Gamma Function, Beta Distribution and Dirichelet Distribution

Zengchang Qin

Intelligent Computing and Machine Learning Lab
School of ASEE, Beihang University
zengchang.qin@gmail.com

November 4, 2011

Abstract

Introduction to Dirichelet Distribution and Maximum Entropy Principal.

1 Gamma Function

Gamma function, which is usually represented by the capital Greek letter Γ, is an extension of factorial function [1] with its argument n shifted down by 1. When n is a real positive integer:

$$\Gamma(n) = (n-1)! \tag{1}$$

Although the Gamma function is defined for all complex numbers except the non-positive integers, it is defined via an improper integral that converges only for complex numbers with a positive real part:

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \tag{2}$$

The basic motivation of Gamma function is to solve the interpolation [2] problem of finding a smooth curve that connects the points (x, y) given by $y = (x-1)!$ at the positive integer values (which are discrete given n is a positive integer) for x. It is possible to find a general formula for factorials using tools such as integrals and limits from calculus. A good solution to this is the gamma function.

[1] Factorial function for an integer n is defined by: $f(n) = n! = \prod_{k=1}^{n} k$. In computing factorial, we usually use its recursive definition:

$$n! = \begin{cases} 1 & \text{if } n = 1; \\ (n-1)! \times n & \text{if } n > 0. \end{cases}$$

[2] In the mathematical field of numerical analysis, interpolation is a method of constructing new data points within the range of a discrete set of known data points. In engineering and science, one often has a number of data points, obtained by sampling or experimentation, which represent the values of a function for a limited number of values of the independent variable. It is often required to interpolate (i.e. estimate) the value of that function for an intermediate value of the independent variable. This may be achieved by curve fitting or regression analysis.

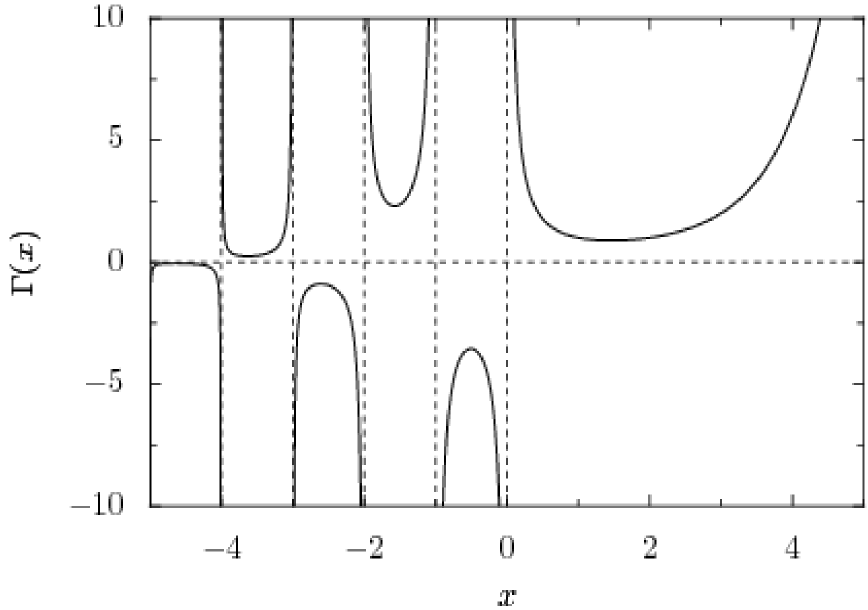

Figure 1: Gamma function.

2 Beta Distribution

The Beta function, also called the Euler integral of the first kind, is a special function defined by:

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)} \tag{3}$$

Beta distribution has two positive parameters α, β defined on the interval $[0, 1]$. The probability density function is defined based on Beta function:

$$f(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1 - x)^{\beta-1} \tag{4}$$

In Bayesian statistics, since beta distributions provide a family of conjugate prior distributions for binomial (including Bernoulli) and geometric distributions. In details, the beta distribution can be seen as the posterior probability of the parameter p of a binomial distribution after observing $\alpha - 1$ successes (with probability p of success) and $\beta - 1$ failures (with probability $1 - p$ of failure). Another way to express this is that placing a prior distribution of $Beta(\alpha, \beta)$ on the parameter p of a binomial distribution is equivalent to adding α pseudo-observations of "success" and β pseudo-observations of "failure" to the actual number of successes and failures observed, then estimating the parameter p by the proportion of successes over both real and pseudo-observations.

