MS 09/10/24

MODELLI BAYESIANI COMPLESSI

Le ultime due volte:

$\theta = \mu$

(A) $$\left.\begin{aligned} &X_1, \dots, X_n \overset{c.i.i.d}{\sim} N(\mu, \sigma^2), \quad \sigma^2 \text{ noto} \\ &\mu \sim N(\mu_0, \sigma_0^2) \end{aligned}\right\} \text{CONIUGATO}$$

parametri incogniti

e

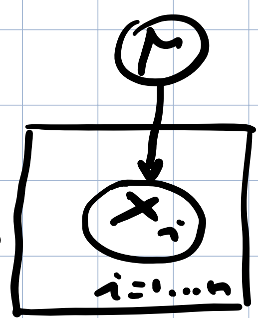
$\theta = p$

(B) $$\left.\begin{aligned} &X_1, \dots, X_n \overset{c.i.i.d}{\sim} \text{Bernoulli}(p) \\ &p \sim \text{Beta}(a, b) \end{aligned}\right\} \text{CONIUGATO}$$

In termini di reti bayesiane:

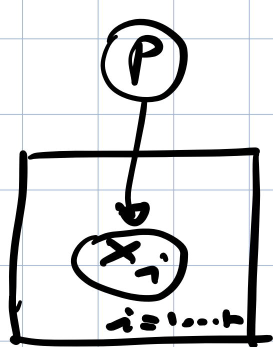
(A)

normale coniugato



(B)

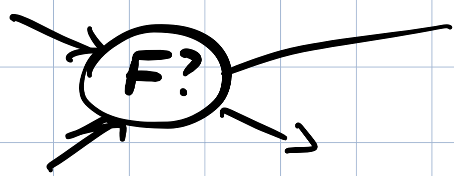
binario coniugato



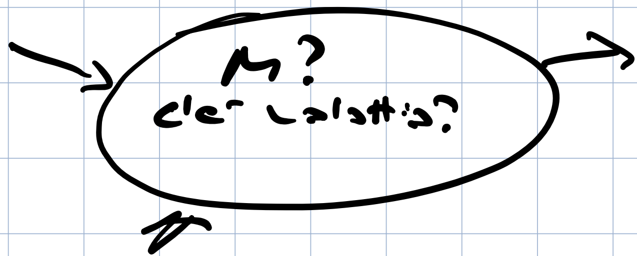
Il fine ultimo è calcolare l'a posteriori tramite l'aggiornamento bayesiano.

D'altra parte, nelle reti bayesiane elementari viste nel primo giorno, un nodo può rappresentare il parametro incognito.

parametro binario
$\theta = F?$



c'è stata frode?
(cybersecurity, statistica forense)



diagnosi

La differenza tecnica con A e B è che in A, B ci sono nodi continui.

Diverse generalizzazioni:

1) nodi delle reti continui o discreti n-ari (non necessariamente binari)

2) strumenti di calcolo per a posteriori nel caso non coniugato

3) modularità = costruire una rete pezzo per pezzo, magari con componenti temporali o gerarchiche

Spieghiamo meglio:

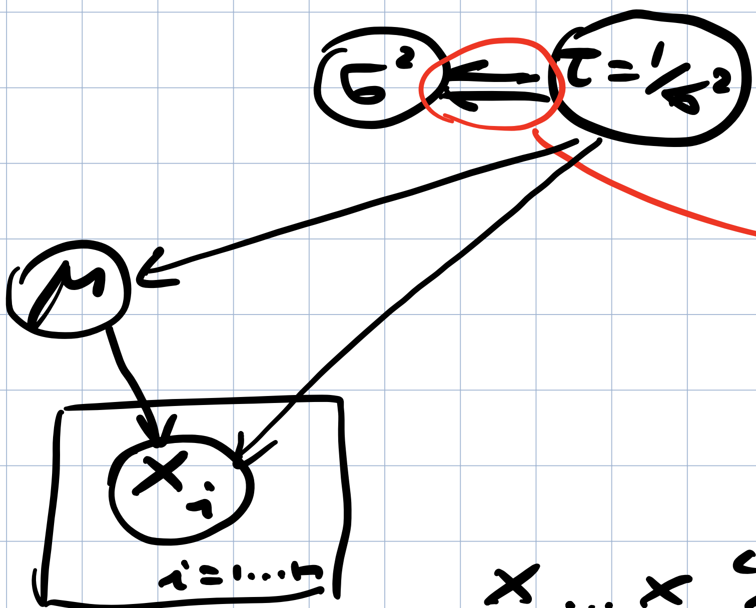
1) dobbiamo studiare variabili qualitative più che binarie (oggi)

