Assignment 1. Due February 6

Corrections: none yet.

1. (*This exercise has two purposes. One is to understand why a sampler might not work well in high dimensions. Part of the "curse of dimensionality" is that low dimensional intuition may be wrong in high dimensions. Another is to understand why some functions have good Gaussian approximations. The analytical method is called Laplace's method.*) Let C_n be the cube in n dimensions, symmetric about the origin, whose side is length 2. This may be written $C_n = [-1, 1]^n$. The n dimensional volume of C_n is $\text{vol}_n(C_n) = 2^n$. If $x \in \mathbb{R}^n$, then $x \in C_n$ if $|x_k| \leq 1$ for all k. You can generate a "random" point $X \in C_n$ by taking $X_k = 2U_k - 1$, where the U_k are independent standard uniforms. This X has uniform probability density inside C_n, which is $f(x) = 2^{-n} = 1/\text{vol}_n(C_n)$ if $x \in C_n$, and $f(x) = 0$ if $x \notin C_n$. Let B_n be the unit ball in n dimensions, so $x \in B_n$ if $(x_1^2 + \cdots + x_n^2)^{1/2} \leq 1$. Clearly $B_n \subset C_n$. We can generate X uniformly distributed in B_n by generating X uniformly distributed in C_n and accepting it if $X \in B_n$. The efficiency of this algorithm is the ratio of the volumes

$$Z_n = \frac{\text{vol}_n(B_n)}{\text{vol}_n(C_n)} \; .$$

This exercise derives an approximate formula for Z_n. The formula shows that shows that $Z_n \to 0$ as $n \to \infty$, exponentially. Therefore the sampling method impractical for large n.

Note: It is possible to find the large n behavior of $I(n)$ (below) using the change of variables $r^2/2 = s$ to express it in terms of the Γ function, whose asymptotics are available on wikepedia – *Stirling's formula*. Please don't do it this way. The asymptotics of Γ are found using the method of this problem, so that approach is not actually easier.

Note: The *unit sphere* in n dimensions is $S_{n-1} = \{|x| = 1\}$. A point x in n dimensional space may be written $x = r\omega$, where $r \geq 1$ is the length, and $\omega \in S_{n-1}$. The $n-1$ dimensional surface element on S_{n-1} is written $d\omega$. The $n-1$ dimensional "surface area" ($n-1$ dimensional volume) of S_{n-1} is called ω_{n-1} and is given by

$$\omega_{n-1} = \int_{S_{n-1}} d\omega \; .$$

Some values are $\omega_1 = 2\pi$ (the circle in $2D$) and $\omega_2 = 4\pi$ (the sphere in $3D$). For the unit sphere S_2 in $3D$, you may have seen a formula like $d\omega = \cos(\phi) d\theta d\phi$ in terms of spherical polar angles. We don't need formulas like this here. We only need the general polar coordinate integration rule $dx = r^{n-1} d\omega dr$, which means that

$$\int_{\mathbb{R}^n} f(x)\, dx = \int_0^\infty \int_{S_{n-1}} f(r\omega) r^{n-1} d\omega dr \ .$$

(a) Show that

$$\mathrm{vol}(B_n) = \frac{\omega_{n-1}}{n} \ .$$

You can do this by

$$\mathrm{vol}(B_n) = \int_{x \in B_n} dx \ .$$

(b) Show that

$$\omega_{n-1} = \frac{(2\pi)^{n/2}}{I(n)} \ ,$$

where

$$I(n) = \int_0^\infty r^{n-1} e^{-r^2/2}\, dr \ . \tag{1}$$

Hint: integrate

$$(2\pi)^{n/2} = \int_{x \in \mathbb{R}^n} e^{-|x|^2/2}\, dx$$

in polar coordinates.

(c) Write $I(n) = \int e^{-\phi(r)}\, dr$ and identify ϕ. Show that ϕ has a unique maximum value achieved at r_*. Calculate $\phi''(r_*)$, $\phi'''(r_*)$, and possibly one more. Let $q(r)$ be the quadratic Taylor approximation to $\phi(r)$ about r_*, which is

$$q(r) = \phi(r_*) + \tfrac{1}{2}\phi''(r_*)(r - r_*)^2 \ . \tag{2}$$

Write the formula for

$$J(n) = \int_{-\infty}^\infty e^{-q(r)}\, dr \ .$$

(d) $J(n)$ is an approximation of $I(n)$. The error is written $K(n) = I(n) - J(n)$. Show that

$$\frac{K(n)}{I(n)} \to 0 \quad \text{as} \quad n \to \infty \ .$$

Hint: there are two kinds of r values: those where the quadratic approximation (2) is accurate, and those where ϕ and q are much smaller than values that matter. For this exercise, you can take the "values that don't matter" set to be $|r - r_*| > n^p$ with $0 < p < \frac{1}{6}$. When $|r - r_*| = n^p$, then e^{-q} does not matter, and $q(r)$ is still relatively close to $\phi(r)$ (use ϕ''' to verify this).