3 Dirichlet Distribution

Dirichlet Distribution (after Johann Peter Gustav Lejeune Dirichlet) is a continuous multivariate probability distribution. The Dirichlet distribution is the multivariate generalization of the beta

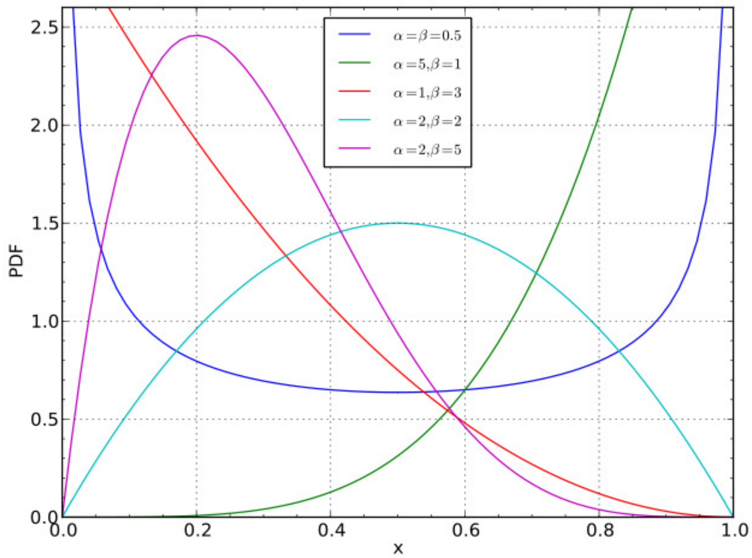

Figure 2: Beta distribution.

distribution (Eq. 4). The PDE for the Dirichlet distribution of order K is a function of a K-dimensional vector $\mathbf{x} = \{x_1, \ldots, x_K\}$:

$$f(\mathbf{x}; \alpha) = Dir(\alpha_1, \ldots, \alpha_K) = \frac{\Gamma(\sum_j \alpha_j)}{\prod_j \Gamma(\alpha_j)} \prod_{j=1}^{K} x_j^{\alpha_j - 1} \tag{5}$$

where $\sum_{i=1}^{K} x_i = 1$ and $\Gamma(\cdot)$ is Gamma function. The first term of eq. 5 is a normalizing constant: which is a multi-nomial beta function expressed in terms of the gamma function:

$$\frac{\prod_j \Gamma(\alpha_j)}{\Gamma(\sum_j \alpha_j)} = B(\alpha) \tag{6}$$

Hence, eq. 5 can be re-written as:

$$f(\mathbf{x}; \alpha) = \frac{1}{B(\alpha)} \prod_{j=1}^{K} x_j^{\alpha_j - 1} \tag{7}$$

4 Maximum Entropy

In Bayesian probability, the principle of *maximum entropy* is a postulate which states that, subject to known constraints, the probability distribution which best represents the current state of knowledge is the one with largest entropy. Given no testable information, the only constraint is that the sum of probabilities must be one. The maximum entropy discrete probability distribution in this situation is the uniform distribution:

