

Cosa accadrà domani?

Max Pierini

[*info@maxpierini.it](mailto:info@maxpierini.it)

June 4, 2020

Che tempo farà domani? Come andranno i miei investimenti? Quanti nuovi casi di COVID-19 in Italia ci dobbiamo aspettare nelle prossime settimane?

Le scienze statistiche cercano, attraverso modelli e calcoli più o meno complessi, di rispondere a domande come queste.

Uno dei metodi utilizzati, forse tra i più affascinanti (e sempre più usati), è l'**inferenza bayesiana**. Nonostante le sue basi affondino fino alla metà del '700, è stata applicata estesamente solo grazie all'uso di computer sempre più potenti, in grado di svolgere in pochi minuti calcoli (spesso su matrici multidimensionali) che sarebbero quasi impossibili da fare a mano. Minuti, certo, e un minuto è un'eternità per un computer. Per renderci conto della proporzione, immaginiamo di dover scrivere le "tabelline" delle moltiplicazioni di tutti i numeri da 1 a 100. Ci metteremmo delle ore usando decine di fogli. Uno smartphone di ultima generazione ci può impiegare 60 milionesimi di secondo nel palmo di una mano (figura 1).

```
In [1]:  
  
import numpy as np  
import datetime as dt  
a = np.arange(1, 101, 1)  
iter = 100000  
res = []  
for _ in range(iter):  
    i = dt.datetime.now()  
    a * a[:, None]  
    z = dt.datetime.now()  
    res.append((z-i).microseconds)  
  
print(f"uSec: {np.mean(res).round(2)}")  
  
uSec: 57.78
```

Figure 1: Esempio di calcolo su uno smartphone (con Jupyter Notebook). Questo semplice script calcola 100'000 volte tutte le moltiplicazioni a due a due di tutti i numeri da 1 a 100 e ritorna il tempo medio di esecuzione in microsecondi (milionesimi di secondo).

1 La strada bagnata

Ci alziamo una mattina di sole e scopriamo che il tratto di strada visibile dalla nostra finestra è bagnato, ma il giorno passato non ha piovuto. Quante probabilità ci sono che abbia piovuto mentre dormivamo la scorsa notte?

Per rispondere a questa domanda possiamo affidarci alla *statistica inferenziale bayesiana* il cui scopo è determinare, con teoremi e formule più o meno complesse, di rispondere a domande simili a questa nei campi più disparati.

La categoria di “domanda” che abbiamo posto è chiamata probabilità **a posteriori** o **condizionata**: ci chiediamo infatti quale sia la probabilità dell’evento “pioggia la scorsa notte” (che chiameremo parametro θ) condizionata dal, ovvero a posteriori del, fatto noto “tratto di strada bagnato” (che chiameremo variabile X). Scriveremo

$$P(\theta|X) = ?$$

e leggeremo “probabilità di θ dato X ”. Il parametro θ rappresenta la *verità* che non conosciamo e che vogliamo scoprire (ha piovuto la scorsa notte?) e la variabile X ciò che abbiamo osservato (la strada è bagnata).

Le probabilità vanno da 0 a 1, ovvero da 0% a 100%. Sarà questo pertanto il tipo di risultato che otterremo.

Analizzando il problema, ci renderemo facilmente conto del fatto che la risposta dipende da una serie di fattori. Vediamoli uno per uno:

1. che probabilità c’è che una strada si bagni se piove?
2. che probabilità c’è, in generale, che piova una sola notte?
3. che probabilità c’è, in generale, che una strada si bagni?

La domanda 1. è ancora una probabilità condizionata (o a posteriori) ma è l’opposto della domanda iniziale: ci chiediamo infatti che probabilità ha una strada di bagnarsi a condizione (o a posteriori del fatto) che piova. Senza entrare troppo nel dettaglio, possiamo facilmente stabilire che sia piuttosto elevata: per non bagnarsi dovrebbe essere “coperta”, ma allora ci troveremmo di fronte ad esempio a un porticato e non più ad un tratto di strada, oppure dovrebbe essere un tratto molto ristretto ad esempio coperto da un’auto parcheggiata che poi si è spostata, ecc (vedremo più avanti che in realtà il tipo di “domanda” posta in questo caso non è esattamente uguale alla precedente).

Assegniamo pertanto a questa probabilità un valore prossimo al 100%, per esempio 99%. La probabilità a posteriori della variabile X (la strada è bagnata?) dato lo specifico valore del parametro θ (ha piovuto di notte) sarà

$$P(X|\theta) = 99.00\%$$

che leggeremo “probabilità di X dato θ ”.

La domanda 2. è invece una probabilità non condizionata ovvero **a priori**: ci si chiede qual’è la probabilità a priori di una sola notte di pioggia. Per rispondere a questa domanda potremmo controllare gli archivi del servizio meteorologico del posto in cui viviamo e segnarci tutte le singole notti di pioggia, diciamo, in un anno: saranno molto frequenti se viviamo, ad esempio, in Irlanda o Scozia ma molto meno se viviamo in Italia, dato che le condizioni climatiche portano più spesso a giornate intere di pioggia (magari consecutive). Supponiamo di scoprire che ci siano state 40 singole notti di pioggia nello scorso anno; la probabilità a priori della variabile pioggia di notte θ sarà pertanto

$$P(\theta) = 40/365 = 10.96\%$$

La domanda 3. è, come la precedente, una probabilità a priori ma più difficile da determinare. Perché? Ci chiediamo quale sia la probabilità a priori che una strada sia bagnata, dovremmo quindi

tener conto di tutte le possibilità tali per cui una strada si bagni, dalle più probabili (la pioggia stessa, un camion delle pulizie stradali passato da poco ecc) alle più improbabili (il ribaltamento di un carretto di consegne di bottiglie di acqua con rottura di sufficienti bottiglie, l'erosione di un canale di irrigazione abbastanza vicino, l'atterraggio di un'astronave aliena che usi acqua come propellente... ecc). Sembra quasi impossibile da determinare, ma esiste un metodo che può venirci in aiuto: possiamo semplicemente sommare tra loro

- la probabilità che una strada si bagni se piove, che abbiamo già stabilito infatti $P(X|\theta) = 99.00\%$, moltiplicato per la probabilità che piova una notte, che conosciamo già $P(\theta) = 10.96\%$
- la probabilità che una strada si bagni se non piove $P(X|\bar{\theta})$, che leggeremo “probabilità di X dato non θ ” moltiplicato per la probabilità che non piova solo una notte $P(\bar{\theta})$, che leggeremo “probabilità di non θ ”

del primo termine conosciamo già tutto, ma come facciamo a determinare i valori del secondo termine? Grazie alle regole delle probabilità, la probabilità “non piova una sola notte” è complementare alla probabilità che “piova una sola notte”, di cui abbiamo già stabilito il valore:

$$P(\bar{\theta}) = 1 - P(\theta) = 1 - 10.96\% = 89.04\%$$

Ci serve dunque assegnare un valore a $P(X|\bar{\theta})$. Osserviamo che, per le regole delle probabilità

$$P(X|\bar{\theta}) = 1 - P(\bar{X}|\bar{\theta})$$

ovvero: la probabilità che una strada *si bagni se non piove* è complementare alla probabilità che *non si bagni se non piove* che corrisponde alla probabilità che una strada sia asciutta se non piove.

In questo caso specifico possiamo solo assegnare un valore arbitrario a $P(\bar{X}|\bar{\theta})$ risparmiandoci solamente il compito di immaginare le situazioni più improbabili (come l'astronave aliena) e considerando che “è molto probabile” che una strada sia asciutta in assenza di pioggia, ad esempio 97%:

$$P(X|\bar{\theta}) = 1 - P(\bar{X}|\bar{\theta}) = 1 - 97.00\% = 3.00\%$$

in altre situazioni invece è un valore noto (come nel caso dei test diagnostici in campo medico che vedremo in seguito).

