Contents

1 Notazione

(il punto e virgola `';'` vuol dire `Parametrizzato Da`)

$$P(x;\theta)$$

PARAMETRIZZATO DA

θ INCOGNITA, MA NON ALEATORIA

COSTANTE

INCOGNITA

puoi anche farlo con

$$P(x|\theta)$$

SCRITTO ANCHE COSÌ

MA QUI x È UNA

VARIABILE ALEATORIA?

siccome è esso stesso[1] una variabile aleatoria

2 Lezione

abemus funzione di verosmiglianza

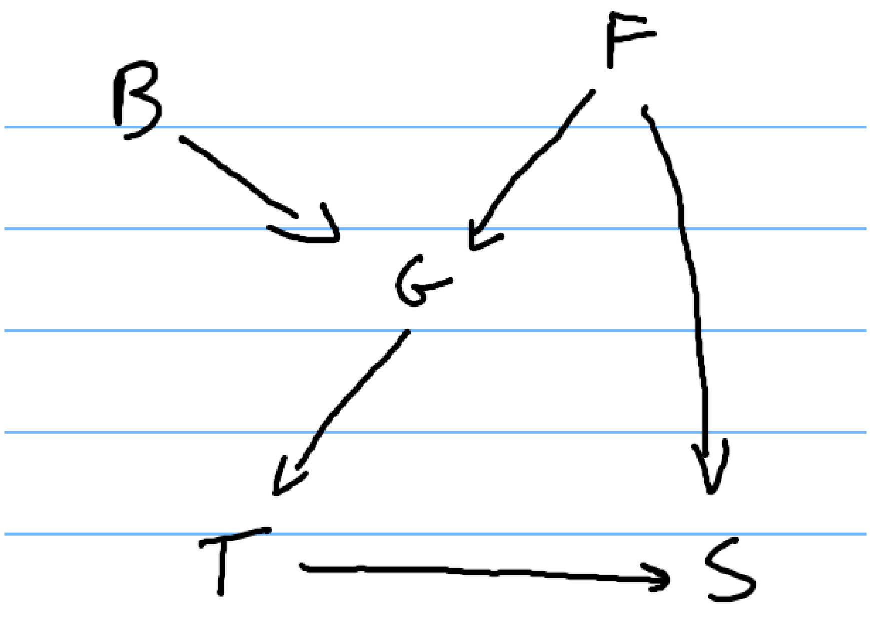
[1] ~~il piacere~~

$$\mathcal{L}(\theta) = \mathbb{P}(D;\theta) = \prod_{i=1}^{n} \mathbb{P}(x^{(i)};\theta)$$

TurnOver	Fuel	Start	Battery	Gauge
tante	belle	cose	brutto	stronzo

abemus un caso in cui non parte la macchina, con una cosa del genere, è comunque un applicazoine della diagnostica

il grafo del cazzo ha questo schema



TurnOver	Fuel	Start	Battery	Gauge
Normal	Yes	Yes	Yes	Yes
Slow	Yes	Yes	Yes	No

in questo casi si assume che siano tutte osservate un caso del genere non può verificarsi

TurnOver	Fuel	Start	Battery	Gauge
Slow	No	?	Yes	No

riprendendo la formula di sopra

$$\mathcal{L}(\theta) = \mathbb{P}(D;\theta) = \prod_{i=1}^{n} \mathbb{P}(x^{(i)};\theta) = \prod_{i=i}^{n} \prod_{c=1}^{N} \mathbb{P}(x_c^{(i)} | \mathbb{P}_a(x_c^{(i)};\theta))$$

