



UNIVERSITÀ  
DEGLI STUDI DI MILANO-BICOCCA

## SYLLABUS DEL CORSO

### Inferenza Bayesiana

2627-2-F8205B011-F8205B011-2

---

#### Obiettivi formativi

L'insegnamento rientra nelle aree di apprendimento delle scienze statistiche, dell'informatica e delle scienze sociali e permette allo studente di apprendere le procedure analitiche ed inferenziali nell'ambito dell'inferenza bayesiana. Il ragionamento bayesiano viene presentato in modo integrato con l'approccio all'inferenza statistica sviluppato in senso classico. L'attività formativa permette agli studenti di acquisire solidi elementi di teoria bayesiana, di sviluppare le applicazioni attraverso un approccio di "problem solving" con dati reali e simulati in riferimento a problemi applicativi nell'ambito della biostatistica. Lo studente acquisisce inoltre abilità nella comunicazione scientifica scritta in quanto è richiesta la stesura di testi a completamento dei risultati ottenuti dalle analisi svolte.

#### Conoscenza e comprensione

Lo studente viene introdotto ai principali modelli statistici bayesiani per l'analisi di dati con diverse tipologie di variabili risposta, alle ipotesi sottostanti tali modelli e ai modelli per l'analisi dei dati longitudinali. Acquisisce inoltre la comprensione delle procedure Markov Chain Monte Carlo (MCMC) e dei relativi algoritmi di stima, nonché la capacità di valutarne l'efficacia.

Lo studente viene anche introdotto al linguaggio di programmazione R, nell'ambiente RMarkdown, che consente di creare documenti riproducibili contenenti codice, risultati, commenti e specifiche procedure per l'analisi bayesiana tramite il software SAS.

Gli esempi applicativi riguardano dati reali e simulati provenienti da diversi ambiti di riferimento del corso di studi. Lo studente apprende inoltre a fornire una descrizione scritta dei risultati ottenuti alla luce delle domande di ricerca, insieme a una valutazione critica dei limiti delle analisi svolte. In questo modo lo studente sviluppa autonomia di giudizio e rafforza le proprie abilità comunicative, sia nella presentazione e nella giustificazione delle scelte metodologiche adottate, sia nell'interpretazione e nella comunicazione dei risultati in relazione alle domande di ricerca.

#### Capacità di applicare conoscenza e comprensione

L'insegnamento fornisce competenze nell'utilizzo dei modelli bayesiani con distribuzioni coniugate, nella scelta delle distribuzioni a priori e nell'impiego degli algoritmi di stima per modelli complessi.

Attraverso R, RStudio e RMarkdown gli studenti imparano a impostare in modo organico il ragionamento statistico, mediante l'analisi dei dati e la redazione di relazioni che illustrino il codice, le analisi e i risultati.

Con l'uso del software SAS, gli studenti apprendono a stimare modelli bayesiani complessi tramite algoritmi di simulazione MCMC e a impostare correttamente gli input richiesti dagli algoritmi di stima.

L'insegnamento consente agli studenti di acquisire solide basi teoriche e la capacità di applicare i modelli statistici presentati a dati reali. Gli studenti saranno in grado di valutare il modello più appropriato in base ai dati disponibili e alle domande di ricerca. Imparano inoltre a scrivere e commentare il codice utilizzato per generare i risultati, adottando un approccio orientato alla scienza aperta (open science) che garantisca la riproducibilità e la replicabilità delle analisi.

Al termine dell'insegnamento, grazie al materiale fornito (le dispense del docente corredate da un'ampia bibliografia, i codici per i software R e SAS e l'interfaccia RMarkdown), lo studente è in grado di proseguire in modo autonomo nell'approfondimento di questa disciplina.

L'insegnamento è indispensabile per il successivo percorso universitario in quanto fornisce i concetti essenziali per lo sviluppo dei metodi bayesiani sia attraverso solide basi teoriche e applicative per i contesti professionali di riferimento (biostatistico/statistico/demografico e affini) degli studenti del corso di laurea in Biostatistica.

## **Contenuti sintetici**

Introduzione all'inferenza bayesiana e alla regola di Bayes.

Metodi di specificazione del modello e delle distribuzioni a priori.

Famiglie coniugate: Gaussiana, Poisson-gamma, beta-binomiale, multinomiale-Dirichlet .

Inferenza Bayesiana non parametrica.

Metodi di sintesi della distribuzione a posteriori, regioni di credibilità e intervalli con la massima densità a posteriori.

Introduzione ai processi stocastici di Markov e proprietà delle catene di Markov.