Possiamo pertanto determinare che la probabilità a priori che una strada sia bagnata è pari al

$$P(X) = P(X|\theta)P(\theta) + P(X|\bar{\theta})P(\bar{\theta}) = 99.00\% \cdot 10.96\% + 3.00\% \cdot 89.04\% = 13.52\%$$

Questa operazione appena effettuata

$$P(X) = \sum_i \left(P(X|\theta_i)P(\theta_i) \right)$$

è detta *marginalizzazione della variabile X* e, per questo motivo, la “domanda” 3. è chiamata *probabilità marginale*.

Arrivati a questo punto chiediamoci: quale relazione hanno le tre probabilità determinate con la probabilità che vogliamo ottenere $P(\theta|X)$?

Possiamo dire che

- la probabilità che una strada sia bagnata se ha piovuto $P(X|\theta)$ e la probabilità che piova una notte sola $P(\theta)$ sono direttamente proporzionali alla probabilità che abbia piovuto la scorsa notte data la strada bagnata: entrambe queste probabilità contribuiscono in modo positivo alla risposta che cerchiamo
- la probabilità che una strada sia bagnata in generale $P(X)$ è inversamente proporzionale alla probabilità che abbia piovuto la scorsa notte se la strada è bagnata: ci potrebbero essere una serie di altri motivi per cui la strada sia bagnata e noi vogliamo solo quelli direttamente collegati alla pioggia

riassumendo

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

l'equazione appena scritta è il *Teorema di Bayes*.

Assegniamo ai termini del teorema i valori ricavati in precedenza e calcoliamo

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} = \frac{99.00\% \cdot 10.96\%}{13.52\%} = 80.24\%$$

Siamo giunti pertanto a determinare, grazie alle regole delle probabilità e al Teorema di Bayes, che la probabilità che la scorsa notte abbia piovuto dato il fatto che vediamo un tratto di strada bagnata è pari al 80.24%.

2 Positivo o negativo

Nel caso specifico appena trattato, la risposta sembra piuttosto ovvia e vano l'intero sforzo per arrivare al risultato, ma in altre situazioni il calcolo bayesiano della probabilità a posteriori può dare responsi anche molto lontani da quelli che ci suggerirebbe il semplice intuito.

Supponiamo ad esempio che un paziente venga sottoposto ad un test di screening per una malattia M piuttosto rara, ad esempio con una prevalenza **PV** nella popolazione di appartenenza dello 0.002%, ovvero 2 casi su 100,000.

Il test cui verrà sottoposto è di tipo qualitativo e fornirà un responso dicotomico: positivo \oplus oppure negativo \ominus . Sappiamo, dalla letteratura medica che abbiamo consultato, che il test ha sensibilità **SE** 85% e specificità **SP** 99% (per approfondimenti, si veda [Test Diagnostici](#)).

Dal punto di vista della statistica bayesiana sensibilità e specificità sono due probabilità a posteriori, rispettivamente:

$$\mathbf{SE} = P(\oplus|M) = 85.00\%$$

$$\mathbf{SP} = P(\ominus|\overline{M}) = 99.00\%$$

ovvero la probabilità a posteriori di ottenere un test positivo se si è malati e la probabilità a posteriori di ottenere un test negativo se non si è affetti dalla relativa patologia.

La prevalenza **PV** della malattia in questione invece è, in questo caso, una probabilità a priori: dato che non abbiamo alcuna informazione sul paziente se non quella di appartenere alla popolazione scelta per lo screening, è la probabilità a priori che il paziente sia affetto dalla malattia $P(M)$.

Quello che vogliamo ottenere sottoponendo il paziente al test è la probabilità di malattia a posteriori dato il risultato \odot del test

$$P(M|\odot)$$

dove la probabilità di malattia M è il parametro θ (verità non nota) mentre il risultato \odot del test è la variabile X .

Applicando quindi il Teorema di Bayes otteniamo che

$$P(M|\odot) = \frac{P(\odot|M)P(M)}{P(\odot)}$$

Supponiamo che il test sia risultato positivo \oplus , la nostra formula diventerà pertanto

$$P(M|\oplus) = \frac{P(\oplus|M)P(M)}{P(\oplus)}$$

Osserviamo che

- $P(\oplus|M)$ altro non è che la sensibilità **SE** del test
- $P(M)$ è, come detto in precedenza, la prevalenza della malattia

Come determiniamo $P(\oplus)$ ovvero la probabilità a priori (probabilità *marginale*) di ottenere un test positivo?

Possiamo procedere *marginalizzando* la variabile \oplus come fatto in precedenza per la probabilità a priori di osservare una strada bagnata

$$P(\oplus) = P(\oplus|M)P(M) + P(\oplus|\overline{M})P(\overline{M})$$

come prima, conosciamo il primo termine della somma (sensibilità e prevalenza) e sappiamo che

$$P(\overline{M}) = 1 - P(M) = 1 - 0.002\% = 99.998\%$$

ma, a differenza dell'esempio precedente, possiamo determinare con precisione il secondo termine coi dati a disposizione infatti

$$P(\oplus|\overline{M}) = 1 - P(\ominus|\overline{M}) = 1 - \mathbf{SP} = 1 - 99.00\% = 1.00\%$$

Possiamo quindi ottenere

$$P(\oplus) = P(\oplus|M)P(M) + P(\oplus|\overline{M})P(\overline{M}) = \mathbf{SE} \cdot \mathbf{PV} + (1 - \mathbf{SP})(1 - \mathbf{PV}) =$$

$$85.00\% \cdot 0.002\% + 1.00\% \cdot 99.998\% = 1.002\%$$

Abbiamo perciò tutti i termini necessari a determinare la probabilità di malattia a posteriori dato il risultato positivo del test. Visto il valore molto elevato di sensibilità e specificità ci aspetteremmo di ottenere un risultato altrettanto notevole ma applicando il teorema di Bayes otteniamo

$$P(M|\oplus) = \frac{P(\oplus|M)P(M)}{P(\oplus)} = \frac{85.00\% \cdot 0.002\%}{1.002\%} = 0.17\%$$

ovvero, in seguito a risultato positivo del test, il nostro paziente avrà una probabilità a posteriori di essere malato pari soltanto allo 0.17%...!

Perché è così bassa nonostante sensibilità e specificità fossero piuttosto elevate? Il risultato dipende dal fatto che siamo partiti con un'esigua probabilità di malattia a priori $\mathbf{PV} = 0.002\%$.

Questo esempio spiega più chiaramente in cosa consista l'analisi bayesiana:

L'analisi bayesiana consiste nella rivalutazione del nostro iniziale livello di fiducia circa l'avverarsi di un evento θ in base all'osservazione di un dato X ovvero nella *ridistribuzione* della probabilità *a priori* di un parametro θ condizionato dalla variabile X

Ripensando all'esempio precedente della strada, cosa abbiamo fatto in realtà? Abbiamo *ridistribuito* l'inizialmente remota probabilità a priori di una notte isolata di pioggia in base all'osservazione del tratto di strada bagnata. Se non avessimo osservato $X = \text{"strada bagnata"}$ o se avessimo osservato qualcosa di differente (ad esempio, solo strisce bagnate sulla strada o dei lavori in corso sulla tubazioni, ecc) la probabilità a priori di pioggia la notte scorsa sarebbe stata modificata di conseguenza.

3 La curva delle idee

Supponiamo ora che il parametro θ di cui vogliamo calcolare la probabilità sia R_t ovvero il tasso di riproduzione effettivo R nel tempo t di COVID-19 in Italia e i valori della variabile X osservata siano i nuovi casi giornalieri k di COVID-19 osservati in Italia dal primo all'ultimo giorno dei dati registrati.