(e) Write the large n asymptotic approximation of Z_n that shows that sampling uniformly in the ball by rejection from the cube is an exponentially bad idea.

2. (*Probability distributions usually depend on parameters. It may not be enough that a sampler "works" for each parameter value. It may need to be efficient uniformly over the parameter. This Exercise is an example of such a sampler. This exercise also demonstrates that some careful analysis can lead to good samplers.*) Let S_n for $n = 0, 1, \ldots$, be independent exponential random variables with rate parameter λ. Let these be the *inter-arrival* times for the arrival times T_n, which means that $T_0 = S_0$, and $T_n = T_{n-1} + S_n$ for $n > 0$. The sequence T_n is a *Poisson process* with *arrival rate* parameter λ. The goal is to find a sampler that samples T_n using an amount of work that is bounded as $n \to \infty$. A direct simulation of the Poisson process takes order n work, because you have to generate all the inter-arrival times from S_0 up to S_n. For simplicity, we take $\lambda = 1$.

   (a) Show that the probability density for T_n is $f_n(t) = \frac{t^n}{n!}e^{-t}$ if $t \geq 0$. Hint: $T_n = T_{n-1} + S_n$, with S_n independent of T_{n-1}, allows you to find f_n from f_{n-1}.

   (b) Determine the behavior of $f_n(t)$ for typical T_n values using the method of Exercise 1. Find the most likely value of T_n by maximizing f_n, then make a Gaussian approximation of f_n about this value, t_{n*}.

   (c) You can find the mean and variance of T_n from the representation of T_n as a sum of independent S_k for $k \leq n$. You can estimate the mean and variance of T_n from the Gaussian approximation of part (2b). Show that these ways of getting the mean and variance give (approximately?) the same result.

   (d) Explain why it is not a good idea to use the Gaussian approximation as a proposal distribution for rejection sampling of f_n.

   (e) Explore using a *double exponential* as a proposal distribution. That is $g_n(t) = \frac{1}{Z}e^{-\alpha_n|t-t_{n*}|}$. Calculate the normalization constant Z. To find the optimal Z you need to solve the two maximization problems, one for $t > t_{n*}$ and one for $t < t_{n*}$. Do not worry about negative T values. Those are rare for large n, and can be rejected for any n.

   (f) What formula for α_n is suggested by the Gaussian approximation? You can choose α_n so that the proposal distribution has the same or similar variance as the true distribution.

   (g) Determine wether this α_n leads to a sampler whose efficiency does not go to zero as $n \to \infty$. If so, you are done. If not, can you adjust α_n to make the sampler uniformly efficient?

3. (*Bayesian posterior*). Imagine an experiment measuring radioactive decay at n times $0 \leq t_1 < \cdots < t_n$. The measured value for time t_i is Y_i. We

want to fit these measurements to a simple exponential decay model

$$y = Ae^{-\lambda t} \ .$$

The probability model is that the measured Y_i is equal to the true value $y(t_i)$ plus mean zero variance σ^2 gaussian observation noise. That is

$$Y_i = Ae^{-\lambda t_i} + \sigma \xi_i \ , \quad \text{where} \ \ \xi_i \sim \mathcal{N}(0,1) \ .$$

The model has three parameters (A, λ, σ^2). Suppose the prior for these is $\pi(A, \lambda, \sigma)$. Write an expression for the likelihood function

$$L(A, \lambda, \sigma | Y_1, \ldots, Y_n) \ .$$

Write an expression for the posterior in terms of L and π. Is there a simple closed form formula for the normalization constant? We want to use a prior of the form

$$\pi(A, \lambda, \sigma) \sim \mathbf{1}_{[0,a]}(A) \, \mathbf{1}_{[0,l]}(\lambda) \, \mathbf{1}_{[0,s]}(\sigma) \ .$$

The *indicator function* $\mathbf{1}_{[a,b]}(x)$ is equal to 1 if $a \leq x \leq b$ and zero otherwise. We would like to make π uniformative by taking $a = \infty$, or $l = \infty$ or $s = \infty$. Two of these are possible, in that the posterior is normalizable, but one of them is not possible (the posterior would have infinite mass for any non-zero normalization constant). Which choice is not possible?

4. (*Programming exercise.*)

   Let $f(x)$ be the probability density $f(x) = \frac{1}{Z} x(1-x)$ for $0 \leq x \leq 1$ and $f(x) = 0$ otherwise. Write a program that generates N independent samples of f and makes a histogram of the results. You should plot this histogram with the theoretical PDF on the same plot to see quantitative agreement for large N. Choose bin size Δx so that the bins with the most hits get around 100 hits (between 20 and 2000, you should experiment with this). Draw one standard deviation error bars on the histogram values. Note how many of the error bars contain the true value.