Simulation Techniques & Estimation in ML

Vivek Sivaramakrishnan

## Introduction

In this workshop, we introduce some simple-to-comprehend statistical problems that have some *bizarre* results. Through the lens of these problems, we introduce the powerful idea of **Simulation** which can help to approximately solve the problem, and in most cases, shed insight on the analytical solutions to these problems. Codes for simulation are incredibly easy and quick to write, and they’re generally preferred over numerical methods to solve complex problems in practice.

Discussion of ideas from simulations naturally extends to the realm of estimation and Machine Learning. We discuss some powerful ideas in this context, and implement them on real-life datasets.

## Abstract

The contents of this workshop are interspersed with hands-on coding exercises that the participants are encouraged to try. Python is recommended, and is used by the instructor while coding along in the workshop. The implementation of simulation techniques and the theoretical solutions in the context of these problems provides an enjoyable medium to integrate theory with practice; analytical probability and simulation code.

## Duration

There are 2 sessions, 3 hours each.

| Session | Topics Covered |
| --- | --- |
| 1 | Randomness, 0-1 Sampling methods, Approximating , Approximating , Inverse-Transform methd, Monty Hall Problem, Bobo Lineage Problem, Snake & Ladders (with a twist) |
| 2 | Maximum Likelihood estimation, Continuous Inverse Transform and Box-Muller Transform, Concavity, Majorization-Minimization framework, EM Algorithm, Mixture Modelling, and EM for missing/censored data. |

## Pre-requisites

This workshop caters to participants of diverse backgrounds and assumes no prior academic expertise. However, a background in programming in Python will help with the coding exercises. Statistics at the High-School level is more than sufficient for the contents covered in the workshop.

## Detailed Plan

### Session 1

* Approximating
  + Grid vs. Simulation (Introduce Law of Large Numbers)
  + How to sample uniform random numbers? Enter Mixed Congruential Method
  + Simulation to estimate
* Approximating by simulation.
* Analysis of distribution of distances between local maximas of any continuous random variable.
* Simulation to find optimal strategy for the Monty Hall Problem
  + How to toss a coin? Enter Discrete Inverse transform method.
* Simulation to solve the Bobo Amoeba extinction Problem
* Simulation to find expected number of moves to finish the circular Snake & Ladders Game.

### Session 2

* Maximum likelihood estimation.
* Example MLE problem involving estimating parameters of a poisson distribution.
* Using discrete inverse transform to sample from Poisson.
* Example MLE problem wrt exponential distribution.
* Sample exponential using continuous inverse transform.
* Example MLE problem involving Gaussian distribution.
* Box-Muller transform to sample from Normal.
* Example MLE on mixture of Gaussians (no closed form)
* Composition method to sample from mixture of Gaussians
* Discussion on concavity, and the MLE not being concave
* Intro to the Majorization-Minimization framework.
* Discussion of the EM algorithm, and latent variables
* EM algorithm for GMMs, and obtaining MLE estimates
* Example MLE on *censored* data
* Explanation of EM in this context