I calcoli saranno più complessi visto che non ci stiamo più occupando di eventi dicotomici, come malattia o salute, strada bagnata o asciutta, notte di pioggia o non, risultato positivo o negativo, bensì una scala continua dei possibili valori di R_t da un minimo di 0 ad un massimo da definire e una scala discreta di valori dei nuovi casi k osservati da un minimo di 0 ad un massimo virtualmente illimitato, ma ciò di cui trattiamo sono comunque probabilità. Infatti, ciò che vogliamo ottenere è sempre $P(R_t|k)$ ovvero, a partire da un livello di fiducia a priori circa le probabilità di ottenere ogni possibile valore di R_t , ridistribuire per ogni giorno in archivio le probabilità dei valori di R_t in base all'osservazione dei k nuovi casi registrati. Usando la notazione ormai nota e il Teorema di Bayes:

$$P(R_t|k) = \frac{P(k|R_t)P(R_t)}{P(k)}$$

dove

- $P(k|R_t)$ sono le probabilità a posteriori di osservare i nuovi casi k per ciascuno dei possibili valori di R_t
- $P(R_t)$ sono le probabilità a priori per ogni valore possibile del numero di riproduzione effettivo
- $P(k)$ è la probabilità a priori (*marginale*) di osservare il numero di nuovi casi giornalieri misurato per ogni giorno in archivio

È più chiara ora la differenza tra la “domanda iniziale” $P(R_t|k)$ e il termine $P(k|R_t)$ del teorema di Bayes: nel primo caso infatti ci chiediamo per ciascun valore possibile di R_t quale sia la probabilità di ottenerlo data l’osservazione nota di k nuovi casi; nel secondo invece ci si chiede quanto sia probabile osservare i k nuovi casi effettivamente misurati per ognuno dei possibili valori di R_t . Non conosciamo direttamente $P(k|R_t)$ ma possiamo dedurlo in base al ragionamento e all’osservazione: essendo il tasso di riproduzione R il numero medio di soggetti suscettibili che un infetto può contagiare, i nuovi casi k sono verosimilmente proporzionali ad esso. Possiamo dunque stabilire una relazione (matematica) tra le due quantità, sempre restando nell’ambito delle probabilità.

La probabilità a posteriori di R_t calcolata per ciascun giorno diventa la probabilità a priori per il giorno successivo, che verrà nuovamente ridistribuita in base ai nuovi casi giornalieri osservati, ecc... fino all’ultimo giorno (presumibilmente oggi).

Come si diceva, l’inferenza bayesiana consiste nella rivalutazione del livello di fiducia iniziale (a priori) su un parametro θ data l’osservazione di un valore della variabile X , osserviamo perciò che il teorema di Bayes può essere riscritto così

$$P(\theta|X) = P(\theta) \cdot \frac{P(X|\theta)}{P(X)}$$

dove $P(\theta)$ è il grado di fiducia iniziale (probabilità a priori) dell’avverarsi di θ e il secondo termine rappresenta il coefficiente che determina “l’entità della rivalutazione”

$$\frac{P(X|\theta)}{P(X)} = \mathcal{L}(\theta|X)$$

che leggeremo **verosimiglianza** \mathcal{L} (dall’inglese *Likelihood*) del parametro θ data l’osservazione di X .

Quindi possiamo dire che

$$P(\theta|X) = P(\theta) \cdot \mathcal{L}(\theta|X)$$

La probabilità a posteriori di θ data l’osservazione di X è direttamente proporzionale alla probabilità a priori di θ e alla verosimiglianza di θ dato X

Tornando all’esempio del numero di riproduzione effettivo possiamo pertanto dire che

$$P(R_t|k) = P(R_t) \cdot \frac{P(k|R_t)}{P(k)}$$

e che

$$P(R_t|k) = P(R_t) \cdot \mathcal{L}(R_t|k)$$

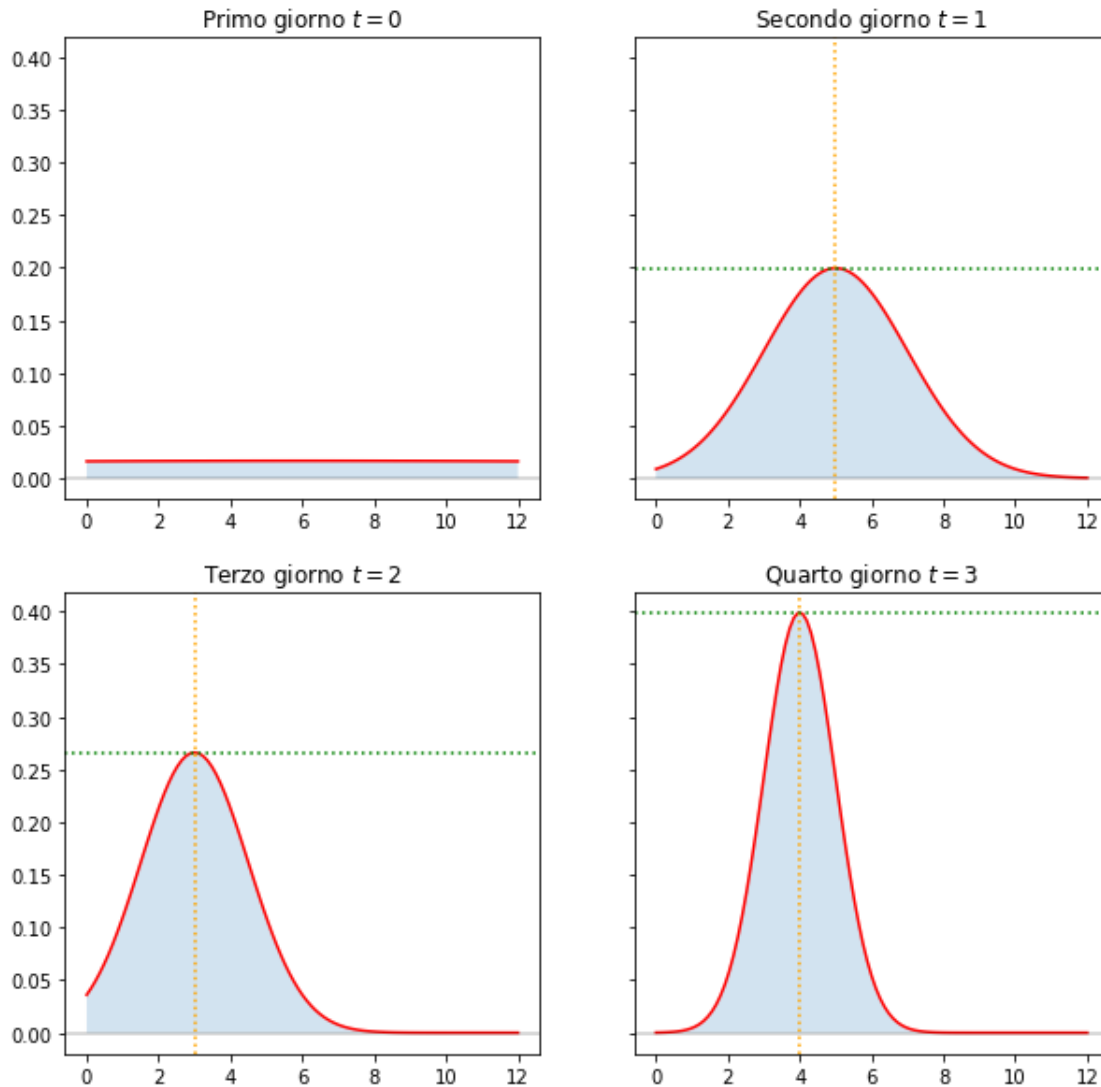


Figure 2: Esempio di ridistribuzione di probabilità. Le curve mostrano la densità di probabilità (pdf) per i valori di R_t .