1) dobbiamo studiare vettori aleatori, per esempio gaussiani (oggi)

2)

ESEMPIO ANCORA CONIUGATO EVOLUZIONE DI (A):

caso della varianza non nota, ricordando che $\tau = 1/\sigma^2$ è la parametrizzazione preferita



$$\theta = \begin{pmatrix} \mu \\ \tau \end{pmatrix}$$

collegamento deterministico (riparametrizzazione)

$$X_1, \dots, X_n \overset{c.i.i.d.}{\sim} N(\mu, \tau^{-1})$$

$$\mu | \tau \sim N(\mu_0, (\tau \tau_0)^{-1})$$

$$\tau \sim Gamma(a, \lambda)$$

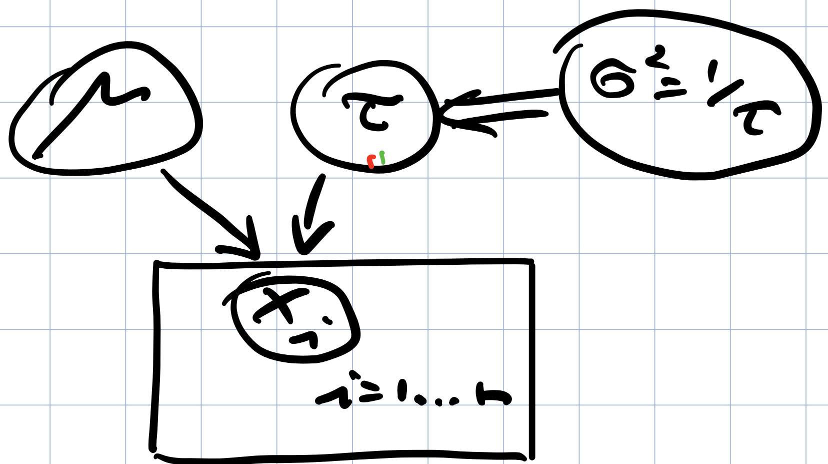
allora si dimostra (esercizio, esempio 6.4 in Gasparini (2013)) che tale scelta è coniugata in quanto, a posteriori

$$\mu \mid \tau, \overbrace{x_1 \ldots x_n}^{\text{dati}} \sim \mathcal{N}\left(\frac{\tau_0\mu_0 + n\bar{x}}{\tau_0+n}, \left(\tau(\tau_0+n)\right)^{-1}\right)$$

$$\tau \mid x_1 \ldots x_n \sim \text{Gamma}\left(\alpha + \frac{n}{2}, \lambda + \frac{\sum (x_i-\bar{x})^2}{2} + \frac{n\tau_0(\bar{x}-\mu_0)^2}{2(\tau_0+n)}\right)$$

che ha la stessa struttura dell'a priori.
Ma se non ci piacesse questa a priori e volessimo dare una distribuzione a priori in cui μ e τ sono indipendenti?
Possiamo farlo, ma non abbiamo più la coniugatezza.

ESEMPIO NON CONIUGATO EVOLUZIONE DI (A)



$$\theta = \begin{pmatrix} \mu \\ \tau \end{pmatrix}$$

$$X_1 \ldots X_n \overset{\text{ciid}}{\sim} \mathcal{N}(\mu, 1/\tau)$$

$$\mu \sim \mathcal{N}(\mu_0, \tau_0)$$

$$\tau \sim \text{Gamma}(a, \lambda)$$

allora la densità a posteriori è

$$\pi(\mu,\tau \mid x_1 \ldots x_n) = \frac{\overbrace{\tau^{n/2} \exp\left\{-\frac{\tau}{2}\sum(x_i-\mu)^2\right\}}^{\text{verosimiglianza}} \overbrace{\tau^{a-1} e^{-\lambda\tau}}^{\text{a priori su } \tau} \overbrace{\exp\left\{-\frac{\tau_0}{2}(\mu-\mu_0)^2\right\}}^{\text{a priori su } \mu}}{\int_{t=0}^{+\infty}\int_{m=-\infty}^{\infty} t^{n/2} \exp\left\{-\frac{t}{2}\sum(x_i-m)^2\right\} t^{a-1} e^{-\lambda t} \exp\left\{-\frac{\tau_0}{2}(m-\mu_0)^2\right\} dm\, dt}$$

rimane una densità propria su $\mathbb{R}^2$
senza identificazione particolare e
senza l'indipendenza di μ e τ a posteriori.

