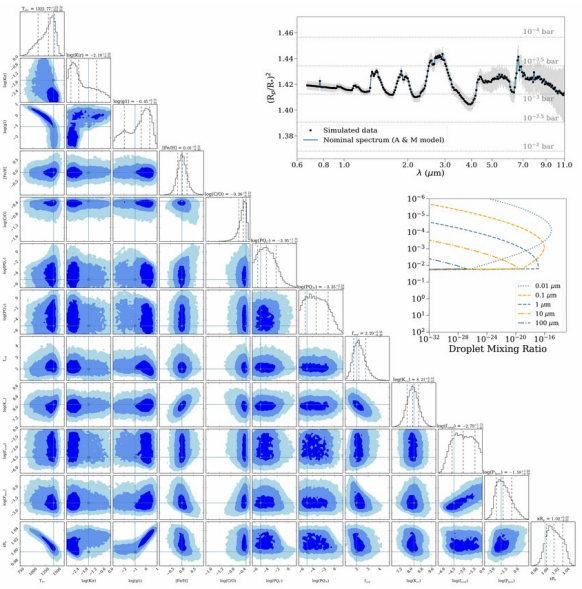

Università degli studi di Milano
Dipartimento di Fisica

Ongoing notes

Computational Methods in Physics

Andrea Clini[1]

winter 2026

Version 0.0.0

[1]PhD student in Physics, University of Milan, andrea.clini@unimi.it,webpage

Abstract

These are never-ending notes where I collect basic statistical ideas while learning them. Hopefully they will grow into something useful (also for others). We should cite something [**?**]

Statement of Non-Originality

These notes are far from being original. They simply rearrange material from courses, books and conversations.

Joining the project

If you find these notes useful, if you find typos or if you would like to collaborate or suggest changes, do get in touch!

README: Collaborating, conventions and examples (in progress)

Conventions for Notes, Questions, TODOs, Problems etc

- Use `\answertag{...}`, `\notetag{...}`, `\questiontag{...}`, `\todotag{...}` for page margin notes (green, yellow, orange, red, respectively). As a rule of thumb
  - `\answertag{...}` for answer
  - `\notetag{...}` for notes or minor problems
  - `\notetag{...}` for questions or problems
  - `\notetag{...}` for things to be done or big problems
- Use `\answerinline{...}`, `\noteinline{...}`, `\questioninline{...}`, `\todoinline{...}` for inline notes (for equations or wherever you like).
- For brevity, use (cite) to denote missing citations and (*ref) to denote missing references to labels.

Examples

This is some text that needs some note. More text that raises some questions.
This sentence contains an inline note Check this value! and continues without breaking the line. And now some more text with some answers. Next we have an example in math mode

$$E = mc^2. \quad \text{Explain this equation is often false.}$$

A margin note /minor problem

A question or problem

An answer.

Conventions for Labelling

Labeling is in snake_case in the form `\label{TYPE:name_in_snake_case}` with some meaningful name.

- Sections, subsections etc are all labelled with `\label{sec:name_of_section`
- Equations are labelled with `\label{eq:name_of_equation}`
- Figures are labelled with `\label{fig:name_of_equation}`
- And so on ...

a Todo or big problem

Something else that should be decided beforehand

I might very well have forgotten something...

Introduction

These notes collect various materials regarding computational methods in Physics... They are mainly for self-learning and rearranged from different sources...

Structure of the work.

The work is organized as follows...

Contents

Part I

Statistics

1 Basics

1 Frequentist vs Bayesian statistics

A frequentist treats parameters as fixed and uncertainty as arising from hypothetical repetitions of the data. Estimators are random variables, and properties such as unbiasedness or coverage are assessed over repeated samples.

A Bayesian treats parameters as random, encoding prior beliefs through a distribution and updating them via Bayes' rule

$$\underbrace{p(\theta \mid x)}_{\text{posterior}} = \frac{p(x \mid \theta)\, p(\theta)}{p(x)} \propto p(x \mid \theta)\ \underbrace{p(\theta)}_{\text{prior}}.$$

The posterior combines prior information with the likelihood and turns point estimates into full distributions.

Example 1.1. Consider flips of a biased coin with heads count k out of n.

- Frequentist: the maximum-likelihood estimate is $\hat{p} = k/n$ and a 95% confidence interval is $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n}$.
- Bayesian: with a Beta(a, b) prior, the posterior is Beta$(a + k, b + n - k)$, summarised by its mean or credible interval.

Both views often agree for large n, but diverge when data are scarce or priors encode substantial structure.

Why cosmologists use Bayesian & CERN people use frequentist statistics?
The answer is simple. Collider and other high-energy experiments can be repeated many times under controlled conditions, allowing frequentist methods to shine through repeated sampling properties. Cosmological observations are unique and non-repeatable, making Bayesian methods more suitable for incorporating prior knowledge and handling uncertainties in a coherent way.

Profiled vs marginalized likelihoods
Frequentist analyses often use profiled likelihood: nuisance parameters, and parameters of interest that we temporarily want to be inclusive for, are simply set to their best-fit values i.e. to their maximum-likelihood estimates. This is computationally simpler but can underestimate uncertainties by ignoring the full distribution of nuisance parameters.

Bayesian analyses marginalize over nuisance parameters and PoI to be 'ignored' by actually integrating over all their possible values weighted by their posterior distributions. This captures the full uncertainty but can be computationally intensive, especially in high-dimensional parameter spaces. Alternatively, we could also integrate them over the prior only, but this is less common and accurate.