Quale sarà, per ogni valore possibile di R_t , la nostra probabilità a priori per il primo giorno $t = 0$? Non abbiamo alcuna informazione su ciò che sia accaduto precedentemente, quindi ogni valore è equamente probabile: la *distribuzione di probabilità* di R_t il primo giorno sarà una linea retta, ovvero una *distribuzione uniforme* (riquadro in alto a sinistra in figura 2). Dopo l'osservazione di k nuovi casi il secondo giorno, la distribuzione a priori di R_t cambierà di conseguenza e diverrà la nuova probabilità a priori per il giorno successivo e così via (riquadri successivi in figura 2). In questo modo, ogni giorno, l'osservazione di k nuovi casi *ridistribuisce* la probabilità a priori in una probabilità a posteriori, che potrà essere usata per il giorno successivo fino all'ultimo giorno in archivio.

È ora più chiaro perchè si parlava di *ridistribuzione* di probabilità: quando abbiamo a che fare

con più valori (variabili discrete come i k nuovi casi o continue come i valori di R_t) la probabilità assegnata a ciascun valore è “riassunta” in una *distribuzione di probabilità* che, graficamente, può essere rappresentata da una curva (come le densità di probabilità in figura 2): ad ogni osservazione la curva si modifica ovvero la probabilità viene ridistribuita in base alle osservazioni.

Supponiamo ora di essere giunti a determinare la probabilità a posteriori di R_t dell'ultimo giorno $t = t^*$ (oggi) ovvero R_{t^*} . Non abbiamo più k nuovi casi da osservare, come determiniamo la probabilità dei valori di R_t per il giorno successivo $t = t^* + 1$ (domani) ovvero R_{t^*+1} ?

Come può il ragionamento seguito finora portare ad una stima di previsione riguardo la probabilità un determinato evento θ nel futuro?

4 Predire il passato

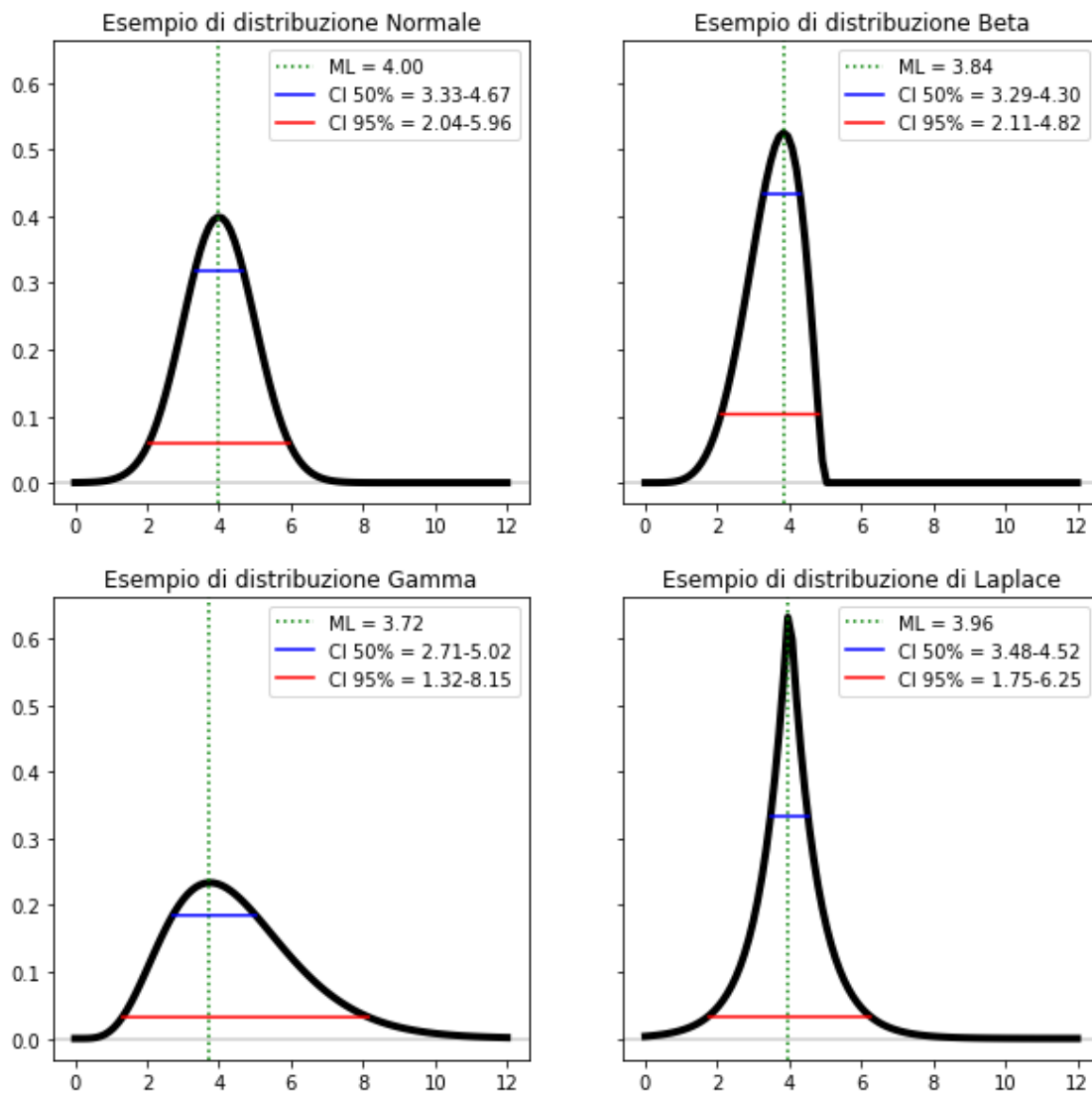


Figure 3: Esempi di distribuzioni di probabilità e di relative valore di massima verosimiglianza o moda (ML, *Maximum Likelihood*) e intervalli di credibilità (CI, *Credible Interval* detti anche HDI, *Highest Density Interval*)

Esistono diversi tipi di distribuzione (vedi figura 3) ognuna descritta da una specifica funzione matematica con ben note proprietà. Per ogni variabile, la scelta della distribuzione a priori è effettuata in base all'analisi della variabile stessa e alla valutazione delle possibilità che supponiamo.

Per ogni distribuzione, possiamo definire il valore centrale della massima verosimiglianza (ML, *Maximum Likelihood* corrispondente alla *moda* della distribuzione) che rappresenta il valore maggiormente probabile e una serie di intervalli di credibilità (CI, *Credible Interval*) che descrivono i valori minimi e massimi entro una probabilità definita (ad esempio, 95% o 50%).

Ripensando all'esempio della ridistribuzione a posteriori della probabilità di R_t notiamo come ogni giorno il valore della massima verosimiglianza (moda) sia sempre più probabile e gli intervalli di credibilità sempre più stretti: dopo giorno dopo, grazie alla rivalutazione delle probabilità guadagniamo credibilità (figura 5).

Valore centrale e intervalli di credibilità, possono poi essere calcolati per ciascun giorno e rendere così graficamente più chiaro l'andamento nel tempo della probabilità del parametro che abbiamo calcolato (figura 4).

