moulines and Kuhn, Lavielle) :

One component case : Iteration $l \to l+1$ of the algorithm :

- Simulation step : $\beta^{l+1} \sim \Pi_{\theta_l}(\beta^l, \cdot)$
  where $\Pi_{\theta_l}(\beta^l, \cdot)$ is a transition probability of a convergent Markov Chain having the posterior distribution as stationary distribution,
- Stochastic approximation :
  $Q_{l+1}(\theta) = Q_l(\theta) + \Delta_l[\log q(y, \beta^{l+1}, \theta) - Q_l(\theta)]$ where (Δ_l) is a decreasing sequence of positive step-sizes.
- Maximization step : $\theta_{l+1} = \arg\max Q_{l+1}(\theta)$

[$*$] $\Pi_{\theta_l}(\beta^l, \cdot)$ given by an hybrid Gibbs sampler

Stochastic version of the EM algorithm (2) :

Since our model is an Exponential Model,

$$q(y, \beta^{l+1}, \theta) = \exp\{-\psi(\theta) + \langle S(y, \beta), \phi(\theta)\rangle\}$$

the stochastic approximation can be done on the sufficient statistics S so that the algorithm is done via :

$$s_{l+1} = s_l + \Delta_l \left(S(y, \beta^{l+1}) - s_l \right)$$

Let $L(s, \theta) = -\psi(\theta) + \langle s, \phi(\theta)\rangle$, $l(\theta) = \log q(y, \theta)$ and $\hat{\theta}(s) \; = \arg\max_s L(s, \theta(s))$ then

$$\theta_{k+1} = \hat{\theta}(s_{l+1}) \; .$$

Stochastic approximation with truncation on random boundaries :

Set $\kappa_0 = 0$, $s_0 \in \mathcal{K}_0$ and $\beta_0 \in \mathrm{K}$.

$\forall k \geq 1$ compute $\bar{s} = s_{k-1} + \Delta_{k-1}(S(\bar{\beta}) - s_{k-1})$
where $\bar{\beta}$ is sampled from a transition kernel $\Pi_{\theta_{k-1}}(\beta_{k-1}, .)$.

If $\bar{s} \in \mathcal{K}_{\kappa_{k-1}}$ and $|\bar{s} - s_{k-1}| \leq \varepsilon_{k-1}$
set $(s_k, \beta_k) = (\bar{s}, \bar{\beta})$ and $\kappa_k = \kappa_{k-1}$,

else set $(s_k, \beta_k) = (\tilde{s}, \tilde{\beta}) \in \mathcal{K}_0 x \mathrm{K}$ and $\kappa_k = \kappa_{k-1} + 1$, where $(\tilde{s}, \tilde{\beta})$ can be chosen through different ways.

$\theta_k = \arg\max_{\theta} L(s_k, \theta)$

Stochastic version of the EM algorithm for the multicomponent model :

Intuitive generalization :

- problem of "trapping states".
- Image analysis interpretation : each iteration tries to deform the data so it is closer to its current component and will not tend to move toward another one
  $\longrightarrow$ high dimensional hidden variable β

Solution proposed : Consider another simulation method based on the Gibbs sampler for the deformation and on another law for the class of a given image.

The new algorithm :

Transition step $l \rightarrow l+1$ using a hybrid Gibbs sampler on (β, τ) :

- for each τ : Run N_l times the hybrid Gibbs Sampler on β given τ.

$$\hat{\beta}_{\tau}^{(l+1)} = \Pi^{N_l}(\beta|\tau)$$

- draw $\tau^{(l+1)}$ through the discrete law with weights :

$$p_{N_l}(\tau) \propto \left(\frac{1}{N} \sum_{i_{mc}=1}^{N_l} \left[\frac{f(\hat{\beta}_{\tau,(i_{mc})})}{q(y, \hat{\beta}_{\tau,(i_{mc})}, \tau|\theta, \rho)} \right] \right)^{-1}$$

