Empirical Normal

1 Why

What is a reasonable normal density to associate with some data?

2 Definition

Let $x^1, \ldots, x^n$ be a dataset in $\mathbf{R}^d$. The *empirical mean* of the dataset is

$$\frac{1}{n}\sum_{k=1}^{n} x^k.$$

The *empirical covariance* of the dataset is

$$\frac{1}{n}\sum_{i=1}^{n}(x^k - \frac{1}{n}\sum_{k=1}^{n} x^k)(x^k - \frac{1}{n}\sum_{k=1}^{n} x^k)^\top.$$

The *empirical normal* associated with the dataset is the normal density whose mean is the empirical mean and whose covariance is the empirical covariance.