Kernel function and KDE estimate

The method uses a Fisher-distribution Kernel, which is a special case of the more general Fisher-Von-Mises distribution. It describes the distribution of points on a unit sphere or equivalently unit vectors in three dimensions. The kernel function centered on unit vector $\mathbf{X}_i$ is

$$K(\mathbf{x}; \mathbf{X}_i, \kappa) = C_0(\kappa) \exp(\mathbf{x} \cdot \mathbf{X}_i), \tag{1}$$

where

$$C_0(\kappa) = \frac{\kappa}{4\pi \sinh(\kappa)}. \tag{2}$$

Both $\mathbf{x}$ and $\mathbf{X}_i$ are unit vectors in three dimensions. The full KDE estimate for n data points $\mathbf{X}_i$ is then

$$\hat{f}(\mathbf{x}; \kappa) = \frac{1}{n} \sum_{i}^{n} K(\mathbf{x}; \mathbf{X}_i, \kappa). \tag{3}$$

Unit test: Check normalization

A basic test of the correctness of the implementation is to check the kernel is actually normalized correctly. Integration of the kernel over the whole sphere should yield a value of unity, and if it does not, then there is an error in the code.

For the purpose of the test, it is necessary to use numerical quadrature to check the normalization, so the truncation error of the quadrature scheme comes into play. Here we will use Gauss-Legendre integration. This approach provides high precision with only a few quadrature points.

There are two approaches to integrating the kernel to check normalization. Firstly, it is possible to apply the quadrature directly to the kernel function. The result should be close to unity for a correct implementation. This approach is straightforward, but it comes with a downside – the numerical quadrature will struggle when the concentration parameter (κ) is very large, and the kernels are very narrow. Departure from unity will occur, not due to error in the implementation of the kernel, but due to the breakdown of the integration scheme.

An alternative approach is to work with the log of the kernel. For a correctly normalized Fisher distribution,

$$\int_{S^2} \log(K) d\Omega = -4\pi \log C_0(\kappa), \tag{4}$$

where $S^2 = \{\mathbf{x} \in \mathbb{R}^3 : ||x|| = 1\}$. Therefore, the second test of normalization involves computing the integral of the log of the kernel and comparing it to Equation (4). The log of the kernel is generally smoother and doesn't encounter the same problems for large κ.

Figure 1 shows the absolute error in the normalization for the direct and log methods. A fixed number of quadrature points are used for all values of κ. When κ is small, the error is consistent with rounding error in both cases. When κ is large, however, the direct method indicates a large error: the normalization departs from unity. There is no similar problem for the log method. Therefore, the departure error is due to a failure of the numerical quadrature not a problem with the implementation of the kernel.

Unit test: Uniform distribution

On the sphere, the Mean Integrated Square Error (MISE), is defined as

$$\text{MISE} = \int_{S^2} (\hat{f} - f)^2 d\Omega. \tag{5}$$

It measures the average difference between the KDE estimate and the actual distribution f. Since it depends on f, which is generally unknown, the MISE cannot be directly computed for most problems. However, for the purpose of testing the KDE code, we can chose a known distribution for f and compute the MISE analytically. A comparison between the analytic MISE and that returned by the KDE code then serves as a basic test on the implementation.

Even with a known distribution, the MISE is difficult to compute analytically. However, for the particularly simple case of a uniform distribution, it can be done. On a unit sphere, the uniform distribution is

$$f(\theta, \phi) d\theta d\phi = \frac{1}{4\pi} \sin(\theta) d\theta d\phi, \tag{6}$$

where θ is the polar angle in the range $[0, \pi]$ and ϕ is the azimuthal angle in the range $[0, 2\pi)$. For this distribution and a Fisher KDE,

$$\text{MISE} = \frac{1}{n} \left(\gamma(\kappa) - \frac{1}{4\pi} \right), \tag{7}$$

where

$$\gamma(\kappa) = \frac{C_0^2(\kappa)}{C_0(2\kappa)}. \tag{8}$$

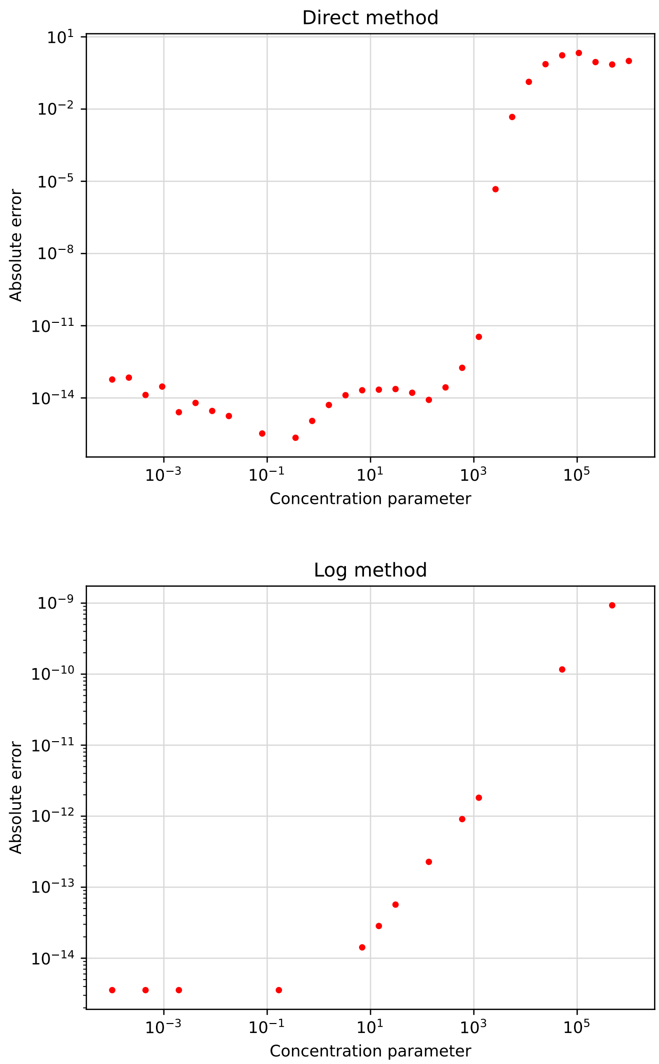

Figure 1: Left: Absolute error from directly integrating the kernel at a fixed quadrature resolution. Eventually κ becomes too big and the error grows because the quadrature is inaccurate for the narrow kernel. Right: Absolute error from integrating the log of the kernel. The error is small over a large range of κ values, suggesting the kernel is correctly normalized. Quadrature error is not a problem in this case.