

Proposte di tesi di Statistica Bayesiana

Alessandra Guglielmi

13 gennaio 2016



Recurrent events

Eventi ricorrenti relativi a più pazienti: scompenso cardiaco, epilessia, infezioni ricorrenti

$0 < T_1 \leq T_2 \leq \dots \leq T_J$ per ogni paziente

Ci interessa l'approccio attraverso i **gap times** $W_j := T_j - T_{j-1}$

Se $\{W_j\}$ sono (condiz) iid, stiamo assumendo un **renewal process**;

ci interessa invece costruire dei modelli per i gap times, indebolendo l'ipotesi di renewal process.

Dati su Heart Failure, collaborazione con dott.ssa leva e prof.ssa Paganoni



Modelli bayesiani nonparametrici per modelli ricorrenti

Condizionatamente ai parametri e alle covariate:

$$\mathcal{L}(W_1, \dots, W_J) = \mathcal{L}(W_1) \times \mathcal{L}(W_2|W_1) \times \dots \times \mathcal{L}(W_J|W_1, \dots, W_{J-1})$$

e ogni $\mathcal{L}(W_J|\mathbf{x}, W_1, \dots, W_{J-1}, \beta)$ è caratterizzata da una distribuzione con **ingredienti nonparametrici**



Modelli bayesiani nonparametrici per modelli ricorrenti

Condizionatamente ai parametri e alle covariate:

$$\mathcal{L}(W_1, \dots, W_J) = \mathcal{L}(W_1) \times \mathcal{L}(W_2|W_1) \times \dots \times \mathcal{L}(W_J|W_1, \dots, W_{J-1})$$

e ogni $\mathcal{L}(W_J|\mathbf{x}, W_1, \dots, W_{J-1}, \beta)$ è caratterizzata da una distribuzione con **ingredienti nonparametrici**

Problemi di regressione con dati longitudinali e modelli bayesiani nonparametrici !!



Modelli bayesiani nonparametrici per modelli ricorrenti

Condizionatamente ai parametri e alle covariate:

$$\mathcal{L}(W_1, \dots, W_J) = \mathcal{L}(W_1) \times \mathcal{L}(W_2|W_1) \times \dots \times \mathcal{L}(W_J|W_1, \dots, W_{J-1})$$

e ogni $\mathcal{L}(W_J|\mathbf{x}, W_1, \dots, W_{J-1}, \beta)$ è caratterizzata da una distribuzione con **ingredienti nonparametrici**

Problemi di regressione con dati longitudinali e modelli bayesiani nonparametrici !!

Un possibile modello:

$$Y_{ij} := \log W_{ij} \text{ e } Y_{ij} = x_i \beta_j + \alpha_{ij1} Y_{ij-1} + \dots + \alpha_{ijj-1} Y_{i1} + \alpha_{ijj} + \sigma \epsilon_{ij}$$

con $\epsilon_{ij} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$

$$\alpha_{ij} = (\alpha_{ij1}, \dots, \alpha_{ijj}) | G_j \stackrel{\text{ind}}{\sim} G_j$$

$$G_1, \dots, G_J \sim \text{prior nonparametrica}$$



Clustering con Bayesian Principal Curves

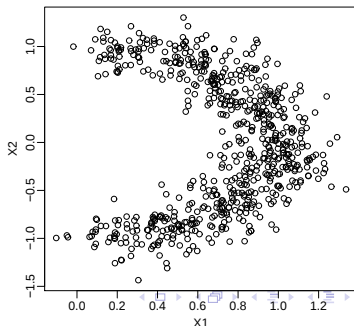
introdurre un modello BNP (mixture di opportuni nuclei di densità in cui la *mixing distribution* è una misura di probabilità aleatoria) in cui il nucleo $f(\cdot)$ sia in grado di rappresentare dati che provengono da **cluster con forma non-standard !**



Clustering con Bayesian Principal Curves

introdurre un modello BNP (mixture di opportuni nuclei di densità in cui la *mixing distribution* è una misura di probabilità aleatoria) in cui il nucleo $f(\cdot)$ sia in grado di rappresentare dati che provengono da **cluster con forma non-standard !**

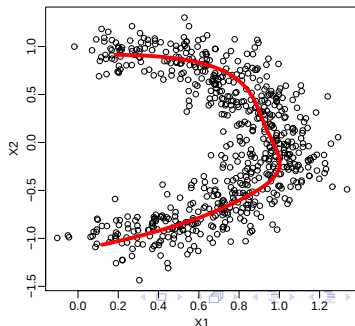
represent data in the same *curved cluster* using one single distribution $f(\cdot)$



Clustering con Bayesian Principal Curves

introdurre un modello BNP (mixture di opportuni nuclei di densità in cui la *mixing distribution* è una misura di probabilità aleatoria) in cui il nucleo $f(\cdot)$ sia in grado di rappresentare dati che provengono da **cluster con forma non-standard !**

represent data in the same *curved cluster* using one single distribution $f(\cdot)$ centered around a curve $m(\cdot)$; the curve here is a **low-rank thin plate spline**



Clustering con Bayesian Principal Curves

Prior 1

- ✓ number of cluster

$$\alpha = 0.137 \Rightarrow \mathbb{E}(J) = 2$$

- ✓ mixed effect parameter

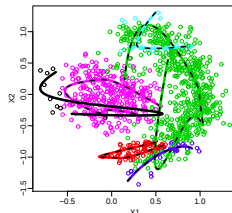
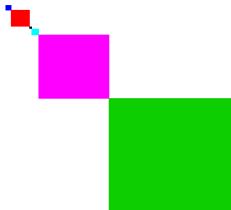
$$\sigma_u, \sigma_v \sim \text{Inv-gamma}(2, 0.01)$$

- ✓ error variance

$$\sigma_1^2, \sigma_2^2 \sim \text{Inv-gamma}(2, 0.01)$$

$$\mathbb{E}(\sigma_f^2) = 0.01, \text{Var}(\sigma_f^2) = +\infty$$

Incidence Matrix



Prior 2

- ✓ number of cluster

$$\alpha = 0.137 \Rightarrow \mathbb{E}(J) = 2$$

- ✓ mixed effect parameter

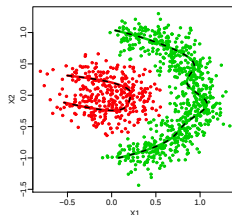
$$\sigma_u, \sigma_v \sim \text{Inv-gamma}(2, 0.01)$$

- ✓ error variance

$$\sigma_1^2, \sigma_2^2 \sim \text{Inv-gamma}(20, 2)$$

$$\mathbb{E}(\sigma_f^2) \simeq 0.1, \text{Var}(\sigma_f^2) \simeq 0.0006$$

Incidence Matrix



Modelli per dati longitudinali - glaucoma patients

Collaborazione con la prof.ssa Giovanna Guidoboni, IUPUI (Indiana University and Purdue University at Indianapolis), USA e i suoi collaboratori (tra cui il nostro prof. Riccardo Sacco) per stabilire se esiste, e può essere determinata, una **progressione della malattia (glaucoma)** in pazienti malati.

Si tratta di cercare opportuni modelli bayesiani per dati longitudinali che abbiano un buon adattamento (predittivo) ai dati in nostro possesso.

