Société de Calcul Mathématique, S. A.
Algorithmes et Optimisation

Uniform law and normalization

- a warning -

by Bernard Beauzamy

May 2008

Warning : if you take a uniform law on the components of a vector, and then normalize the vector, you do not get a uniform law on the normalized vector.

Quite often, people do the following :

- They draw a sample for a random vector $(x_1,\ldots,x_N)$; each x_n follows a uniform law, for instance on the interval $[0,1]$;

- And then they normalize, that is they replace each x_n by

$$y_n = \frac{x_n}{\sqrt{\sum_{i=1}^{N} x_i^2}}, \text{ so that indeed } \sum_{n=1}^{N} y_n^2 = 1.$$

Here are the two main situations where such a random sampling, followed by normalization, are met :

1. Taking a random direction in a many-dimensional space : starting from a given point, one wants to move in any direction in the space. But the "direction" is defined by a normalized vector. For instance, you are in a space with 40 parameters, you have done some measurement at a given point and you want to explore the neighborhood of this point. For this aim, you have to choose a direction at random.

2. Defining random proportions. For instance, you have 40 goods, and you want to define something which is a combination of these goods : $\alpha_1 g_1 + \cdots + \alpha_N g_N$, with $\alpha_i \geq 0$ and $\sum \alpha_i = 1$. You want to produce something which is a random combination, namely the α_i should be random, still satisfying the normalization $\sum \alpha_i = 1$.

In fact, these two examples are the same : in the first case, the normalization is made using the l_2 norm, and in the second the l_1 norm.

People often think that,if you take the vector $X=(x_1,...,x_N)$ with a uniform law, and then normalize it, the normalized vector $Y=\left(y_1,...,y_n\right)$ will have a uniform law on the unit sphere, or, if one prefers, that all portions of the unit sphere have the same probability. This is wrong.

In mathematical terms, the radial projection of the uniform law on the unit hypercube is not the uniform law on the unit sphere.

Let us see this on a very simple example, dimension 2.

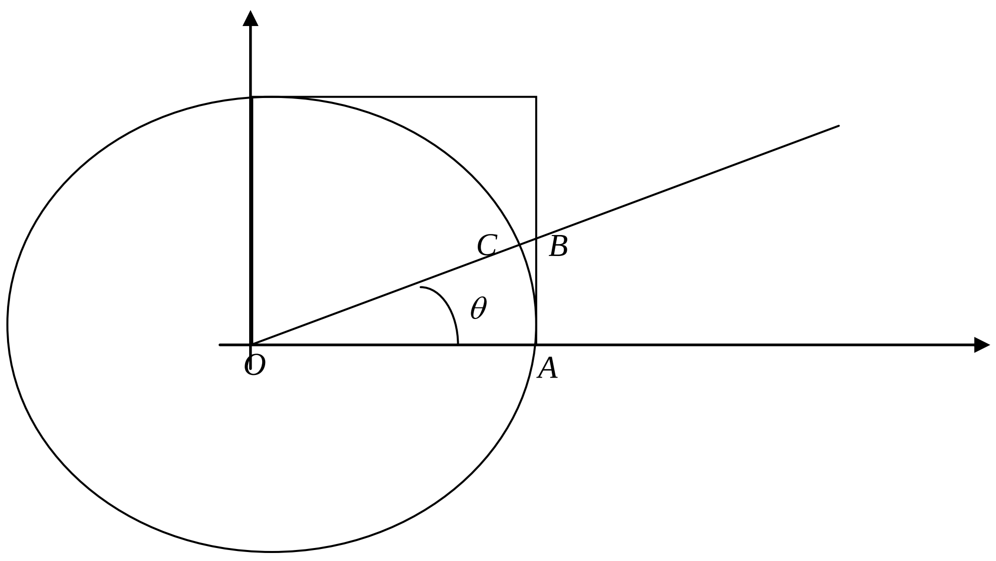
One takes two random numbers x_1 and x_2 with uniform law between 0 and 1. This means that we have a uniform law in the unit square $\left[0,1\right]^2$.

For each x_1 and x_2, we consider the normalized vector :

$$y_1=\frac{x_1}{\sqrt{x_1^2+x_2^2}}$$
$$y_2=\frac{x_2}{\sqrt{x_1^2+x_2^2}}$$

and we want to find the law of this vector, which is on the unit circle.