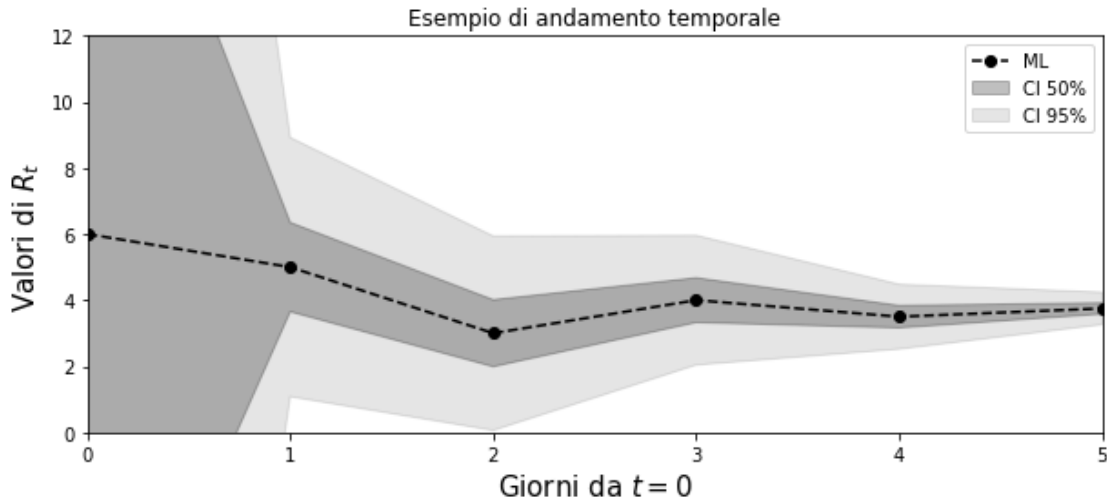


Figure 4: Esempio di visualizzazione nel tempo dei valori centrali di massima verosimiglianza (moda) e intervalli di credibilità della probabilità a posteriori del parametro R_t .

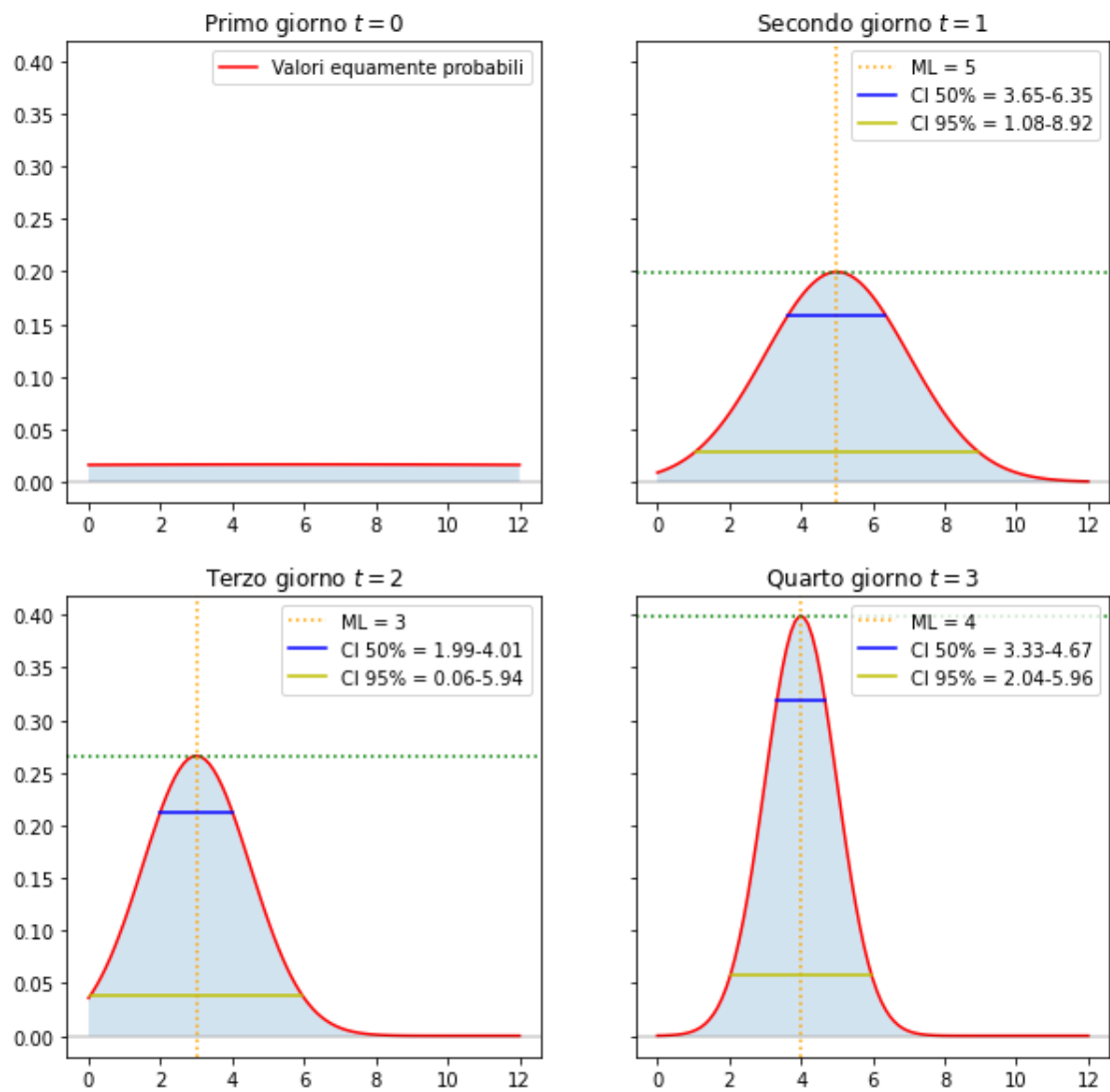


Figure 5: Esempio di ridistribuzione di probabilità a posteriori. Ogni giorno il valore centrale di massima verosimiglianza (moda) è sempre più probabile e gli intervalli di credibilità sempre più stretti.

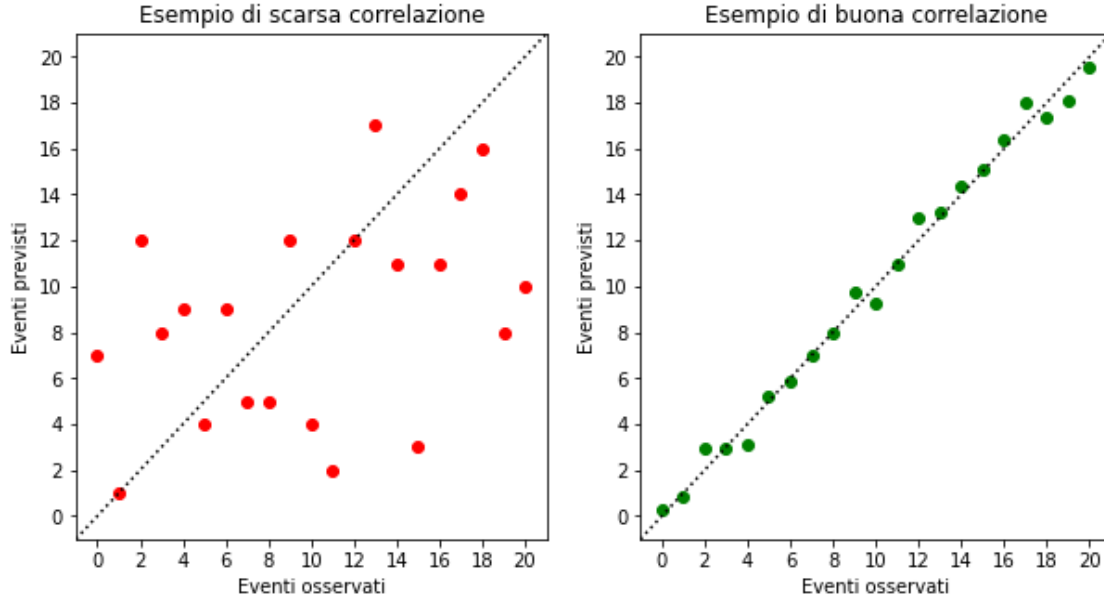


Figure 6: Esempi di correlazione tra valori osservati e previsti dal modello.