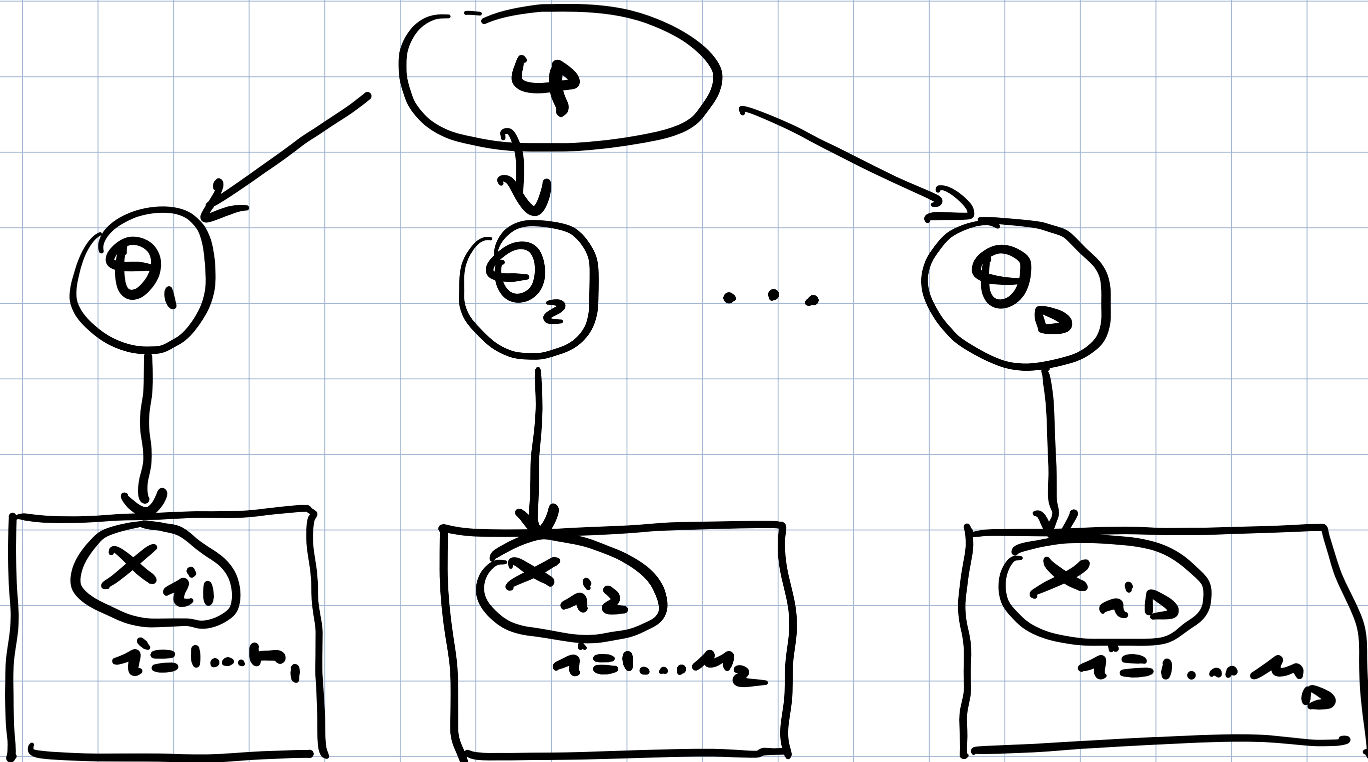
2) problemi di calcolo, per esempio nel caso non coniugato ora visto.

Soluzione:
simulare la a posteriori, spesso tramite metodi:
MCMC (Markov Chain Monte Carlo)
oppure altre approssimazioni (per esempio inferenza variazionale).

In questo corso vedremo software (JAGS oppure pgmpy oppure Bayes Server / HUGIN)
mentre approfondimenti sono nel corso STATISTICA COMPUTAZIONALE

3) modularità: importanti strutture possono essere aggiunte alle reti bayesiane, per esempio
3.1 gerarchica
3.2 temporale