- $$\beta^{(l+1)} = \hat{\beta}_{\tau^{(l+1)},(N_l)}$$

Theoretical Results :

With these models and algorithms we have proved some important asymptotic results :

- Consistency of the MAP estimator
- Convergence of both stochastic algorithms

MCMC-SAEM :

Template estimation :

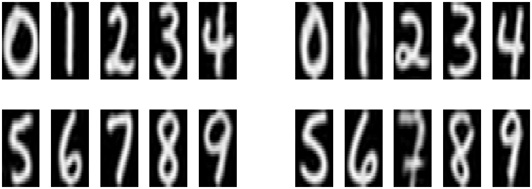


Fig.: Left : one component prototype. Right : 2 component prototypes

MCMC-SAEM :

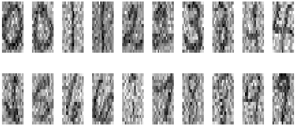
The Geometric Distribution :

FIG.: Left : 20 examples of the training set. Right : 20 examples drawn from the prior geometric distribution.

MCMC-SAEM :

Geometry :

In presence of noise :



Medical Images : Splenium of the Corpus Callossum

47 images of the corpus callosum (and part of the cerebellum)

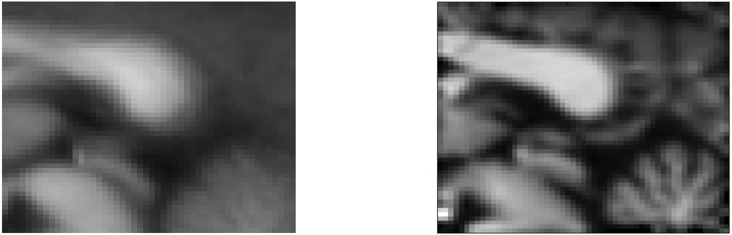


FIG.: Left : Gray level mean of the 47 images. Right : template estimated with the stochastic algorithm.

Medical Images : Splenium of the Corpus Callossum

47 images of the corpus callosum (and part of the cerebellum) clustered into 2 components by the multicomponent model.

FIG.: Results from the estimation with the stochastic algorithm. Left : component 1. Right : component 2.

Robustness of the algorithm :

Same hyper-parameters as the previous gray level images

Outline

Three generative statistical models and stochastic algorithms

Bayesian Mixed Effect DTI Template :

Goals of our approach :

* Estimate a Template of Diffusion Tensor Image on a given region of the anatomy
* Just use the Diffusion Weight Images [DWIs] (real vector corresponding to the response to some different gradients)

Previous approach

Subjects	Least square approx. Min of energy		mean	
Subject 1	DWI $G1$ ... DWI Gm	$\rightarrow$	DTI 1	$\rightarrow$
Subject 2	DWI $G1$ ... DWI Gm	$\rightarrow$	DTI 2	$\rightarrow$
...				
Subject n	DWI $G1$ ... DWI Gm	$\rightarrow$	DTI n	$\rightarrow$

DTI template

Our approach

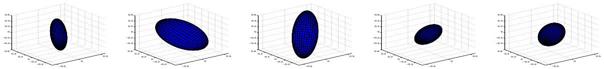
Subjects	observed	(hidden)	Max likelihood	
Subject 1	DWI $G1$		$\longrightarrow$	
	...	(DTI 1)	$\longrightarrow$	
	DWI Gm		$\longrightarrow$	
Subject 2	DWI $G1$		$\longrightarrow$	
	...	(DTI 2)	$\longrightarrow$	DTI template
	DWI Gm		$\longrightarrow$	
...				
Subject n	DWI $G1$		$\longrightarrow$	
	...	(DTI n)	$\longrightarrow$	
	DWI Gm		$\longrightarrow$	

Results on synthetic data :

