Homework 3 - numpy, scipy and stats

1. Rejection Sampling

Rejection sampling is a basic Monte Carlo technique to generate observations from a specified distribution with probability density function (pdf) $f(x)$. In practice, it is used to draw random variables from a target distribution, $f(x)$, when directly sampling from that distribution is difficult. The basic idea is to instead sample from a reference distribution, $g(x)$, which a) is easy to draw from and b) satisfies the condition that $g(x) > 0$ whenever $f(x) > 0$, and to then selectively accept samples such that sampling from f is achieved.

The basic rejection sampling algorithm is the following:

- Begin with a reference distribution, g and a scalar M such that $f(x) < Mg(x)$ for all x.
- Until a specified number of samples is achieved:
  1. Sample a candidate, x_c, from $g(x)$ and a value u from a uniform distribution over $[0, 1]$ (Hint: See `scipy.stats.uniform`)
  2. if $u < \frac{f(x_c)}{Mg(x_c)}$ then accept x_c as a sample from $f(x)$, else reject x_c.

(a) Write a 1-D rejection sampler in `python` using tools from `scipy.stats`. The sampler should take as input: the functional form of the target distribution, a reference distribution in the form of a scipy.stats object (e.g. `scipy.stats.norm(0,1)`), and the number of samples desired. It should output the samples, the value of M, and the proportion of samples that was accepted.

(b) The Laplace distribution (a.k.a. double exponential distribution) has the pdf

$$f(x) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right) \tag{1}$$

Use your rejection sampler to draw a sample of size 1000 from the Laplace distribution with parameters $\mu = 0, b = 1$. Use a Cauchy distribution as the reference distribution. Plot a histogram of your samples, over-plotting the true Laplace pdf. Use a Kolmogorov-Smirnov test to test the hypothesis that your sample was drawn from a Laplace(0,1) distribution.

(c) Again use your sampler to draw a sample of size 1000 from the Laplace distribution, but this time use a Student's t distribution with 2 degrees of freedom as the reference distribution. What is the acceptance rate? How does it compare to the acceptance rate from part (b)?

(d) Generate a sample of size 5000 from a continuous distribution of your choosing. (Hint: Make sure to use an appropriate reference distribution). Plot a histogram of your samples, over-plotting the true pdf.