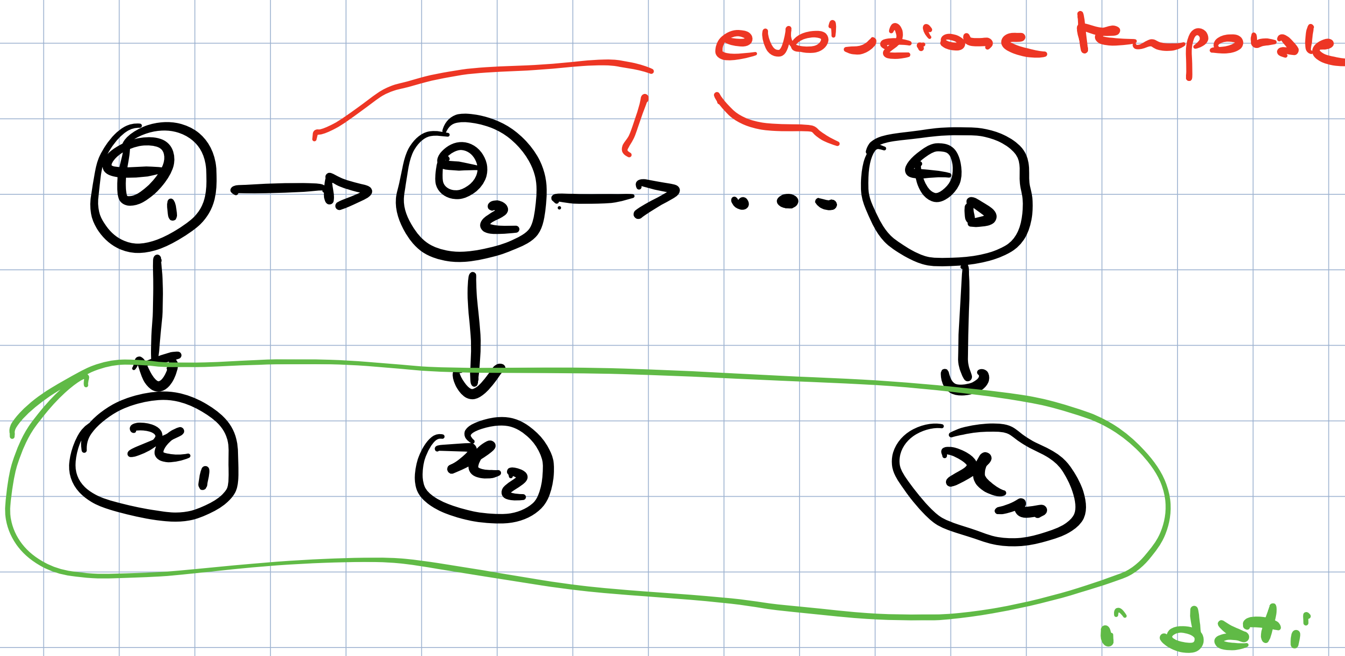
Per dare un'idea di 3.1:



per esempio, D scuole, ciascuna con il suo parametro di qualità θ_j e ciascuna con i voti dei suoi studenti: θ_{ij} $i=1\ldots n_j$

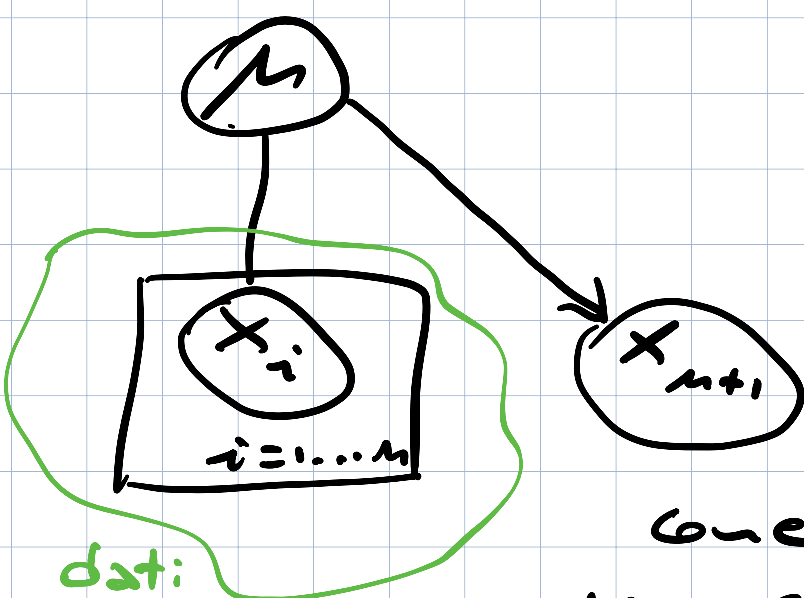
Le scuole hanno in comune una distribuzione dei parametri di secondo livello caratterizzata da un iperparametro φ.

Per dare un'idea di 3.2:



Struttura HMM (Hidden Markov Model)

Discorso simile per la previsione.
Prendiamo un esempio semplice
evoluzione di A



non essendo x_{n+1} ancora osservato (a differenza di $x_1 \ldots x_n$), si comporta come se fosse un parametro.

Se la struttura e' semplice come nel DAG, allora si può aggiornare la distribuzione su μ e usare questa per aggiornare la distribuzione di x_{n+1}.

→ ESEMPI in laboratorio 17/10/24