

[Dirichlet Distribution](#) [Latent Dirichlet Allocation](#) [Intuitive Explanations](#) [Probability \(statistics\)](#)+1 

What is an intuitive explanation of the Dirichlet distribution?

[Answer](#) [Follow](#) · 83 [Request](#) 1   


5 Answers


**Max Sklar**, Data Scientist with a Math Background

Answered August 16, 2017 · Upvoted by

Yinan Mao, Ph.D Statistics, National University of Singapore


I like to draw an analogy between the Dirichlet Distribution and the normal distribution, since most people understand the normal distribution.

The [normal distribution](#)  is a probability distribution over all the real numbers. It is described by a mean and a variance. The mean is the expected value of this distribution, and the variance tells us how much we can expect samples to deviate from the mean. If the variance is very high, then you're going to see values that are both much smaller than the mean and much larger than the mean. If the variance is small, then the samples will be very close to the mean. If the variance goes close to zero, all samples will be almost exactly at the mean.

The [Dirichlet distribution](#)  is a probability distribution as well - but it is not sampling from the space of real numbers. Instead it is sampling over a probability simplex.

And what is a probability simplex? It's a bunch of numbers that add up to 1. For example:

(0.6, 0.4)
(0.1, 0.1, 0.8)
(0.05, 0.2, 0.15, 0.1, 0.3, 0.2)

These numbers represent probabilities over K distinct categories. In the above examples, K is 2, 3, and 6 respectively. That's why they are also called [categorical distributions](#) .

When we are dealing with categorical distributions and we have some uncertainty over what that distribution is, simplest way to represent that uncertainty as a probability distribution is the Dirichlet.

A K-dimensional Dirichlet distribution has K parameters. These parameters can be any positive number. For example, a 4-dimensional Dirichlet may look like this:






(23, 6, 32, 39)

In the normal case, the mean and the variance tell us what kind of samples to expect. What do the above parameters tell us? Note that these 4 parameters can be normalized (divided by their sum) to form a probability distribution times a normalization constant:


$100 * (0.23, 0.06, 0.32, 0.39)$

The probabilities that come out of it (23%, 6%, 32%, 39%) just happen to be the mean value of the Dirichlet! So, all samples from it will center around that simplex. The normalization constant - 100 in this case - isn't the variance but it's related. The higher it is, the closer samples will be to the mean. 100 is a fairly high weight, so most samples from this distribution will be close to (23%, 6%, 32%, 39%).

When the normalization constant gets very low (close to zero) the variance gets higher and higher. The furthest you can get from a point in the simplex is usually in one of the corners, for example (0, 0, 1, 0). When the normalization constant gets low, not only do we expect to be far away from the mean. We actually expect to flip to one of the corners of the simplex with probabilities as described by the mean.

To learn more, check out my presentation "Digging into the Dirichlet" that's posted on [Hakka Labs](#)  with slides, or on [YouTube](#)  where you can follow slides [here](#) . Also check out [my paper](#)  and accompanying [python code](#)  on how to compute the most likely Dirichlet that fits a given dataset.

41.5K views · View Upvoters · View Sharers

 319  1  5   

5 comments from Jose Tomas Mendez and more

Related Questions

[Latent Dirichlet Allocation \(LDA\): What is the best way to determine k \(number of topics\) in...](#)

[What is a good explanation of Latent Dirichlet Allocation?](#)

[Latent Dirichlet Allocation \(LDA\): What does it mean by "Generating a word from a..."](#)

[What's the difference between Latent Semantic Indexing \(LSI\) and Latent Dirichlet Allocation...](#)

[What is an intuitive explanation of Dirichlet process clustering? How do Polya's Urn or Sti...](#)

[What is the difference between Latent Semantic Analysis \(LSA\) and Latent Dirichlet...](#)

[What is the best way to explain topic modeling to a layman?](#)

[What is the difference between keyword extraction and Topic modeling \(such as Latent...](#)

[What is an intuitive explanation of Gradient Boosting?](#)

[What is the k-Nearest Neighbour algorithm?](#)

[What type of problems can be solved by this...](#)

[Ask Question](#)