MCMC Sampling in High Dimensions

The Normal Distribution

The Normal distribution is a continuous probability distribution with two parameters:

- the mean ("average") of the distribution which defines which its *location*, usually written μ (`mu`)
- the standard deviation ("variability") which defines the *scale*, usually written σ (`sigma`). (Note: standard deviation is σ, variance is σ^2).

The *standard normal distribution* is the normal distribution with a mean of zero and a variance of one.

Normal distributions are often used to represent a real-valued random variable whose true distribution isn't known.

Stan provides the function `normal_rng` which takes two real-values arguments `mu` and `sigma` which specify the location and scale, respectively. (Note: `sigma` must be positive). This function may only be used in generated quantities block.

Other "Bell Curve" Distributions: Cauchy, Logistic, Student's t

The shape of the probability densities functions for these distributions are all bell-shaped, however they have more area under the tail of the curve than the Normal does, thus the extreme quantiles are farther from the mean. For the Cauchy distribution, the area under the tail is infinite! These distributions are useful when the amount of variance is unknown.

Stan provides functions `cauchy_rng`, `student_t_rng`, and `logistic_rng`, all of which take location and scale parameters `mu` and `sigma`.

The MultiNormal Distribution

The multivariate normal distribution is the extension of the normal distribution of a single random variable to a vector of random variables (a "random vector"). Each element of a random vector is a variable which is normally distributed, i.e., each element is drawn from its own normal distribution with parameters `mu` and `sigma`.

The following code snippet shows how to generate a standard random vector of length N in Stan:

```
vector[N] rv;
for (i in 1:N) rv[i] = normal(0, 1);
```

- What is the Expectation of a random variable generated from a standard normal distribution?
- What is the Expectation of a random vector generated from a standard multinormal distribution?

Exercise 1: Choice of location and scale parameters for the Normal distribution

- Write a program `gen_norm_variate.stan` which generates a random variate `x` for some choice of `mu` and `sigma`. You can either specify `mu` and `sigma` directly in the program or pass these values in as data.
- Use this program to generate a sample, as before, then create a histogram plot to show the resulting density of `x` when `mu` is 0 and `sigma` is 1.
- Repeat this several times, choosing both positive and negative values of `mu`, and values for `sigma` less than 1 as well as greater than 1.

Exercise 2: Choice of location and scale parameters for the Cauchy, Student's t, and logistic distributions

- Write a program `gen_cauchy_variate.stan` which generates a random variate `x` for some choice of `mu` and `sigma`.
- Use this program to generate a sample
- Using the same values of `mu` and `sigma`, generate a sample using program `gen_norm_variate.stan`
- Create histogram plots for both, compare and contrast.

Exercise 3: Expectations of the MultiNormal distribution and the curse of dimensionality

- Write a program `gen_std_norm_vector.stan` which generates a standard random vector `X` of length N for some choice of N *and* which computes the distance between X and the origin as quantity of interest `dist_to_origin`.
- Generate an x,y plot for x in 1:256 where x is the length of the random vector X and y is the expectation of the distance to the origin when X is a standard MultiNormal random vector. (You should be able to write a Stan program which computes the distance to origin for vectors of length 1:256 all in one go.)