Modello di passeggiata casuale. Modello di transizione per dati longitudinali.

Modelli di Markov a variabili latenti per dati longitudinali ed estensioni con covariate sia nel modello osservato che nel modello latente.

Metodi Markov Chain Monte Carlo (MCMC): Algoritmo Metropolis-Hastings e campionamento Gibbs.

Test diagnostici per la convergenza.

Esercitazioni svolte in relazione a specifici problemi applicativi utilizzando l'ambiente R, RStudio con l'editor di testo RMarkdown ed il software SAS.

## **Programma esteso**

Durante l'insegnamento viene ripresa la regola di Bayes e la regola delle probabilità totali. Vengono sviluppati gli aspetti riguardanti la specificazione delle distribuzioni a priori, la stima esatta delle distribuzioni a posteriori e l'interpretazione dei modelli. Viene introdotto il modello beta-binomiale e altre famiglie coniugate: Gaussiana, modello Poisson-gamma, multinomiale-Dirichlet e particolare enfasi è posta anche sulla distribuzione predittiva.

L'inferenza puntuale bayesiana viene confrontata con l'inferenza classica. Vengono illustrate le caratteristiche di scelta e di determinazione della distribuzione a priori, sia informativa sia non informativa. La nozione di scambiabilità viene illustrata attraverso il teorema di rappresentazione di De Finetti. La distribuzione a posteriori viene sintetizzata attraverso le regioni di credibilità, e gli intervalli con la massima densità a posteriori.

Vengono introdotti i processi stocastici markoviani enunciando le proprietà e le caratteristiche delle catene di Markov. La passeggiata casuale viene illustrata attraverso le simulazioni delle traiettorie per matrici stocastiche con diverse dimensioni. Viene introdotto il modello di transizione per dati longitudinali, ed il modello di Markov latente. Vengono illustrati anche da un punto di vista computazionale gli algoritmi di stima utilizzati nell'ambito del metodo MCMC per approssimare la distribuzione a posteriori: l'algoritmo Metropolis-Hastings e l'algoritmo di campionamento Gibbs (Gibbs sampling). Vengono discusse diverse misure riferite sia alle analisi grafiche che ai test statistici che permettono la valutazione diagnostica della convergenza.

La teoria viene affiancata da numerosi esempi di applicazione dei modelli bayesiani nell'ambito della biostatistica attraverso dati reali e simulati riguardanti l'epidemiologia, la farmacoepidemiologia, la medicina e la biologia oltre che l'ecologia e le scienze ambientali. Si intende inoltre facilitare lo sviluppo della conoscenza della semantica in ambiente R e del software SAS. Gli esempi sono svolti in RStudio con l'ausilio di RMarkdown. Lo studente durante le esercitazioni è incoraggiato, anche tramite l'apprendimento cooperativo, ad elaborare documenti riproducibili concernenti anche il commento critico ai risultati delle analisi. Vengono utilizzati i seguenti pacchetti: probBayes, learnBayes, stan, LMest, LaplaceDemon, Nimble, RMarkdown attraverso il pacchetto knitr, per integrare il codice, i risultati delle analisi ed i commenti. L'analisi con il software SAS viene svolta attraverso la libreria PROC MCMC.

## **Prerequisiti**

Si consiglia di aver acquisito e di riprendere le nozioni impartite nei seguenti insegnamenti: Statistica, Probabilità e inferenza statistica e Modelli Statistici II.

## **Metodi didattici**

Sono previste lezioni frontali, svolte in presenza in modalità erogativa presso il laboratorio informatico, relative agli aspetti teorici dell'insegnamento e affiancate da esercitazioni pratiche. Le lezioni si svolgono presso il laboratorio informatico. Durante le attività formative con l'ausilio di R, RStudio, dell'ambiente RMarkdown e del software SAS, gli studenti imparano ad analizzare i dati, a stimare i modelli bayesiani e a corredare le analisi con i commenti e interpretazioni, elaborando documenti che consentono di riprodurre e replicare le analisi svolte. Settimanalmente vengono assegnati esercizi di riepilogo, basati su dati reali o simulati, con i quali gli studenti vengono incoraggiati ad affrontare problemi applicativi coerenti con gli argomenti teorici illustrati a lezione, favorendo anche lo sviluppo dell'apprendimento cooperativo. L'insegnamento si svolge in 35 ore di didattica erogativa, dedicate alla presentazione degli aspetti teorici e metodologici, e 12 ore di didattica interattiva, dedicate a esercitazioni guidate, applicazioni con R/RStudio/RMarkdown e SAS, discussione dei risultati e attività di problem solving. Le videoregistrazioni asincrone rese disponibili sulla piattaforma e-learning costituiscono materiale integrativo di supporto allo studio e non sostituiscono le attività didattiche in presenza.