Fix any θ, $0\le\theta\le\frac{\pi}{2}$:



The point Y is on the arc AC if and only if the point X is inside the triangle OAB. The area of this triangle is :

If $0\le\theta\le\frac{\pi}{4}$, $area(OAB)=\frac{\tan(\theta)}{2}$

If $\frac{\pi}{4} \le \theta \le \frac{\pi}{2}$, $area(OAB) = 1 - \frac{\tan(\frac{\pi}{2} - \theta)}{2} = 1 - \frac{1}{2\tan(\theta)}$

Let $f(\theta)$ be the density function of the vector Y and $F(\theta)$ its repartition function. We get :

If $0 \le \theta \le \frac{\pi}{4}$, $F(\theta) = \frac{\tan(\theta)}{2}$

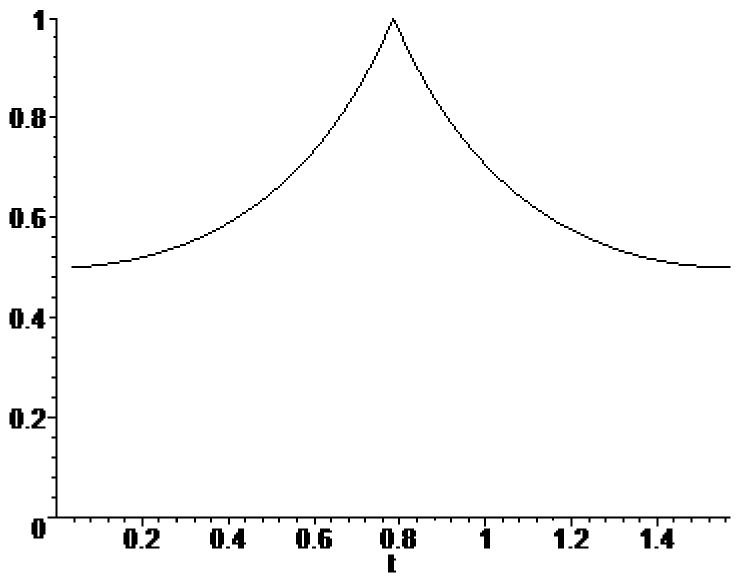
If $\frac{\pi}{4} \le \theta \le \frac{\pi}{2}$, $F(\theta) = 1 - \frac{1}{2\tan(\theta)}$

Since $f(\theta) = F'(\theta)$, we obtain :

If $0 \le \theta \le \frac{\pi}{4}$, $f(\theta) = \frac{1}{2\cos^2(\theta)}$

If $\frac{\pi}{4} \le \theta \le \frac{\pi}{2}$, $f(\theta) = \frac{1}{2\sin^2(\theta)}$

Here is the graph of this density of probability :



We see that it is not at all uniform !

Let us compute the probability of the intervals

$$I_1 = \left\{ 0 \le \theta \le \frac{\pi}{10} \right\}$$

$$I_2 = \left\{ \frac{2\pi}{10} \le \theta \le \frac{3\pi}{10} \right\}$$

Both have the same width, namely $\frac{\pi}{10}$, and the second one is centered around $\frac{\pi}{4}$. We get:

$$P\{I_1\} = F\left(\frac{\pi}{10}\right) = \frac{\tan(\pi/10)}{2} \approx 0.1625$$

$$\begin{aligned} P\{I_2\} &= F\left(\frac{\pi}{4}+\frac{\pi}{20}\right) - F\left(\frac{\pi}{4}-\frac{\pi}{20}\right) \\ &= 2\left(F\left(\frac{\pi}{4}\right) - F\left(\frac{\pi}{4}-\frac{\pi}{20}\right)\right) \\ &= 1 - \tan\left(\frac{\pi}{4}-\frac{\pi}{10}\right) \approx 0.4905 \end{aligned}$$

So there is a considerable difference between both probabilities, despite the fact that both intervals have same width.

The same mistake would occur if we would project upon a hyperplane, instead of the unit circle, and would occur as well with other laws (not uniform) : the ratio between the area of the sector (on the circle or on a hyperplane) and the area of the triangle is not constant.

Uniform law on the unit sphere

Communicated by Paul Deheuvels, Professor, University Paris 6, member French Academy of Sciences.

Let $N \geq 1$ and let $X_1,...,X_N$ be independent normal variables (mean 0, variance 1). Then the variables :

$$\frac{X_1}{\sqrt{X_1^2+\cdots+X_N^2}},\ldots,\frac{X_N}{\sqrt{X_1^2+\cdots+X_N^2}}$$

are independent and follow a uniform law on the unit sphere of the $N-$dimensional Euclidean space.

In other words, in order to obtain a uniform law on the sphere, one should not start with uniform variables, but with Gaussian.

Reference :

Muirhead, R. J. (1982). Aspects of Multivariate Statistical Theory. Wiley, New York, ISBN 0-471-09442-0, page 37.