# ETC 2420/5242 Lab 11 2016

*Souhaib Ben Taieb*

*Week 11*

## Purpose

This lab is to practice Monte Carlo sampling methods.

## Question 1

Suppose we want to approximate the number $\pi$. Imagine throwing lots of darts at a 1 meter by 1 meter square board. Draw a circle of diameter 1 on the board. If you throw the darts uniformly so that you are equally likely to hit any point on the square, then after throwing many many darts, you would expect the proportion of darts landing inside the circle to be equal to its area, $\pi$. The more darts you throw, the better the approximation of $\pi$.

Simulate this experiment by generating many points on the unit square $[-1, 1] \times [-1, 1]$, then counting how many landed in the circle. How good is your approximation if you throw 100 darts? 1,000 darts? 10,000 darts? 100,000 darts?

## Question 2

Suppose it is hard to sample from $f(x)$. Rejection sampling uses random samples from another density $g(x)$ we know how to sample from. First find a constant $c$ such that $f(x) \leq cg(x)$ for all $x \in \mathcal{X}$, then follow these steps:

1. Generate a random sample x with the density $g(x)$.
2. Generate a uniformly distributed random sample $u$ on the interval $\mathcal{X}$. If $u \leq \frac{f(x)}{cg(x)}$, then output $x$; otherwise reject $x$ and return to step 1.

We will use rejection sampling to generate random samples from the density function $f(x) = \frac{x(1-x)e^x}{3-e}$ with $x \in [0, 1]$ using a uniform proposal, i.e. $g(x) = 1$ for $x \in [0, 1]$.

a. Write a function that takes $c$ as argument and returns a random sample $x$ from $f(x)$ if the sample is accepted or $-1$ if it is rejected.
b. Run your functions 1000 times for $c \in \{1.56, 2, 5, 20\}$. Compute the percentage of rejected samples. Plot an histogram of the accepted samples (using freq=FALSE so it plots the density) together with the density of $f$. What do you observe when you increase the value of $c$? Why?

## Question 3

Suppose we cannot easily sample from the probaiblity density $f$. The Markov Chain Monte Carlo (MCMC) method allows us to sample from $f$ by constructing a Markov chain $X_1, X_2, \ldots$, whose stationary distribution is $f$.

The Metropolis-Hastings (M-H) algorithm is a specific MCMC method that works as follows. Let $q(y|x)$ be an arbitrarly, friendly distribution (i.e. we know how to sample from $q(y|x)$), also called the *proposal distribution*. The M-H algorithm creates a sequence of observations $X_0, X_1, \ldots$, as follows.

Choose $X_0$ arbitrarily. Suppose we have generated $X_0, X_1, \ldots, X_i$. To generate $X_{i+1}$, do the following:

a. Generate a *proposal* or *candidate* value $Y \sim q(y|X_i)$.
b. Evaluate $r(X_i, Y)$ where

$$r(x,y) = min\left\{1, \frac{f(y)q(x|y)}{f(x)q(y|x)}\right\}$$

.

c. Set

$$X_i = \begin{cases} Y & \text{with probability } r \\ X_i & \text{with probaiblity } 1-r \end{cases}.$$

Remark 1: A common choice for $q(y|x)$ is $\mathcal{N}(x, b^2)$ for some $b > 0$. In that case, because $q(y|x) = q(x|y)$, $r = min\left\{\frac{f(y)}{f(x)}, 1\right\}$.

Remark 2: A simple way to execute (c) is to generate $U \sim \text{Uniform}(0,1)$. If $U < r$, set $X_{i+1} = Y$ otherwise $X_{i+1} = X_i$.

We want to generate samples from the Cauchy distribution that has density

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

a. What is $r(x, y)$ in this case?

b. Using $X_0 = 0$, run the simulator for $b \in \{0.1, 1, 10\}$. Plot an histogram of the samples you obtain for each value of $b$, as well as the samples over time, i.e. $(i, X_i)$.

c. Using the previous plots, explain the behavior of the chain for the different values of $b$.

## TURN IN

- Your `.Rmd` file
- Your Word (or pdf) file that results from knitting the Rmd.
- Make sure your group members are listed as authors, one person per group will turn in the report
- DUE: Wednesday after the lab, by 7am, loaded into moodle

## Resources

- Lecture slides on Bayesian reasoning