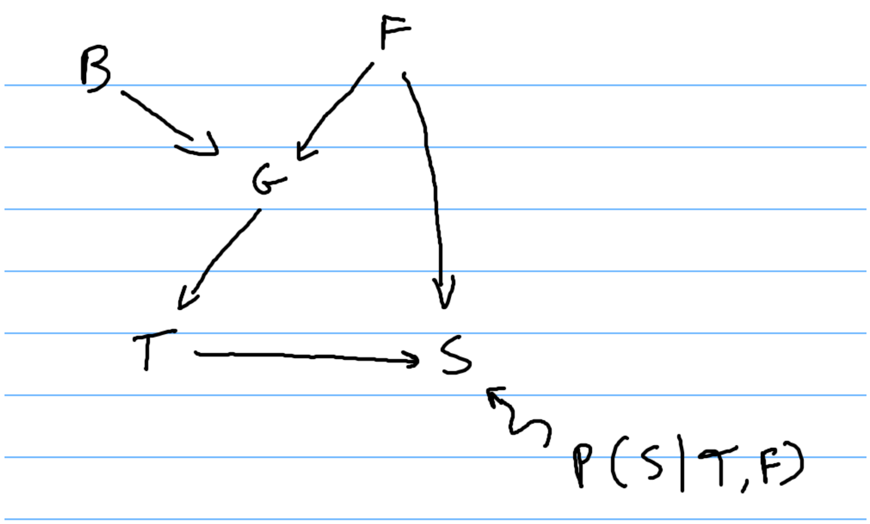
(dove c è l'indice della colonna)

la cosa bella adesso è che posso scambiare l'ordine delle produttorie

$$\begin{aligned} & \prod_{i=i}^{n} \prod_{c=1}^{N} \mathbb{P}(x_c^{(i)} | \mathbb{P}_a(x_c^{(i)};\theta)) \\ = & \prod_{c=1}^{N} (\prod_{i=i}^{n} \mathbb{P}(x_c^{(i)} | \mathbb{P}_a(x_c^{(i)};\theta))) \\ = & \prod_{c=1}^{N} (\mathcal{L}_c(\theta)) \end{aligned}$$

quindi adesso posso risovere tutti i problemi di likelihood separatamente i problemi di massima verosimiglianza non comunicano tra di loro

per risolvere ad esempio il problema di massima verosimiglianza per S



allora basta prendere la tabella, proiettarla sulle variabili S, T, V, e risolvere il problema di massima verosimiglianza localmente

la stima di massima verosimiglianza

$$\hat{O}_{cjk} = \frac{N_{cjk}}{\sum_r N_{cjr}}$$

dove le N_{cjk} sono le cosiddette *statistiche sufficienti* visto che bastano, sono *sufficienti*, per stimare i parametri del modello

è come una relazione d'indipendenza condizionali, i parametri diventano condizionalmente indipendenti dal dataset se conosco le statistiche sufficienti

3 Se ho `Maybe`

se ho `Maybe` i problemi smettono di essere indipendenti quando invece ho dei dati `?` la cosa diventa un problema, visto che i problemi cominciano a comunicare l'uno con l'altro

diventa un problema che la probabilità del dataset diventa una variabile aleatoria visto che ho dei pezzi non conosciuti

si utilizza un cosiddetto *expectation maximization* algoritmo facile da spiegare, difficile da giustificare[2]

siano

x : i dati osservati

z : i dati non osservati

allora

$$\mathcal{L}(\theta) = \mathbb{P}(x, z; \theta)$$

ci s'ha da gestire il prodotto di tutti, visto che siamo stronzi famo il logaritmo per avere la somma, quindi

$$l(\theta) = \log \sum_z \mathbb{P}(x, z; \theta)$$

ok, abbiamo una somma di funzioni convese, somma questa non è necessariamente convessa, quindi qui il problema di ottimizzazione globale diventa difficile

allora

1. Likelihood non è convessa
2. Ricerca di un massimo locale, si usa Expectation Maximization
3. Repeat

[2]il link non c'entra un cazzo, era per flexare che sapevo esistesse questa cosa

inizializza a cazzo

- si riempono i dati mancanti coi parametri presi dall' inverenza
- calcola le Expected Sufficient Statistics
- Aggiorna i parametri usando le Expected Sufficient Statistics

quindi

- si fa un'expectation
- si fa una maximization
- repeat