An emcee example

11/29/16

Project 4 Relevance...

Project 4 - Bayesian Calculations

Exercise 1:

Create a Bayesian Regression class that takes a log posteriori, data points, number of walkers, etc as initilization and provides all required methods and attributes required for Bayesian Regression Problems. The class should also be able to provide an integration method that allows you to integrate with respect to the posteriori distribution. Additionally, the class should allow to calculate all probabilities including the predictive distribution. Also include methods for visualization such as corner maps. Even though this class is part of a project you should consider it as your first self build Data Science tool.

The class should also be able to provide an integration method that allows you to integrate with respect to the posteriori distribution.

Bayesian Theorem

We can interpret the Bayes' Rule

$$prob(\Theta|\mathcal{X}) \quad = \quad \frac{prob(\mathcal{X}|\Theta) \;\cdot\; prob(\Theta)}{prob(\mathcal{X})}$$

as

$$posterior \quad = \quad \frac{likelihood \;\cdot\; prior}{evidence}$$

This formula looks pretty innocent. There are 3 terms that we have to divide. The problem is the evidence which is the normalization constant

$$P(\mathcal{D}) = \int_{\theta \in \Theta} d\pi(\theta) f(x|\mathcal{D})$$

This integral is already analytically dificult to calculate for the simplest models. Even numerical integration blows up for simple models in low dimensions.

- So what do we do if we can't calculate the integral?
- We remember that the integral of a function $f(X)$ from a measure theoretic standpoint is the weighted average of that function.

Enter Monte Carlo Integration, and eventually MCMC….

What’s the difference between a Monte Carlo Simulation and a Monte Carlo Integration?

Comparing and Contrasting: MC Simulation vs Integration

Both are types of Monte Carlo Methods

Monte Carlo Method: A technique that can be used to solve a mathematical or statistical problem

Monte Carlo Simulation:

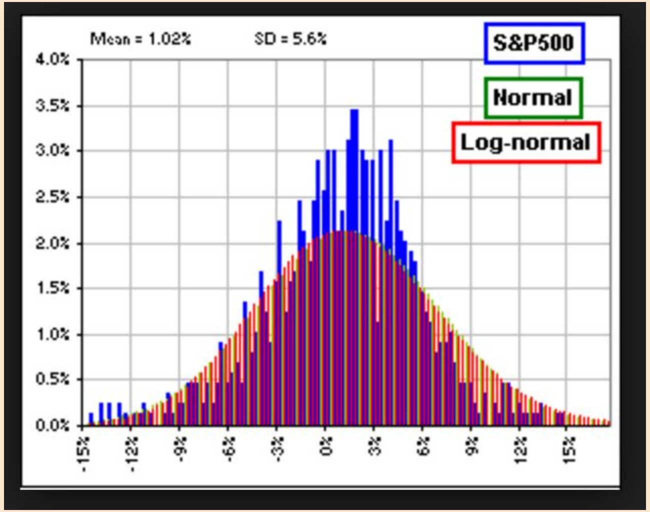
- A type of Monte Carlo method that uses repeated sampling to determine the properties of some phenomenon (or behavior).
- This method provides a random sample from the posterior distribution in Bayesian inference.
  - This sample, then approximates and summarizes all essential features of the posterior.

Monte Carlo Integration:

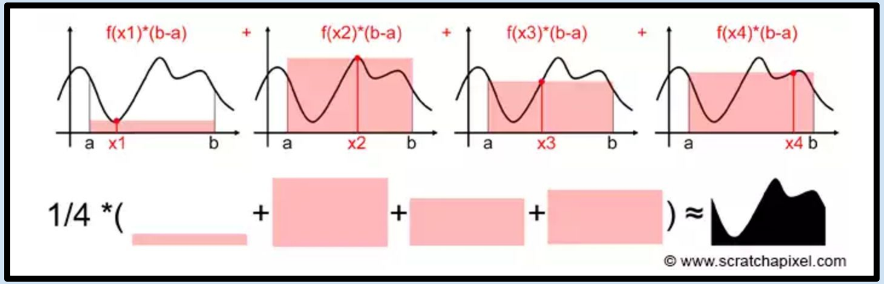
- A type of Monte Carlo method used for numerical integration. It uses random sampling to calculate the numerical value of a definite integral.
- There are different methods to perform a monte carlo integration, such as uniform sampling and importance sampling.

Comparing and Contrasting: MC Simulation vs Integration

Monte Carlo Simulation:



Monte Carlo Integration:



What’s the difference between a Monte Carlo Integration and Markov Chain Monte Carlo (MCMC)?

Comparing and Contrasting: MC Integration vs Markov Chain MC

- MCMC takes MC Integration one step further...or rather many steps
- Key term to think about when using either: **convergence**

Monte Carlo Integration:

- The random samples of the integrand are statistically independent.

MCMC:

- The random samples of the integrand are correlated.
    - A Markov Chain is constructed in such a way as to have the integrand as its equilibrium distribution.