Nell'esempio precedente del calcolo di R_t , non abbiamo ottenuto informazioni sul futuro, ma abbiamo molte più informazioni sul passato. Integrando queste informazioni con altri dati utili (ad esempio, periodi di lockdown, mobilità della popolazione, tempi stimati tra contagio e ospedalizzazione, ecc) possiamo creare un **modello** ovvero un'insieme di funzioni che descrivono le distribuzioni di probabilità a priori e le presupposte relazioni di causalità di tutti gli elementi utili che possono influire sul risultato che vogliamo ottenere: il tasso di riproduzione effettivo R_t .

Il modello creato non dovrà “semplicemente” essere in grado di descrivere il passato ma, generando valori casuali dalle distribuzioni impostate e calcolandone le relazioni, produrre un percorso simile alla realtà.

Analizziamo un semplice esempio, preso dal libro *Doing Bayesian Analysis* [ref](#).

Un politico vive su una delle isole di un arcipelago. Essendo in campagna elettorale vuole visitare tutte le isole che lo compongono e passare più tempo dove la popolazione è maggiore ma non conosce i dati demografici di ciascuna isola. Grazie però al suo team può sapere

- la popolazione dell'isola su cui si trova
- la popolazione della prossima isola che intende visitare

Supponiamo che l'arcipelago sia costituito da 7 isole che chiameremo θ_1, θ_2 ecc. La popolazione di ciascuna sarà $P(\theta_i)$ con i da 1 a 7 e, per semplicità, diremo che $P(\theta_i) < P(\theta_{i+1})$ pertanto la popolazione dell'isola θ_5 sarà maggiore di θ_4 ma minore rispetto a θ_6 e così via. Il nostro politico potrà spostarsi solamente in isole contigue ovvero trovandosi sull'isola θ_i potrà decidere se

- restare su θ_i un altro giorno
- spostarsi su θ_{i+1}
- spostarsi su θ_{i-1}

Supponiamo che parta ad esempio dall'isola θ_2 . Come può decidere se restare o spostarsi e dove? Un metodo potrebbe essere quello di lanciare una moneta: se uscisse testa otterrebbe la *proposta* di

muoversi sull'isola θ_{i+1} (quindi θ_3 nell'esempio) se uscisse croce su θ_{i-1} . Comunicata la proposta, il suo team potrà chiedere all'isola scelta quale sia la sua popolazione. Se la popolazione fosse maggiore dell'isola su cui si trova ora si sposterà sicuramente altrimenti sarà nuovamente in dubbio sul da farsi e si sposterà solamente su base probabilistica generando ad esempio una percentuale casuale da 0% a 100% con una roulette numerata: si sposterebbe solamente se la percentuale random generata fosse minore della proporzione tra la popolazione dell'isola proposta e dell'isola attuale

$$p_{\text{spostarsi}} = \frac{P(\theta_{\text{proposta}})}{P(\theta_{\text{attuale}})}$$

Se ad esempio il giorno $t = 1$ uscisse testa, la proposta di spostarsi da θ_2 a θ_3 verrà immediatamente accettata perché la popolazione è maggiore. Il tempo successivo $t = 2$ sarà su θ_3 e immaginiamo che il lancio della moneta dia croce. La popolazione su θ_2 è minore che sull'attuale θ_3 . Cosa fare? Il rapporto tra le due popolazioni sarà

$$p_{\text{spostarsi}} = \frac{P(\theta_2)}{P(\theta_3)} = \frac{2}{3} = 66, \bar{6}\%$$

se dunque il giro di roulette risultasse in una percentuale inferiore si muoverebbe sull'isola proposta altrimenti resterebbe sull'attuale. E così via giorno per giorno (per completare il quadro, immaginiamo due isole “virtuali” θ_0 e θ_8 con popolazione nulla: le proposte verso queste isole inesistenti saranno sicuramente rifiutate).

Per quanto possa sembrare del tutto casuale, grazie alle relazioni predeterminate e dopo un sufficiente numero di spostamenti, questo metodo (il cui nome tecnico è *algoritmo Metropolis della catena Markov Monte Carlo*) porterà il nostro politico ad aver passato in proporzione più tempo sulle isole più popolate. La dimostrazione matematica del perché il metodo funzioni è molto complessa e al di là delle finalità di questa semplice introduzione all'argomento. Si rimanda alla bibliografia per approfondimenti specifici.

Cosa abbiamo fatto dunque? Abbiamo impostato un *modello* e, grazie alla generazioni di valori casuali (il lancio della moneta e il giro di roulette) e alle relazioni determinate (in quali casi spostarsi), abbiamo **simulato** un possibile percorso del politico tra le isole in n spostamenti. Potremmo ripetere la simulazione partendo da un'isola diversa in $t = 1$ e per più o meno spostamenti. Se il risultato delle simulazioni fosse sempre che il politico abbia passato più tempo sulle isole più popolate, avremo ottenuto il risultato cercato.

Proviamo una simulazione in **python** del modello impostato. Definiamo anzitutto il lancio della moneta:

```
1 def lancio_della_moneta():
2     return (-1) ** np.random.randint(2)
```

questa semplice funzione ritorna in modo casuale +1 (testa) o -1 (croce). Useremo `lancio_della_moneta` per determinare su quale isola θ proporre di spostarsi. Definiamo il giro della roulette:

```
1 def giro_di_roulette():
2     return np.random.random()
```

la funzione ritorna un numero reale casuale tra 0 e 1, ovvero una percentuale tra 0% e 100%. Useremo `giro_di_roulette` per calcolare la probabilità di spostarci in caso la moneta ritorni -1 (croce). Definiamo infine la funzione di simulazione del percorso effettuato in base alle isole e al numero di spostamenti, iniziando da un'isola scelta casualmente

```
1 def simulazione(isole, spostamenti):
2     percorso = [np.random.choice(list(isole))]
3     for spostamento in range(spostamenti-1):
4         prossima_isola = percorso[-1] + lancio_della_moneta()
5         if (
6             not prossima_isola or
7             prossima_isola > len(isole)
8         ):
9             percorso.append(percorso[-1])
10            continue
11        if prossima_isola > percorso[-1]:
12            percorso.append(prossima_isola)
13            continue
14        p_spostamento = prossima_isola / percorso[-1]
15        if giro_di_roulette() < p_spostamento:
16            percorso.append(prossima_isola)
17        else:
18            percorso.append(percorso[-1])
19    return percorso
```

e le isole dell'arcipelago con la loro popolazione

```
1 isole = {}
2 for isola in range(7):
3     isole.update({isola+1: (isola+1)*1000})
```