2 Monte Carlo Methods

Monte Carlo methods approximate expectations

$$\mathbb{E}_\pi[f(X)] = \int f(x)\,\pi(x)\,dx$$

by averaging samples from a target distribution π. When drawing independent samples is hard, we build a Markov chain whose stationary distribution is π. If the chain is ergodic, long runs approximate the desired expectation.

1 Monte Carlo Markov Chains

1.1 Metropolis–Hastings algorithm

Suppose our target distribution has pdf $\propto \pi(x)$ with unknown normalization. The Metropolis–Hastings algorithm constructs a Markov chain with stationary distribution π, that will eventually converge and sample from it. Given a current state x_t and a proposal pdf $q(\cdot \mid x_t)$ of your choice, which obeys

- it is centered around the current state x_t;
- it is NOT required to be symmetric around x_t (the symmetric case is called just Metropolis);
- further properties of $q(\cdot \mid x_t)$ might depend on x_t e.g. the variance;
- it is easy to sample from.

The algorithm proceeds as follows:

1. Draw $x' \sim q(\cdot \mid x_t)$ and propose it as the new state.
2. Accept with probability $A(x_t \to x_{t+1}) = \min\left(1, \frac{\pi(x')\,q(x_t|x')}{\pi(x_t)\,q(x'|x_t)}\right)$; otherwise stay at x_t. The transition probability is then
$$T(x_{t+1} \mid x_t) = q(x_{t+1} \mid x_t)\,A(x_t \to x_{t+1}).$$
3. Set $x_{t+1} = x'$ if accepted, else $x_{t+1} = x_t$ and iterate.

The acceptance step enforces detailed balance, so that it leaves a stationary distribution π invariant. Indeed we have

$$\frac{A(x \to y)}{A(y \to x)} = \frac{\pi(y)\,q(x \mid y)}{\pi(x)\,q(y \mid x)} \Rightarrow \pi(x)\,T(y \mid x) = \pi(y)\,T(x \mid y).$$

Then, by definition of transition probability if $q(\cdot|x_i) =: \pi_i \sim \pi$ at step i, we have

$$\pi_{i+1}(x) = \sum_y \pi_i(y)\,T(x \mid y) = \sum_y \pi(y)\,T(x \mid y) = \sum_y \pi(x)\,T(y \mid x) = \pi(x) \underbrace{\sum_y T(y \mid x)}_{=1} = \pi(x).$$

Practical remarks

Choose proposals that mix well, discard early burn-in, and thin only if storage is a concern. Autocorrelation reveals whether the chain is exploring the space or getting stuck.

Why does Metropolis–Hastings work? It is easier to understand in the symmetric case (Metropolis). Then, we have $q(x|y) = q(y|x)$ and the acceptance ratio simplifies to

$$A(x \to y) = \min\left(1, \frac{\pi(y)}{\pi(x)}\right).$$

This means that we always accept moves to higher probability states, and accept moves to lower probability states with a probability proportional to how much lower the probability is. This allows the chain to explore the state space while still favoring high-probability regions, leading to convergence to the target distribution π over time.

1.1.1 Metropolis-Hastings for bayesian statistics

Suppose we have a prior distribution $p(\theta)$ over parameters θ and a likelihood $p(x \mid \theta)$ for data x. The posterior distribution is given by Bayes' theorem

$$p(\theta \mid x) = \frac{p(x \mid \theta)\, p(\theta)}{p(x)} \propto p(x \mid \theta)\, p(\theta).$$

The issue is that we do not know the normalization constant $p(x)$. We might just want to integrate over all the paramater space and divide by the volume, but this is often intractable. Indeed, Monte Carlo methods are precisely designed to approximate such integrals, ofetn intractable otherwise because of high dimensionality.

Then we define $\pi(\theta) := p(\theta \mid x) = \frac{1}{N} p(x \mid \theta)\, p(\theta)$ for some unknown normalization constant N and use Metropolis-Hastings to sample from the posterior. Note that in the acceptance ratio, the normalization constant cancels out

$$A(\theta \to \theta') = \min\left(1, \frac{\pi(\theta')\, q(\theta \mid \theta')}{\pi(\theta)\, q(\theta' \mid \theta)}\right) = \min\left(1, \frac{p(x \mid \theta')\, p(\theta')\, q(\theta \mid \theta')}{p(x \mid \theta)\, p(\theta)\, q(\theta' \mid \theta)}\right).$$

This allows us to sample from the posterior distribution without knowing the normalization constant, enabling Bayesian inference through MCMC methods.

Once we have the normalization, we can draw level lines and use again Monte Carlo and Metropolis-Hastings to compute confidence regions. Namely, we take the maximum of the posterior, and then draw level lines at decreasing values until we enclose the desired probability mass. The probability mass inside a region R, e.g. $R = \{p(\theta \mid x) \geqslant L\}$ for some level L, is

$$P(R) = \int_R p(\theta \mid x)\, d\theta.$$

This is approximately computed by Monte Carlo using again e.g. Metropolis-Hastings to sample from the posterior and counting the fraction of samples that fall inside R.

Part II

Machine Learning & Data Analysis

3 Machine Learning

1 Basics of machine learning

Part III

Numerical Methods & Optimization

4 Numerical Methods for PDEs

5 Numerical Linear Algebra

1 Matrix decompositions

6 Optimization

1 Convex Optimization

Part IV

Computational & Statisical Softwares

7 Mathematica

1 Symbolic computation