- 2 different template tensors

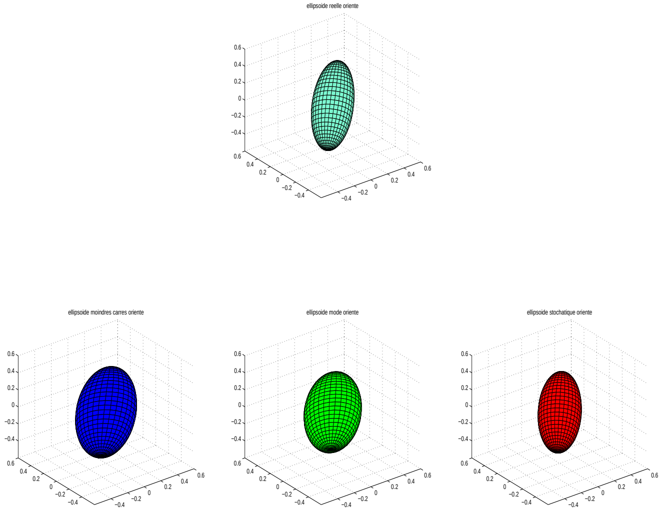
$$FA_1 = 0.6791, \quad FA_2 = 0.6918$$
$$ADC_1 = 0.5038, \quad ADC_2 = 0.4329$$

- 50 random samples with 15 subjects each



	LS	FAM-EM.	SAEM	LS	FAM-EM	SAEM
bias	0.1960	0.7386	0.1409	0.2095	0.5655	0.1269
var	0.7033	0.3489	0.7351	0.4853	0.2648	0.4969
mse	0.8993	1.0875	0.8760	0.6949	0.8303	0.6238
FA	0.6534	0.6193	0.6683	0.6683	0.6332	0.6814
ADC	0.5752	0.5627	0.5460	0.5044	0.4932	0.4747

BME-DTI Template



Outline

Three generative statistical models and stochastic algorithms

Noisy ICA Model

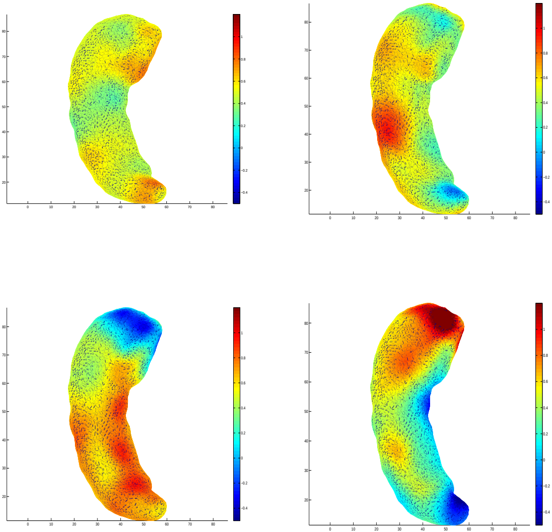
- Observations : X_1^n such as $X_i = AY_i + \sigma\varepsilon$,
- A : source matrix
- σ^2 : variance of the Gaussian noise
- Y_1^n hidden variables.
- Model :

$$\begin{cases} Y_{1,1}^{n,p} & \sim \quad \otimes_{i=1}^{n} \otimes_{j=1}^{p} \nu_\eta \mid \eta \,, \\ X_1^n & \sim \quad \otimes_{i=1}^{N} \mathcal{N}(AY_i, \sigma^2 Id) \mid A,\ \sigma^2,\ Y_1^N \,. \end{cases}$$

- Various choice of the distribution ν_η

Same MCMC-SAEM algorithm to treat this estimation problem.

Experiments : 101 subjects, 20 I.C.



Conclusion

- Generative statistical model = proper statistical framework for designing and inferring population average

- Stochastic algorithm of multiple uses even in difficult conditions

Thank you !

MCMC-SAEM :

The Geometric Distribution :

Between 2 classes :

Between 2 components in the same class :