Facciamo ora “girare” il modello nelle simulazioni A, B e C con rispettivamente 1000, 2000 e 3000 spostamenti

```
1 simulazione_A = simulazione(isole, spostamenti=1000)
2 simulazione_B = simulazione(isole, spostamenti=2000)
3 simulazione_C = simulazione(isole, spostamenti=3000)
```

e vediamone il risultato in figura 7

Esempio di simulazioni algoritmo Metropolis della Markov chain Monte Carlo

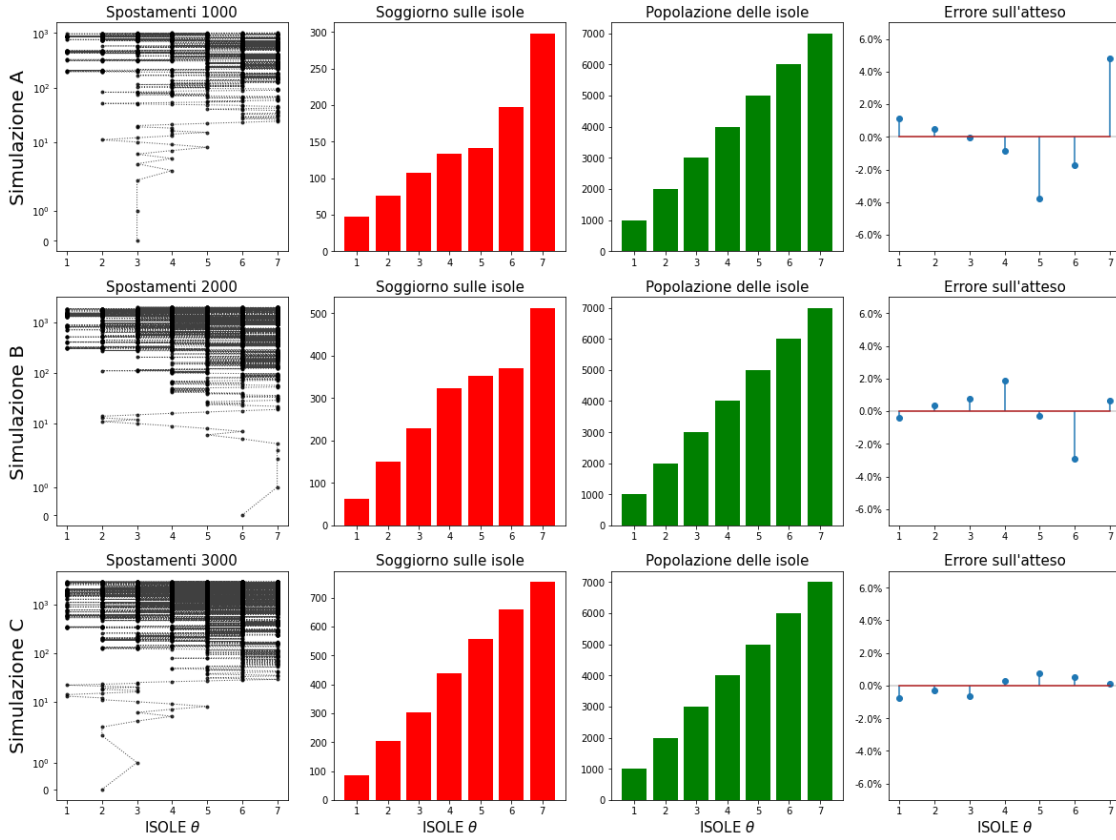


Figure 7: Esempio di simulazioni di un semplice algoritmo Metropolis del metodo Markov chain Monte Carlo.

Notiamo come effettivamente, dopo un numero sufficiente di spostamenti, la distribuzione del tempo passato in ciascuna isola si avvicini sempre più alla sua popolazione e gli errori percentuali sul tempo atteso siano via via minori.

Al di là della simulazione in sé stessa possiamo però anche calcolare la probabilità di restare o spostarsi su un'isola diversa. Supponiamo di partire dall'isola centrale θ_4 . La probabilità di spostarsi sull'isola θ_5 dipende solo dal lancio della moneta, dunque è del 50%. La probabilità di spostarsi sull'isola θ_3 dipende sia dal lancio della moneta 50% che dalla proporzione tra le popolazioni di θ_3 e θ_4 quindi $50\% \cdot \frac{3}{4} = 37.5\%$. La restante $100\% - 50\% - 37.5\% = 12.5\%$ è la probabilità di restare sull'isola θ_4 .

Si procederà allo stesso modo per ogni tempo successivo. Come vediamo in figura 8, già al tempo $t = 15$ la distribuzione di probabilità del modello (steli) si avvicina alla distribuzione voluta (barre verdi) basata sulla popolazione delle isole.

Esempio di distribuzione di probabilità con algoritmo Metropolis (MCMC)

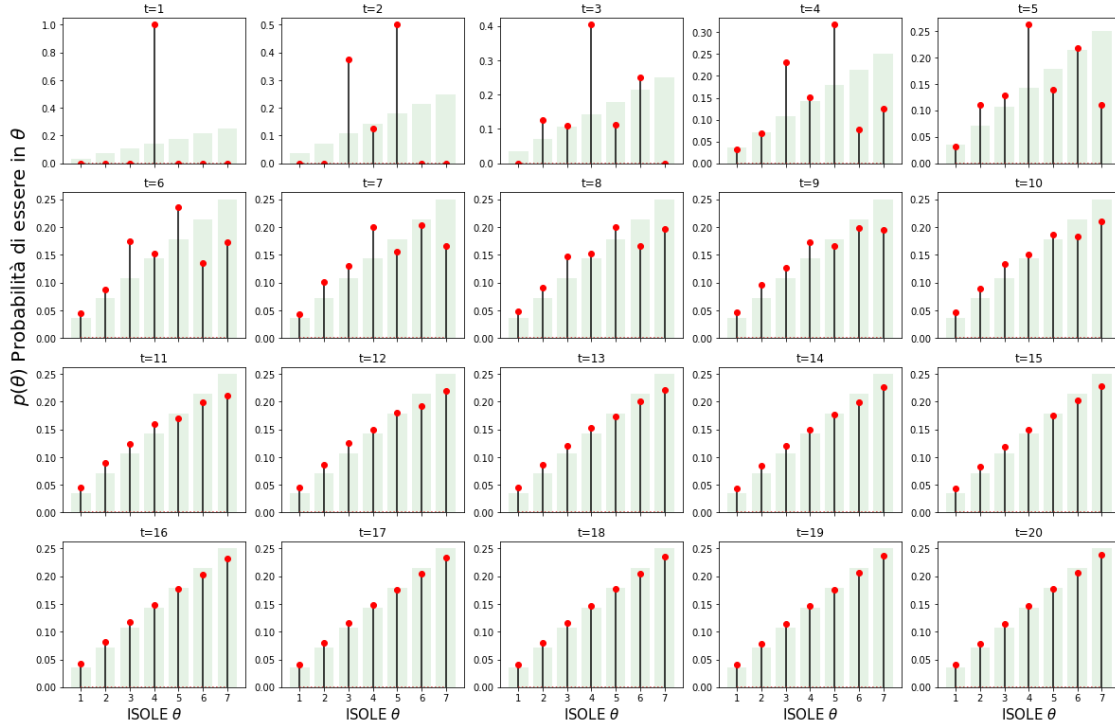


Figure 8: Esempio di distribuzione di probabilità un semplice algoritmo Metropolis del metodo Markov chain Monte Carlo.

L'algoritmo funziona anche con distribuzioni target differenti, in Appendice un esempio per distribuzione target casuale (figure 10 e 11)

Con il modello appena creato però abbiamo solo simulato uno o più possibili percorsi oppure calcolato la probabilità di trovarsi in un dato luogo in un dato tempo, come possiamo capire se un modello è in grado descrivere la realtà?

In verità, abbiamo già tutto ciò che ci serve perché anche nell'esempio precedente avevamo una distribuzione *target* da raggiungere, dipendente in quel caso dalla popolazione delle isole, ma potrebbe per esempio dipendere dal numero di nuovi casi di una malattia infettiva o dalla temperatura registrata in una determinata area geografica o ancora dall'andamento di determinati titoli in borsa, ecc.

Tornando all'esempio del calcolo di R_t , impostiamo tutti i parametri del modello (probabilità a priori, variabili random e relazioni tra variabili) in un apposito strumento di calcolo (ad esempio [Stan](#)) sul nostro desktop computer e cerchiamo di... “predire il passato” ovvero simulare gli eventi già accaduti. Fermiamo il calcolo ad un determinato giorno $t = t'$ e controlliamo se i dati generati sono simili a quelli realmente misurati. Del passato conosciamo tutto perché è già accaduto, possiamo pertanto valutare se e quanto le distribuzioni a priori di probabilità e le presunte relazioni tra gli eventi che abbiamo definito siano valide. In qualche minuto (ricordiamoci cosa può fare uno smartphone in pochi milionesimi di secondo) avremo i risultati delle simulazioni e potremo confrontarli con i dati realmente osservati tramite un'analisi della loro correlazione (figura 6).