## **Modalità di verifica dell'apprendimento**

Le seguenti modalità di verifica dell'apprendimento si applicano sia agli studenti frequentanti sia a quelli non frequentanti le lezioni frontali. L'esame consiste in una prova scritta, con domande aperte ed esercizi applicativi,

alla quale può seguire, su richiesta dello studente o del docente, un colloquio orale facoltativo. Non sono previste prove intermedie.

La prova scritta ha una durata massima di due ore e si svolge in laboratorio informatico. Durante l'esame gli studenti devono rispondere a domande teoriche formulate in modo aperto e risolvere esercizi applicativi relativi agli argomenti trattati durante le lezioni e alle esercitazioni svolte.

Le domande teoriche sono finalizzate a verificare la comprensione dei concetti fondamentali dell'inferenza statistica bayesiana e dei relativi metodi avanzati. Gli esercizi applicativi svolti in R, RStudio, RMarkdown e il software SAS, consentono di valutare la capacità dello studente di applicare modelli statistici bayesiani a dati reali o simulati, nonché produrre report riproducibili, contenenti codice, analisi e interpretazione dei risultati.

La prova scritta è valutata in trentesimi secondo una griglia di punteggio comunicata agli studenti prima della prova. La valutazione tiene conto di: correttezza teorica delle risposte; appropriatezza della scelta del modello; correttezza dell'implementazione computazionale; interpretazione critica dei risultati; chiarezza e completezza della comunicazione scritta. Il colloquio orale facoltativo, se richiesto dallo studente o dal docente, consiste in una discussione della prova scritta e degli argomenti teorici dell'insegnamento e può confermare, integrare o modificare la valutazione finale. Durante la prova è consentito l'utilizzo del materiale didattico e di quello fornito dal docente, compresi i codici sviluppati in R e SAS nel corso delle attività didattiche. L'esame si intende superato con una votazione minima di 18/30.

## Testi di riferimento

Il materiale didattico principale è costituito dalle dispense preparate dal docente, che trattano gli argomenti teorici del corso, le applicazioni sviluppate con il software R, gli esercizi e le relative soluzioni. Le dispense sono rese disponibili nella pagina dell'insegnamento sulla piattaforma e-learning dell'Ateneo. Al termine di ogni lezione, sono rese disponibili le slides utilizzate, i programmi di calcolo, i dataset impiegati durante le esercitazioni e il materiale di supporto. Con cadenza settimanale vengono assegnati esercizi, corredati dalle relative soluzioni. Nella medesima pagina della piattaforma e-learning sono inoltre disponibili esempi di prove d'esame degli anni precedenti.

I principali riferimenti bibliografici sono riportati nelle dispense del corso. Di seguito sono elencati i testi di riferimento fondamentali, disponibili presso la biblioteca di Ateneo, anche in formato e-book:

Albert, J. (2009). *Bayesian computation with R*. Springer Science & Business Media.

Albert, J., Hu, J. (2019). *Probability and Bayesian modeling*. Chapman and Hall/CRC.

Bartolucci, F., Farcomeni, A., Pennoni, F. (2013). *Latent Markov Models for longitudinal data*, Chapman and Hall/CRC, Boca Raton.

Migon, H. S., Gamerman, D., Louzada, F. (2014). *Statistical inference: an integrated approach*. Chapman & Hall.

Pennoni, F. (2026). *Dispensa di Inferenza Bayesiana: Teoria e applicazioni con R e SAS*. Dipartimento di Statistica e Metodi Quantitativi, Università degli Studi di Milano-Bicocca.

Reich, B. J. (2026). *Bayesian Statistical Methods: With Applications to Machine Learning*. CRC Press.

Robert, C., Casella, G. (2004). *Monte Carlo Statistical Methods* (2nd ed.). Springer-Verlag, New York.

SAS Institute. (2012). *SAS/STAT® 12.1 User's Guide: The MCMC Procedure (PROC MCMC)*.

R Core Team (2026). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

## **Periodo di erogazione dell'insegnamento**

1° semestre, Ciclo II, Novembre 2026 - Gennaio 2027

## **Lingua di insegnamento**

L'insegnamento è erogato in lingua italiana. Gli studenti Erasmus possono utilizzare il materiale didattico disponibile in lingua inglese, fornito dal docente su richiesta. Possono inoltre richiedere di sostenere la prova d'esame in lingua inglese.

## **Sustainable Development Goals**

SALUTE E BENESSERE

---