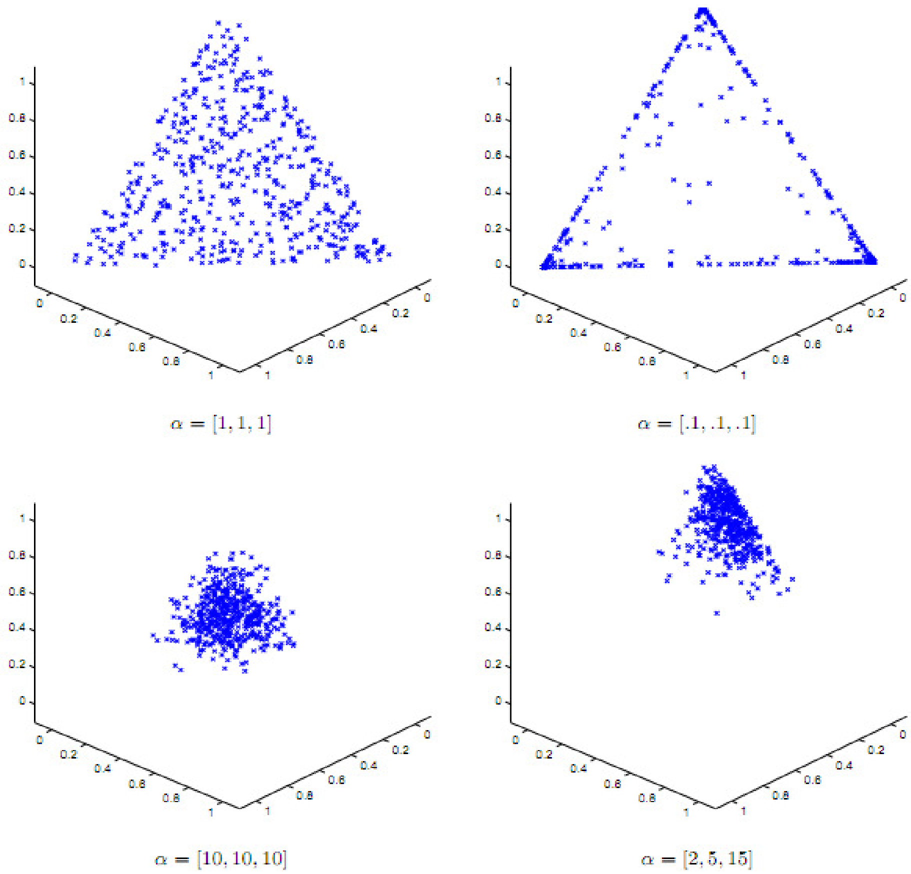

Figure 3: Sampling from given Dirichlet distributions with different parameterizations.

$$p(x_i) = \frac{1}{n} \ for \ i = 1, \ldots, n \tag{8}$$

This can be proved as the following: given the probabilities of all possible outcomes of an event $\{p_1, p_2, \ldots, p_n\}$, the entropy is:

$$H(p) = -\sum_i p_i \log p_i \tag{9}$$

Therefore, the maximum entropy can be obtained to solve the maximization problem of:

$$\arg\max_p H(p) \ s.t. : \sum_i p_i = 1 \tag{10}$$

Using Lagrange multiplier:

$$\mathcal{L}(\lambda, p_1, \ldots, p_n) = -\sum_i p_i \log p_i + \lambda(\sum_i p_i - 1) \tag{11}$$

Let:

$$\frac{\partial \mathcal{L}}{\partial p_i} = -\frac{\partial(p_i \log p_i - \lambda p_i)}{\partial p_i} = -\log p_i - 1 + \lambda = 0 \tag{12}$$

We then can yield that $p_1 = p_2 = \ldots = p_n$. Considering the constraint $\sum_i p_i = 1$, we can obtain: $p_1 = p_2 = \ldots = p_n = \frac{1}{n}$ for achieving the maximum value for $H(p)$.

References

WikiPedia: http://en.wikipedia.org/wiki/Gamma_function

D. J. C. Mackay (2003), *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press.

C. E. Shannon, A mathematical theory of communication, *Bell System Technical Journal*, vol. 27, pp. 379-423 and 623-656, July and October.

J. Wu (2005), Some properties of the normal distribution, Tutorial, source from: *http://www.cc.gatech.edu/~wujx/paper/Gaussian.pdf*, Georgia Institute of Technology.