Se non abbiamo ottenuto una buona correlazione, dovremo rivedere il modello, modificarlo e rifinirlo, fino ad ottenere una buona simulazione degli eventi accaduti in precedenza e infine potremo

applicarlo a partire dall'ultimo giorno, oggi, supponendo differenti *scenari* e prevedendo così le probabilità dei futuri valori del parametro di nostro interesse.

Ad esempio, possiamo supporre che variazioni della mobilità della popolazione influiscano verosimilmente sul tasso di riproduzione di un'epidemia, aumentando o diminuendo la possibilità di contatti tra possibili infetti. Si possono perciò definire alcuni possibili scenari al termine di un lockdown

1. scenario A: aumento della mobilità del 50%
2. scenario B: aumento della mobilità del 25%
3. eccetera...

Calcolati infine valore di massima verosimiglianza (moda) e intervalli di credibilità delle probabilità a posteriori ottenute per i giorni futuri, potremo visualizzare graficamente il risultato delle nostre previsioni per ciascuno degli scenari ipotizzati (figura 9).

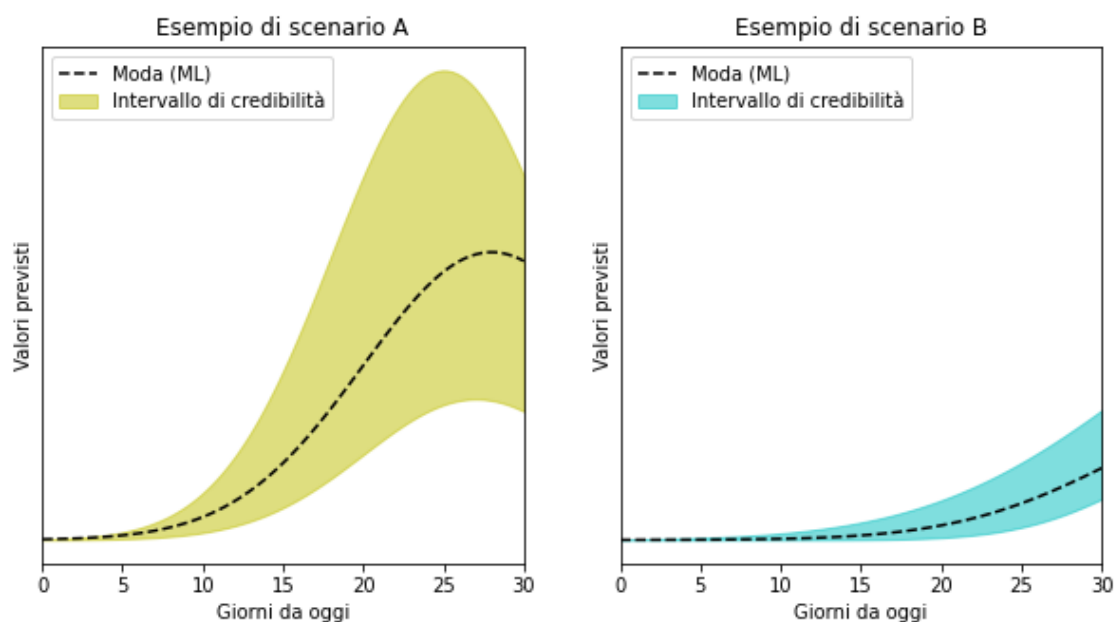


Figure 9: Esempio di previsioni per differenti scenari.

5 Una semplice complessità

Abbiamo visto come, nonostante i calcoli dell'inferenza bayesiana possano essere talmente complessi da poter essere svolti solo da un computer sufficientemente potente con l'ausilio di strumenti informatici appositi, le sue basi si fondino su principi semplici che applichiamo, spesso inconsciamente, nella vita di tutti i giorni (pensiamo all'esempio iniziale della strada bagnata).

Non dimentichiamoci che il computer più potente che conosciamo siamo noi stessi e gli apparentemente semplici ma profondamente complessi meccanismi del pensiero umano.

6 Appendice

Esempio di simulazioni algoritmo Metropolis della Markov chain Monte Carlo

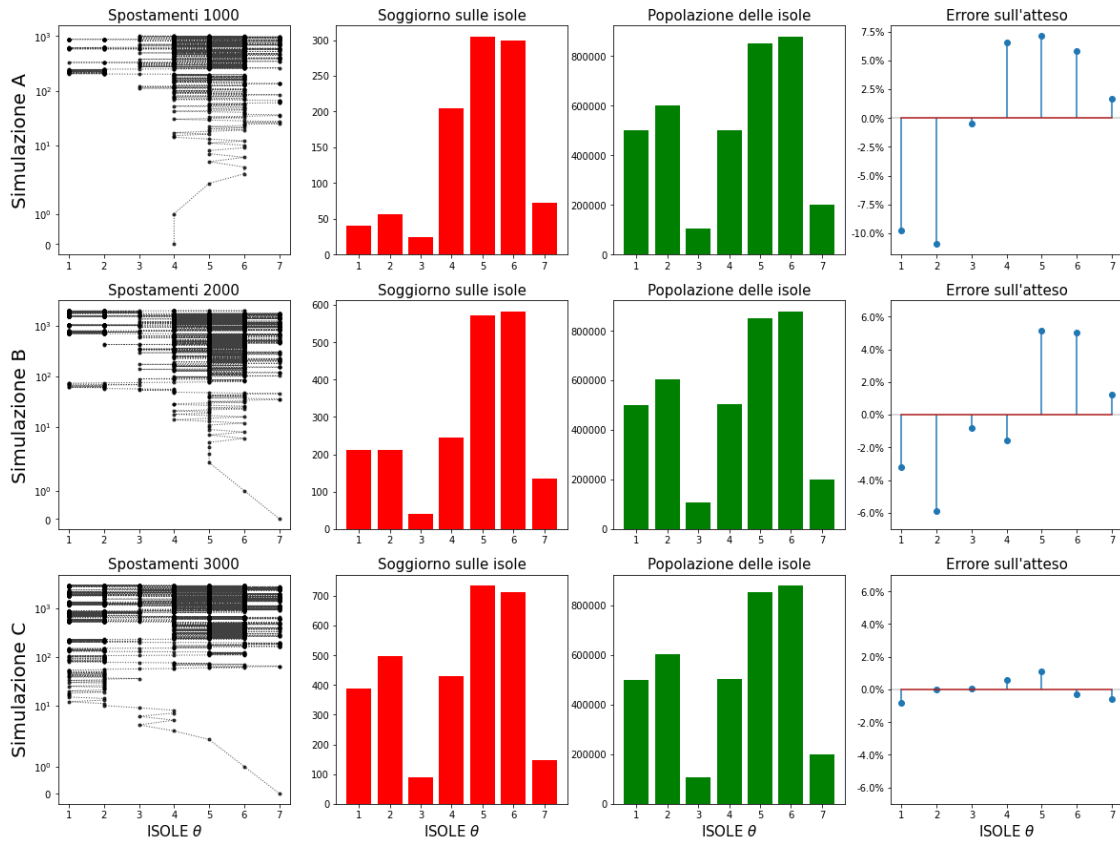


Figure 10: Esempio di simulazioni di un algoritmo Metropolis del metodo Markov chain Monte Carlo con distribuzione target random.

Esempio di distribuzione di probabilità con algoritmo Metropolis (MCMC)



Figure 11: Esempio di distribuzione di probabilità un algoritmo Metropolis del metodo Markov chain Monte Carlo con distribuzione target random.