1 Introduction

What's the best way to space intermediates for NCMC or an alchemical free energy calculations, or mixture sampling in general? Say, over λ.

Given some metric tensor g_{ij}, an infinitesimal length can be defined as

$$ds^2 = g_{11}dx_1^2 + g_{12}dx_1dx_2 + g_{22}dx_2^2 + ... \tag{1}$$

With the total length given between the initial and final points $\mathbf{x}_i$ and $\mathbf{x}_f$, respectively, is given by

$$L = \int_{\mathbf{x}_i}^{\mathbf{x}_f} ds \tag{2}$$

For probability distributions, an appropriate choice of metric tensor is the Fisher information metric

$$g_{ij} = \int p(x|\lambda)\frac{\partial \ln(p(x|\lambda)}{\partial \lambda_i}\frac{\partial \ln(p(x|\lambda)}{\partial \lambda_j}dx \tag{3}$$

so that

$$L = \int \sqrt{\sum_{i,j} g_{ij}dx_idx_j} \tag{4}$$

2 One dimension

2.1 Mixture of Gaussians

Looking at a mixture of Gaussian distributions centered on zero

$$p(x|\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-x^2}{2\sigma^2}\right) \tag{5}$$

Let's say we have two σs we wish to sample between. We can place $N-2$ intermediate Guassians between the two end points, but how should they be optimally spaced to ensure maximal overlap over each other? In the notation about, it'll be easier if we choose σ^2 rather than σ to be our λ.

To evaluating g_{ij}, it's easiest to first compute

$$\begin{aligned}\frac{\partial \ln(p(x|\sigma)}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma^2}\left[-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{x^2}{2\sigma}\right]\\ &= \frac{x^2}{\sigma^4} - \frac{1}{\sigma^2}\end{aligned}$$

So that

$$\begin{aligned}\left(\frac{\partial \ln(p(x|\sigma)}{\partial \sigma^2}\right)^2 &= (\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2})^2 \qquad (6)\\ &= \frac{x^4}{4\sigma^8} - \frac{x^2}{2\sigma^6} + \frac{1}{4\sigma^4} \qquad (7)\end{aligned}$$

The score is tricky to evaluate and involves knowing the moments of the Gaussian distribution. In any case, one finds that

$$g_\sigma = \frac{1}{2\sigma^4} \tag{8}$$

The thermodynamic length is therefore given by

$$
\begin{aligned}
L &= \int_{\sigma_i^2}^{\sigma_f^2} \sqrt{\frac{1}{2\sigma^4}(d\sigma^2)^2} && (9)\\
&= \frac{1}{\sqrt{2}} \int_{\sigma_i^2}^{\sigma_f^2} \frac{1}{\sigma^2} d\sigma^2 && (10)\\
&= \frac{1}{\sqrt{2}} \int_{\gamma_i}^{\gamma_f} \frac{1}{\gamma} d\gamma && (11)\\
&= \sqrt{2} \ln\left(\frac{\sigma_f}{\sigma_i}\right), && (12)
\end{aligned}
$$

where the change of variables $\gamma = \sigma^2$ was used in the third line. If we have a total number of N distributions, we want to space them at equal distances from each other, which means that each distribution has a distance $l = L/N$ from the previous.

If $f(\sigma)$ is a function that maps σ to a distance away from the distribution with σ_i then

$$
\begin{aligned}
& L = f(\sigma) \\
\implies & \sigma_n = f^{-1}\left(\frac{L}{n}\right), \qquad n = 1, 2, ..., N
\end{aligned}
$$

is how we should choose σ.

[graph with log spacing]