Markov Chain Monte Carlo

- There are several MCMC methods. These methods are a class of algorithms for sampling from a probability distribution based on constructing a Markov Chain that has the desired distribution of its equilibrium distribution.
  - The state of the chain after a number of steps is then used as a sample of the desired distribution.
  - The quality of the sample improves as a function of the number of steps (in other words, more steps = better).
- When an MCMC method is used for approximating a multi-dimensional integral, an ensemble of “walkers” move around randomly.
  - At each point, where a walker steps, the integrand value at that point is counted towards the integral.
  - The walker then may make a number of tentative steps around the area, looking for a place with a reasonably high contribution to the integral.

MARKOV CHAIN MONTE CARLO

Random walks on a graph

You implicitly pick a graph G

You implicitly choose transition probabilities for the edges to make the stationary distribution P.

Project 4 Relevance...

Project 4 - Bayesian Calculations

Exercise 1:

Create a Bayesian Regression class that takes a log posteriori, data points, number of walkers, etc as initilization and provides all required methods and attributes required for Bayesian Regression Problems. The class should also be able to provide an integration method that allows you to integrate with respect to the posteriori distribution. Additionally, the class should allow to calculate all probabilities including the predictive distribution. Also include methods for visualization such as corner maps. Even though this class is part of a project you should consider it as your first self build Data Science tool.

The class should also be able to provide an integration method that allows you to integrate with respect to the posteriori distribution.

Bayesian Theorem

We can interpret the Bayes' Rule

$$prob(\Theta|\mathcal{X}) \quad = \quad \frac{prob(\mathcal{X}|\Theta) \ \cdot \ prob(\Theta)}{prob(\mathcal{X})}$$

as

$$posterior \quad = \quad \frac{likelihood \ \cdot \ prior}{evidence}$$

After all, inferring the value of the parameters given our model and the observed data, is very simple and can be accomplished in 3 steps:

1. Specify the **priors**;
2. Estimate the **likelihood**;
3. Compute the **posterior**.

...but, how can we compute the posterior?

2. Computing posterior PDFs with Markov Chain Monte Carlo (MCMC)

2.1 Purpose of MCMC:

Build a series of points in the space of parameters, i.e. a *chain*, $ (0), (1),, (t), (t+1),, (M) $ such that the position $\vec\theta(t+1)$ only depends on the position of $\vec\theta(t)$.

2.2 Crucial property:

The **density** of the samples in the chain is directly proportional to the posterior PDF.

2.3 Result:

The chain converges to a *stationary* state where successive elements of the chain are samples of the target posterior distribution $p(\vec\theta|d)$.

This means that, once we have obtained the chain of M samples, we have everything we need. We can compute:

- the marginalized distribution of each parameter θ_i by simply approximating it with the histogram of the samples projected into the parameter space spanned by θ_i (see below for details);
- the **expectation value** of a function f of the model parameters: $$ f(\vec\theta)\, =\, \int p(\vec\theta|d)\ f(\vec\theta)\ \mathrm{d}\vec\theta \approx \frac{1}{M}\sum_{t=0}^{M-1} f(\vec\theta(t)) $$

emcee

The MCMC Hammer

- emcee is a python module that implements a very cool MCMC sampling algorithm called an ensemble sampler. In order to more efficiently sample the parameter space, many samplers (called walkers) run in parallel and periodically exchange states.

- The MCMC Ensemble Sampler in Python (AKA: emcee) is designed for Bayesian parameter estimation.
- The emcee library is setup to where you have to write the entire probability function yourself. Following the example on emcee's site, we do this by writing log-prior, log-likelihood, and log-probability functions.

Let’s follow along a simple example...

Go to site -- they take the example further:

http://eso-python.github.io/ESOPythonTutorials/ESOPythonDemoDay8_MCMC_with_emcee.html

Let’s reverse engineer this!

Jupyter notebook with example:

https://github.com/kgracia44/DSI-TA-Reviews/blob/master/DSI_3_Keplers/Week_05/Example_emcee_Keplers.ipynb

Other sources

Other example using emcee:
https://users.obs.carnegiescience.edu/cburns/ipynbs/Emcee.html
Example of monte carlo integration:
http://code.activestate.com/recipes/577263-numerical-integration-using-monte-carlo-method/
Markov info (go to the answer):
http://stats.stackexchange.com/questions/165/how-would-you-explain-markov-chain-monte-carlo-mcmc-to-a-layperson
Markov info:
https://jeremykun.com/2015/04/06/markov-chain-monte-carlo-without-all-the-bullshit/
Monte Carlo Integration in Bayesian Estimation:
https://engineering.purdue.edu/kak/Tutorials/MonteCarloInBayesian.pdf
Monte Carlo Methods in Practice:
http://www.scratchapixel.com/lessons/mathematics-physics-for-computer-graphics/monte-carlo-methods-in-practice/monte-carlo-methods